Quantum variational learning for quantum error-correcting codes

Chenfeng Cao[1], Chao Zhang[1], Zipeng Wu[1], Markus Grassl[2], and Bei Zeng[1]

[1]Department of Physics, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China
[2]International Centre for Theory of Quantum Technologies, University of Gdansk, 80-309 Gdansk, Poland

Quantum error correction is believed to be a necessity for large-scale fault-tolerant quantum computation. In the past two decades, various constructions of quantum error-correcting codes (QECCs) have been developed, leading to many good code families. However, the majority of these codes are not suitable for near-term quantum devices. Here we present VarQEC, a noise-resilient variational quantum algorithm to search for quantum codes with a hardware-efficient encoding circuit. The cost functions are inspired by the most general and fundamental requirements of a QECC, the Knill-Laflamme conditions. Given the target noise channel (or the target code parameters) and the hardware connectivity graph, we optimize a shallow variational quantum circuit to prepare the basis states of an eligible code. In principle, Var-QEC can find quantum codes for any error model, whether additive or non-additive, degenerate or non-degenerate, pure or impure. We have verified its effectiveness by (re)discovering some symmetric and asymmetric codes, e.g., $((n, 2^{n-6}, 3))_2$ for n from 7 to 14. We also found new $((6, 2, 3))_2$ and $((7, 2, 3))_2$ codes that are not equivalent to any stabilizer code, and extensive numerical evidence with VarQEC suggests that a $((7, 3, 3))_2$ code does not exist. Furthermore, we found many new channel-adaptive codes for error models involving nearest-neighbor correlated errors. Our work sheds new light on the understanding of QECC in general, which may also help to enhance near-term device performance with channel-adaptive error-correcting codes.

Bei Zeng: zengb@ust.hk

1 Introduction

Fault-tolerant quantum computers promise to solve some computational problems much faster than classical machines, such as quantum chemistry simulation [1], prime factorization [2], solving linear systems of equations [3]. However, quantum information carried by current noisy intermediate-scale quantum (NISQ) systems is highly fragile and can be easily altered by the environment. The aforementioned tasks are so far out of reach.

The most promising technique to maintain coherence and protect the quantum information from noise is *quantum error-correcting codes* [4–8]. The main idea of quantum error correction is to encode the low-dimensional quantum state in a larger system such that errors occurring during the computation can be corrected due to the physical redundancy. As long as the noise rate p is below a specific threshold, QECCs can correct the error and reduce the error probability from $\mathcal{O}(p)$ to higher orders. In recent years, the intrinsic connections between QECCs and other areas of physics, such as quantum gravity [9], have also been noticed.

Knill and Laflamme devised sufficient and necessary conditions (known as the *Knill-Laflamme conditions*) for quantum error correction [10]. In principle, we can find any QECC as long as we find solutions to the Knill-Laflamme conditions. However, solving these systems of equations is extremely difficult in the general case. Therefore many open problems in this field remain unsolved, e.g., do all degenerate QECCs obey the Hamming bound? Which QECC has the highest error threshold? In practice, researchers usually analyze QECCs under the Pauli framework and have developed various QECC families, such as surface codes [11, 12], Calderbank-Shor-Steane (CSS) codes [13, 14], stabilizer codes [5], code-

word stabilized (CWS) codes [15, 16], quantum low-density parity-check codes [17, 18].

Up till now, no logical qubit/operation with useful fidelity was realized in experiments since current gate noise rates are still much larger than the requirements. Very recently, Egan *et al.* [19] prepared a Bacon–Shor logical qubit with 13 trapped-ion qubits and demonstrated a logical single-qubit Clifford gate. Further, Postler *et al.* [20] demonstrated a logical T-gate based on the 7-qubit color code. However, the fidelities of these state-of-the-art logical qubits are even lower than those of the physical qubits. From a theoretical perspective, the codes used in those experiments are not device-tailored and may not be optimal for the system. The noise channels on different physical platforms differ significantly [21–23]. Symmetric QECC constructions under the Pauli framework can not be directly adapted to non-Hermitian/non-unitary noise channels. It is highly desirable to design asymmetric or channel-adaptive QECCs with a hardware-efficient encoder. Such device-tailored codes can protect logical information more efficiently.

Besides analytical constructions, researchers have been trying to find QECCs with computational methods for a long time. Refs. [16, 24–26] designed classical algorithms for finding quantum codes associated with graphs. Ref. [27] used numerical greedy search for finding stabilizer codes. These algorithms, however, cannot find arbitrary codes and are extremely time-consuming. With the popularity of artificial intelligence, researchers also started to design and optimize quantum codes with neural networks [28–31]. These classical black-box models perform pretty well for certain problems. In this work, we add a new general method to this toolbox. We devise a hybrid quantum-classical algorithm called VarQEC for finding quantum error-correcting codes. The cost functions therein are based on the Knill-Laflamme conditions. We iteratively update the parameters in a variational quantum circuit (VQC) with stochastic gradient descent. If the final cost functions are sufficiently small, we obtain an approximate quantum code whose inaccuracy is bounded. Compared with the classical iterative algorithm introduced in Ref. [32], our method yields the encoding circuit, not merely the encoding isometry. After finding a QECC and its encoder, the de-

coding operation can be found via various methods like semidefinite programming [33], convex optimization [34], or classical/quantum machine learning [35–37].

VarQEC allows for non-Hermitian or non-unitary errors and is surprisingly effective. We numerically verify its effectiveness up to 14 qubits. For symmetric Pauli errors, we successfully rediscover many good quantum codes, e.g., $((5,2,3))_2$, $((5,6,2))_2$, $((6,2,3))_2$, $((7,2,3))_2$, $((8,8,3))_2$, $((9,8,3))_2$, $((10,2^4,3))_2$, $((11,2^5,3))_2$, $((12,2^6,3))_2$, $((13,2^7,3))_2$, $((14,2^8,3))_2$, $((10,4,4))_2$. Some of the $((6,2,3))_2$, $((7,2,3))_2$ codes we find are not locally equivalent to any CWS code. It is an open question of whether there is a quantum code with parameters $((7,3,3))_2$, our numerical evidence suggests that it is non-existent. Then we apply VarQEC to search for asymmetric codes (which detect more Pauli-X/Y errors than Pauli-Z errors or vice versa) and make new discoveries. Furthermore, we search for channel-adaptive codes for nearest-neighbor collective amplitude damping and nearest-neighbor collective phase-flips, and find eligible new codes with a hardware-efficient encoding circuit for various connectivity graphs. Since VarQEC is capable to find a QECC with the shallowest possible encoding circuit, it is promising to design codes with sufficient fidelity that can be tested and implemented on near-term devices. Although only relatively small systems were investigated in this paper, hierarchical concatenation can construct good quantum codes with large code lengths and distances [5, 38, 39].

The paper is organized as follows. In Sec. 2, we introduce some background of quantum error correction. In Sec. 3, we introduce our cost functions and present propositions to support our definitions. In Sec. 4, we explain the VarQEC algorithm in detail. In Sec. 5, we show quantum codes (re)discovered thereby, including symmetric, asymmetric, and channel-adaptive codes for nearest-neighbor collective amplitude damping and nearest-neighbor collective phase-flips. Sec. 6 discusses the noise resilience feature of VarQEC. Sec. 7 discusses the barren plateaus and the noise-induced barren plateaus in VarQEC optimization. In Sec. 8, we verify our algorithm by an experiment on an IBM quantum device. The conclusions and future directions are sum-

marized and discussed in Sec. 9. The appendices give some proof details, a discussion on overparameterization, an alternative variational ansatz, some non-CWS quantum codes, and a list of the quantum weight enumerators of quantum codes discovered by VarQEC.

2 Preliminaries

In classical computation and communication, redundancy is added when encoding a message such that the errors can be detected and corrected. Although each bit may flip with some probability, the encoded message can be recovered with high probability. The philosophy behind quantum error correction is the same. We use several low-fidelity physical qudits (e.g., qubits) to encode the logical quantum information redundantly and nonlocally. Then quantum errors can be detected through syndrome measurements and corrected through a unitary operation. A q-ary QECC $\mathcal{C}$ is a K-dimensional subspace of the q^n-dimensional Hilbert space $(\mathbb{C}^q)^{\otimes n}$, where n is the number of physical qudits (referred to as the *code length*). For qubit systems, $q = 2$, $\mathcal{C} \subset (\mathbb{C}^2)^{\otimes n}$. When $K = 1$, the code is a fixed quantum state without computational use. Throughout this paper, we only discuss $K \geq 2$.

Knill and Laflamme developed a general theory of quantum error correction. They obtained the sufficient and necessary conditions for an exact QECC [10]: a quantum code with orthonormal basis states $\{|\psi_j\rangle\}$ corrects the error set $\mathcal{E} = \{E_\alpha\}$ if and only if

$$P_c E_\alpha^\dagger E_\beta P_c = \lambda_{\alpha\beta} P_c, \tag{1}$$

holds for all $E_\alpha, E_\beta \in \mathcal{E}$. Here, $P_c = \sum_j |\psi_j\rangle\langle\psi_j|$ is the orthogonal projector onto the code space, and each $\lambda_{\alpha\beta}$ is a complex number. Moreover, we say the quantum code is *non-degenerate* if the matrix $\lambda_{\alpha\beta}$ has full rank [40].

We can understand these conditions intuitively. When $i \neq j$, $\langle\psi_i|E_\alpha^\dagger E_\beta|\psi_j\rangle = 0$ for any error product $E_\alpha^\dagger E_\beta$. This means orthogonal logical states remain orthogonal after the noise channel, the logical information is not corrupted. When $i = j$, $\langle\psi_j|E_\alpha^\dagger E_\beta|\psi_j\rangle = \lambda_{\alpha\beta}$ with $\lambda_{\alpha\beta}$ being a constant only determined by the error product. This indicates that the projections between subspaces induced by different errors are

information-preserving, the errors have an orthogonal decomposition. Therefore, we can correct the error without knowing or destroying the quantum superposition state.

The *quantum error detection conditions* have a similar form: a quantum code with code space projector P_c can detect the error set $\mathcal{E} = \{E_\mu\}$ if and only if

$$P_c E_\mu P_c = \lambda_\mu P_c \tag{2}$$

holds for all $E_\mu \in \mathcal{E}$.

In experiments, most quantum errors are uncorrelated single-qudit errors. A natural measure of the capability of a QECC is the number of single-qudit errors that it can detect. This motivated the concept of "code distance": the distance of a QECC is the largest possible integer d such that the code can detect any error non-trivially acting on at most $d - 1$ qudits. Researchers usually denote the code parameters of a q-ary QECC with code length n, code dimension K, and code distance d as $((n, K, d))_q$.

Comparing the Knill-Laflamme conditions and quantum error detection conditions, we know that a distance-d QECC can correct any error set $\mathcal{E} = \{E_\alpha\}$ with each E_α non-trivially acting on at most $\lfloor (d-1)/2 \rfloor$ qudits.

For convenience, 2-ary quantum codes are usually constructed and analyzed in the Pauli framework. Consider an n-fold Pauli tensor product

$$O_\alpha \in \{X, Y, Z, I\}^{\otimes n}. \tag{3}$$

Denote the number of X factors, Y factors and Z factors in O_α as $\mathrm{wt}_X(O_\alpha)$, $\mathrm{wt}_Y(O_\alpha)$, and $\mathrm{wt}_Z(O_\alpha)$. The weight of O_α is

$$\mathrm{wt}(O_\alpha) = \mathrm{wt}_X(O_\alpha) + \mathrm{wt}_Y(O_\alpha) + \mathrm{wt}_Z(O_\alpha). \tag{4}$$

An equivalent definition of the code distance of a QECC with projector P_c is the largest possible integer d such that

$$P_c O_\alpha P_c = \lambda_\alpha P_c \tag{5}$$

holds for all Pauli tensor product O_α with $\mathrm{wt}(O_\alpha) < d$.

In practical scenarios, Pauli-Z errors are usually more prevalent than Pauli-X and Pauli-Y [41]. Accordingly, we use a parameter c_Z to characterize this noise bias and define the following c_Z-effective weight and c_Z-effective distance.

Definition 1 *The c_Z-effective weight of a Pauli tensor product O_α is*

$$\mathrm{wt}_e(O_\alpha, c_Z) = \mathrm{wt}_X(O_\alpha) + \mathrm{wt}_Y(O_\alpha) + c_Z \mathrm{wt}_Z(O_\alpha), \quad (6)$$

where $c_Z > 0$. The c_Z-effective distance of a quantum code with projector P_c is the largest possible integer $d_e(c_Z)$ such that

$$P_c O_\alpha P_c = \lambda_\alpha P_c \quad (7)$$

holds for all Pauli tensor product O_α with $\mathrm{wt}_e(O_\alpha, c_Z) < d_e(c_Z)$.

This definition is a generalization of the concept of "effective distance" introduced in Ref. [42]. An asymmetric code with code parameters $((n, K, d_e(c_Z)))_2$ can correct arbitrary Pauli error with c_Z-effective weight smaller than $d_e(c_Z)/2$, and detect arbitrary Pauli error with c_Z-effective weight smaller than $d_e(c_Z)$. When Pauli-Z errors occur more frequently than Pauli-X/Y errors, $0 < c_Z < 1$; when the relaxation times (T_1) are much smaller than the dephasing times (T_2), Pauli-X/Y errors occur more frequently, $c_Z > 1$.

Quantum codes with relatively small distances can be concatenated to construct a code with large code length and distance, as illustrated in Fig. 1. Suppose we have an outermost code with parameters $((n_1, K, d_1))_q$, other outer codes with parameters $((n_2, q, d_2))_q$, $((n_3, q, d_3))_q$, ..., $((n_{l-1}, q, d_{l-1}))_q$, and an inner code with parameters $((n_l, q, d_l))_q$. We can construct a large code through several levels of concatenation: the logical data is first encoded using the outermost code, each physical qudit therein is further encoded using the $((n_2, q, d_2))_q$ code, and so forth. The hierarchically concatenated quantum code has parameters

$$((\prod_j n_j, K, \prod_j d_j))_q. \quad (8)$$

Likewise, we can concatenate asymmetric codes. A distance lower bound is given as follows.

Theorem 1 *Consider asymmetric outer codes with parameters $((n_1, K, d_e(c_Z) = \delta_1))_2$, $((n_2, 2, d_e(c_Z) = \delta_2))_2$, $((n_3, 2, d_e(c_Z) = \delta_3))_2$, ..., $((n_{l-1}, 2, \delta_{l-1}))_2$, and an inner code with parameters $((n_l, 2, \delta_l))_2$. Concatenating these codes yields a new code with parameters*

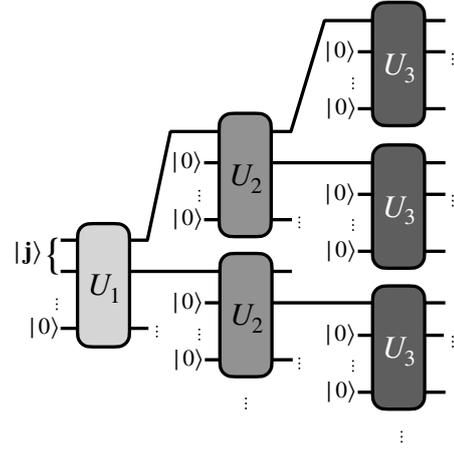$$((\prod_j n_j, K, d_e(c_Z) = \delta))_2 \quad (9)$$

Figure 1: Schematic illustration of quantum code concatenation. After finding quantum codes with encoders $U_1, U_2, U_3, \ldots$, we hierarchically concatenate these encoders to obtain a large-distance quantum code.

where

$$\delta \geq \min\{1, c_Z\} \prod_j \left\lceil \frac{\delta_j}{\max\{1, c_Z\}} \right\rceil. \quad (10)$$

Proof Assume the concatenated code cannot detect a Pauli tensor product O_α. For the outer code, errors occur on at least $\lceil \delta_1 / \max\{1, c_Z\} \rceil$ qubits. Each of these qubits is connected to a block of the first inner code $((n_2, 2, d_e(c_Z) = \delta_2))_2$ and for every such block, errors occur on at least $\lceil \delta_2 \max\{1, c_Z\} \rceil$ qubits. From similar arguments, errors occur on at least $\lceil \delta_j \max\{1, c_Z\} \rceil$ qubits in the j-th block. The weight of O_α is bounded by

$$\mathrm{wt}(O_\alpha) \geq \prod_j \lceil \delta_j / \max\{1, c_Z\} \rceil. \quad (11)$$

Hence, the c_Z-effective weight of O_α is at least

$$\min\{1, c_Z\} \prod_j \lceil \delta_j / \max\{1, c_Z\} \rceil. \quad (12)$$

The concatenated code can detect any Pauli tensor product with c_Z-effective weight smaller than this value, we conclude

$$\delta \geq \min\{1, c_Z\} \prod_j \left\lceil \frac{\delta_j}{\max\{1, c_Z\}} \right\rceil. \quad (13)$$

$\blacksquare$

3 Theoretical Basis

A lot of methods for constructing QECCs are using the stabilizer formalism, but there are not

that many outside the Pauli framework. This work aims to search for quantum codes based on the most fundamental principle, i.e., the Knill-Laflamme conditions and the quantum error detection conditions. A crucial tool in our scheme is the variational quantum circuit which consists of multiple layers of parameterized quantum gates.

The primary ingredient of a variational algorithm is the cost function(s). We define the cost functions of VarQEC as follows.

Definition 2 (Cost functions) *Consider an error set $\mathcal{E} = \{E_\mu\}$ and a length-n quantum code with parameterized orthogonal basis states*

$$\{|\psi_1(\boldsymbol{\theta})\rangle, |\psi_2(\boldsymbol{\theta})\rangle, \ldots, |\psi_K(\boldsymbol{\theta})\rangle\}. \qquad (14)$$

We define the ℓ_1-norm cost function

$$
\begin{aligned}
C^{\ell_1}_{n,K,\mathcal{E}}(\boldsymbol{\theta}) \equiv \sum_{E_\mu \in \mathcal{E}} \Big(&\sum_{1 \le i < j \le K} |\langle\psi_i|E_\mu|\psi_j\rangle| \\
&+ \sum_{j=1}^{K} |\langle\psi_j|E_\mu|\psi_j\rangle - \overline{\langle E_\mu\rangle}|/2 \Big)
\end{aligned}
\qquad (15)
$$

and the ℓ_2-norm cost function

$$
\begin{aligned}
C^{\ell_2}_{n,K,\mathcal{E}}(\boldsymbol{\theta}) \equiv \sum_{E_\mu \in \mathcal{E}} \Big(&\sum_{1 \le i < j \le K} |\langle\psi_i|E_\mu|\psi_j\rangle|^2 \\
&+ \sum_{j=1}^{K} |\langle\psi_j|E_\mu|\psi_j\rangle - \overline{\langle E_\mu\rangle}|^2/4 \Big)
\end{aligned}
\qquad (16)
$$

where $\overline{\langle E_\mu\rangle} = \sum_{j=1}^{K}\langle\psi_j|E_\mu|\psi_j\rangle/K$.

Clearly, $C^{\ell_1}_{n,K,\mathcal{E}}$ and $C^{\ell_2}_{n,K,\mathcal{E}}$ are always non-negative and have the same zero-points. When $C^{\ell_1}_{n,K,\mathcal{E}} \le 1$, $C^{\ell_2}_{n,K,\mathcal{E}} \le (C^{\ell_1}_{n,K,\mathcal{E}})^2$. When $C^{\ell_1}_{n,K,\mathcal{E}} = 0$, the quantum code can perfectly detect the error set $\mathcal{E}$.

To find symmetric codes with code parameters $((n,K,d))_2$, we use the Pauli error model and choose

$$\mathcal{E} = \{O_\alpha | \operatorname{wt}(O_\alpha) < d\}, \qquad (17)$$

where O_αs are Pauli tensor products. Likewise, when searching for asymmetric codes with code parameters $((n,K,d_e(c_Z)))_2$, we choose

$$\mathcal{E} = \{O_\alpha | \operatorname{wt}_e(O_\alpha, c_Z) < d_e(c_Z)\}. \qquad (18)$$

To find channel-adaptive codes for a general noise channel $\mathcal{N}(\rho) = \sum_\alpha E_\alpha \rho E_\alpha^\dagger$, we choose

$$\mathcal{E} = \{E_\alpha^\dagger E_\beta | E_\alpha, E_\beta \text{ are Kraus operators of } \mathcal{N}\}. \qquad (19)$$

Note that the error set $\mathcal{E}$ in principle can include non-unitary and non-Hermitian errors. For such errors, we can either twirl them to Pauli errors or simulate them directly by adding ancilla qubits and performing *positive-operator valued measures* (POVMs).

In practice, due to the inexact realization of an encoding isometry, quantum error correction/detection conditions are not exactly satisfied, and QECCs cannot protect the information from errors perfectly. Nevertheless, QECCs can still detect and correct most errors. Such approximate quantum error correction schemes hold great promise [43, 44]. A parameter ε characterizes the *inaccuracy* of an approximate code. If a QECC is ε-correctable for a noise channel $\mathcal{N}$, its worst-case entanglement fidelity is greater than $1 - \varepsilon$ with appropriate recovery [45]. Bény *et al.* proposed an approximate version of the Knill-Laflamme conditions for such approximate codes.

Lemma 1 (Corollary 2, Ref. [46]) *A code defined by the projector P_c is ε-correctable under a noise channel $\mathcal{N}(\rho) = \sum_\alpha E_\alpha \rho E_\alpha^\dagger$, if and only if*

$$P_c E_\alpha^\dagger E_\beta P_c = \lambda_{\alpha\beta} P_c + P_c B_{\alpha\beta} P_c, \qquad (20)$$

where $\lambda_{\alpha\beta}$ are the components of a non-negative Hermitian operator with trace one, $B_{\alpha\beta}$ is a Hermitian operator, and the Bures distance [47] between two channels $\Lambda(\rho) = \sum_{\alpha\beta}\lambda_{\alpha\beta}\operatorname{Tr}(\rho)|\alpha\rangle\langle\beta|$ and $(\Lambda + \mathcal{B})(\rho) = \Lambda(\rho) + \sum_{\alpha\beta}\operatorname{Tr}(\rho B_{\alpha\beta})|\alpha\rangle\langle\beta|$ satisfies

$$d(\Lambda + \mathcal{B}, \Lambda) \le \varepsilon. \qquad (21)$$

Based on this lemma, we modify Corollary 5 of Ref. [45] and give a proposition to support our definitions.

Proposition 2 *Consider an n-qubit noise channel $\mathcal{N}(\rho) = \sum_\alpha E_\alpha \rho E_\alpha^\dagger$, and a quantum error-correcting code*

$$\mathcal{C} = span\{|\psi_1\rangle, |\psi_2\rangle, \ldots, |\psi_K\rangle\}. \qquad (22)$$

We choose the error product set $\mathcal{E} = \{E_\alpha^\dagger E_\beta | E_\alpha, E_\beta \text{ are Kraus operators of } \mathcal{N}\}$. Denote the cost function Eq. (15) of the basis states as $C^{\ell_1}_{n,K,\mathcal{E}}$. Then the code $\mathcal{C}$ is ε-correctable under $\mathcal{N}$ with ε bounded by

$$\varepsilon \le K\sqrt{2C^{\ell_1}_{n,K,\mathcal{E}}}. \qquad (23)$$

Proof Let $\lambda_{\alpha\beta} = \sum_j \langle\psi_j|E_\alpha^\dagger E_\beta|\psi_j\rangle/K$.
To satisfy Eq. (20), we set

$$B_{\alpha\beta} = \sum_{i\neq j}\langle\psi_i|E_\alpha^\dagger E_\beta|\psi_j\rangle|\psi_i\rangle\langle\psi_j|$$
$$+ \sum_j (\langle\psi_j|E_\alpha^\dagger E_\beta|\psi_j\rangle - \lambda_{\alpha\beta})|\psi_j\rangle\langle\psi_j|. \quad (24)$$

Then

$$d(\Lambda + \mathcal{B}, \Lambda) \leq K\|\mathcal{B}\|_1^{1/2}$$
$$\leq K\Big(\sum_{\alpha,\beta}\|B_{\alpha\beta}\|_1\Big)^{1/2}$$
$$\leq K\Big(\sum_{\alpha,\beta}\Big(\sum_{i\neq j}|\langle\psi_i|E_\alpha^\dagger E_\beta|\psi_j\rangle|+$$
$$\sum_j|\langle\psi_j|E_\alpha^\dagger E_\beta|\psi_j\rangle - \lambda_{\alpha\beta}|\Big)\Big)^{1/2}$$
$$= K\sqrt{2C_{n,K,\mathcal{E}}^{\ell_1}}. \quad (25)$$

According to Lemma 1, the inaccuracy ε of code $\mathcal{C}$ is upper bounded by $K\sqrt{2C_{n,K,\mathcal{E}}^{\ell_1}}$. ∎

In short, given a noise channel $\mathcal{N}$, as long as we minimize the channel-adaptive cost function to a sufficiently small value, we rigorously find an approximate channel-adaptive code with small inaccuracy. Similar bounds for symmetric or asymmetric quantum codes are given as follows.

Proposition 3 *Consider an n-qubit noise channel $\mathcal{N}(\rho) = \sum_\alpha E_\alpha\rho E_\alpha^\dagger$ where each E_α non-trivially acts on no more than $\lfloor(d-1)/2\rfloor$ qubits, and a quantum error-correcting code*

$$\mathcal{C} = span\{|\psi_1\rangle, |\psi_2\rangle, \ldots, |\psi_K\rangle\}. \quad (26)$$

We choose $\mathcal{E} = \{O_\alpha| \text{wt}(O_\alpha) < d\}$, where O_αs are Pauli tensor products. Denote the cost function Eq. (15) of the basis states as $C_{n,K,\mathcal{E}}^{\ell_1}$, the number of Kraus operators of $\mathcal{N}$ as m. Then the code $\mathcal{C}$ is ε-correctable under $\mathcal{N}$ with ε bounded by

$$\varepsilon \leq 2^{n/4+d/2}K\sqrt{mC_{n,K,\mathcal{E}}^{\ell_1}}. \quad (27)$$

Proof The proof is given in Appendix A. ∎

Proposition 4 *Consider an n-qubit noise channel $\mathcal{N}(\rho) = \sum_\alpha E_\alpha\rho E_\alpha^\dagger$ with each E_α proportional to a Pauli error with c_Z-effective weight smaller than $d_e(c_Z)/2$, and a quantum error-correcting code*

$$\mathcal{C} = span\{|\psi_1\rangle, |\psi_2\rangle, \ldots, |\psi_K\rangle\}. \quad (28)$$

We choose $\mathcal{E} = \{O_\alpha| \text{wt}_e(O_\alpha, c_Z) < d_e(c_Z)\}$, where O_αs are Pauli tensor products. Denote the cost function Eq. (15) of the basis states as $C_{n,K,\mathcal{E}}^{\ell_1}$, the number of Kraus operators of $\mathcal{N}$ as m. Then the code $\mathcal{C}$ is ε-correctable under $\mathcal{N}$ with ε bounded by

$$\varepsilon \leq K\sqrt{2mC_{n,K,\mathcal{E}}^{\ell_1}}. \quad (29)$$

Proof The proof is given in Appendix B. ∎

Note that Propositions 3 and 4 give pretty loose bounds. The true code inaccuracy, which depends on the particular noise channel, is usually significantly smaller.

4 Algorithm

Variational quantum circuits (VQCs) have been widely used in near-term quantum algorithms for various tasks [48, 49], such as ground state preparation [50, 51], eigenenergy estimation [52, 53], quantum data compression [54, 55], quantum circuit compiling [56, 57]. Give a pure product state as input, one iteratively updates the circuit parameters based on measurement results, and finally outputs the desired state. In VarQEC, the output states serve as the basis states of a quantum code, and its encoder is given by the quantum circuit. The structure of our algorithm is illustrated in Fig. 2.

Suppose we have a NISQ device with a *hardware connectivity graph* G. The vertices denote qubits, and the edges denote adjacent qubit pairs. One can apply single-qubit rotations to each qubit and two-qubit gates to adjacent qubits. We aim to find a K-dimensional QECC that can detect an error set $\mathcal{E} = \{E_\mu\}$, and the encoding circuit should be as shallow as possible.

Before running the algorithm, we design a multilayered VQC which is hardware-efficient for the connectivity graph. Denote the number of VQC layers as L, the maximum acceptable number of layers as $L_{\max}$, the evolution of the VQC as $U(\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ are the circuit parameters. We start from $L = 1$ and sample the initial $\boldsymbol{\theta}$ randomly. Also, we delicately select $k = \lceil\log(K)\rceil$ physical qubits to prepare the logical data, where the logarithm is with respect to base 2. These k qubits should be scattered instead of concentrated since we hope the remaining qubits are connected to them by very few edges.
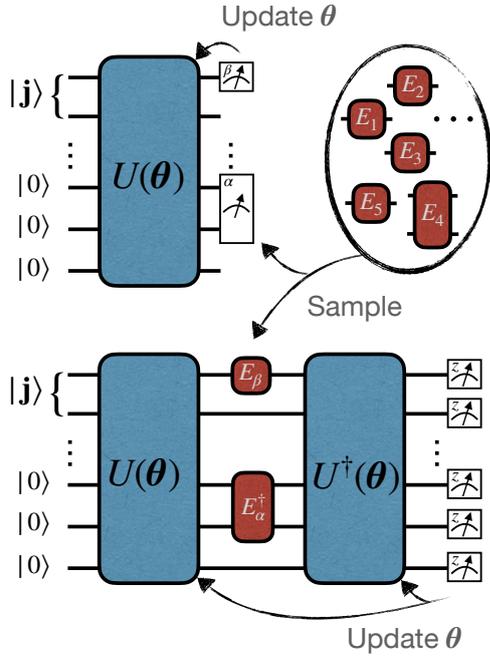
Figure 2: Schematic illustration of VarQEC. The encoder is trained via mini-batch learning: we iteratively sample errors from an error set, run the variational quantum circuit $U(\boldsymbol{\theta})$ and do measurements, then update $\boldsymbol{\theta}$.

First, we initialize the selected qubits to one of the K binary strings $|\mathbf{0}\rangle$, $|\mathbf{1}\rangle$, …, $|\mathbf{K-1}\rangle$, and initialize the remaining qubits to $|0\rangle^{\otimes(n-k)}$. These product states span the input code space

$$\mathcal{C}_{\mathrm{in}} = \mathrm{span}\{|\mathbf{0}\rangle|0\rangle^{\otimes(n-k)}, \ldots, |\mathbf{K-1}\rangle|0\rangle^{\otimes(n-k)}\}. \tag{30}$$

$\mathcal{C}_{\mathrm{in}}$ is a QECC with distance $d = 1$.

The cost functions can be estimated by running specific circuits and doing measurements. To estimate $\langle\psi_j|E_\mu|\psi_j\rangle$, we prepare the initial state $|\mathbf{j-1}\rangle|0\rangle^{\otimes(n-k)}$, evolve the system with the VQC $U(\boldsymbol{\theta})$, then measure the local observable E_μ. If errors $E_{\mu 1}$, $E_{\mu 2}$, … commute, they can be measured simultaneously in a single shot. To estimate $|\langle\psi_i|E_\mu|\psi_j\rangle|$, we start from $|\mathbf{j-1}\rangle|0\rangle^{\otimes(n-k)}$, then sequentially evolve the system with $U(\boldsymbol{\theta})$, E_μ, and $U^\dagger(\boldsymbol{\theta})$, then measure the final state in the computational basis. The measurements are assisted by *post-selection*: we first measure the $n-k$ auxiliary qubits, and if the result is $|0\rangle^{n-k}$, we measure the remaining k qubits. Denote the probability of obtaining the binary string $|\mathbf{i-1}\rangle|0\rangle^{\otimes(n-k)}$ as p_{ij}, $|\langle\psi_i|E_\mu|\psi_j\rangle| = \sqrt{p_{ij}}$. Theoretically, this step will also yield $\langle\psi_j|E_\mu|\psi_j\rangle$. However, since VarQEC is a NISQ algorithm, we prefer to use a shallower circuit to estimate cost function terms whenever possible.

In the above description, we assume the error set $\mathcal{E}$ only consists of Pauli errors. It does not matter if $\mathcal{E}$ includes non-unitary or non-Hermitian terms. Adding ancilla qubits or Pauli twirling can handle it. See Sec. 5.3.1 for a detailed example.

The optimization of $\boldsymbol{\theta}$ consists of two stages. The first and the main stage is *mini-batch learning*. After sampling the initial $\boldsymbol{\theta}$, we minimize $C_{n,K,\mathcal{E}}^{\ell_2}$ with mini-batch gradient descent. The schematic is shown in Fig. 2. Within each iteration, we sample a subset $\mathcal{E}_S \subset \mathcal{E}$, estimate the corresponding partial ℓ_2-norm cost function

$$C_{n,K,\mathcal{E}_S}^{\ell_2}(\boldsymbol{\theta}) \equiv \sum_{E_\mu \in \mathcal{E}_S} \Big(\sum_{1 \le i < j \le K} |\langle\psi_i|E_\mu|\psi_j\rangle|^2$$
$$+ \sum_{j=1}^{K} |\langle\psi_j|E_\mu|\psi_j\rangle - \overline{\langle E_\mu\rangle}|^2/4\Big) \tag{31}$$

and its gradient

$$\nabla\boldsymbol{\theta} = \frac{\partial C_{n,K,\mathcal{E}_S}^{\ell_2}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \tag{32}$$

via measurements, then perform a single gradient descent step with a learning rate η:

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \eta\nabla\boldsymbol{\theta}. \tag{33}$$

The required number of measurements to estimate $C_{n,K,\mathcal{E}_S}^{\ell_2}(\boldsymbol{\theta})$ up to additive error ϵ is of order $\mathcal{O}(K^2|\mathcal{E}_S|^2/\epsilon^2)$. The gradient can be estimated through finite-differencing or by combining the chain rule and the parameter shift rule [58]. Mini-batch gradient descent allows for a more robust convergence and avoids being trapped in a local minimum. We repeat sampling and gradient descent until convergence. The reason that we minimize $C_{n,K,\mathcal{E}}^{\ell_2}$ first is because it converges much faster than $C_{n,K,\mathcal{E}}^{\ell_1}$. In addition, $C_{n,K,\mathcal{E}}^{\ell_2}$ is differentiable but $C_{n,K,\mathcal{E}}^{\ell_1}$ is not.

If the error set $\mathcal{E}$ consists of too many terms, a promising alternative method is to construct "classical shadows" [59] for each basis state $|\psi_j\rangle$, then use the shadows to estimate the cost functions classically. The shadow tomography technique can help us implement large-batch optimization with a smaller measurement overhead.

After adequate mini-batch learning, if $C_{n,K,\mathcal{E}}^{\ell_2}$ is relatively small (e.g., $C_{n,K,\mathcal{E}}^{\ell_2} < 0.01$), we estimate $C_{n,K,\mathcal{E}}^{\ell_1}$ and *fine-tune* the parameters $\boldsymbol{\theta}$ with

respect to it since $C_{n,K,\mathcal{E}}^{\ell_1}$ is directly related to the inaccuracy of the code (see Propositions 2, 3, 4). In this work, we use Powell's method [60], a gradient-free optimizer, for fine-tuning. If $C_{n,K,\mathcal{E}}^{\ell_1}$ is smaller than an acceptable cost tolerance $C_{\text{tol}}^{\ell_1}$, we stop the optimization and output the final parameters $\boldsymbol{\theta}_{\text{opt}}$. Throughout this paper, we set the tolerance as

$$C_{\text{tol}}^{\ell_1} \equiv 1 \times 10^{-6}. \tag{34}$$

In the ideal case, we obtain the optimal parameters

$$\boldsymbol{\theta}_{\text{opt}} = \arg\min_{\boldsymbol{\theta}} C_{n,K,\mathcal{E}}^{\ell_1}(\boldsymbol{\theta}). \tag{35}$$

The output QECC

$$
\begin{aligned}
\mathcal{C}_{\text{out}}(\boldsymbol{\theta}_{\text{opt}}) = \text{span}\{ \\
|\psi_1\rangle = U(\boldsymbol{\theta}_{\text{opt}})|\mathbf{0}\rangle|0\rangle^{\otimes(n-k)}, \\
|\psi_2\rangle = U(\boldsymbol{\theta}_{\text{opt}})|\mathbf{1}\rangle|0\rangle^{\otimes(n-k)}, \\
\cdots, \\
|\psi_K\rangle = U(\boldsymbol{\theta}_{\text{opt}})|\mathbf{K}-\mathbf{1}\rangle|0\rangle^{\otimes(n-k)} \\
\}.
\end{aligned}
\tag{36}
$$

is the target approximate quantum code with small inaccuracy. The variational quantum circuit $U(\boldsymbol{\theta}_{\text{opt}})$ serves as the encoding circuit. Further, we can remove redundant gates from the VQC.

If $C_{n,K,\mathcal{E}}^{\ell_1}$ is greater than $C_{\text{tol}}^{\ell_1}$, we increase the circuit depth L and repeat the optimization steps. If $C_{n,K,\mathcal{E}}^{\ell_1}$ is always greater than the tolerance even when $L = L_{\max}$, we fail to find an eligible code. The detailed procedure is illustrated in Algorithm 1.

A natural question arises: can a fixed-depth VQC find any $((n,K))_2$ quantum code? Haug *et al.* used the quantum Fisher information matrix to assess the expressive power of a VQC with a fixed input state $|0\rangle^{\otimes n}$ [61]. We generalize this notion to multiple inputs to assess the expressive power of a VQC in VarQEC. If a VQC is capable of finding any $((n,K))_2$ quantum code, we say it is *overparameterized* with respect to code parameters $((n,K))_2$. See Appendix C for more details.

When the VQC $U(\boldsymbol{\theta})$ is underparameterized for $((n,K))_2$, the set of reachable output codes forms a low-dimensional submanifold of the complex Grassmannian $\mathbf{Gr}(K, 2^n)$,

$$\{\mathcal{C}_{\text{out}}(\boldsymbol{\theta})|\boldsymbol{\theta}\} \subseteq \mathbf{Gr}(K, 2^n). \tag{37}$$

Algorithm 1: VarQEC

Input: Error set $\mathcal{E}$, hardware-efficient VQC $U(\boldsymbol{\theta})$ with L layers, acceptable number of layers $L_{\max}$, acceptable cost tolerance $C_{\text{tol}}^{\ell_1}$.

Output: An approximate quantum code with a hardware-efficient encoder that detects $\mathcal{E}$.

$L \leftarrow 1$.
while $L \leq L_{\max}$ *and* $C_{n,K,\mathcal{E}}^{\ell_1}(\boldsymbol{\theta}) > C_{\text{tol}}^{\ell_1}$ **do**
  while $C_{n,K,\mathcal{E}}^{\ell_2}(\boldsymbol{\theta})$ *has not converged* **do**
    Sample a subset $\mathcal{E}_S \subset \mathcal{E}$.
    Prepare the K input strings.
    Run $U(\boldsymbol{\theta})$, output $\{|\psi_j\rangle\}$.
    Measure observables $E_\mu \in \mathcal{E}_S$.
    Prepare the K input strings.
    Run $U^\dagger(\boldsymbol{\theta})E_\mu U(\boldsymbol{\theta})$ for $E_\mu \in \mathcal{E}_S$.
    Do projective measurements.
    Estimate $C_{n,K,\mathcal{E}_S}^{\ell_2}(\boldsymbol{\theta})$.
    Vary $\boldsymbol{\theta}$, repeat the above steps to estimate $\partial C_{n,K,\mathcal{E}}^{\ell_2}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$.
    Perform a gradient descent step, update $\boldsymbol{\theta}$.
  end
  if $C_{n,K,\mathcal{E}}^{\ell_2}(\boldsymbol{\theta}) < 0.01$ **then**
    while $C_{n,K,\mathcal{E}}^{\ell_1}(\boldsymbol{\theta})$ *has not converged* **do**
      Prepare the K input strings.
      Run $U(\boldsymbol{\theta})$, output $\{|\psi_j\rangle\}$.
      Measure observables $E_\mu \in \mathcal{E}$.
      Prepare the K input strings.
      Run $U^\dagger(\boldsymbol{\theta})E_\mu U(\boldsymbol{\theta})$ for $E_\mu \in \mathcal{E}$.
      Do projective measurements.
      Estimate $C_{n,K,\mathcal{E}_S}^{\ell_1}(\boldsymbol{\theta})$.
      Vary $\boldsymbol{\theta}$, repeat the above steps to estimate $\partial C_{n,K,\mathcal{E}}^{\ell_1}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$.
      Minimize $C_{n,K,\mathcal{E}}^{\ell_1}(\boldsymbol{\theta})$, update $\boldsymbol{\theta}$.
    end
  end
  $L \leftarrow L + 1$.
end
$\boldsymbol{\theta}_{\text{opt}} \leftarrow \boldsymbol{\theta}$.
Return: $\boldsymbol{\theta}_{\text{opt}}, C_{n,K,\mathcal{E}}^{\ell_1}(\boldsymbol{\theta}_{\text{opt}})$.

The VarQEC algorithm searches this submanifold for an eligible code. When $U(\boldsymbol{\theta})$ is overparameterized for $((n,K))_2$, it can explore all relevant directions and the set of reachable output codes is equivalent to $\mathbf{Gr}(K, 2^n)$, i.e., VarQEC is capable to find arbitrary $((n,K))_2$ quantum code. The required number of periodic bounded

real parameters to overparameterize a VQC is at least $2K(2^n - K)$ since the complex dimension of $\mathbf{Gr}(K, 2^n)$ is $K(2^n - K)$.

In Ref. [62], Johnson *et al.* proposed a related algorithm named QVECTOR, which samples a random 2-design unitary and optimizes parameterized encoding and decoding circuits simultaneously to improve the quantum average fidelity. Compared with QVECTOR, VarQEC can find not only channel-adaptive codes but also quantum codes with specific code parameters. VarQEC does not need a deep random circuit, which is a daunting challenge on NISQ devices, to sample a bunch of input states. We train the encoder without considering the decoder. The optimization is less likely to be trapped in a local minimum. The cost functions are estimated by measuring some local observables. We can rigorously obtain an ε-correctable approximate QECC with arbitrarily small ε. The noise models in our methods are also flexible and can be artificially assigned.

5 Results

5.1 Symmetric codes

We verify the validity of our algorithm by rediscovering some symmetric codes with well-known code parameters. The cost functions are defined in Eqs. (15), (16). For code parameters $((n, K, d))_2$, the total number of Pauli errors O_α to consider is

$$|\{O_\alpha\}| = \sum_{j=0}^{d-1} \binom{n}{j} 3^j. \qquad (38)$$

Without loss of generality, we use the complete bipartite connectivity graph: denote the qubits selected for the input as $\{Q_0, Q_1, \ldots, Q_{k-1}\}$ ($k = \lceil \log(K) \rceil$), the unselected qubits as $\{Q_k, Q_{k+1}, \ldots, Q_{n-1}\}$, the graph consists of $k(n - k)$ edges that connect every selected qubit and every unselected qubit, qubits in the same set are not directly connected. For such graphs, the initial logical data can spread to each qubit rapidly since the graph diameter is only 2. The variational quantum circuit has alternating layers of single-qubit rotations R_x-R_z acting on all qubits and Ising-type interactions R_{zz} acting on adjacent qubits. Denote the number of layers as

L. The VQC evolution is of the form

$$U(\boldsymbol{\theta}) = U_E(\boldsymbol{\theta}_E) \prod_{l=1}^{L} U_l(\boldsymbol{\theta}_l), \qquad (39)$$

where $\boldsymbol{\theta}_l$ and $\boldsymbol{\theta}_E$ are elements in $\boldsymbol{\theta}$, $U_l(\boldsymbol{\theta}_l)$ denotes the l-th layer evolution, $U_E(\boldsymbol{\theta}_E)$ denotes the rightmost R_x-R_z rotations which are used to search the manifold of locally equivalent quantum codes. Since R_z and R_{zz} gates in the last layer commute, and R_z-R_x-R_z rotations can realize arbitrary single-qubit unitary, locally equivalent QECCs can be found by the same VQC. In principle, any n-qubit unitary evolution can be realized by this ansatz with a sufficiently large number of layers since $\{R_x, R_z, R_{zz}\}$ is a universal quantum gate set. The connectivity graph and the periodic-structured VQC ansatz for $n = 5, k = 2$ are shown in Fig. 3(a,b). In general, with the increase of L, the achievable quantum codes form a higher dimensional submanifold of $\mathbf{Gr}(K, 2^n)$, as shown in Fig. 3(c). When the VQC is overparameterized (L is no less than a critical number L_{crit}) for code parameters $((n, K))_2$, VarQEC can explore the whole $\mathbf{Gr}(K, 2^n)$ manifold.

An alternative variational circuit for finding additive quantum codes is discussed in Appendix D.

We apply our algorithm to search for $((n, K, d))_2$ codes where the code length n ranges from 3 to 12, the code dimension K ranges from 2 to 8, the code distance d ranges from 2 to 4. Fig. 4 shows $C_{n,K,\mathcal{E}}^{\ell_1}(\boldsymbol{\theta}_{\text{opt}})$ as a function of code parameters n, K, d. We rediscover quantum codes with parameters

$$\begin{aligned}
&((4, 4, 2))_2, ((5, 6, 2))_2, \\
&((5, 2, 3))_2, ((6, 2, 3))_2, \\
&((7, 2, 3))_2, ((8, 8, 3))_2, \qquad (40) \\
&((9, 8, 3))_2, ((10, 4, 4))_2, \\
&((11, 4, 4))_2.
\end{aligned}$$

The $((5, 2, 3))_2$ and $((8, 8, 3))_2$ codes are non-degenerate, the $((6, 2, 3))_2$ codes are degenerate, and both cases are possible for $((7, 2, 3))_2$. There is an obvious phase transition between achievable and (probably) non-achievable code parameters. Further, we fix $d = 3$ and rediscover larger codes with parameters

$$\begin{aligned}
&((10, 2^4, 3))_2, ((11, 2^5, 3))_2, \\
&((12, 2^6, 3))_2, ((13, 2^7, 3))_2, \qquad (41) \\
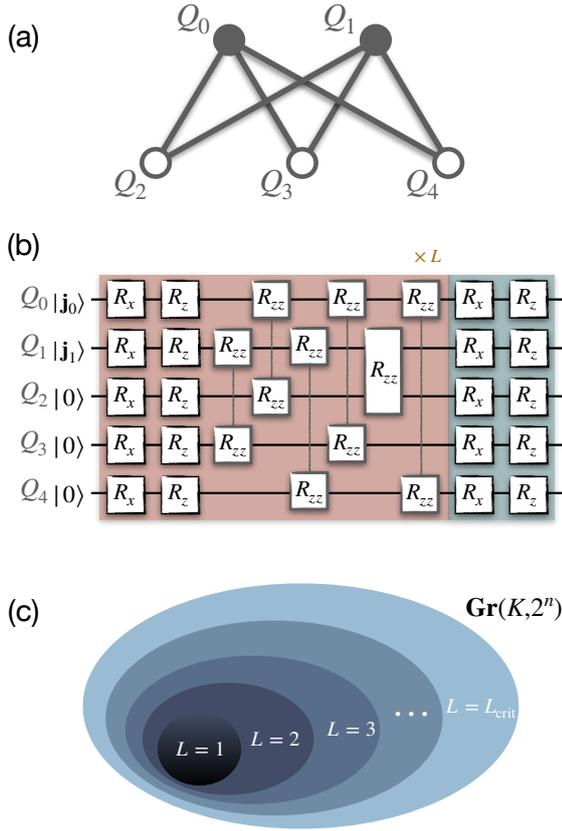&((14, 2^8, 3))_2.
\end{aligned}$$

Figure 3: Schematic of a connectivity graph, the periodic-structured variational ansatz, and the achievable quantum codes. (a) The bipartite connectivity graph with five physical qubits. Gray lines connect adjacent qubits. $\{Q_0, Q_1\}$ are the selected qubits to prepare the logical data. (b) The corresponding variational quantum circuit (VQC) with L layers. Within each layer, we apply R_x-R_z rotations to each qubit and apply R_{zz} gates to adjacent qubits. (c) With the increase of L, VarQEC is capable to find quantum codes in a higher-dimensional manifold until overparameterization ($L = L_{\mathrm{crit}}$).

All these codes can be found by a shallow VQC, e.g., a 5-layer VQC can encode $((5, 2, 3))_2$, $((12, 2^6, 3))_2$, $((14, 2^8, 3))_2$; a 4-layer VQC can encode $((6, 2, 3))_2$, $((8, 8, 3))_2$; a 3-layer VQC suffices to encode $((7, 2, 3))_2$. In our experiments, either $C_{n,K,\mathcal{E}}^{\ell_1}(\boldsymbol{\theta}_{\mathrm{opt}}) < 1 \times 10^{-6}$ or $C_{n,K,\mathcal{E}}^{\ell_1}(\boldsymbol{\theta}_{\mathrm{opt}}) \geq 1$ holds. Fig. 5 shows some cost curves at the minibatch learning stage. Within each iteration, we sample 20% of $\{O_\alpha\}$ as the batch and perform a stochastic gradient descent with learning rate $\eta = 1 \times 10^{-2}$.

In the following, we verify the *local equivalence* (LE) between two quantum codes with K-dimensional projectors P_c and P_c' by sampling permutations of qubits Π_q and numerically min-
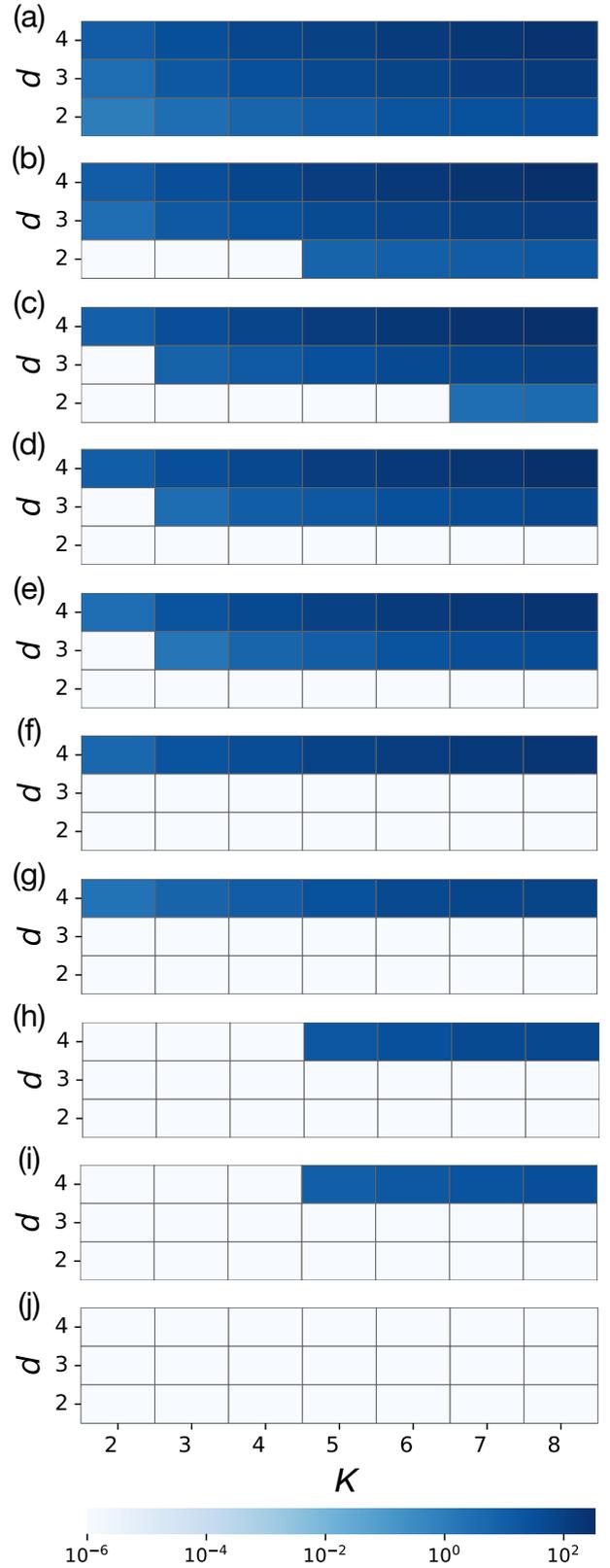

Figure 4: Minimum cost function $C_{n,K,\mathcal{E}}^{\ell_1}(\boldsymbol{\theta}_{\mathrm{opt}})$ for different code length n, code dimension K, and code distance d. (a) $n = 3$. (b)$n = 4$. (c)$n = 5$. (d)$n = 6$. (e)$n = 7$. (f)$n = 8$. (g)$n = 9$. (h)$n = 10$. (i)$n = 11$. (j)$n = 12$.
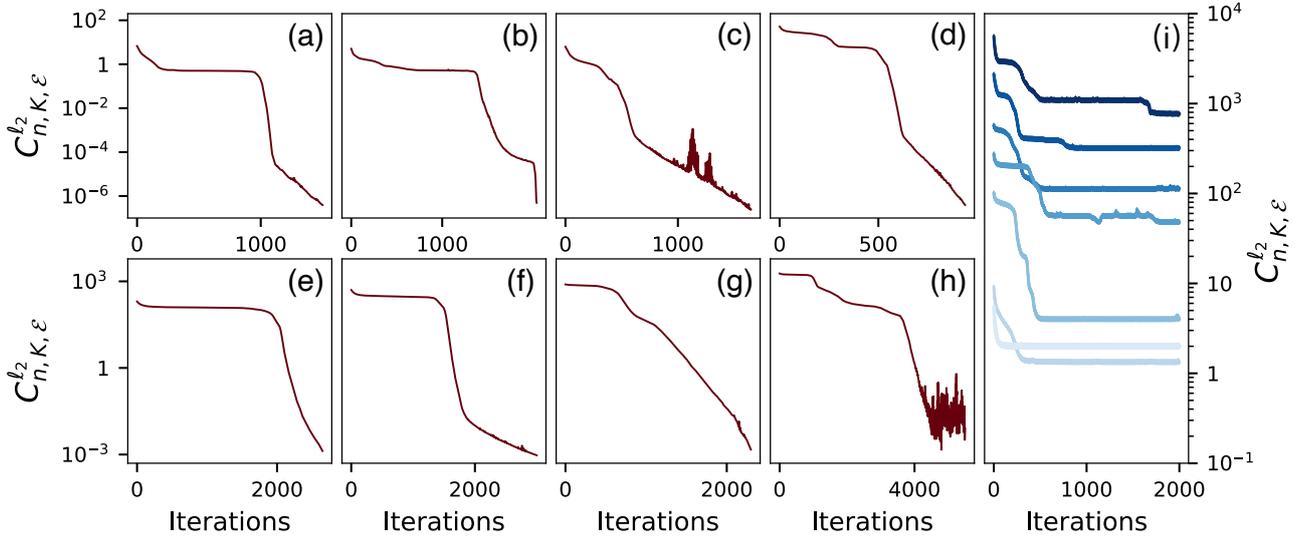
Figure 5: Learning curves of VarQEC for finding symmetric codes with achievable code parameters (a-h)$((5,2,3))_2$, $((6,2,3))_2$, $((7,2,3))_2$, $((8,8,3))_2$, $((11,2^5,3))_2$, $((12,2^6,3))_2$, $((13,2^7,3))_2$, $((14,2^8,3))_2$, and (probably) non-achievable code parameters (i) from bottom to top: $((7,3,3))_2$, $((4,2,3))_2$, $((9,2^4,3))_2$, $((10,2^5,3))_2$, $((11,2^6,3))_2$, $((12,2^7,3))_2$, $((13,2^8,3))_2$.

imizing the LE-cost function

$$C_{LE} = |\operatorname{Tr}(U(\phi)\Pi_q P_c \Pi_q^{-1} U^\dagger(\phi) P_c') - K|^2, \quad (42)$$

where

$$U(\phi) = \exp(-i\sum_{j=1}^n \phi_{j,1} Z_j) \exp(-i\sum_{j=1}^n \phi_{j,2} X_j)$$

$$\cdot \exp(-i\sum_{j=1}^n \phi_{j,3} Z_j) \quad (43)$$

is a product of single-qubit unitaries with $3n$ parameters ϕ. If there exist Π_q and ϕ such that $C_{LE} < 1 \times 10^{-10}$, we say P_c and P_c' are (locally) equivalent.

The minimum code length that protects a logical qubit against arbitrary one-qubit errors is $n = 5$. The $((5,2,3))_2$ code we rediscover is equivalent to the perfect code devised in Ref. [63]. This code is unique and translational invariant. The $((5,6,2))_2$ code we rediscover is equivalent to the original non-additive CWS code devised in Ref. [64]. For parameters $((6,2,3))_2$, we sample different initial VQC parameters $\boldsymbol{\theta}$ and find a mass of non-additive codes that are not mutually equivalent. This is consistent with our observation that an infinite family of non-equivalent $((6,2,3))_2$ codes exist. For parameters $((7,2,3))_2$, we find non-equivalent quantum codes and some of them are not equivalent to CWS codes. See Appendix E for more discussions on

$((6,2,3))_2$ and $((7,2,3))_2$. The $((8,8,3))_2$ code we rediscover is equivalent to the additive code stabilized by

$$
\begin{array}{ccccccccc}
g_1 & = & X & X & X & X & X & X & X & X \\
g_2 & = & Z & Z & Z & Z & Z & Z & Z & Z \\
g_3 & = & I & X & Y & Z & Z & Y & X & I \\
g_4 & = & Z & Y & Z & Y & X & I & X & I \\
g_5 & = & X & Y & Y & X & I & Z & Z & I
\end{array}
\quad (44)
$$

up to permutation of qubits.

It is an open question whether a $((7,3,3))_2$ QECC exists. We have not yet found such a code with VarQEC, even if using an overparameterized VQC ($L = 31$) that is capable of finding any $((7,3))_2$ quantum code and sampling 20000 optimization starting points. This strongly indicates that a quantum code with parameters $((7,3,3))_2$ is nonexistent.

5.2 Asymmetric codes

In quantum experiments, the decoherence time of a physical qubit is mainly influenced by two factors: the relaxation time T_1 and the dephasing time T_2. Relaxation leads to all Pauli errors, whereas dephasing only leads to phase-flips (Pauli-Z errors). Denote the probabilities of X, Y, and Z errors as p_x, p_y, p_z respectively. Usually, $p_x = p_y \neq p_z$. The asymmetry between X/Y and Z errors motivates people to construct *asymmetric* QECCs that han-

dle them differently [65, 66]. Asymmetric codes are more resource-efficient since they can detect/correct more Pauli-X/Y errors than Pauli-Z errors or vice versa in response to demand. Researchers have extended several constructions from symmetric codes to asymmetric codes [66–72]. Note that the classification of symmetric and asymmetric codes depends on the error detecting/correcting capability instead of the code construction method.

For a system with X/Y-error probabilities $p_x = p_y$ and Z-error probability p_z. We set the bias parameter c_Z as

$$c_Z = \frac{\log p_z}{\log p_x}. \tag{45}$$

such that $p_z = p_x^{c_Z}$.

In most scenarios, dephasing is dominating and phase-flip errors are more prevalent than X/Y errors. Accordingly, $0 < c_Z < 1$. First, we fix $c_Z = 1/2$ (i.e., $p_z \approx p_x^{1/2}$) and apply VarQEC to find $((n, K, d_e(1/2)))_2$ codes that encodes one logical qubit ($K = 2$) or one logical qutrit ($K = 3$). We discover asymmetric codes

$$\begin{aligned} ((6, 2, d_e(\tfrac{1}{2}) = 2))_2, \\ ((7, 3, d_e(\tfrac{1}{2}) = 2))_2. \end{aligned} \tag{46}$$

They can detect more Z errors than X/Y errors, specifically, detect the error set

$$\mathcal{E}_{1/2}^{\{2\}} = \{I\} \cup \{X_j, Y_j, Z_j, X_i Z_j, Y_i Z_j, Z_i Z_j Z_k\} \tag{47}$$

with indices $i, j, k \in [1, n]$.

We now consider the opposite situation where X/Y errors are more prevalent than Z. In the extreme case, $T_2 \to +\infty$, the only source of decoherence is qubit relaxation. This process at finite temperature is modeled by the generalized amplitude damping channel. Its Kraus representation has operators

$$A_0 = \sqrt{p} \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix}$$
$$= \sqrt{p} I - \frac{\sqrt{p}\gamma}{4}(I - Z) + \mathcal{O}\left(\sqrt{p}\gamma^2\right),$$
$$A_1 = \sqrt{p} \begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix} = \frac{\sqrt{p}\gamma}{2}(X + iY),$$
$$A_2 = \sqrt{1-p} \begin{pmatrix} \sqrt{1-\gamma} & 0 \\ 0 & 1 \end{pmatrix}$$

$$= \sqrt{1-p} I - \frac{\sqrt{1-p}\gamma}{4}(I + Z)$$
$$+ \mathcal{O}\left(\sqrt{1-p}\gamma^2\right),$$
$$A_3 = \sqrt{1-p} \begin{pmatrix} 0 & 0 \\ \sqrt{\gamma} & 0 \end{pmatrix} = \frac{\sqrt{\gamma - p\gamma}}{2}(X - iY), \tag{48}$$

where γ is the damping rate, p is a constant determined by the temperature. A_0 and A_2 introduce Pauli-Z errors of order $\mathcal{O}(\gamma)$, A_1 and A_3 introduce Pauli-X and -Y errors of order $\mathcal{O}(\sqrt{\gamma})$. When γ is small, $c_Z = \log p_z / \log p_x \approx 2$.

Now we fix $c_Z = 2$ and apply VarQEC to find asymmetric codes with 2-effective distance 3. We rediscover codes with parameters

$$\begin{aligned} ((5, 2, d_e(2) = 3))_2, \\ ((6, 4, d_e(2) = 3))_2, \\ ((7, 8, d_e(2) = 3))_2. \end{aligned} \tag{49}$$

These codes were introduced in Ref. [42]. They can detect more X/Y errors than Z errors, i.e., the error set

$$\mathcal{E}_2^{\{3\}} = \{I\} \cup \{X_j, Y_j, Z_j, X_i X_j, X_i Y_j, Y_i Y_j\} \tag{50}$$

with indices $i, j \in [1, n]$.

Furthermore, we find new codes with 2-effective distance 4 with $K = 2$ or $K = 3$, i.e.,

$$\begin{aligned} ((6, 2, d_e(2) = 4))_2, \\ ((8, 3, d_e(2) = 4))_2. \end{aligned} \tag{51}$$

Some $((6, 2, d_e(2) = 4))_2$ codes are equivalent to the additive $((6, 2, 3))_2$ code stabilized by

$$\begin{array}{ccccccc} g_1 & = & X & I & X & Y & Z & X \\ g_2 & = & Z & I & I & I & I & Z \\ g_3 & = & I & X & X & X & X & I \\ g_4 & = & I & Z & I & Y & X & Z \\ g_5 & = & I & I & Z & X & Y & Z. \end{array} \tag{52}$$

They can detect the error set

$$\mathcal{E}_2^{\{4\}} = \mathcal{E}_2^{\{3\}} \cup \{X_i Z_j, Y_i Z_j, X_i X_j X_k, X_i X_j Y_k, \\ X_i Y_j Y_k, Y_i Y_j Y_k\}. \tag{53}$$

with indices $i, j, k \in [1, n]$. For the generalized amplitude damping channel, $((6, 2, d_e(2) = 4))_2$ and $((8, 3, d_e(2) = 4))_2$ can detect up to three A_1/A_3 errors or one A_0/A_2 error, and correct one A_1/A_3 error. Assisted by post-selection [73], these codes hold the promise to achieve lower logical error rate than codes with $d_e(2) = 3$.

5.3 Channel-adaptive codes

In the previous sections, we have only discussed uncorrelated errors, symmetric or asymmetric. This section considers quantum channels with correlated noise. We apply VarQEC to find the corresponding channel-adaptive codes.

Correlated errors are ubiquitous in quantum computing experiments. When two adjacent qubits are not sufficiently separated, the errors occurring on them can be highly correlated [22]. These spatially correlated errors invalidate many well-known quantum codes and dim the hope of fault-tolerant quantum computing. Suppose we ignore the exact connectivity graph and the noise type. In that case, we need at least 11 physical qubits to protect one qubit of information from general correlated errors (i.e., the double error-correcting $((11, 2, 5))_2$ code) [74]. Even so, the encoding isometry may not be hardware-efficient. In the following, we investigate two correlated noise channels in detail and introduce channel-adaptive codes discovered by VarQEC.

5.3.1 Nearest-neighbor collective amplitude damping

The first testbed is the *nearest-neighbor collective amplitude damping channel*. Suppose we have n qubits in a ring, as shown in Fig. 6. Every two neighboring qubits collectively interact with a single environment and exhibit collective dynamics of amplitude damping [75, 76]. The corresponding Kraus operators are

$$
\begin{aligned}
K_0 =& \sqrt{\frac{\gamma_{01}}{2}} |00\rangle(\langle 01| + \langle 10|) \\
&+ \sqrt{\frac{\gamma_{12}}{2}}(|01\rangle + |10\rangle)\langle 11|, \\
K_1 =& \sqrt{\gamma_{02}}|00\rangle\langle 11|, \\
K_2 =& \sqrt{\frac{1 - \gamma_{01}}{4}}(|01\rangle + |10\rangle)(\langle 01| + \langle 10|) \\
&+ \sqrt{1 - \gamma_{02} - \gamma_{12}}|11\rangle\langle 11| \\
&+ \frac{1}{2}(|01\rangle - |10\rangle)(\langle 01| - \langle 10|) \\
&+ |00\rangle\langle 00|,
\end{aligned} \tag{54}
$$

where $\gamma_{01}, \gamma_{02}, \gamma_{12}$ are damping rates. For a short decay time τ, γ_{01} and γ_{12} are of order $\mathcal{O}(\tau)$, γ_{02} is of order $\mathcal{O}(\tau^2)$. Each error acts on two neighbouring qubits Q_j-Q_{j+1}. To find quantum codes that approximately correct one nearest-neighbor collective amplitude damping error, we expand the
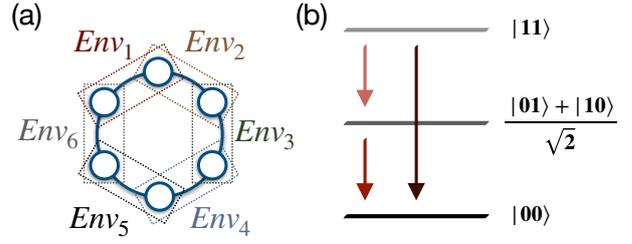

Figure 6: Schematic of nearest-neighbor collective amplitude damping. (a) Qubits in a ring, neighboring two qubits collectively interact with a single environment. (b) Decay processes.

above Kraus operators with respect to τ, abandon trivial/high-order terms and obtain

$$
\begin{aligned}
K_0' =& \frac{1}{\sqrt{2}}|00\rangle(\langle 01| + \langle 10|) + \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)\langle 11|, \\
K_1' =& |00\rangle\langle 11|, \\
K_2' =& \frac{1}{2}(|01\rangle + |10\rangle)(\langle 01| + \langle 10|) + |11\rangle\langle 11|.
\end{aligned} \tag{55}
$$

Each K_0' contributes one factor of $\sqrt{\tau}$, each K_1'/K_2' contributes one factor of τ. Suppose E_α and E_β are products of the identity, K_0', K_1', and K_2'. The target error set (in VarQEC)

$$
\mathcal{E} = \left\{ E_\alpha^\dagger E_\beta \right\} \tag{56}
$$

consists of terms with total order less than $\tau^{3/2}$. Note that here, some error products $E_\alpha^\dagger E_\beta$ are non-unitary and non-Hermitian. To compute cost functions $C_{n,K,\mathcal{E}}^{\ell_1}$ and $C_{n,K,\mathcal{E}}^{\ell_2}$ (Eqs. (15), (16)), we need to estimate

$$
\langle \psi_j | E_\alpha^\dagger E_\beta |\psi_j\rangle \tag{57}
$$

and

$$
|\langle \psi_i | E_\alpha^\dagger E_\beta |\psi_j\rangle| \tag{58}
$$

for various i, j, α, β $(i \neq j)$.

$\langle \psi_j | E_\alpha^\dagger E_\beta |\psi_j\rangle$ is a complex number that can be obtained as follows: we prepare the state $|\psi_j\rangle$ and measure two Hermitian observables $(E_\alpha^\dagger E_\beta + E_\beta^\dagger E_\alpha)/2$ and $(E_\alpha^\dagger E_\beta - E_\beta^\dagger E_\alpha)/2i$. The first expectation value gives the real part of $\langle \psi_j | E_\alpha^\dagger E_\beta |\psi_j\rangle$ and the second expectation value gives its imaginary part.

We estimate $|\langle \psi_i | E_\alpha^\dagger E_\beta |\psi_j\rangle|$ with $i \neq j$ using POVMs. Specifically, we prepare the state $|\psi_j\rangle$, add an ancilla qubit and implement the operation

$$
\begin{aligned}
\Lambda_{\alpha\beta}(|\psi_j\rangle) =& E_\alpha^\dagger E_\beta|\psi_j\rangle\langle\psi_j|E_\beta^\dagger E_\alpha \otimes |0\rangle\langle 0|_{\text{anc}} \\
&+ E_{\text{aux}}|\psi_j\rangle\langle\psi_j|E_{\text{aux}}^\dagger \otimes |1\rangle\langle 1|_{\text{anc}},
\end{aligned} \tag{59}
$$

where E_{aux} is an auxiliary Kraus operator such that $E_\beta^\dagger E_\alpha E_\alpha^\dagger E_\beta + E_{\text{aux}}^\dagger E_{\text{aux}} = I$. Then we measure the ancilla qubit in the computational basis and post-select the cases of $|0\rangle$. Measuring the ancilla qubit in the $|0\rangle$ state indicates that the error $E_\alpha^\dagger E_\beta$ has occurred. The corresponding probability is

$$p_0 = \text{Tr}(E_\alpha^\dagger E_\beta |\psi_j\rangle\langle\psi_j| E_\beta^\dagger E_\alpha) \qquad (60)$$

and the corresponding state is

$$\frac{E_\alpha^\dagger E_\beta |\psi_j\rangle\langle\psi_j| E_\beta^\dagger E_\alpha}{p_0}. \qquad (61)$$

For these postselected states, we apply the inverse of the VQC and do projective measurements. The conditional probability of obtaining the binary string $|\mathbf{i-1}\rangle|0\rangle^{\otimes(n-k)}$ is

$$p_{ij} = \frac{|\langle\psi_i|E_\alpha^\dagger E_\beta|\psi_j\rangle|^2}{p_0}. \qquad (62)$$

Therefore, we can estimate $|\langle\psi_i|E_\alpha^\dagger E_\beta|\psi_j\rangle|$ by $\sqrt{p_{ij}p_0}$.

Suppose 0 or 1 error occurs during a short decay time τ, we apply VarQEC and find quantum codes with length and dimension

$$\begin{aligned}
&((4,3))_2, ((5,2))_2 \\
&((6,5))_2, ((7,8))_2 \\
&((8,9))_2, ((9,16))_2.
\end{aligned} \qquad (63)$$

These codes can reduce the error from $\mathcal{O}(\sqrt{\tau})$ to $\mathcal{O}(\tau)$.

5.3.2 Nearest-neighbor collective phase-flips

The second noise channel we consider is a combined channel of *nearest-neighbor collective phase-flips* and *single-qubit errors*. The channel consists of two stages. In the first stage, a local depolarizing error with noise rate p

$$\mathcal{N}_{\text{DP}_j}(\rho) = (1 - \frac{3p}{4})\rho + \frac{p}{4}(X_j\rho X_j + Y_j\rho Y_j + Z_j\rho Z_j) \qquad (64)$$

occurs on each qubit. In other words, Pauli errors X, Y, Z occur on each qubit with probability $p/4$. Different local errors act independently. We denote the corresponding global noise channel as

$$\mathcal{N}_1 = \prod_{j=1}^n \mathcal{N}_{\text{DP}_j} \qquad (65)$$

In the second stage, nearest-neighbor collective phase-flip errors ZZ with noise rate p_{zz}

$$\mathcal{N}_{\text{Z}_i\text{Z}_j}(\rho) = (1-p)\rho + p_{zz}Z_iZ_j\rho Z_iZ_j \qquad (66)$$

occur on adjacent qubit pairs Q_i-Q_j. We denote the corresponding global noise channel as

$$\mathcal{N}_2 = \prod_{\langle i,j\rangle} \mathcal{N}_{\text{Z}_i\text{Z}_j}. \qquad (67)$$

The overall process is

$$\mathcal{N} = \mathcal{N}_2 \circ \mathcal{N}_1. \qquad (68)$$

Directly applying VarQEC to $\mathcal{N}$ is not resource efficient since the Kraus representation of $\mathcal{N}$ consists of $\mathcal{O}(\exp(n))$ operators. For practical purposes, we apply our algorithm to the following channel instead,

$$\begin{aligned}
\mathcal{N}'(\rho) =& (1 - \sum_j \frac{3p}{4} - \sum_{\langle i,j\rangle} p_{zz})\rho \\
&+ \sum_j \frac{p}{4}(X_j\rho X_j + Y_j\rho Y_j + Z_j\rho Z_j) \\
&+ \sum_{\langle i,j\rangle} p_{zz}Z_iZ_j\rho Z_iZ_j.
\end{aligned} \qquad (69)$$

The second term takes summation over all qubits. The last term takes summation over all adjacent qubit pairs $\langle i,j\rangle$. Its Kraus operators are

$$\begin{aligned}
\mathcal{E}' = \Big\{ &\sqrt{(1 - \sum_j \frac{3p}{4} - \sum_{\langle i,j\rangle} p_{zz})}I, \\
&\sqrt{\frac{p}{4}}X_j, \sqrt{\frac{p}{4}}Y_j, \sqrt{\frac{p}{4}}Z_j, \sqrt{p_{zz}}Z_iZ_j \Big\}.
\end{aligned} \qquad (70)$$

where qubit-i and qubit-j are adjacent. $\mathcal{N}'$ is the first-order approximation of $\mathcal{N}$ with respect to the error parameters p and p_{zz}. The Kraus representation of $\mathcal{N}'$ consists of only poly(n) operators. $\mathcal{N}$ and $\mathcal{N}'$ are equivalent in the zero-noise limit,

$$\lim_{p,p_{zz}\to 0} \mathcal{N}' = \mathcal{N}. \qquad (71)$$

Suppose for $\mathcal{N}'$, we find an ε-correctable approximate code with $\varepsilon \ll 1$. Namely, with appropriate recovery $\mathcal{R}$, the entanglement fidelity is

$$F_e(\mathcal{R}\mathcal{N}') \geq 1 - \varepsilon. \qquad (72)$$

Then for the original noise channel $\mathcal{N}$, the entanglement fidelity naturally has the form

$$\begin{aligned}
F_e(\mathcal{R}\mathcal{N}) =& 1 - \mathcal{O}(\varepsilon p) - \mathcal{O}(\varepsilon p_{zz}) \\
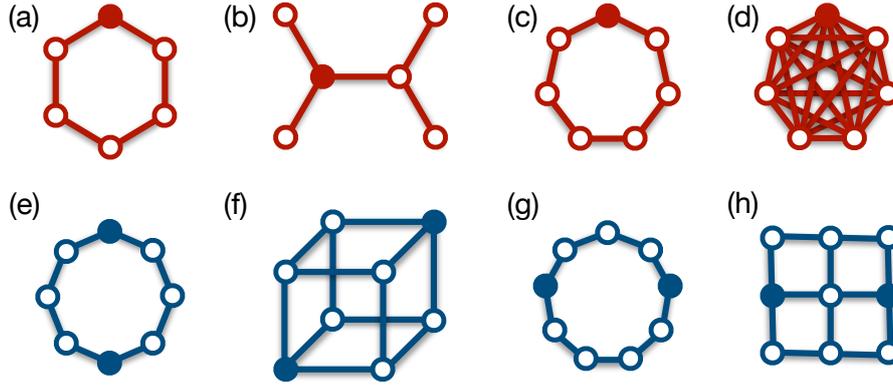&- \mathcal{O}(p^2) - \mathcal{O}(p_{zz}^2).
\end{aligned} \qquad (73)$$

Figure 7: Hardware connectivity graphs with $6 \sim 9$ vertices (physical qubits). Filled circles represent the initial k qubits to prepare the logical data. For these graphs, there exist channel-adaptive codes to protect k qubit(s) of information from general one-qubit errors and nearest-neighbor collective phase-flips. The code length n, the code dimension K, and the number of VQC layers L are (a) $n = 6, K = 2, L = 5$; (b) $n = 6, K = 2, L = 6$; (c) $n = 7, K = 2, L = 2$; (d) $n = 7, K = 2, L = 2$; (e) $n = 8, K = 4, L = 4$; (f) $n = 8, K = 4, L = 4$; (g) $n = 9, K = 4, L = 3$; (h) $n = 9, K = 4, L = 2$.

The QECC can push the first-order errors down to an extremely small level. To find quantum codes that correct multiple errors, we can choose a higher-order approximation and similarly implement VarQEC.

Given a connectivity graph G with edge number $|E(G)|$ and maximum vertex degree $\Delta(G)$. In the following, we set

$$p_{zz} = 0.99/(3n + |E(G)|), \quad p = 4p_{zz}. \quad (74)$$

For a generic input state ρ, the probability of receiving the same state after going through the noise channel is about ~ 0.01. The target error list to detect in VarQEC is

$$\mathcal{E} = \left\{ E_\alpha^\dagger E_\beta | E_\alpha, E_\beta \in \mathcal{E}' \right\}. \quad (75)$$

Still, we use the VQC ansatz with alternating layers of single-qubit rotations R_x-R_z acting on all qubits and Ising-type interactions R_{zz} acting on adjacent qubits. The circuit depth of a VQC with L layers is of order $\mathcal{O}(L\Delta(G))$.

After adequate optimization, we find approximate channel-adaptive codes for $\mathcal{N} = \mathcal{N}_2 \circ \mathcal{N}_1$ with hardware connectivity graphs shown in Fig. 7. The codes for graphs (a,b,h) are degenerate, and the others are non-degenerate. Six physical qubits suffice to encode one logical qubit, and eight physical qubits suffice to encode two logical qubits.

Note that up to a local unitary transformation, these codes can correct an arbitrary single-qubit error followed by an adjacent $U \otimes U$ error for any fixed $U \in \mathrm{U}(2)$ with eigenvalues $\{-1, 1\}$.

We investigated the codes for graphs (c,d) in more detail. Clearly, they have code parameters $((7, 2, 3))_2$. We calculate their *quantum weight enumerators* [77], which were defined by

$$A(z) = \sum_{j=0}^{n} A_j z^j, \quad B(z) = \sum_{j=0}^{n} B_j z^j \quad (76)$$

with coefficients

$$A_j = \frac{1}{K^2} \sum_{\mathrm{wt}(O_\alpha)=j} \mathrm{Tr}(O_\alpha P_c) \mathrm{Tr}(O_\alpha^\dagger P_c), \quad (77)$$

$$B_j = \frac{1}{K} \sum_{\mathrm{wt}(O_\alpha)=j} \mathrm{Tr}(O_\alpha P_c O_\alpha^\dagger P_c). \quad (78)$$

These two codes are locally equivalent and therefore have the same weight enumerators, i.e.,

$$A(z) = 1 + 2z^3 + 9z^4 + 24z^5 + 22z^6 + 6z^7, \quad (79)$$

$$B(z) = 1 + 17z^3 + 45z^4 + 78z^5 + 82z^6 + 33z^7. \quad (80)$$

Further, we verified that they are locally equivalent to a non-degenerate additive code stabilized by

$$
\begin{aligned}
g_1 &= X \quad I \quad Z \quad X \quad X \quad I \quad X \\
g_2 &= Z \quad I \quad I \quad X \quad X \quad X \quad Z \\
g_3 &= I \quad X \quad Z \quad X \quad Z \quad Z \quad Z \\
g_4 &= I \quad Z \quad Z \quad I \quad Z \quad Y \quad Z \\
g_5 &= I \quad I \quad Y \quad X \quad Z \quad I \quad X \\
g_6 &= I \quad I \quad I \quad Z \quad Y \quad Y \quad X
\end{aligned} \quad (81)
$$

up to permutation of qubits. This additive code can correct arbitrary single-qubit errors and 2-qubit collective phase-flips occurring on any qubit pairs, i.e., the error set

$$\mathcal{E} = \left\{ I, X_j, Y_j, Z_j, Z_i Z_j \right\}. \qquad (82)$$

with indices $i, j \in [1, n]$.

According to the quantum Hamming bound, for one logical qubit, no non-degenerate quantum code with code length $n < 7$ can correct arbitrary single-qubit errors as well as 2-qubit collective phase-flips since

$$2^n \geq K(3n + \frac{n(n-1)}{2}) \qquad (83)$$

with $K = 2$ only holds when $n \geq 7$.

Two $((7,2,3))_2$ stabilizer codes were investigated in detail. One is the famous Steane code [14] based on the Calderbank-Shor-Steane (CSS) construction. The other is a non-CSS code found by numerical greedy search, called the bare code [27]. Their weight enumerators are as follows,

$$A^{\{\text{Steane}\}}(z) = 1 + 21z^4 + 42z^6,$$
$$B^{\{\text{Steane}\}}(z) = 1 + 21z^3 + 21z^4$$
$$+ 126z^5 + 42z^6 + 45z^7, \qquad (84)$$

and

$$A^{\{\text{bare}\}}(z) = 1 + 5z^2 + 11z^4 + 47z^6,$$
$$B^{\{\text{bare}\}}(z) = 1 + 5z^2 + 36z^3$$
$$+ 11z^4 + 96z^5 + 47z^6. \qquad (85)$$

QECCs with different weight enumerators are not locally and translationally equivalent. Our code is different from the Steane and the bare $((7,2,3))_2$ codes. See Appendix F more weight enumerators.

The Steane and the bare codes cannot correct nearest-neighbor collective phase-flips. For the combined channel of nearest-neighbor collective phase-flips with noise rate p_{zz} and single-qubit errors with noise rate p, the entanglement fidelity of our code is of the form

$$F_e(\mathcal{RN}) = 1 - \mathcal{O}(p^2) - \mathcal{O}(p_{zz}^2) \qquad (86)$$

whereas the entanglement fidelity of the Steane and the bare codes is of the form

$$F_e(\mathcal{RN}) = 1 - \mathcal{O}(p_{zz}) - \mathcal{O}(p^2) - \mathcal{O}(p_{zz}^2). \qquad (87)$$

Although our code was written in a quantum simulator that has not been open-sourced yet, we rewrote some example implementations with Qiskit, an open-source software development kit. They are available on GitHub [78].

6 Noise Resilience

Although the previously introduced results are obtained by numerical simulation, VarQEC is a hybrid quantum-classical algorithm meant to be run on NISQ devices where quantum gates are inevitably noisy. In this section, we demonstrate that VarQEC is pretty resilient to random gate errors. As long as the error rate p_{gate} is below a reasonable threshold, VarQEC can find an efficient encoding circuit that prepares the correct code. This resilience is essentially analogous to the noise resilience in variational quantum compiling [56].

We start from the simplest noise model, global depolarizing, and introduce the following theorem.

Theorem 5 *Suppose the variational quantum circuit in VarQEC is accompanied by global depolarizing noise acting continuously throughout the circuit. If the ideal circuit is capable of finding an eligible quantum code, after adequate optimization with the noisy circuit, the output parameters $\boldsymbol{\theta}'_{\text{opt}}$ are still correct.*

Proof Consider the cost functions in Eqs. (15), (16). Due to the global depolarizing noise, when we run the VQC $U(\boldsymbol{\theta})$ to prepare a basis state $|\psi_j\rangle$, we instead obtain

$$\rho_j = (1 - \epsilon_1)|\psi_j\rangle\langle\psi_j| + \epsilon_1 \frac{I}{2^n}; \qquad (88)$$

when we apply $U^\dagger(\boldsymbol{\theta})E_\mu U(\boldsymbol{\theta})$ to an initial binary string $|\mathbf{j} - \mathbf{1}\rangle|\mathbf{0}\rangle$ to prepare the output state $|\psi_{j,\mu}\rangle$, we instead obtain

$$\rho_{j,\mu} = (1 - \epsilon_2)|\psi_{j,\mu}\rangle\langle\psi_{j,\mu}| + \epsilon_2 \frac{I}{2^n}. \qquad (89)$$

Noise rates ϵ_1 and ϵ_2 are unknown constants determined by the circuit depth. After adequate optimization with the noisy VQC, we obtain the
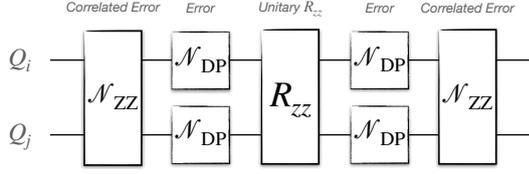
Figure 8: Gate error model. R_{zz} is accompanied by local depolarizing noise $\mathcal{N}_{\mathrm{DP}}$ and collective phase-flip error $\mathcal{N}_{\mathrm{ZZ}}$ both before and after.

pseudo-optimal parameters

$$\boldsymbol{\theta}'_{\mathrm{opt}} = \arg\min_{\boldsymbol{\theta}}$$

$$\sum_{E_\mu \in \mathcal{E}} \Big(\sum_{1 \leq i < j \leq K} \sqrt{\langle \mathbf{i}-\mathbf{1}|\langle \mathbf{0}|\rho_{j,\mu}|\mathbf{i}-\mathbf{1}\rangle|\mathbf{0}\rangle}$$

$$+ \sum_{j=1}^{K} |\operatorname{Tr}(\rho_j E_\mu) - \sum_{i=1}^{K} \operatorname{Tr}(\rho_i E_\mu)/K|/2 \Big)$$

$$= \arg\min_{\boldsymbol{\theta}}$$

$$\sum_{E_\mu \in \mathcal{E}} \Big(\sum_{1 \leq i < j \leq K} \sqrt{(1-\epsilon_2)|\langle \psi_i|E_\mu|\psi_j\rangle|^2 + \frac{\epsilon_2}{2^n}}$$

$$+ \sum_{j=1}^{K} \frac{1-\epsilon_1}{2} |\langle \psi_j|E_\mu|\psi_j\rangle - \overline{\langle E_\mu\rangle}| \Big).$$

$$(90)$$

Since the ideal variational quantum circuit is capable of finding an eligible quantum code, each term in the cost function Eq. (15) can be minimized to 0 (i.e., $|\langle \psi_i|E_\mu|\psi_j\rangle| = 0$, $|\langle \psi_j|E_\mu|\psi_j\rangle - \overline{\langle E_\mu\rangle}| = 0$). Comparing Eq. (15) and Eq. (90), we conclude that

$$\boldsymbol{\theta}'_{\mathrm{opt}} = \boldsymbol{\theta}_{\mathrm{opt}}. \qquad (91)$$

∎

To sum up, VarQEC is perfectly resilient to global depolarizing noise, i.e., it can find the correct encoding circuit in the presence of global depolarizing.

In practical scenarios, circuit noise is more complicated and single-qubit errors dominate. Now we consider a more realistic model. Suppose each 2-qubit R_{zz} gate in the VQC is accompanied by local depolarizing noise and collective phase flips, as illustrated in Fig. 8. Before the ideal unitary R_{zz}, the two qubits goes through $\mathcal{N}_{\mathrm{DP}}^{\otimes 2} \circ \mathcal{N}_{\mathrm{ZZ}}$, after the ideal R_{zz}, the system goes through $\mathcal{N}_{\mathrm{ZZ}} \circ \mathcal{N}_{\mathrm{DP}}^{\otimes 2}$. In the following, for gate error rate p_{gate}, we set the error rate of each $\mathcal{N}_{\mathrm{DP}}$ as $p_{\mathrm{gate}}/2$, the error rate of each $\mathcal{N}_{\mathrm{ZZ}}$ as $p_{\mathrm{gate}}/8$.

Still, we use VarQEC to find channel-adaptive codes for noise channel $\mathcal{N}$ (Eq. (68)) with hard-

ware connectivity graphs (c,d) shown in Fig. 7. The difference is that this time the VQC is noisy. After optimization, we obtain the pseudo-optimal parameters $\boldsymbol{\theta}'_{\mathrm{opt}}$. It is interesting to note that if we transfer $\boldsymbol{\theta}'_{\mathrm{opt}}$ to an ideal VQC, the corresponding cost function $C_{n,K,\mathcal{N}}^{\ell_1}(\boldsymbol{\theta}'_{\mathrm{opt}})$ can be much smaller than the one we estimated with the noisy VQC. Namely, we find a roughly correct encoder even if we use a noisy VQC in our algorithm. The comparison of cost functions for different gate error rates is given in Fig. 9(a). The cost reduction for both graphs is obvious. Two-qubit gate error rates on state-of-the-art NISQ computers are about $\sim 10^{-2}$ [79]. One can run our algorithm on current hardware directly.

Suppose the input state of a quantum circuit is $|\psi_{\mathrm{in}}\rangle$, the target unitary evolution is U_{ideal}. The ideal output state is

$$|\psi_{\mathrm{ideal}}\rangle = U_{\mathrm{ideal}}|\psi_{\mathrm{in}}\rangle. \qquad (92)$$

However, due to quantum gate errors, the output state ρ_{out} is a mixed state. We express ρ_{out} as a summation of three terms,

$$\rho_{\mathrm{out}} = \mathcal{N}_{\mathrm{circuit}}(|\psi_{\mathrm{in}}\rangle)$$
$$= \lambda_0 |\psi_{\mathrm{ideal}}\rangle\langle\psi_{\mathrm{ideal}}| + \lambda_1 \frac{I}{2^n} + \lambda_2 \rho_2, \qquad (93)$$

where $\mathcal{N}_{\mathrm{circuit}}$ denotes that channel of the noisy quantum circuit, λ_1 is the smallest eigenvalue of ρ_{out} multiplied by 2^n, $I/2^n$ is the maximally mixed state, ρ_2 is a density operator orthogonal to $|\psi_{\mathrm{ideal}}\rangle$, i.e.,

$$\operatorname{Tr}(\rho_1 |\psi_{\mathrm{ideal}}\rangle\langle\psi_{\mathrm{ideal}}|) = 0. \qquad (94)$$

The latter two terms of Eq. (93) are both induced by gate errors, but they have different effects on the noise resilience of our algorithm. The second term is a global white noise, as we analyzed in Theorem 5, it does not affect the optimal parameters. However, the third term $\lambda_2 \rho_2$ non-trivially alters the optimization landscape and introduces some local minima. Usually, both the second term and the third term are not negligible. Nevertheless, we are certain about the trend: with the increase of circuit depth, the second term will dominate the third term eventually [80, 81].

For VQCs corresponding to graphs (c,d), we fix gate error rate 0.01, and try different numbers of layers with randomly sampled $\boldsymbol{\theta}$. The average value of λ's are shown in Fig. 9(b,c). Each
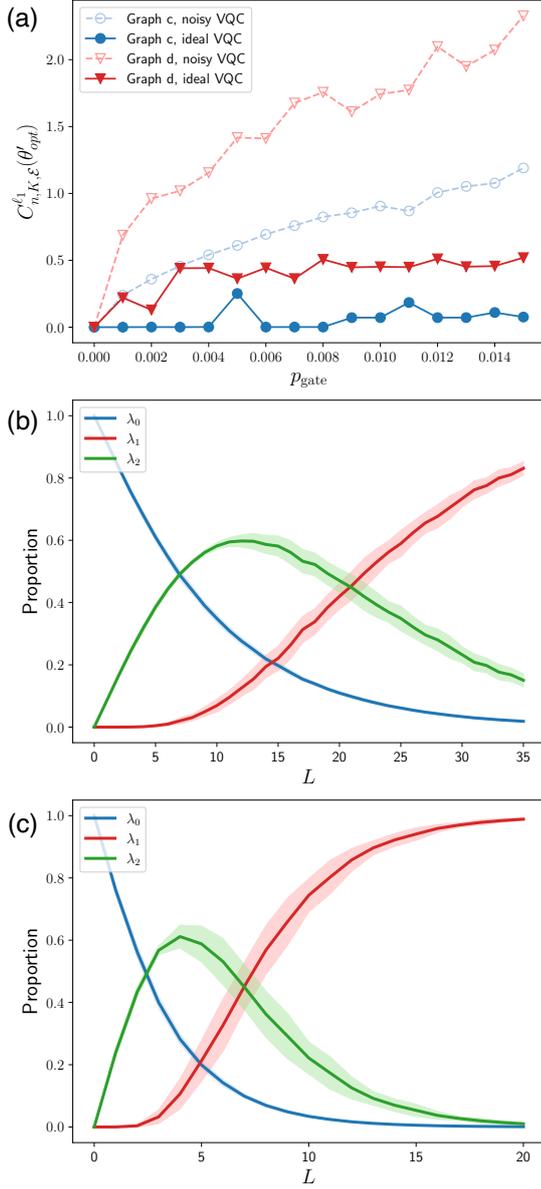
Figure 9: Noise resilience of VarQEC using VQCs with connectivity graphs (c,d) in Fig. 7. (a) The noisy- and ideal-VQC cost functions versus gate error rate. $\boldsymbol{\theta}'_{\mathrm{opt}}$ were obtained by optimizing a noisy VQC. (b) Average λs (see Eq. (93)) versus the number of VQC layers for graph (c). (c) Average λs versus the number of VQC layers for graph (d). The shaded areas represent the standard deviation of 100 samples.

point is averaged over 100 samples. Compared with the one-dimensional ring (graph (c)), vertices in the complete graph (graph (d)) are more tightly connected, local errors can be transformed into global white noise more rapidly. For both graphs, $\lambda_1 \ll \lambda_2$ when L is relatively small and $\lambda_1 \gg \lambda_2$ when L is relatively large. With the decrease of gate error rate and the increase of cir-

cuit depth, VarQEC will become more resilient to noise. Additionally, one might also consider estimating cost functions in VarQEC more precisely with error mitigation techniques like virtual distillation [82, 83].

7 Barren Plateaus

The *barren plateau* (BP) [84, 85] and the *noise-induced barren plateau* (NIBP) [86] are two daunting challenges in variational quantum optimization. In this section, we numerically investigate their effects in the VarQEC algorithm.

The barren plateau is a phenomenon where the gradients vanish exponentially with the increasing number of qubits [84]. It occurs when the VQCs form a unitary 2-design, regardless of whether the VQC is noisy or noiseless. Ref. [85] connected the locality of the cost function and the trainability of the corresponding VQC. If the cost function is local and the circuit depth is of order $\mathcal{O}(\log(n))$, the BP does not occur (i.e., the VQC is trainable). However, if the cost function is global or the circuit depth is of order $\mathcal{O}(\mathrm{poly}(n))$, a BP occurs in the optimization landscape, and the VQC is untrainable.

In near-term quantum computation, dominating errors always only act on several local qubits. Accordingly, the cost functions $C_{n,K,\mathcal{E}}^{\ell_1}$ (Eq. (15)) and $C_{n,K,\mathcal{E}}^{\ell_2}$ (Eq. (16)) are merely influenced by local errors E_μ. Therefore, we expect the same conclusion to hold for VarQEC: the BP does not occur when the circuit depth is of order $\mathcal{O}(\log(n))$ and occurs when the circuit depth is of order $\mathcal{O}(\mathrm{poly}(n))$.

Without loss of generality, here we use the star connectivity graph S_{n-1} (S_7 is illustrated in Fig. 10(b) inset) and focus on $C_{n,2,\mathcal{E}}^{\ell_2}$ with $\mathcal{E} = \{O_\alpha \mid \mathrm{wt}(O_\alpha) < 3\}$, i.e., searching for QECCs that encode one logical qubit information and correct an arbitrary single-qubit error. Fig. 10(a) plots the partial derivative of the off-diagonal cost

$$\sum_{E_\mu \in \mathcal{E}} \sum_{1 \le i < j \le K} |\langle \mathbf{i} | \langle \mathbf{0} | U^\dagger(\boldsymbol{\theta}) E_\mu U(\boldsymbol{\theta}) | \mathbf{j} \rangle | \mathbf{0} \rangle|^2 \quad (95)$$

and the diagonal cost

$$\sum_{E_\mu \in \mathcal{E}} \sum_{j=1}^{K} |\langle \mathbf{j} | \langle \mathbf{0} | U^\dagger(\boldsymbol{\theta}) E_\mu U(\boldsymbol{\theta}) | \mathbf{j} \rangle | \mathbf{0} \rangle - \overline{\langle E_\mu \rangle}|^2 / 4. \quad (96)$$
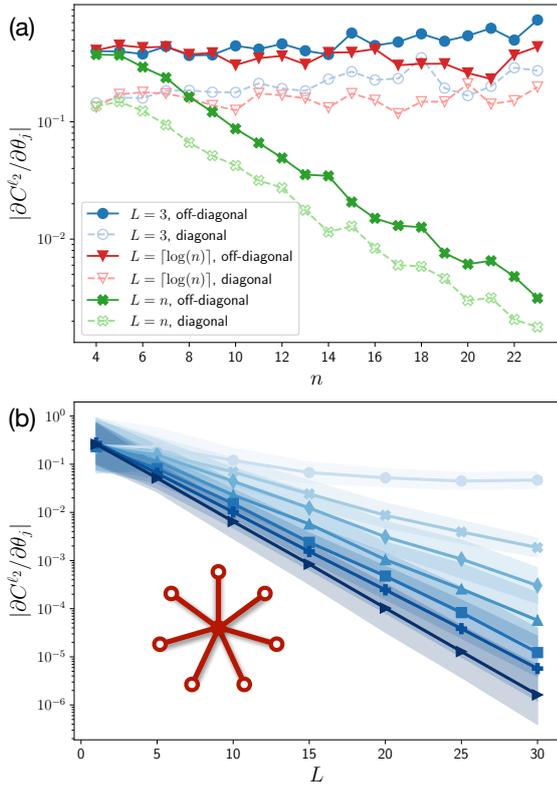
Figure 10: Barren plateaus and noise-induced barren plateaus in VarQEC. (a) Partial derivatives of the off-diagonal and diagonal parts of $C_{n,K,\mathcal{E}}^{\ell_2}$ with respect to a random circuit parameter for different system sizes and circuit depths. Each point is averaged over 1000 samples. (b) Partial derivatives of $C_{n,K,\mathcal{E}}^{\ell_2}$ with respect to a random parameter for gate noise rates (from top to bottom) $p = 0, 5 \times 10^{-3}, 0.01, 0.02, 0.03, 0.04, 0.05$. Each point is averaged over 1000 samples, and the shaded areas represent the standard deviations. Inset: connectivity graph of S_7.

with respect to a randomly selected circuit parameter θ_j. When the number of VQC layers is $L = 3$ or $L = \lceil \log(n) \rceil$, the circuit is trainable. However, when $L = n$, both off-diagonal and diagonal gradients decay exponentially with the increasing number of qubits.

The noise-induced barren plateau refers to a conceptually different phenomenon where cost gradients vanish exponentially with L due to hardware noise accumulation [86]. Consequently, the gradients vanish exponentially with n if L grows linearly with n. Unlike the noise-free BP, NIPB only occurs when the VQC is noisy, regardless of whether the circuits form a unitary 2-design. Still, we consider the local noise model illustrated in Fig. 8, system size $n = 8$, number of layers $L = 1, 5, 10, 15, 20, 25, 30$, noise rate

$p = 0, 5 \times 10^{-3}, 0.01, 0.02, 0.03, 0.04, 0.05$. The numerical results for the gradients are shown in Fig. 10(b). With the increase of L, the partial derivatives of $C_{n,K,\mathcal{E}}^{\ell_2}$ with respect to a random parameter decay exponentially, and the decay factor is determined by the noise rate. This illustrates that although VarQEC can find a roughly correct encoder after adequate training with a noisy VQC (noise resilience), the required training time grows exponentially with the number of circuit layers.

BPs and NIBPs manifest themselves in VarQEC when the circuit depth gets large. Nevertheless, we do not need to worry too much about them. From a practical standpoint, we are more interested in QECCs with a shallow (even constant depth) encoding circuit. The gradients of cost functions tend to be large when searching for these codes. In addition, there are more and more effective strategies to mitigate BPs, e.g., cost function partitioning and meta-learning [87] as well as optimization guided by classical shadows [88]. These protocols can be applied to VarQEC reasonably.

8 Experiment on an IBM machine

Now we experimentally demonstrate VarQEC with a real superconducting quantum machine, *ibm_quito* [89]. The connectivity graph of *ibm_quito* is shown in Fig. 11(a). Our goal is to find a 4-qubit approximate QECC to correct one amplitude damping error [43] using physical qubits Q_0, Q_1, Q_2, Q_3.

The Kraus operators of the amplitude damping channel are

$$A_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix}$$
$$= I - \frac{\gamma}{4}(I - Z) + \mathcal{O}\left(\gamma^2\right), \tag{97}$$
$$A_1 = \begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix} = \frac{\sqrt{\gamma}}{2}(X + iY).$$

Each $(I - Z)$ term contributes a factor of γ and each $(X + iY)$ term contributes a factor of $\sqrt{\gamma}$. To correct a single amplitude damping error, we only need to consider error products with total order less than $\gamma^{3/2}$:

$$\mathcal{E} = \{I, X_j + iY_j, X_j - iY_j, (X_i - iY_i)(X_j + iY_j),$$
$$I_j - Z_j\}. \tag{98}$$

The variational quantum circuit we use is illustrated in Fig. 11(b). When the rotation angle $\theta = \pm\pi/2$, the VQC serves as an exact encoder. Since Q_2 and Q_3 are not directly connected, the IBM compiler adds 2 additional SWAP gates (each realized by 3 CNOT gates) to implement CNOT between Q_2 and Q_3). The hardware-efficient VQC after compiling is shown in Fig. 11(c).

Due to hardware constraints, we slightly modify the VarQEC algorithm and enhance it with quantum error mitigation (EM) as follows. Suppose the initial parameter $\theta = 0.1$, we iteratively apply the VQC to input states $|0000\rangle$ and $|0010\rangle$, do quantum state tomography on the output mixed states and record their density matrices ρ_1 and ρ_2. Then we classically extract their dominating eigenstates

$$|\tilde{\psi}_1\rangle = \lim_{M\to\infty} \frac{\rho_1^M}{\mathrm{Tr}(\rho_1^M)}, |\tilde{\psi}_2\rangle = \lim_{M\to\infty} \frac{\rho_2^M}{\mathrm{Tr}(\rho_2^M)}, \tag{99}$$

estimate the cost functions $C_{n,K,\mathcal{E}}^{\ell_2}$ and $C_{n,K,\mathcal{E}}^{\ell_1}$ of logical basis states $\{|\tilde{\psi}_1\rangle, (|\tilde{\psi}_2\rangle - \langle\tilde{\psi}_1|\tilde{\psi}_2\rangle|\tilde{\psi}_1\rangle)/c\}$, where c is a normalization factor.

The cost gradients are estimated by finite differencing:

$$\frac{\partial C_{4,2,\mathcal{E}}(\theta)}{\partial\theta} \approx \frac{C_{4,2,\mathcal{E}}(\theta + \delta\theta) - C_{4,2,\mathcal{E}}(\theta - \delta\theta)}{2\delta\theta} \tag{100}$$

with $\delta\theta = 0.05$. In the first stage (first 15 iterations), we minimize $C_{n,K,\mathcal{E}}^{\ell_2}$ with learning rate $\eta = 1$ until $C_{n,K,\mathcal{E}}^{\ell_2} < 0.01$. Then we switch to $C_{n,K,\mathcal{E}}^{\ell_1}$ and minimize it with a smaller learning rate $\eta = 0.05$. The training curves of the estimated $C_{n,K,\mathcal{E}}^{\ell_2}$ with/without error mitigation and its real value are shown in Fig. 11(d). After adequate training (25 iterations), the parameter θ converges to about 1.63, slightly greater than the ideal angle $\pi/2$ (indicated by the dashed line in the inset). Nevertheless, this difference is acceptable, the VQC still encodes an approximate amplitude damping code.

We implement a total of 152 quantum circuits for this experiment: 100 for estimating the gradients and 52 for estimating the cost functions.

9 Conclusions and Outlooks

In this work, we proposed VarQEC, an effective variational quantum algorithm for finding
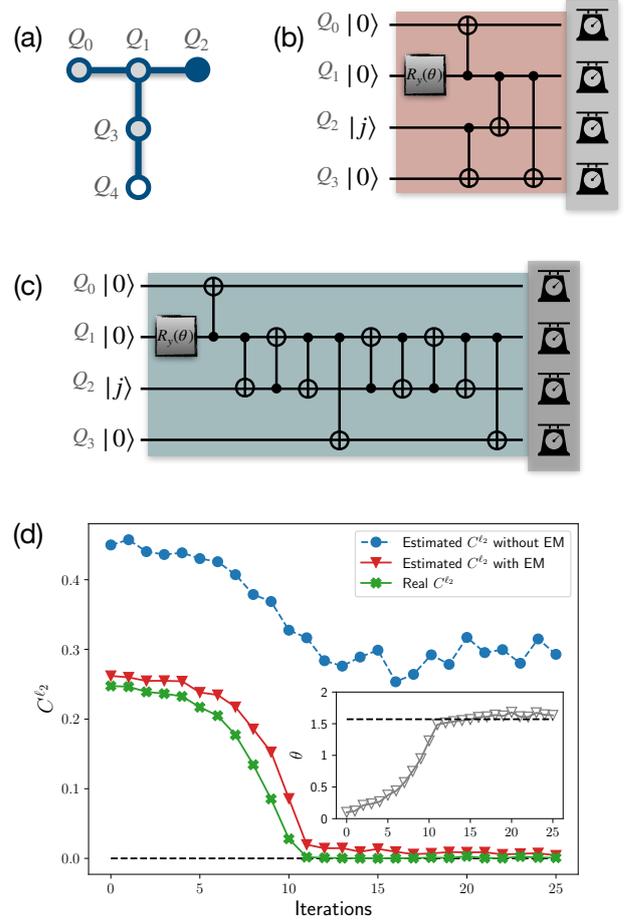

Figure 11: Experimental results of VarQEC for finding an approximate amplitude damping code. (a) The hardware connectivity of ibm_quito. Q_0, Q_1, Q_2, Q_3 are used in the experiment. Q_2 encodes the initial logical information. (b-c) The original and the compiled variational quantum circuits. (d) The estimated ℓ_2-norm cost functions with/without error mitigation (EM) and its real value during variational learning. Inset: the rotation angle θ during learning.

various quantum error-correcting codes. VarQEC is capable of finding arbitrary quantum codes since the cost functions therein are based on the most general requirement of a QECC, the Knill-Laflamme conditions. We demonstrated its efficacy by discovering/rediscovering some symmetric, asymmetric, and channel-adaptive codes, e.g., $((5,2,3))_2$, $((5,6,2))_2$, $((6,2,3))_2$, $((7,2,3))_2$, $((12,2^6,3))_2$, $((14,2^8,3))_2$, $((10,4,4))_2$, $((6,2,d_e(2) = 4))_2$, $((8,3,d_e(2) = 4))_2$. Some discovered codes are equivalent to stabilizer ones and some are not. We investigated them in detail. In particular, VarQEC provided numerical evidence that a quantum code with parameters $((7,3,3))_2$ does

not exist. It is worth mentioning that the channel-adaptive codes with optimized encoding circuits found by our method can then be used as inner codes on the physical level in a concatenation scheme. Stabilizer QECCs for qudits can be used as outer codes.

VarQEC is robust to hardware noise; therefore, it is particularly promising in the NISQ era. A problem worth studying further is how to choose the most resource-efficient variational quantum circuit in VarQEC. There is reason to believe that the optimal VQC ansatz is code-dependent. For example, when the target quantum code is translational-invariant, one may use a VQC with a certain amount of symmetry, where different gates can share the same parameter. If we slightly modify the cost functions, VarQEC can be used for finding some QECC variants like the hybrid quantum-classical codes [90], estimating the zero-error capacity of noisy quantum channels [91], and solving quantum marginal problems [92].

VarQEC can also be directly revised to a classical algorithm. When a NISQ processor is not accessible, one can replace the VQCs with classical variational ansatzes like tensor networks [93–95] or neural network quantum states [96], and then similarly implement optimization and search for eligible quantum codes merely with a classical computer. However, the encoding circuits can not be naturally obtained.

Acknowledgments

References

[1] N. Cody Jones, James D. Whitfield, Peter L. McMahon, Man-Hong Yung, Rodney Van Meter, Alán Aspuru-Guzik, and Yoshihisa Yamamoto. "Faster quantum chemistry simulation on fault-tolerant quantum computers". New Journal of Physics **14**, 115023 (2012).

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

[3] Aram W. Harrow, Avinatan Hassidim, and Seth Lloyd. "Quantum algorithm for linear systems of equations". Phys. Rev. Lett. **103**, 150502 (2009).

[4] Peter W. Shor. "Scheme for reducing decoherence in quantum computer memory". Phys. Rev. A **52**, R2493–R2496 (1995).

[5] Daniel Gottesman. "Stabilizer codes and quantum error correction" (1997). arXiv:quant-ph/9705052.

[6] Daniel A. Lidar and Todd A. Brun. "Quantum error correction". Cambridge University Press. (2013).

[7] Bei Zeng, Xie Chen, Duan-Lu Zhou, and Xiao-Gang Wen. "Quantum information meets quantum matter: From quantum entanglement to topological phases of many-body systems". Springer. (2019).

[8] Steven M. Girvin. "Introduction to quantum error correction and fault tolerance" (2021). arXiv:2111.08894.

[9] Fernando Pastawski, Beni Yoshida, Daniel Harlow, and John Preskill. "Holographic quantum error-correcting codes: toy models for the bulk/boundary correspondence". Journal of High Energy Physics **2015**, 149 (2015).

[10] Emanuel Knill and Raymond Laflamme. "Theory of quantum error-correcting codes". Phys. Rev. A **55**, 900–911 (1997).

[11] A. Yu Kitaev. "Quantum computations: algorithms and error correction". Russian Mathematical Surveys **52**, 1191–1249 (1997).

[12] Austin G. Fowler, Matteo Mariantoni, John M. Martinis, and Andrew N. Cleland. "Surface codes: Towards practical large-scale quantum computation". Phys. Rev. A **86**, 032324 (2012).

[13] A. R. Calderbank and Peter W. Shor. "Good quantum error-correcting codes exist". Phys. Rev. A **54**, 1098–1105 (1996).

[14] Andrew Steane. "Multiple-particle interference and quantum error correction". Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences **452**, 2551–2577 (1996).

[15] Andrew Cross, Graeme Smith, John A. Smolin, and Bei Zeng. "Codeword stabilized quantum codes". In 2008 IEEE International Symposium on Information Theory. Pages 364–368. (2008).

[16] Isaac Chuang, Andrew Cross, Graeme Smith, John Smolin, and Bei Zeng. "Codeword stabilized quantum codes: Algorithm and structure". Journal of Mathematical Physics **50**, 042109 (2009).

[17] Nikolas P. Breuckmann and Jens Niklas Eberhardt. "Quantum low-density parity-check codes". PRX Quantum **2**, 040101 (2021).

[18] Pavel Panteleev and Gleb Kalachev. "Asymptotically good quantum and locally testableclassical ldpc codes". In Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing. Pages 375–388. Association for Computing Machinery (2022).

[19] Laird Egan, Dripto M. Debroy, Crystal Noel, Andrew Risinger, Daiwei Zhu, Debopriyo Biswas, Michael Newman, Muyuan Li, Kenneth R. Brown, Marko Cetina, and Christopher Monroe. "Fault-tolerant control of an error-corrected qubit". Nature **598**, 281–286 (2021).

[20] Lukas Postler, Sascha Heußen, Ivan Pogorelov, Manuel Rispler, Thomas Feldker, Michael Meth, Christian D. Marciniak, Roman Stricker, Martin Ringbauer, Rainer Blatt, Philipp Schindler, Markus Müller, and Thomas Monz. "Demonstration of fault-tolerant universal quantum gate operations". Nature **605**, 675–680 (2022).

[21] Christopher M. Dawson, Henry L. Haselgrove, and Michael A. Nielsen. "Noise thresholds for optical quantum computers". Phys. Rev. Lett. **96**, 020501 (2006).

[22] C. D. Wilen, S. Abdullah, N. A. Kurinsky, C. Stanford, L. Cardani, G. D'Imperio, C. Tomei, L. Faoro, L. B. Ioffe, C. H. Liu, A. Opremcak, B. G. Christensen, J. L. DuBois, and R. McDermott. "Correlated charge noise and relaxation errors in superconducting qubits". Nature **594**, 369–373 (2021).

[23] Qihao Guo, Yuan-Yuan Zhao, Markus Grassl, Xinfang Nie, Guo-Yong Xiang, Tao Xin, Zhang-Qi Yin, and Bei Zeng. "Testing a quantum error-correcting code on various platforms". Science Bulletin **66**, 29–35 (2021).

[24] Sixia Yu, Qing Chen, and C. H. Oh. "Graphical quantum error-correcting codes" (2007). arXiv:0709.1780.

[25] Dan Hu, Weidong Tang, Meisheng Zhao, Qing Chen, Sixia Yu, and C. H. Oh. "Graphical nonbinary quantum error-correcting codes". Phys. Rev. A **78**, 012306 (2008).

[26] Akshaya Jayashankar, Anjala M. Babu, Hui Khoon Ng, and Prabha Mandayam. "Finding good quantum codes using the cartan form". Phys. Rev. A **101**, 042307 (2020).

[27] Muyuan Li, Mauricio Gutiérrez, Stanley E. David, Alonzo Hernandez, and Kenneth R. Brown. "Fault tolerance with bare ancillary qubits for a [[7,1,3]] code". Phys. Rev. A **96**, 032341 (2017).

[28] Thomas Fösel, Petru Tighineanu, Talitha Weiss, and Florian Marquardt. "Reinforcement learning with neural networks for quantum feedback". Phys. Rev. X **8**, 031084 (2018).

[29] Paul Baireuther, Thomas E. O'Brien, Brian Tarasinski, and Carlo W. J. Beenakker. "Machine-learning-assisted correction of correlated qubit errors in a topological code". Quantum **2**, 48 (2018).

[30] Philip Andreasson, Joel Johansson, Simon Liljestrand, and Mats Granath. "Quantum error correction for the toric code using deep reinforcement learning". Quantum **3**, 183 (2019).

[31] Hendrik Poulsen Nautrup, Nicolas Delfosse, Vedran Dunjko, Hans J. Briegel, and Nicolai Friis. "Optimizing quantum error correction codes with reinforcement learning". Quantum **3**, 215 (2019).

[32] M. Reimpell and R. F. Werner. "Iterative optimization of quantum error correcting codes". Phys. Rev. Lett. **94**, 080501 (2005).

[33] Andrew S. Fletcher, Peter W. Shor, and Moe Z. Win. "Optimum quantum error recovery using semidefinite programming". Phys. Rev. A **75**, 012338 (2007).

[34] Andrew S. Fletcher. "Channel-adapted quantum error correction" (2007). arXiv:0706.3400.

[35] Ryan Sweke, Markus S. Kesselring, Evert P. L. van Nieuwenburg, and Jens Eisert. "Reinforcement learning decoders for fault-tolerant quantum computation". Machine Learning: Science and Technology **2**, 025005 (2020).

[36] Ye-Hua Liu and David Poulin. "Neural belief-propagation decoders for quantum error-correcting codes". Phys. Rev. Lett. **122**, 200501 (2019).

[37] David F. Locher, Lorenzo Cardarelli, and Markus Müller. "Quantum error correction with quantum autoencoders" (2022). arXiv:2202.00555.

[38] Emanuel Knill and Raymond Laflamme. "Concatenated quantum codes" (1996). arXiv:quant-ph/9608012.

[39] Markus Grassl, Peter Shor, Graeme Smith, John Smolin, and Bei Zeng. "Generalized concatenated quantum codes". Phys. Rev. A **79**, 050306 (2009).

[40] Daniel Gottesman. "An introduction to quantum error correction". In Proceedings of Symposia in Applied Mathematics. Volume 58, pages 221–236. (2002).

[41] P. Aliferis, F. Brito, D. P. DiVincenzo, J. Preskill, M. Steffen, and B. M. Terhal. "Fault-tolerant computing with biased-noise superconducting qubits: a case study". New Journal of Physics **11**, 013061 (2009).

[42] Tyler Jackson, Markus Grassl, and Bei Zeng. "Concatenated codes for amplitude damping". In 2016 IEEE International Symposium on Information Theory (ISIT). Pages 2269–2273. (2016).

[43] Debbie W. Leung, M. A. Nielsen, Isaac L. Chuang, and Yoshihisa Yamamoto. "Approximate quantum error correction can lead to better codes". Phys. Rev. A **56**, 2567–2573 (1997).

[44] Benjamin Schumacher and Michael D. Westmoreland. "Approximate quantum error correction". Quantum Information Processing **1**, 5–12 (2002).

[45] Fernando G. S. L. Brandão, Elizabeth Crosson, M. Burak Şahinoğlu, and John Bowen. "Quantum error correcting codes in eigenstates of translation-invariant spin chains". Phys. Rev. Lett. **123**, 110502 (2019).

[46] Cédric Bény and Ognyan Oreshkov. "General conditions for approximate quantum error correction and near-optimal recovery channels". Phys. Rev. Lett. **104**, 120501 (2010).

[47] Donald Bures. "An extension of Kakutani's theorem on infinite product measures to the tensor product of semifinite w*-algebras". Transactions of the American Mathematical Society **135**, 199–212 (1969).

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

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

[50] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J. Love, Alán Aspuru-Guzik, and Jeremy L. O'Brien. "A variational eigenvalue solver on a photonic quantum processor". Nature Communications **5**, 4213 (2014).

[51] Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M. Chow, and Jay M. Gambetta. "Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets". Nature **549**, 242–246 (2017).

[52] Yunseong Nam, Jwo-Sy Chen, Neal C. Pisenti, Kenneth Wright, Conor Delaney, Dmitri Maslov, Kenneth R. Brown, Stewart Allen, Jason M. Amini, Joel Apisdorf, Kristin M. Beck, Aleksey Blinov, Vandiver Chaplin, Mika Chmielewski, Coleman Collins, Shantanu Debnath, Kai M. Hudek, Andrew M. Ducore, Matthew Keesan, Sarah M. Kreikemeier, Jonathan Mizrahi, Phil Solomon, Mike Williams, Jaime David Wong-Campos, David Moehring, Christo-

Accepted in ⟨ quantum 2022-09-30, click title to verify. Published under CC-BY 4.0.

23

pher Monroe, and Jungsang Kim. "Ground-state energy estimation of the water molecule on a trapped-ion quantum computer". npj Quantum Information **6**, 33 (2020).

[53] Chenfeng Cao, Yunlong Yu, Zipeng Wu, Nic Shannon, Bei Zeng, and Robert Joynt. "Mitigating algorithmic errors in quantum optimization through energy extrapolation". Quantum Science and Technology (2022).

[54] Jonathan Romero, Jonathan P Olson, and Alan Aspuru-Guzik. "Quantum autoencoders for efficient compression of quantum data". Quantum Science and Technology **2**, 045001 (2017).

[55] Chenfeng Cao and Xin Wang. "Noise-assisted quantum autoencoder". Phys. Rev. Applied **15**, 054012 (2021).

[56] Kunal Sharma, Sumeet Khatri, M. Cerezo, and Patrick J. Coles. "Noise resilience of variational quantum compiling". New Journal of Physics **22**, 043006 (2020).

[57] Xiaosi Xu, Simon C. Benjamin, and Xiao Yuan. "Variational circuit compiler for quantum error correction". Phys. Rev. Applied **15**, 034068 (2021).

[58] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii. "Quantum circuit learning". Phys. Rev. A **98**, 032309 (2018).

[59] Hsin-Yuan Huang, Richard Kueng, and John Preskill. "Predicting many properties of a quantum system from very few measurements". Nature Physics **16**, 1050–1057 (2020).

[60] M. J. D. Powell. "An efficient method for finding the minimum of a function of several variables without calculating derivatives". The Computer Journal **7**, 155–162 (1964).

[61] Tobias Haug, Kishor Bharti, and M. S. Kim. "Capacity and quantum geometry of parametrized quantum circuits". PRX Quantum **2**, 040309 (2021).

[62] Peter D. Johnson, Jonathan Romero, Jonathan Olson, Yudong Cao, and Alán Aspuru-Guzik. "QVECTOR: an algorithm for device-tailored quantum error correction" (2017). arXiv:1711.02249.

[63] Raymond Laflamme, Cesar Miquel, Juan Pablo Paz, and Wojciech Hubert Zurek. "Perfect quantum error correcting code". Phys. Rev. Lett. **77**, 198–201 (1996).

[64] Eric M. Rains, R. H. Hardin, Peter W. Shor, and N. J. A. Sloane. "A nonadditive quantum code". Phys. Rev. Lett. **79**, 953–954 (1997).

[65] A. M. Steane. "Simple quantum error-correcting codes". Phys. Rev. A **54**, 4741–4751 (1996).

[66] Lev Ioffe and Marc Mézard. "Asymmetric quantum error-correcting codes". Phys. Rev. A **75**, 032345 (2007).

[67] Pradeep Kiran Sarvepalli, Andreas Klappenecker, and Martin Rotteler. "Asymmetric quantum LDPC codes". In 2008 IEEE International Symposium on Information Theory. Pages 305–309. (2008).

[68] Pradeep Kiran Sarvepalli, Andreas Klappenecker, and Martin Rötteler. "Asymmetric quantum codes: constructions, bounds and performance". Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **465**, 1645–1672 (2009).

[69] Martianus Frederic Ezerman, San Ling, and Patrick Sole. "Additive asymmetric quantum codes". IEEE Transactions on Information Theory **57**, 5536–5550 (2011).

[70] Martianus Frederic Ezerman, Somphong Jitman, San Ling, and Dmitrii V. Pasechnik. "CSS-like constructions of asymmetric quantum codes". IEEE Transactions on Information Theory **59**, 6732–6754 (2013).

[71] Tyler Jackson, Markus Grassl, and Bei Zeng. "Codeword stabilized quantum codes for asymmetric channels". In 2016 IEEE International Symposium on Information Theory (ISIT). Pages 2264–2268. (2016).

[72] J. Pablo Bonilla Ataides, David K. Tuckett, Stephen D. Bartlett, Steven T. Flammia, and Benjamin J. Brown. "The xzzx surface code". Nature Communications **12**, 2172 (2021).

[73] Prithviraj Prabhu and Ben W. Reichardt. "Distance-four quantum codes with combined postselection and error correction" (2021). arXiv:2112.03785.

[74] A. R. Calderbank, E. M. Rains, P. M. Shor, and N. J. A. Sloane. "Quantum error correction via codes over GF(4)". IEEE Transactions on Information Theory **44**, 1369–1387 (1998).

Accepted in ⟨ ⟩uantum 2022-09-30, click title to verify. Published under CC-BY 4.0.

24

[75] Yusuke Hama. "Quantum circuits for collective amplitude damping in two-qubit systems" (2020). arXiv:2012.02410.

[76] Markus Grassl, Linghang Kong, Zhaohui Wei, Zhang-Qi Yin, and Bei Zeng. "Quantum error-correcting codes for qudit amplitude damping". IEEE Transactions on Information Theory 64, 4674–4685 (2018).

[77] Peter Shor and Raymond Laflamme. "Quantum analog of the macwilliams identities for classical coding theory". Phys. Rev. Lett. 78, 1600–1602 (1997).

[78] Chenfeng Cao. "VarQEC GitHub repository". https://github.com/caochenfeng/VarQEC-public (2022).

[79] Zijun Chen, Kevin J. Satzinger, Juan Atalaya, Alexander N. Korotkov, Andrew Dunsworth, Daniel Sank, Chris Quintana, Matt McEwen, Rami Barends, Paul V. Klimov, Sabrina Hong, Cody Jones, Andre Petukhov, Dvir Kafri, Sean Demura, Brian Burkett, Craig Gidney, Austin G. Fowler, Alexandru Paler, Harald Putterman, Igor Aleiner, Frank Arute, Kunal Arya, Ryan Babbush, Joseph C. Bardin, Andreas Bengtsson, Alexandre Bourassa, Michael Broughton, Bob B. Buckley, David A. Buell, Nicholas Bushnell, Benjamin Chiaro, Roberto Collins, William Courtney, Alan R. Derk, Daniel Eppens, Catherine Erickson, Edward Farhi, Brooks Foxen, Marissa Giustina, Ami Greene, Jonathan A. Gross, Matthew P. Harrigan, Sean D. Harrington, Jeremy Hilton, Alan Ho, Trent Huang, William J. Huggins, L. B. Ioffe, Sergei V. Isakov, Evan Jeffrey, Zhang Jiang, Kostyantyn Kechedzhi, Seon Kim, Alexei Kitaev, Fedor Kostritsa, David Landhuis, Pavel Laptev, Erik Lucero, Orion Martin, Jarrod R. McClean, Trevor McCourt, Xiao Mi, Kevin C. Miao, Masoud Mohseni, Shirin Montazeri, Wojciech Mruczkiewicz, Josh Mutus, Ofer Naaman, Matthew Neeley, Charles Neill, Michael Newman, Murphy Yuezhen Niu, Thomas E. O'Brien, Alex Opremcak, Eric Ostby, Bálint Pató, Nicholas Redd, Pedram Roushan, Nicholas C. Rubin, Vladimir Shvarts, Doug Strain, Marco Szalay, Matthew D. Trevithick, Benjamin Villalonga, Theodore White, Z. Jamie Yao, Ping Yeh, Juhwan Yoo, Adam Zalcman, Hartmut Neven, Sergio Boixo, Vadim Smelyanskiy, Yu Chen, Anthony Megrant, Julian Kelly, and Google Quantum AI. "Exponential suppression of bit or phase errors with cyclic error correction". Nature 595, 383–387 (2021).

[80] Alexander M. Dalzell, Nicholas Hunter-Jones, and Fernando G. S. L. Brandão. "Random quantum circuits transform local noise into global white noise" (2021). arXiv:2111.14907.

[81] Abhinav Deshpande, Bill Fefferman, Alexey V. Gorshkov, Michael J. Gullans, Pradeep Niroula, and Oles Shtanko. "Tight bounds on the convergence of noisy random circuits to uniform" (2021). arXiv:2112.00716.

[82] William J. Huggins, Sam McArdle, Thomas E. O'Brien, Joonho Lee, Nicholas C. Rubin, Sergio Boixo, K. Birgitta Whaley, Ryan Babbush, and Jarrod R. McClean. "Virtual distillation for quantum error mitigation". Phys. Rev. X 11, 041036 (2021).

[83] Bálint Koczor. "Exponential error suppression for near-term quantum devices". Phys. Rev. X 11, 031057 (2021).

[84] Jarrod R. McClean, Sergio Boixo, Vadim N. Smelyanskiy, Ryan Babbush, and Hartmut Neven. "Barren plateaus in quantum neural network training landscapes". Nature Communications 9, 4812 (2018).

[85] M. Cerezo, Akira Sone, Tyler Volkoff, Lukasz Cincio, and Patrick J. Coles. "Cost function dependent barren plateaus in shallow parametrized quantum circuits". Nature Communications 12, 1791 (2021).

[86] Samson Wang, Enrico Fontana, M. Cerezo, Kunal Sharma, Akira Sone, Lukasz Cincio, and Patrick J. Coles. "Noise-induced barren plateaus in variational quantum algorithms". Nature Communications 12, 6961 (2021).

[87] Taylor L. Patti, Khadijeh Najafi, Xun Gao, and Susanne F. Yelin. "Entanglement devised barren plateau mitigation". Phys. Rev. Research 3, 033090 (2021).

[88] Stefan H. Sack, Raimel A. Medina, Alexios A. Michailidis, Richard Kueng, and Maksym Serbyn. "Avoiding barren plateaus

using classical shadows". PRX Quantum **3**, 020365 (2022).

[89] 5 qubit backend: IBM Q team. "IBM Q 5 Quito backend specification v1.1.34". Retrieved from `https://quantum-computing.ibm.com` (2022).

[90] Markus Grassl, Sirui Lu, and Bei Zeng. "Codes for simultaneous transmission of quantum and classical information". In 2017 IEEE International Symposium on Information Theory (ISIT). Pages 1718–1722. (2017).

[91] Runyao Duan. "Super-activation of zero-error capacity of noisy quantum channels" (2009). arXiv:0906.2527.

[92] Xiao-Dong Yu, Timo Simnacher, Nikolai Wyderka, H. Chau Nguyen, and Otfried Gühne. "A complete hierarchy for the pure state marginal problem in quantum mechanics". Nature Communications **12**, 1012 (2021).

[93] Román Orús. "Tensor networks for complex quantum systems". Nature Reviews Physics **1**, 538–550 (2019).

[94] J. Ignacio Cirac, David Pérez-García, Norbert Schuch, and Frank Verstraete. "Matrix product states and projected entangled pair states: Concepts, symmetries, theorems". Rev. Mod. Phys. **93**, 045003 (2021).

[95] Song Cheng, Chenfeng Cao, Chao Zhang, Yongxiang Liu, Shi-Yao Hou, Pengxiang Xu, and Bei Zeng. "Simulating noisy quantum circuits with matrix product density operators". Phys. Rev. Research **3**, 023005 (2021).

[96] Giuseppe Carleo and Matthias Troyer. "Solving the quantum many-body problem with artificial neural networks". Science **355**, 602–606 (2017).

[97] Carl W. Helstrom. "Quantum detection and estimation theory". Journal of Statistical Physics **1**, 231–252 (1969).

[98] Dominik Šafránek. "Simple expression for the quantum Fisher information matrix". Phys. Rev. A **97**, 042322 (2018).

[99] Jing Liu, Haidong Yuan, Xiao-Ming Lu, and Xiaoguang Wang. "Quantum fisher information matrix and multiparameter estimation". Journal of Physics A: Mathematical and Theoretical **53**, 023001 (2019).

[100] Johannes Jakob Meyer. "Fisher Information in Noisy Intermediate-Scale Quantum Applications". Quantum **5**, 539 (2021).

[101] John Milnor and James D Stasheff. "Characteristic classes. annals of mathematics studies, volume 76". Princeton University Press. (2016).

Appendices

A Proof of Proposition 3

Proof   From the completeness relation of the Kraus operators $\{E_\alpha\}$,

$$\sum_{\alpha=1}^m E_\alpha^\dagger E_\alpha = I, \tag{101}$$

we know

$$
\begin{aligned}
\operatorname{Tr}(\sum_{\alpha=1}^m E_\alpha E_\alpha^\dagger) &= \sum_{\alpha=1}^m \operatorname{Tr}(E_\alpha E_\alpha^\dagger) \\
&= \sum_{\alpha=1}^m \operatorname{Tr}(E_\alpha^\dagger E_\alpha) \\
&= \operatorname{Tr}(\sum_{\alpha=1}^m E_\alpha^\dagger E_\alpha) \\
&= 2^n.
\end{aligned}
\tag{102}
$$

Each $E_\alpha E_\alpha^\dagger$ and $E_\alpha^\dagger E_\alpha$ are positive semidefinite. Denote the eigenvalues of $E_\alpha E_\alpha^\dagger$ as $\xi_1^\alpha \geq \xi_2^\alpha \geq \cdots \geq \xi_{2^n}^\alpha \geq 0$, the eigenvalues of $E_\beta E_\beta^\dagger$ as $\xi_1^\beta \geq \xi_2^\beta \geq \cdots \geq \xi_{2^n}^\beta \geq 0$, $\sum_{\alpha,j} \xi_j^\alpha = \sum_{\beta,j} \xi_j^\beta = 2^n$.

Then we have

$$
\begin{aligned}
&\operatorname{Tr}\Big(\sum_{\beta=1}^m \sum_{\alpha=1}^m (E_\alpha^\dagger E_\beta)^\dagger E_\alpha^\dagger E_\beta\Big) \\
&= \sum_{\beta=1}^m \sum_{\alpha=1}^m \operatorname{Tr}(E_\alpha E_\alpha^\dagger E_\beta E_\beta^\dagger) \\
&\leq \sum_{\beta=1}^m \sum_{\alpha=1}^m \sum_{j=1}^{2^n} \xi_j^\alpha \xi_j^\beta \\
&\leq \sum_{\beta=1}^m \sum_{\alpha=1}^m \sum_{i=1}^{2^n} \sum_{j=1}^{2^n} \xi_i^\alpha \xi_j^\beta \\
&= 2^{2n}.
\end{aligned}
\tag{103}
$$

The first inequality uses von Neumann's trace in-

equality.

Each E_α non-trivially acts on no more than $\lfloor (d-1)/2 \rfloor$ qubits, therefore, each error product $E_\alpha^\dagger E_\beta$ non-trivially acts on no more than $(d-1)$ qubits. We expand $E_\alpha^\dagger E_\beta$ in the Pauli basis,

$$E_\alpha^\dagger E_\beta = \sum_\gamma \chi_\gamma^{\alpha\beta} O_\gamma^{\alpha\beta}, \qquad (104)$$

where each $O_\gamma^{\alpha\beta}$ is a Pauli tensor product with weight less than d, $O_\gamma^{\alpha\beta\dagger} O_\gamma^{\alpha\beta} = I$, $|\{O_\gamma^{\alpha\beta}\}| \le 4^{d-1} m^2$. Then

$$\begin{aligned}
\mathrm{Tr}\left((E_\alpha^\dagger E_\beta)^\dagger E_\alpha^\dagger E_\beta\right) &= \sum_\gamma \mathrm{Tr}\left(|\chi_\gamma^{\alpha\beta}|^2 O_\gamma^{\alpha\beta\dagger} O_\gamma^{\alpha\beta}\right) \\
&= \sum_\gamma |\chi_\gamma^{\alpha\beta}|^2 \,\mathrm{Tr}(O_\gamma^{\alpha\beta\dagger} O_\gamma^{\alpha\beta}) \\
&= 2^n \sum_\gamma |\chi_\gamma^{\alpha\beta}|^2,
\end{aligned} \qquad (105)$$

$$\mathrm{Tr}\left(\sum_{\beta=1}^m \sum_{\alpha=1}^m (E_\alpha^\dagger E_\beta)^\dagger E_\alpha^\dagger E_\beta\right) = 2^n \sum_{\alpha,\beta,\gamma} |\chi_\gamma^{\alpha\beta}|^2. \qquad (106)$$

According to Eq. (103), we have

$$\sum_{\alpha,\beta,\gamma} |\chi_\gamma^{\alpha\beta}|^2 \le 2^n. \qquad (107)$$

For the basis states $\{|\psi_1\rangle, |\psi_2\rangle, \ldots, |\psi_K\rangle\}$,

$$\begin{aligned}
&\sum_{\alpha,\beta} \sum_{1\le i < j \le K} |\langle \psi_i | E_\alpha^\dagger E_\beta |\psi_j\rangle| \\
&= \sum_{\alpha,\beta,\gamma} \sum_{1\le i < j \le K} |\chi_\gamma^{\alpha\beta}| |\langle \psi_i | O_\gamma^{\alpha\beta} |\psi_j\rangle| \\
&\le \sum_{\alpha,\beta,\gamma} |\chi_\gamma^{\alpha\beta}| \sum_{1\le i < j \le K} \sum_{\mathrm{wt}(O_{\alpha'}) < d} |\langle \psi_i | O_{\alpha'} |\psi_j\rangle| \\
&\le 2^{n/2+d-1} m \sum_{1\le i < j \le K} \sum_{\mathrm{wt}(O_{\alpha'}) < d} |\langle \psi_i | O_{\alpha'} |\psi_j\rangle|.
\end{aligned} \qquad (108)$$

Similarly, we have

$$\begin{aligned}
&\sum_{\alpha,\beta} \sum_{j=1}^K |\langle \psi_j | E_\alpha^\dagger E_\beta |\psi_j\rangle - \overline{\langle E_\alpha^\dagger E_\beta \rangle}|/2 \\
&\le 2^{n/2+d-1} m \sum_{\mathrm{wt}(O_{\alpha'}) < d} \sum_{j=1}^K |\langle \psi_j | O_{\alpha'} |\psi_j\rangle - \overline{\langle O_{\alpha'} \rangle}|/2.
\end{aligned} \qquad (109)$$

Denote $\mathcal{E}' = \{E_\alpha^\dagger E_\beta | E_\alpha, E_\beta \text{ are Kraus operators of } \mathcal{N}\}$, we have

$$C_{n,K,\mathcal{E}'}^{\ell_1} \le 2^{n/2+d-1} m C_{n,K,\mathcal{E}}^{\ell_1}. \qquad (110)$$

According to Proposition 2, the code is ε-correctable with ε bounded by

$$\varepsilon \le K\sqrt{2 C_{n,K,\mathcal{E}'}^{\ell_1}} \le 2^{n/4+d/2} K \sqrt{m C_{n,K,\mathcal{E}}^{\ell_1}}. \qquad (111)$$

■

B Proof of Proposition 4

Proof Since each E_α is proportional to a Pauli error, we have

$$E_\alpha^\dagger E_\alpha = E_\alpha E_\alpha^\dagger. \qquad (112)$$

Further, from the completeness relation of the Kraus operators $\{E_\alpha\}$,

$$\sum_{\alpha=1}^m E_\alpha^\dagger E_\alpha = I, \qquad (113)$$

we obtain the completeness relation of the error products, $\{E_\alpha^\dagger E_\beta\}$

$$\sum_{\beta=1}^m \sum_{\alpha=1}^m (E_\alpha^\dagger E_\beta)^\dagger E_\alpha^\dagger E_\beta = \sum_{\beta=1}^m E_\beta^\dagger E_\beta = I. \qquad (114)$$

The c_Z-effective weight of each E_α smaller than $d_e(c_Z)/2$, therefore, each error product $E_\alpha^\dagger E_\beta$ is proportional to a Pauli tensor product with c_Z-effective weight

$$\mathrm{wt}_e(E_\alpha^\dagger E_\beta, c_Z) < d_e(c_Z). \qquad (115)$$

Denote

$$E_\alpha^\dagger E_\beta = \chi^{\alpha\beta} O^{\alpha\beta}, \qquad (116)$$

where $O^{\alpha\beta}$ is a Pauli tensor product.

According to Eq. (114), we have the normalization condition

$$\sum_{\alpha,\beta} |\chi^{\alpha\beta}|^2 = 1. \qquad (117)$$

For the basis states $\{|\psi_1\rangle, |\psi_2\rangle, \ldots, |\psi_K\rangle\}$,

$$\begin{aligned}
&\sum_{\alpha,\beta} \sum_{1\le i < j \le K} |\langle \psi_i | E_\alpha^\dagger E_\beta |\psi_j\rangle| \\
&= \sum_{\alpha,\beta} \sum_{1\le i < j \le K} |\chi^{\alpha\beta}| |\langle \psi_i | O^{\alpha\beta} |\psi_j\rangle| \\
&\le \sum_{\alpha,\beta} |\chi^{\alpha\beta}| \sum_{1\le i < j \le K} \sum_{\mathrm{wt}_e(O_{\alpha'},c_Z) < d_e(c_Z)} |\langle \psi_i | O_{\alpha'} |\psi_j\rangle| \\
&\le m \sum_{1\le i < j \le K} \sum_{\mathrm{wt}_e(O_{\alpha'},c_Z) < d_e(c_Z)} |\langle \psi_i | O_{\alpha'} |\psi_j\rangle|.
\end{aligned} \qquad (118)$$

Similarly, we have

$$\sum_{\alpha,\beta} \sum_{j=1}^{K} |\langle\psi_j|E_\alpha^\dagger E_\beta|\psi_j\rangle - \overline{\langle E_\alpha^\dagger E_\beta\rangle}|/2$$

$$\leq m \sum_{\mathrm{wt}_e(O_{\alpha'},c_{\mathbf{Z}})<d_e(c_{\mathbf{Z}})} \sum_{j=1}^{K} |\langle\psi_j|O_{\alpha'}|\psi_j\rangle - \overline{\langle O_{\alpha'}\rangle}|/2. \tag{119}$$

Denote
$\mathcal{E}' = \{E_\alpha^\dagger E_\beta | E_\alpha, E_\beta$ are Kraus operators of $\mathcal{N}\}$, we have

$$C_{n,K,\mathcal{E}'}^{\ell_1} \leq m C_{n,K,\mathcal{E}}^{\ell_1}. \tag{120}$$

According to Proposition 2, the code is ε-correctable with ε bounded by

$$\varepsilon \leq K\sqrt{2C_{n,K,\mathcal{E}'}^{\ell_1}} \leq K\sqrt{2mC_{n,K,\mathcal{E}}^{\ell_1}}. \tag{121}$$

■

C Parameter Dimension and Overparameterization

The quantum Fisher information matrix (QFIM) is an essential concept in quantum metrology [97–99]. In recent years, its applications in NISQ algorithms and quantum machine learning have also been noticed [61, 100]. Ref. [61] uses the QFIM to assess the expressive power of a VQC with the fixed input state $|0\rangle^{\otimes n}$. In VarQEC, however, we use K orthogonal input states to find an $((n,K))_2$ quantum code. In this section, we generalize the notion of QFIM to multiple input states to quantify the expressive power of a VQC for preparing an $((n,K))_2$ quantum code. Based on that, we discuss the parameter dimension and the overparameterization of a VQC encoder.

Suppose the VQC encoder has parameters

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_N). \tag{122}$$

For a fixed pure input state, one relates the QFIM $\mathcal{F}(\boldsymbol{\theta})$ to the distance in the space of pure quantum states by

$$\mathrm{Dist}(|\psi(\boldsymbol{\theta})\rangle, |\psi(\boldsymbol{\theta}+d\boldsymbol{\theta})\rangle)^2 = \sum_{l,m} \mathcal{F}_{l,m}(\boldsymbol{\theta}) d\theta_l d\theta_m, \tag{123}$$

where $\mathrm{Dist}(|\psi(\boldsymbol{\theta})\rangle, |\psi(\boldsymbol{\theta}')\rangle) = 1 - |\langle\psi(\boldsymbol{\theta})|\psi(\boldsymbol{\theta}')\rangle|^2$. The QFIM is an N by N matrix

$$\mathcal{F}_{lm}(\boldsymbol{\theta}) = 4\,\mathrm{Re}\left[\langle\partial_l\psi|\partial_m\psi\rangle - \langle\partial_l\psi|\psi\rangle\langle\psi|\partial_m\psi\rangle\right]. \tag{124}$$

where $|\partial_l\psi\rangle$ denotes $\partial|\psi(\boldsymbol{\theta})\rangle/\partial\theta_l$. In this case, the parameter dimension D_c for a VQC is defined as the number of independent parameters that the VQC can express in the space of output states. Numerical evidence shows that D_c is usually equivalent to the rank of QFIM for hardware-efficient VQCs with periodic and non-correlated random parameters $\boldsymbol{\theta}$ [61].

In VarQEC, the inputs are K orthogonal pure states. Denote the projector onto the output space as P_c. We relate $\mathcal{F}(\boldsymbol{\theta})$ to the distance in the space of K-dimensional projectors,

$$\mathrm{Dist}_K(P_c(\boldsymbol{\theta}), P_c(\boldsymbol{\theta}+d\boldsymbol{\theta}))^2 = \sum_{l,m} \mathcal{F}_{l,m}(\boldsymbol{\theta}) d\theta_l d\theta_m, \tag{125}$$

where the distance

$$\mathrm{Dist}_K(P_c(\boldsymbol{\theta}), P_c(\boldsymbol{\theta}')) = \left(\frac{\mathrm{Tr}\sqrt{\sqrt{P_c(\boldsymbol{\theta})}P_c(\boldsymbol{\theta}')\sqrt{P_c(\boldsymbol{\theta})}}}{K}\right)^2 \tag{126}$$

is defined as the fidelity between the normalized mixed states of projectors $P_c(\boldsymbol{\theta})$ and $P_c(\boldsymbol{\theta}')$. Suppose the projector $P_c(\boldsymbol{\theta})$ has eigen decomposition

$$P_c(\boldsymbol{\theta}) = \sum_{j=1}^{K} |\psi_j\rangle\langle\psi_j| \tag{127}$$

and denote the basis of its orthogonal complement as $\{|\psi_j\rangle\}_{j=K+1,K+2,\ldots,2^n}$, the QFIM under our framework is of the form

$$\mathcal{F}_{lm}(\boldsymbol{\theta}) = 2\frac{\partial^2}{\partial\delta_l\partial\delta_m}\mathrm{Dist}_K(P_c(\boldsymbol{\theta}), P_c(\boldsymbol{\theta}+\boldsymbol{\delta}))|_{\boldsymbol{\delta}=\mathbf{0}}$$

$$= \frac{2}{K^2}\sum_{\min\{i,j\}\leq K} \frac{\mathrm{Re}\left(\langle\psi_i|\partial_l P_c|\psi_j\rangle\langle\psi_j|\partial_m P_c|\psi_i\rangle\right)}{\langle\psi_i|P_c|\psi_i\rangle + \langle\psi_j|P_c|\psi_j\rangle}. \tag{128}$$

We remark that the derivation of QFIM for projectors is the same as for density matrices [98]. Therefore, a similar formula can be used to compute the QFIM of a VQC with mixed inputs/outputs.

Through sampling random parameters $\boldsymbol{\theta}$ from the interval $[0,2\pi)^N$ and computing the QFIM, we can estimate the parameter dimension D_c by $\mathrm{rank}(\mathcal{F}(\boldsymbol{\theta}))$. The VarQEC algorithm searches a D_c-dimensional submanifold of the complex Grassmannian $\mathbf{Gr}(K, 2^n)$.

Without loss of generality, we consider the connectivity graph shown in Fig. 12. For $K = 1, 2, 3, 4$, we randomly sample parameters $\boldsymbol{\theta}$ and
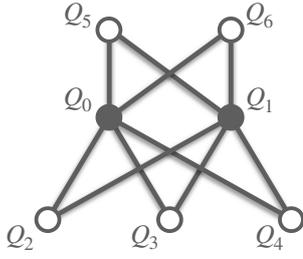
Figure 12: The bipartite connectivity graph for finding a quantum code with parameters $((7,3,3))_2$. Qubits Q_0 and Q_1 are selected to prepare the logical data.
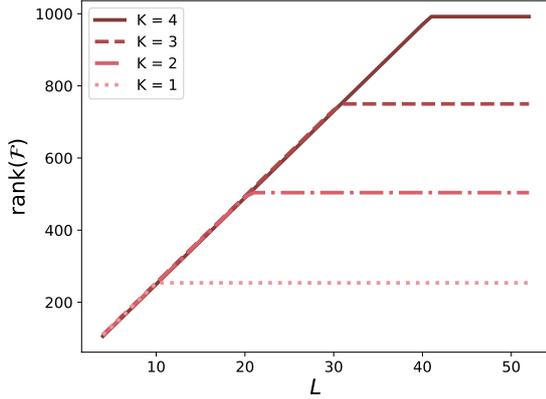

Figure 13: $\mathrm{Rank}(\mathcal{F}(\boldsymbol{\theta}))$ versus the number of VQC layers for the connectivity graph shown in Fig. 12. For code dimensions $K = 1, 2, 3, 4$, the maximum ranks are $D_c^{\max} = 254, 504, 750, 992$, and the required numbers of VQC layers to achieve overparameterization are $L_{\mathrm{crit}} = 10, 21, 31, 41$.

plot $\mathrm{rank}(\mathcal{F}(\boldsymbol{\theta}))$ as a function of the number of VQC layers L in Fig. 13. Almost no parameterized gate is redundant when the circuit is underparameterized ($D_c/N \approx 1$). With the increase of L, $\mathrm{rank}(\mathcal{F}(\boldsymbol{\theta}))$ increases approximately linearly until achieving its maximum $D_c^{\max}$. The maximum parameter dimension for code length n and code dimension K is of the form

$$D_c^{\max} = 2K(2^n - K). \tag{129}$$

This agrees with the fact that the dimension of the complex Grassmannian $\mathbf{Gr}(K, 2^n)$ is $K(2^n - K)$ [101]. When $D_c = D_c^{\max}$, the VQC can explore the whole $\mathbf{Gr}(K, 2^n)$ manifold and prepare arbitrary $((n, K))_2$ quantum code. The required number of layers to saturate the maximum parameter dimension is approximately

$$L_{\mathrm{crit}} = \left\lceil \frac{2K(2^n - K) - 2n}{2n + |E(G)|} \right\rceil, \tag{130}$$

where $|E(G)|$ is the number of edges of the connectivity graph.

To find a quantum code with parameters $((7,3,3))_2$, we sample 100 different initial values of $\boldsymbol{\theta}$ and implement VarQEC with an overparameterized VQC ($L = 31$). However, the cost function $C_{n,K,\mathcal{E}}^{\ell_1}(\boldsymbol{\theta})$ is always greater than 1. A $((7,3,3))_2$ code is improbable to exist.

D A variational quantum encoder for additive codes

The VQC with bipartite connectivity performs well in most cases. However, for some code parameters (e.g., $((10,4,4))_2$), it needs a large bunch of samples of the initial $\boldsymbol{\theta}$ to find an eligible code. Here we propose AC-VQC, another variational quantum circuit with all-to-all connectivity, to complement the bipartite ansatz.

The AC-VQC is especially resource-efficient in finding encoding circuits of additive codes. The structure of an AC-VQC is similar to the circuit of the quantum Fourier transform, as shown in Fig. 14. We start from two physical qubits (Q_0, Q_1) and apply a 2-qubit parameterized unitary operator U_{01} to them. Then, we add another qubit (Q_2), apply 2-qubit parameterized unitary operators U_{02}/U_{12} to the new one and each of the qubits that already exist (Q_0-Q_2, Q_1-Q_2). Repeat the steps until the system size equals n. In the end, we apply single-qubit rotations R_z and R_x to all qubits to explore the manifold of locally equivalent codes. The initial k qubits prepare the logical data. The total circuit depth is of order $\mathcal{O}(\sum_{j=1}^{n-1} j) = \mathcal{O}(n^2)$.
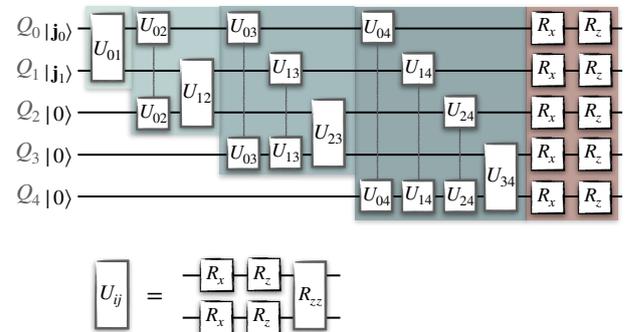

Figure 14: Schematic of AC-variational quantum circuit with $n = 5$, $k = 2$. Physical qubits are added layer by layer.

Accepted in ⟨ quantum 2022-09-30, click title to verify. Published under CC-BY 4.0.

29

E Non-CWS Quantum Codes with Parameters $((6,2,3))_2$, $((7,2,3))_2$

The $((5,2,3))_2$ code is known to be unique. However, the classification and construction of $((6,2,3))_2$ and $((7,2,3))_2$ quantum codes are unclear. Some are said to be "non-CWS" since they are not locally equivalent to CWS codes. Here we present a general construction of non-CWS quantum codes based on stabilizer ones.

Theorem 6 *If there exists a quantum code $\mathcal{C}$ with parameters $((n, 2^k, d))_q$, then there exist degenerate $((n', 2^k, d))_q$ codes $\{\mathcal{C}'\}$ with $n' > n$ that are not locally equivalent to a CWS code.*

Proof For $n' > n+1$, we can directly obtain non-CWS codes by taking the tensor product with a non-stabilizer state.

For $n' = n+1$, we take the tensor product of the code $\mathcal{C}$ with a fixed (stabilizer) state, then apply a non-Clifford entangling unitary operation to one of the original qudits and the additional qudit. This will conjugate the Pauli-stabilizers to non-local stabilizers. The resulting code is a non-CWS degenerate code with parameters $((n', 2^k, d))_q$. ∎

For example, we start from the $((5,2,3))_2$ code $(Q_0, Q_1, \ldots, Q_4)$ and initialize an addition qubit (Q_5) in state $|0\rangle$, then apply the transformation U on Q_5, controlled by Q_4 with

$$U = \frac{1}{5}\begin{pmatrix} 3 & 4 \\ -4 & 3 \end{pmatrix}. \tag{131}$$

The resulting code is a non-CWS degenerate $((6,2,3))_2$ code with basis states

$$
\begin{aligned}
|\psi_1\rangle =& \frac{\sqrt{2}}{20}(5|000000\rangle - 5i|001100\rangle - 3i|010010\rangle \\
& + 4i|010011\rangle + 3|011110\rangle - 4|011111\rangle \\
& - 3|100110\rangle + 4|100111\rangle - 3i|101010\rangle \\
& + 4i|101011\rangle - 5i|110100\rangle - 5|111000\rangle),
\end{aligned}
\tag{132}
$$

$$
\begin{aligned}
|\psi_2\rangle =& \frac{\sqrt{2}}{20}(3|00110\rangle - 4|00111\rangle - 3i|01010\rangle \\
& + 4i|01011\rangle - 5i|10100\rangle + 5|11000\rangle \\
& + 5|100000\rangle + 5i|101100\rangle + 3i|110010\rangle \\
& - 4i|110011\rangle + 3|111110\rangle - 4|111111\rangle).
\end{aligned}
\tag{133}
$$

Its weight enumerators are

$$A(z) = 1 + \frac{9}{25}z + \frac{16}{25}z^2 + \frac{311}{25}z^4 + \frac{391}{25}z^5 + \frac{48}{25}z^6, \tag{134}$$

$$
\begin{aligned}
B(z) =& 1 + \frac{9}{25}z + \frac{16}{25}z^2 + \frac{654}{25}z^3 + \frac{193}{5}z^4 \\
& + \frac{937}{25}z^5 + \frac{594}{25}z^6.
\end{aligned}
\tag{135}
$$

Likewise, non-CWS $((7,2,3))_2$ codes can be constructed based on $((6,2,3))_2$ stabilizer codes.

Another class of non-CWS $((6,2,3))_2$ codes are unitarily related to the additive $((6,2,3))_2$ code stabilized by

$$
\begin{aligned}
g_1 &= Y & I & Z & X & X & Y \\
g_2 &= Z & X & I & X & I & Z \\
g_3 &= I & Z & X & X & X & X \\
g_4 &= I & I & I & I & Z & Z \\
g_5 &= Z & Z & Z & Z & I & I.
\end{aligned}
\tag{136}
$$

Here "unitarily related" means they can be transformed to this code when allowing permutations of qubits and a unitary transformation of the form $U = \otimes_{j=1}^5 U_j$ where U_1, U_2, U_3, U_4 are single-qubit unitaries and U_5 is a 2-qubit unitary. In our numerical experiments, all the $((6,2,3))_2$ codes discovered by VarQEC are unitarily related to this $((6,2,3))_2$ stabilizer code or the $((5,2,3))_2$ perfect code, some are related to both.

Consider a quantum code $\mathcal{C}$ that is capable of correcting an error set $\mathcal{E}$. If linearly independent errors in $\mathcal{E}$ map $\mathcal{C}$ to linearly independent subspaces, we say $\mathcal{C}$ is *non-degenerate* with respect to $\mathcal{E}$. If linearly independent errors in $\mathcal{E}$ map $\mathcal{C}$ to mutually orthogonal subspaces, we say $\mathcal{C}$ is *pure* with respect to $\mathcal{E}$ [74]. A pure code must be non-degenerate. For additive codes and CWS codes, "non-degenerate" and "pure" are equivalent [74]. However, we note that some of our $((7,2,3))_2$ codes are non-degenerate but impure. Here we give a "trivial" construction of non-degenerate and impure $((7,2,3))_2$ codes. Still, we start with the $((5,2,3))_2$ code on qubits $Q_0, Q_1, \ldots, Q_4$ and add two additional qubits Q_5, Q_6 in a fixed state $|00\rangle$. Then we can apply either a three-qubit unitary on, e.g., Q_4, Q_5, Q_6, or two two-qubit unitaries on, e.g., Q_3, Q_5 and Q_4, Q_6. When these unitaries are non-Clifford, the resulting code is a non-CWS degenerate and impure code. We give a detailed example here. The following basis states span a non-degenerate but impure $((7,2,3))_2$ code:

$$|\psi_1\rangle = \frac{1}{20}(-4\omega^3|0000011\rangle + 3\omega^2|0000100\rangle - 3\omega|0001110\rangle - 4\omega|0001111\rangle + 3\omega^3|0010110\rangle$$
$$+ 4\omega^3|0010111\rangle - 4\omega|0011011\rangle + 3|0011100\rangle + 3\omega^3|0100110\rangle + 4\omega^3|0100111\rangle - 4\omega|0101011\rangle$$
$$+ 3\omega|0101100\rangle - 4\omega^3|0110011\rangle + 3\omega^2|0110100\rangle - 3\omega|0111110\rangle - 4\omega|0111111\rangle - 4\omega^3|1000011\rangle$$
$$+ 3\omega^2|1000100\rangle + 3\omega|1001110\rangle + 4\omega|1001111\rangle + 3\omega^3|1010110\rangle - 4\omega^3|1010111\rangle - 4\omega|1011011\rangle$$
$$- 3|1011100\rangle - 3\omega^3|1100110\rangle - 4\omega^3|1100111\rangle) - 4\omega|1101011\rangle + 3|1101100\rangle + 4\omega^3|1110011\rangle$$
$$- 3\omega^2|1110100\rangle - 3\omega|1111110\rangle - 4\omega|1111111\rangle), \tag{137}$$

$$|\psi_2\rangle = \frac{1}{20}(4\omega|0000011\rangle - 3|0000100\rangle + 3\omega^3|0001110\rangle + 4\omega^3|0001111\rangle + 3\omega|0010110\rangle + 4\omega|0010111\rangle$$
$$- 4\omega^3|0011011\rangle + 3\omega^2|0011100\rangle - 3\omega|0100110\rangle - 4\omega|0100111\rangle + 4\omega^3|0101011\rangle - 3\omega^2|0101100\rangle$$
$$- 4\omega|0110011\rangle + 3|0110100\rangle - 3\omega^3|0111110\rangle - 4\omega^3|0111111\rangle - 4\omega|1000011\rangle + 3|1000100\rangle$$
$$+ 3\omega^3|1001110\rangle + 4\omega^3|1001111\rangle - 3\omega|1010110\rangle - 4\omega|1010111\rangle - 4\omega^3|1011011\rangle + 3\omega^2|1011100\rangle$$
$$- 3\omega|1100110\rangle - 4\omega|1100111\rangle - 4\omega^3|1101011\rangle + 3\omega^2|1101100\rangle - 4\omega|1110011\rangle + 3|1110100\rangle$$
$$+ 3\omega^3|1111110\rangle - 4\omega^3|1111111\rangle), \tag{138}$$

where $\omega = \exp(i\pi/4)$. Its weight enumerators are

$$A(z) = 1 + \frac{106}{125}z + z^2 + \frac{144}{125}z^3 + \frac{1299}{125}z^4$$
$$+ \frac{3318}{125}z^5 + \frac{2451}{125}z^6 + \frac{432}{125}z^7, \tag{139}$$

$$B(z) = 1 + \frac{106}{125}z + z^2 + + \frac{606}{25}z^3 + \frac{7071}{125}z^4$$
$$+ \frac{9318}{125}z^5 + \frac{8679}{125}z^6 + \frac{3546}{125}z^7. \tag{140}$$

F Quantum Weight Enumerators

This section lists the quantum weight enumerators of some QECCs discovered/rediscovered by VarQEC.

F.1 Symmetric codes

$$A^{\{5,6,2\}}(z) = 1 + 1.667z^4 + 2.667z^5, \tag{141}$$

$$B^{\{5,6,2\}}(z) = 1 + 20z^2 + 50z^3 + 75z^4 + 46z^5; \tag{142}$$

$$A^{\{5,2,3\}}(z) = 1 + 15z^4, \tag{143}$$

$$B^{\{5,2,3\}}(z) = 1 + 30z^3 + 15z^4 + 18z^5; \tag{144}$$

$$A^{\{6,2,3\}}(z) = 1 + 0.267z + 0.732z^2 + 12.070z^4$$
$$+ 15.732z^5 + 2.197z^6, \tag{145}$$

$$B^{\{6,2,3\}}(z) = 1 + 0.267z + 0.732z^2 + 25.605z^3$$
$$+ 37.676z^4 + 38.126z^5 + 24.591z^6; \tag{146}$$

$$A^{\{7,8,2\}}(z) = 1 + 5z^4 + 2z^5 + 2z^6 + 6z^7, \tag{147}$$

$$B^{\{7,8,2\}}(z) = 1 + 17z^2 + 40z^3 + 195z^4 + 328z^5$$
$$+ 299z^6 + 144z^7; \tag{148}$$

Non-degenerate $((7,2,3))_2$:
$$A^{\{7,2,3\}}(z) = 1 + 1.437z^2 + 18.125z^4 + 43.437z^6, \tag{149}$$

$$B^{\{7,2,3\}}(z) = 1 + 1.437z^2 + 25.311z^3 + 18.125z^4$$
$$+ 117.377z^5 + 43.437z^6 + 49.311z^7; \tag{150}$$

Degenerate $((7,2,3))_2$:
$$A^{\{7,2,3\}}(z) = 1 + 5z^2 + 11z^4 + 47z^6, \tag{151}$$

$$B^{\{7,2,3\}}(z) = 1 + 5z^2 + 36z^3 + 11z^4 + 96z^5$$
$$+ 47z^6 + 60z^7; \tag{152}$$

$$A^{\{8,2,3\}}(z) = 1 + 0.015z^3 + 13.924z^4 + 24.091z^5$$
$$+ 40.030z^6 + 39.893z^7 + 9.046z^8, \tag{153}$$

$$B^{\{8,2,3\}}(z) = 1 + 11.970z^3 + 38.152z^4$$
$$+ 119.817z^5 + 159.939z^6 \tag{154}$$
$$+ 124.213z^7 + 56.909z^8;$$

$$A^{\{8,8,3\}}(z) = 1 + 28z^6 + 3z^8, \tag{155}$$

$$B^{\{8,8,3\}}(z) = 1 + 56z^3 + 210z^4 + 336z^5 + 728z^6$$
$$+ 504z^7 + 213z^8; \tag{156}$$

$$A^{\{9,8,3\}}(z) = 1 + 0.042z^4 + 5.875z^5 + 16.083z^6$$
$$+ 24.083z^7 + 14.875z^8 + 2.042z^9, \tag{157}$$

$$B^{\{9,8,3\}}(z) = 1 + 40z^3 + 162.332z^4 + 479.004z^5$$
$$+ 952.664z^6 + 1224.664z^7$$
$$+ 932.004z^8 + 304.332z^9; \tag{158}$$

$$A^{\{10,16,3\}}(z) = 1 + 0.127z^5 + 8.525z^6 + 20.443z^7$$
$$+ 21.253z^8 + 11.430z^9 + 1.221z^{10}, \tag{159}$$

$$B^{\{10,16,3\}}(z) = 1 + 55.810z^3 + 275.961z^4$$
$$+ 954.241z^5 + 2366.014z^6$$
$$+ 4120.948z^7 + 4622.227z^8$$
$$+ 3061.001z^9 + 926.797z^{10}; \tag{160}$$

$$A^{\{10,4,4\}}(z) = 1 + 90z^6 + 135z^8 + 30z^{10}, \tag{161}$$

$$B^{\{10,4,4\}}(z) = 1 + 90z^4 + 216z^5 + 720z^6 + 720z^7$$
$$+ 1485z^8 + 600z^9 + 264z^{10}. \tag{162}$$

$$A^{\{11,32,3\}}(z) = 1 + 0.160z^6 + 12.522z^7 + 23.316z^8$$
$$+ 18.319z^9 + 7.524z^{10} + 1.158z^{11} \tag{163}$$

$$B^{\{11,32,3\}}(z) = 1 + 75.003z^3 + 443.260z^4$$
$$+ 1729.654z^5 + 5219.456z^6$$
$$+ 11343.613z^7 + 16918.654z^8$$
$$+ 16859.456z^9 + 10185.630z^{10}$$
$$+ 2760.274z^{11} \tag{164}$$

$$A^{\{12,64,3\}}(z) = 1 + 2z^7 + 15z^8 + 24z^9 + 16z^{10}$$
$$+ 6z^{11}, \tag{165}$$

$$B^{\{12,64,3\}}(z) = 1 + 104z^3 + 649z^4 + 2976z^5$$
$$+ 10472z^6 + 27184z^7 + 50691z^8$$
$$+ 67616z^9 + 60952z^{10} + 33192z^{11}$$
$$+ 8307z^{12}. \tag{166}$$

$$A^{\{13,128,3\}}(z) = 1 + 5z^8 + 16z^9 + 24z^{10} + 16z^{11}$$
$$+ 2z^{12}, \tag{167}$$

$$B^{\{13,128,3\}}(z) = 1 + 138z^3 + 929z^4 + 4814z^5$$
$$+ 19592z^6 + 58628z^7 + 131987z^8$$
$$+ 219836z^9 + 263864z^{10}$$
$$+ 215954z^{11} + 107925z^{12}$$
$$+ 24918z^{13}. \tag{168}$$

$$A^{\{14,256,3\}}(z) = 1 + 2z^8 + 4z^9 + 18z^{10} + 28z^{11}$$
$$+ 11z^{12}, \tag{169}$$

$$B^{\{14,256,3\}}(z) = 1 + 180z^3 + 1295z^4 + 7436z^5$$
$$+ 34418z^6 + 117320z^7$$
$$+ 307391z^8 + 616280z^9$$
$$+ 923372z^{10} + 1007300z^{11}$$
$$+ 755921z^{12} + 348636z^{13}$$
$$+ 74754z^{14}. \tag{170}$$

F.2 Asymmetric codes

$$A^{\{6,2,d_e(\frac{1}{2})=2\}}(z) = 1 + 6z^2 + 9z^4 + 16z^6, \tag{171}$$

$$B^{\{6,2,d_e(\frac{1}{2})=2\}}(z) = 1 + 15z^2 + 8z^3 + 39z^4$$
$$+ 24z^5 + 41z^6; \tag{172}$$

$$A^{\{7,3,d_e(\frac{1}{2})=2\}}(z) = 1 + 1.111z^2 + 2.667z^3$$
$$+ 4.778z^4 + 13.333z^5$$
$$+ 15.444z^6 + 53.333z^7, \tag{173}$$

$$B^{\{7,3,d_e(\frac{1}{2})=2\}}(z) = 1 + 3.667z^2 + 24z^3 + 61.667z^4$$
$$+ 120z^5 + 125.667z^6 + 48z^7; \tag{174}$$

$$A^{\{5,2,d_e(2)=3\}}(z) = 1 + 4z^2 + 3z^4 + 8z^5, \tag{175}$$

$$B^{\{5,2,d_e(2)=3\}}(z) = 1 + 12z^2 + 10z^3 + 19z^4$$
$$+ 22z^5; \tag{176}$$

$$A^{\{6,4,d_e(2)=3\}}(z) = 1 + z^3 + 4z^4 + 7z^5 + 3z^6, \tag{177}$$

$$B^{\{6,4,d_e(2)=3\}}(z) = 1 + 7z^2 + 36z^3 + 75z^4 + 92z^5$$
$$+ 45z^6; \tag{178}$$

$$A^{\{7,8,d_e(2)=3\}}(z) = 1 + z^4 + 6z^5 + 6z^6 + 2z^7, \tag{179}$$

$$B^{\{7,8,d_e(2)=3\}}(z) = 1 + 9z^2 + 64z^3 + 179z^4$$
$$+ 312z^5 + 323z^6 + 136z^7; \tag{180}$$

$$A^{\{6,2,d_e(2)=4\}}(z) = 1 + z^2 + 11z^4 + 16z^5 + 3z^6, \tag{181}$$

$$B^{\{6,2,d_e(2)=4\}}(z) = 1 + z^2 + 24z^3 + 35z^4 + 40z^5$$
$$+ 27z^6; \tag{182}$$

$$A^{\{8,3,d_e(2)=4\}}(z) = 1 + 0.224z^2 + 1.770z^3$$
$$+ 4.919z^4 + 17.794z^5$$
$$+ 28.155z^6 + 23.103z^7$$
$$+ 8.368z^8, \tag{183}$$

$$B^{\{8,3,d_e(2)=4\}}(z) = 1 + 2.021z^2 + 18.612z^3$$
$$+ 68.016z^4 + 154.775z^5$$
$$+ 237.904z^6 + 210.612z^7$$
$$+ 75.058z^8. \tag{184}$$

F.3 Channel-adaptive codes

F.3.1 Nearest-neighbor collective amplitude damping

This part lists the quantum weight enumerators of the channel-adaptive codes for the one-dimensional nearest-neighbor collective amplitude damping errors discussed in Sec. 5.3.1.

$$A^{\{4,3\}}(z) = 1 + 0.111z + 1.222z^2 + 0.778z^3$$
$$+ 2.222z^4, \tag{185}$$

$$B^{\{4,3\}}(z) = 1 + 1.667z + 11.667z^2 + 17z^3$$
$$+ 16.667z^4; \tag{186}$$

$$A^{\{5,2\}}(z) = 1 + z + 2z^2 + 2z^3 + 5z^4 + 5z^5, \tag{187}$$

$$B^{\{5,2\}}(z) = 1 + z + 10z^2 + 18z^3 + 21z^4 + 13z^5; \tag{188}$$

$$A^{\{6,5\}}(z) = 1 + 1.307z^2 + 0.032z^3 + 3.890z^4$$
$$+ 0.097z^5 + 6.474z^6, \tag{189}$$

$$B^{\{6,5\}}(z) = 1 + 18.533z^2 + 32.162z^3 + 103.449z^4$$
$$+ 96.485z^5 + 68.371z^6; \tag{190}$$

$$A^{\{7,8\}}(z) = 1 + z^4 + 6z^5 + 6z^6 + 2z^7, \tag{191}$$

$$B^{\{7,8\}}(z) = 1 + 9z^2 + 64z^3 + 179z^4 + 312z^5$$
$$+ 323z^6 + 136z^7; \tag{192}$$

$$A^{\{8,9\}}(z) = 1 + 0.038z^2 + 0.124z^3 + 0.827z^4$$
$$+ 5.282z^5 + 10.035z^6 + 8.816z^7$$
$$+ 2.322z^8, \tag{193}$$

$$B^{\{8,9\}}(z) = 1 + 6.119z^2 + 57.119z^3 + 200.778z^4$$
$$+ 475.985z^5 + 713.867z^6 + 618.896z^7$$
$$+ 230.237z^8; \tag{194}$$

$$A^{\{9,16\}}(z) = 1 + z^5 + 8z^6 + 14z^7 + 7z^8 + z^9, \tag{195}$$

$$B^{\{9,16\}}(z) = 1 + 4z^2 + 80z^3 + 326z^4 + 936z^5$$
$$+ 1924z^6 + 2464z^7 + 1841z^8$$
$$+ 616z^9. \tag{196}$$

F.3.2 Nearest-neighbor collective phase-flips

This part lists the quantum weight enumerators of the channel-adaptive codes for the combined noise channel $\mathcal{N}$ (Eq. (68)) with hardware connectivity graphs shown in Fig. 7.

$$A^{\{a\}}(z) = 1 + z^2 + 11z^4 + 16z^5 + 3z^6, \tag{197}$$

$$B^{\{a\}}(z) = 1 + z^2 + 24z^3 + 35z^4 + 40z^5 + 27z^6; \tag{198}$$

$$A^{\{b\}}(z) = 1 + z^2 + 11z^4 + 16z^5 + 3z^6, \tag{199}$$

$$B^{\{b\}}(z) = 1 + z^2 + 24z^3 + 35z^4 + 40z^5 + 27z^6; \tag{200}$$

$$A^{\{c\}}(z) = 1 + 2z^3 + 9z^4 + 24z^5 + 22z^6 + 6z^7,$$
(201)

$$B^{\{c\}}(z) = 1 + 17z^3 + 45z^4 + 78z^5 + 82z^6 + 33z^7;$$
(202)

$$A^{\{d\}}(z) = 1 + 2z^3 + 9z^4 + 24z^5 + 22z^6 + 6z^7,$$
(203)

$$B^{\{d\}}(z) = 1 + 17z^3 + 45z^4 + 78z^5 + 82z^6 + 33z^7;$$
(204)

$$A^{\{e\}}(z) = 1 + 3.944z^4 + 12.112z^5 + 24z^6 \\ + 19.888z^7 + 3.056z^8,$$
(205)

$$B^{\{e\}}(z) = 1 + 27.888z^3 + 86.336z^4 + 215.776z^5 \\ + 319.776z^6 + 268.336z^7 + 104.888z^8;$$
(206)

$$A^{\{f\}}(z) = 1 + 3.983z^4 + 12.033z^5 + 24z^6 \\ + 19.967z^7 + 3.017z^8,$$
(207)

$$B^{\{f\}}(z) = 1 + 27.967z^3 + 86.100z^4 + 215.933z^5 \\ + 319.933z^6 + 268.100z^7 + 104.967z^8;$$
(208)

$$A^{\{g\}}(z) = 1 + 0.886z^3 + 3.282z^4 + 11.604z^5 \\ + 30.352z^6 + 45.220z^7 + 29.366z^8 \\ + 6.290z^9,$$
(209)

$$B^{\{g\}}(z) = 1 + 21.648z^3 + 78.704z^4 + 232.520z^5 \\ + 482.255z^6 + 618.352z^7 + 462.041z^8 \\ + 151.480z^9;$$
(210)

$$A^{\{h\}}(z) = 1 + 0.004z^3 + 3.996z^4 + 10.009z^5 \\ + 35.974z^6 + 44.004z^7 + 23.030z^8 \\ + 9.983z^9,$$
(211)

$$B^{\{h\}}(z) = 1 + 16.026z^3 + 89.948z^4 + 224.044z^5 \\ + 487.965z^6 + 623.974z^7 + 445.087z^8 \\ + 159.956z^9.$$
(212)