Techniques for learning sparse Pauli-Lindblad noise models

Ewout van den Berg and Pawel Wocjan

IBM Quantum, IBM Thomas J. Watson Research Center, Yorktown Heights, NY, USA

Error-mitigation techniques such as probabilistic error cancellation and zero-noise extrapolation benefit from accurate noise models. The sparse Pauli-Lindblad noise model is one of the most successful models for those applications. In existing implementations, the model decomposes into a series of simple Pauli channels with one- and two-local terms that follow the qubit topology. While the model has been shown to accurately capture the noise in contemporary superconducting quantum processors for error mitigation, it is important to consider higher-weight terms and effects beyond nearest-neighbor interactions. For such extended models to remain practical, however, we need to ensure that they can be learned efficiently. In this work we present new techniques that accomplish exactly this. We introduce twirling based on Pauli rotations, which enables us to automatically generate single-qubit learning correction sequences and reduce the number of unique fidelities that need to be learned. In addition, we propose a basis-selection strategy that leverages graph coloring and uniform covering arrays to minimize the number of learning bases. Taken together, these techniques ensure that the learning of the extended noise models remains efficient, despite their increased complexity.

1 Introduction

An important primitive in many quantum algorithms, such as variational quantum eigensolvers [1–3], is the accurate estimation of quantum observable expectation values. Until fault tolerance is achieved, these estimates will be affected by noise inherent in current quantum processors and the associated control systems. A practical way of reducing the effect of noise is the use of error-mitigation techniques. These techniques can be roughly divided into two complementary groups, depending on what part of the quantum circuit execution is considered. The first group of mitigation algorithms is aimed specifically at reducing state-preparation and measurement errors [4–6]. The second group mostly considers noise associated with the application of gates, and contains techniques such as zero-noise extrapolation [7–9] and probabilistic error cancellation [7]. In zero-noise extrapolation (ZNE), observable estimates are obtained at carefully controlled noise levels, which are then extrapolated to the point were gates are presumed noiseless. Meanwhile, in probabilistic error cancellation (PEC), noiseless operators are implemented as (quasi)-probabilistic mixtures of potentially noisy gates. This technique is leveraged in [10] to mitigate noise associated with Hermitian Clifford operators by learning a sparse Pauli-Lindblad noise model (discussed in more detail in the next section),

and subsequently applying the inverse noise map in a quasi-probabilistic manner to obtain unbiased estimates of the expectation value of desired observables. When noise models associated with the gates in a circuit are available they provide an ideal way of controlling the noise level in the context of ZNE [11–14]. Indeed, noise amplification based on the sparse Pauli-Lindblad noise model was used to great effect in this manner to mitigate noise in a 127-qubit Ising model in [15]. Aside from these applications, noise models can help characterize the performance of quantum processors and identify possible points of improvements.

In this work, we focus on techniques for learning extended versions of the sparse Pauli-Lindblad noise model proposed in [10] and reviewed in Section 2. We introduce Pauli-rotation twirling in Section 3. In Section 4 we show how rotation twirling can be used in shaping the noise and the design of noise-learning sequences. Both techniques are leveraged in Section 5 to improve learning protocols through optimized basis selection. We provide concluding remarks in Section 6.

2 Sparse Pauli-Lindblad noise models

Noisy quantum gates $\tilde{\mathcal{U}}$ can be modeled as an ideal operator $\mathcal{U}$ preceded by a noise channel $\tilde{\Lambda}$. By restricting $\mathcal{U}$ to represent a Clifford operator, we can efficiently implement Pauli twirling [16–20], which turns noise channels into Pauli channels. We can therefore assume that $\tilde{\Lambda}$ is a Pauli channel, namely $\tilde{\Lambda}(\rho) = \sum_i \alpha_i P_i \rho P_i$ where P_i represents a Pauli operator. A general n-qubit Pauli channel is characterized by 4^n coefficients, so, clearly, the representation and learning of the full channel does not scale well in the number of qubits. Moreover, even if there are only a small number of non-zero α_i parameters, the inverse noise map generally still requires the evaluation and inversion of all 4^n Pauli fidelities[1] $f_b = \frac{1}{2^n}\mathrm{Tr}(P_b\tilde{\Lambda}(P_b))$.

The Pauli-Lindblad noise model introduced in [10] is an alternative model that is more scalable and comes with convenient properties. The model is defined as $\Lambda(\rho) = \exp[\mathcal{L}]\rho$, where $\mathcal{L}(\rho) = \sum_{k\in\mathcal{K}} \lambda_k(P_k\rho P_k^\dagger - \rho)$ is a Lindbladian with Pauli jump operators $\{P_k\}_{k\in\mathcal{K}}$. It can be equivalently expressed as the composition of a series of simple Pauli channels:

$$\Lambda(\rho) = \bigcirc_{k\in\mathcal{K}} \left(w_k \cdot + (1-w_k) P_k \cdot P_k^\dagger\right)(\rho),$$

with $w_k = (1+e^{-2\lambda_k})/2$ and $\lambda_k \geq 0$. The expressivity of the model depends on the choice of $\mathcal{K}$, but always remains a strict subset of all possible Pauli channels. By restricting $\mathcal{K}$ to the set of all one- and two-local Pauli terms following the qubit topology, we obtain a sparse Pauli-Lindblad noise model [10].

Learning of the model parameters $\lambda = \{\lambda_k\}_{k\in\mathcal{K}}$ makes use of the following relation to Pauli fidelities:

$$f_b = \tfrac{1}{2^n}\mathrm{Tr}(P_b\Lambda(P_b)) = \exp\left(-2\sum_{k\in\mathcal{K}} \lambda_k \langle b,k\rangle_{\mathrm{sp}}\right), \tag{1}$$

where $\langle b,k\rangle_{\mathrm{sp}}$ is the symplectic inner product between Paulis P_b and P_k, which has a value of 0 when the terms commute and 1 otherwise. For a vector $f = \{f_b\}_{b\in\mathcal{B}}$ of measured fidelities and binary matrix $M(\mathcal{B},\mathcal{K})$ with entries $\langle b,k\rangle_{\mathrm{sp}}$ it follows from Eq. (1) that $M(\mathcal{B},\mathcal{K})\lambda =$

[1] When $\tilde{\Lambda}$ is a Pauli channel, it holds that $\tilde{\Lambda}(P_b) = f_b P_b$ for all Pauli operators P_b. The scalars f_b are therefore commonly referred to as the eigenvalues of the channel.

$-\log(f)/2$, where λ is the vector of model parameters $\{\lambda_k\}_{k\in\mathcal{K}}$. Learning the noise model parameters based on a vector of estimated fidelities $\hat{f}$ can therefore be done using nonnegative least-squares optimization:

$$\underset{\lambda\geq 0}{\text{minimize}} \quad \tfrac{1}{2}\|M(\mathcal{B},\mathcal{K})\lambda + \log(\hat{f})/2\|. \tag{2}$$

The solution of this problem is unique whenever $M(\mathcal{B},\mathcal{K})$ is full column rank, and a convenient result in this context is the following:

Theorem 1 ([10, Theorem SIV.1])**.** *For non-empty support sets $\mathcal{S}_i \subseteq [n]$, denote by $\mathcal{P}_i$ all n-qubit non-identity Paulis supported on $\mathcal{S}_i$, and their union by $\mathcal{P} = \bigcup_i \mathcal{P}_i$. Then $M(\mathcal{P},\mathcal{P})$ is full rank.*

From this we conclude that for the above two-local noise model, it suffices to choose $\mathcal{B} = \mathcal{K}$. Adding more terms to $\mathcal{B}$ does not affect the column rank, but does result in an overcomplete set of equations, which can help prevent overfitting when using noisy fidelity estimates $\hat{f}$.

Estimating the fidelities of Pauli channels can be done using several protocols, for instance [21–24]. To understand how these work, note that we can express the initial quantum state ρ_0 as a weighted sum of Pauli terms $\frac{1}{2^n}\sum_i \alpha_i P_i$. In particular, the zero state $|0\rangle\langle 0|$ can be written as the sum of all Pauli-Z terms, each with $\alpha_i = 1$. The Pauli transfer matrix of any Clifford operator is a signed permutation matrix, while that of a Pauli channel is a diagonal matrix whose entries correspond to the channel Pauli fidelities. Since Pauli terms never split into multiple terms under these transformation, we can track individual Pauli terms and their associated fidelity as successive noisy Clifford gates are applied. Repeated application of a noisy Clifford operator $\tilde{\mathcal{U}}$ therefore leads to a product of fidelities that depends on the transformation of the Pauli term. Whenever the number of applications of the operator, or depth, k is such that $\mathcal{U}^k(P) = P$ we can learn the fidelities free from any state-preparation and measurement (SPAM) errors [4, 22]. To keep learning manageable, we must avoid products of many fidelities and therefore require $\mathcal{U}$ to be Hermitian; that is $\mathcal{U}^2 = I$. More specifically, we allow $\mathcal{U}$ to be a *layer* of simultaneous single- and two-qubit Hermitian Clifford gates. For such layers we can estimate products of at most two unique fidelities, say $(f_1 f_2)^{k/2}$, for any even depth k. The resulting fidelities can then be found by fitting an exponential curve through the fidelity estimates at different depths [21].

With appropriate basis changes, this method allows us to measure the fidelity pair $f_i f_j$ for any Pauli pair P_i and $P_j = \mathcal{U}(P_i)$. However, ideally we would measure individual fidelities, rather than pairs. Whenever the support of P_i and P_j is the same we can interleave applications of $\tilde{\mathcal{U}}$ with appropriate single-qubit correction gates to map P_j back to P_i. For instance, we can change a Pauli X term to a Pauli Z term by applying a single-qubit Hadamard gate. We consider such correction sequences in Section 4.2. Since the identity is invariant under conjugation by any unitary, it is not possible to change an identity term to a non-identity term, and vice versa. In other words, it is not possible to change the support of a Pauli by means of single-qubit gates, and it was shown in [25] that fidelity pairs corresponding to Paulis with different supports cannot be learned in a SPAM-free manner. In such cases, we can write $(\alpha f_i)(f_j/\alpha) = f_i f_j$ for any $\alpha \neq 0$ and there is therefore no unique way to extract the individual fidelities from their product.

As a practical solution, a *symmetry assumption* was introduced in [10], which takes f_i and f_j to be equal whenever disambiguation is not possible. With this we can learn f_i by

measurements in a Pauli basis that contains P_i or its corresponding term P_j. Although measurements obtained from circuits in a single fixed basis can be used to estimate exponentially many fidelities, not all fidelities in our target set $\mathcal{B}$ can be learned in a single basis. It is therefore natural to consider the minimal set of bases from which we can learn all desired fidelities. For the two-local noise model, it was shown in [10] that measurements in nine different bases suffice for the qubit topologies under consideration. In Section 5 we present an efficient basis-selection algorithm that applies for our extended noise models and includes the existing learning algorithm as a special case.

3 Pauli rotation twirls

Twirling is a technique that is used to shape noise channels. Pauli twirling of a noise channel $\tilde{\Lambda}$ associated to a noisy n-qubit Clifford operator $\tilde{\mathcal{O}} = \mathcal{O} \circ \tilde{\Lambda}$, with $\mathcal{O}(\rho) = O\rho O^\dagger$ is given by

$$\mathbb{E}_{i\in\mathcal{I}} \left[O P_i^\dagger \tilde{\Lambda}(P_i \rho P_i^\dagger) P_i O^\dagger \right], \tag{3}$$

where $\mathcal{I}$ represents the n-qubit Pauli group, or an appropriate subset thereof [26]. Since the noise of interest appears in the context of a gate, rather than by itself, we cannot just apply Pauli operators prior to and following the noise channel. Instead, the circuit implementation of a given twirl element P_i relies on the identity $OP_i^\dagger = \sigma_i Q_i^\dagger O$, which follows from the $OP_iO^\dagger = \sigma_i Q_i$ for some Pauli Q_i with sign σ_i. Using this identity we can rewrite Eq. (3) in terms of the noisy operator $\tilde{\mathcal{O}}$ and obtain the equivalent expression

$$\mathbb{E}_{i\in\mathcal{I}} \left[Q_i^\dagger \tilde{\mathcal{O}}(P_i \rho P_i^\dagger) Q_i \right],$$

which shows that Paul twirling can be implemented in practice.

3.1 Pauli rotation twirl

We now generalize the Pauli twirl to a twirl based on Pauli rotation operators. This twirl will help reduce the number of unique fidelities and thereby simplify the noise learning process. The rotation operator $R_P(\theta)$ for Pauli operator P and angle θ is defined as

$$R_P(\theta) = \exp\left(-i\tfrac{\theta}{2}P\right) = \cos(\theta/2)I - i\sin(\theta/2)P. \tag{4}$$

The corresponding twirl, with fixed θ for simplicity, is then given by

$$\mathbb{E}_{j\in\mathcal{J}} \left[O R_{P_j}(-\theta) \tilde{\Lambda}(R_{P_j}(\theta) \rho R_{P_j}(-\theta)) R_{P_j}(\theta) O^\dagger \right], \tag{5}$$

where $\mathcal{J}$ represents (a subset of) the Pauli group. Note that setting $\theta = \pi$ in Eq. (4) yields the identity $R_P(\pi) = (-i)P$. Therefore, applying the rotation twirl with $\theta = \pi$ reduces to the standard Pauli twirl. For the implementation of the twirl, we again need to conjugate the twirl operators by O, which satisfies

$$O R_{P_j}(\theta) O^\dagger = O \exp\left(-i\tfrac{\theta}{2}P_j\right) O^\dagger = \exp\left(-i\tfrac{\theta}{2} O P_j O^\dagger\right) = \exp\left(-i\sigma_j \tfrac{\theta}{2} Q_j\right) = R_{Q_j}(\sigma_j \theta). \tag{6}$$

Post-multiplying by O then gives $OR_{P_j}(\theta) = R_{Q_j}(\sigma_i\theta)O$, which allows us to rewrite Eq. (5) as

$$\mathbb{E}_{j\in\mathcal{J}} \left[R_{Q_j}(-\sigma_j\theta)\ \tilde{\mathcal{O}}(R_{P_j}(\theta)\rho R_{P_j}(-\theta))\ R_{Q_j}(\sigma_j\theta) \right]. \tag{7}$$

Rotation	$\theta = \pi$	$\theta = \pi/2$	$\theta = \pi/4$
$R_X(\theta)$	$(-i)X$	$\left(\frac{1-1j}{\sqrt{2}}\right) S_X$	$\left(e^{-i\pi/8}\right) HTH$
$R_Y(\theta)$	$(-i)Y$	HZ	$\left(e^{-i\pi/8}\right) S_X^\dagger T S_X$
$R_Z(\theta)$	$(-i)Z$	$\left(\frac{1-1j}{\sqrt{2}}\right) S$	$\left(e^{-i\pi/8}\right) T$

Table 1: Expression of single-qubit Pauli rotations in terms of basic gates for three specific θ values. The operations corresponding to rotations by $-\theta$ are given by the adjoint of the listed operations. For instance, $R_Z(-\pi/4) = (e^{i\pi/8})T^\dagger$ and $R_Y(-\pi/2) = ZH$.

Although the twirl is now expressed in terms of the noisy operator $\tilde{\mathcal{O}}$, implementing general rotations for arbitrary Pauli terms P_j and angles θ may require non-trivial circuits that contain two-qubit gates, which themselves may be noisy. We may therefore want to limit $\mathcal{J}$ to a subset of Pauli operators for which the P_j and Q_j rotation operations can be implemented using single-qubit rotations, which are readily implemented on contemporary quantum processors. Since Pauli operators decouple into single-qubit Pauli terms, this holds trivially for all $\mathcal{J}$ in the special case of Pauli twirls. For all other values of θ, it suffices to find values of j for which both P_j and Q_j have weight one. For particular angles θ we can then further simplify the implementation and express the rotation in terms of basic single-qubit gates, as summarized in Table 1.

We now show that for each two-qubit Hermitian Clifford gate O, there always exists a two-qubit Pauli of the form $PI = P \otimes I$ such that its corresponding Pauli $\sigma O(PI)O^\dagger$ is also weight-one, and likewise for a Pauli of the form IP. This demonstrates that non-trivial Pauli rotation twirling exists for these gates for any angle θ. To show that a two-qubit Pauli of the form PI exists, suppose by contradiction that XI, YI, and ZI all map to weight-two Paulis. Because conjugation by a Clifford operator preserves commutation relations, these new Paulis mutually anticommute and must therefore be of the form $S \times \{X, Y, Z\}$ or $\{X, Y, Z\} \times S$. Any remaining weight-two Pauli operator can be verified to anticommute with one of these new operators. Since Pauli operators IX, IY, and IZ commute with all Pauli operators of the form PI, they therefore cannot map to any weight-two Pauli, and must retain their original support. But even with the same support at least one will anticommute with the image of XI, YI, and ZI, which contradicts the definition of Clifford operators. Using an analogous derivation, we can show a similar result for Paulis of the form IP.

3.2 Classes of two-qubit Hermitian Clifford operators

For the remainder of the paper it will be convenient to have a clear characterization of the changes in the support of Pauli operators following conjugation by two-qubit Hermitian Clifford operators. In particular, this classification enables us to determine suitable weight-one Pauli pairs, thereby extending the existence proof from Section 3.1.

Transitions in support happen when an I term changes to an $\{X, Y, Z\}$ term, or vice versa. As an example, consider the conditional-Z (CZ) gate. It maps XI to XZ changing the support, whereas it maps XX to YY maintaining the original support.

We show that Hermitian two-qubit Clifford operators can be partitioned into four classes depending on how their mapping affects the support of Pauli operators. These classes are illustrated in Figure 1, which groups Paulis of the form $\{I, A, B, C\} \otimes \{I, D, E, F\}$ in boxes,

based on their support, where $\{A, B, C\}$ and $\{D, E, F\}$ represent arbitrary permutations of $\{X, Y, Z\}$. We will use three basic properties:

1. For any Clifford operator O, the conjugation of the product of two Paulis is equal to the product of their individual conjugations: $O(P \cdot Q)O^\dagger = OP(O^\dagger O)QO^\dagger = (OPO^\dagger)(OQO^\dagger)$.

2. Commutation relations between Paulis are preserved under conjugation by Clifford operators: if Paulis P and Q commute then so do $OPO^\dagger$ and $OQO^\dagger$.

3. Hermiticity of O implies that whenever P maps to $Q = OPO^\dagger$ then Q maps to P.

With this, we can characterize the four different classes. Throughout the analysis we ignore phase factors in expressions such as $X \cdot Y = Z$.

Class 1 If AI and BI map to Paulis of the form PI and QI, then it follows from property 1 that $CI = (AI) \cdot (BI)$ maps to $(P \cdot Q)I$ and therefore also does not change support. It further follows from property 2 that ID cannot change support, otherwise it would anticommute with one of PI, QI, or $(P \cdot Q)I$. The same applies to IE and IF. This setting occurs whenever O decouples into two single-qubit operators, i.e., $O = O_1 \otimes O_2$. At least one Pauli in each of the top and right weight-one boxes must map to itself; cyclic changes between three terms would imply that $O^2 \neq I$ and therefore contradict Hermiticity of O.

Class 2 Similar to class 1, suppose that AI and BI respectively map to IF and IE, then it follows from property 1 that CI must map to ID, and that all weight-two terms retain their support. Using properties 1 and 3 we find that $AF = (AI) \cdot (IF)$ maps to $(IF) \cdot (AI) = AF$. Applying the same reasoning to the other pairs, we find that BE and CD also map to themselves.

Class 3 Suppose CI maps to PI and AI maps to ST, then, using property 1, we find that BI must map to $(S \cdot P)T$, and therefore also changes support. Hermitian Clifford operators map one Pauli to another and vice versa. Since AI and BI already map to other Pauli terms, we conclude that CI must map to itself, that is, $P = C$. In order for ST and $(S \cdot P)T$ to anticommute with CI, both S and $S \cdot P$ must differ from C, which means that $\{AI, BI\}$ maps to $\{AT, BT\}$ in some order. The terms in the bottom-right box (ID, IE, and IF) must map to Paulis that commute with AT and BT. This excludes terms in the 3×3 weight-two box that are in the same row or column. This leaves exactly two feasible elements and exactly one of ID, IE, IF must therefore retain its support (if multiple elements retain their support, class 1 would apply). Assume this element is ID, then following the same argument as for CI, we conclude that ID must map to itself. Since both CI and ID map to themselves, so does their product CD. Finally, note that the unspecified term T above must be equal to D, since that row is excluded for the two terms from the bottom-right box that change support, as illustrated in Figure 1.

Class 4 Suppose CI maps to ID and that AI maps to weight-two Pauli PQ. Following property 1, $BI = (AI) \cdot (CI)$ maps to $P(Q \cdot D)$, and $CD = (CI) \cdot (ID)$ maps to itself. From property 2 we further have that PQ and $P(Q \cdot D)$ must anticommute with ID, which means

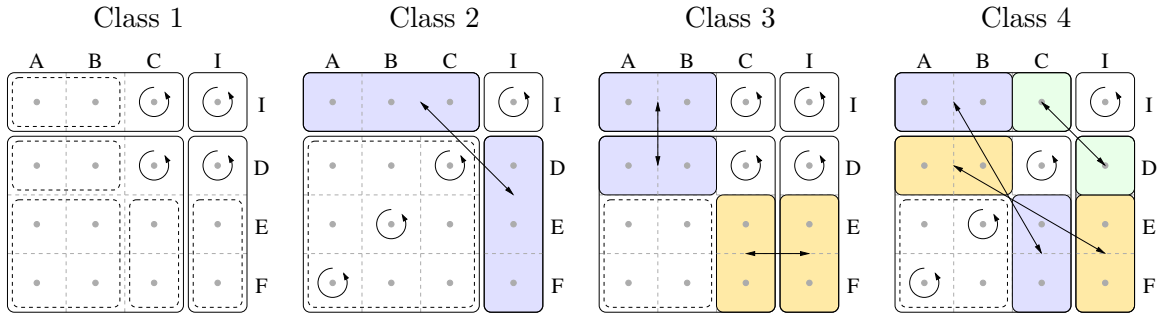

Figure 1: Pauli support transitions for four classes of two-qubit Hermitian Clifford operators, where $\{A, B, C\}$ and $\{D, E, F\}$ are arbitrary permutations of $\{X, Y, Z\}$. Column (row) labels correspond to the Pauli on the first (second) qubit. Pauli terms within each of the dashed regions remain in those regions after conjugation. The cyclic arrow indicates mapping of a Pauli to itself.

that both Q and $Q \cdot D$ both differ from D, so without loss of generality we can set $Q = E$. Since ID and AI commute, so do their images CI and PQ, which implies that $P = C$. We conclude that AI maps to $PQ = CE$. Applying properties 1 and 3 on both terms we find that $(AI) \cdot (CE) = BE$ maps to itself, and likewise for AF through multiplication by CD. Finally, combining the mappings for CI, CE, and ID we find that IE and IF respectively map to AD and BD.

Excluding other classes The only possible class remaining is one in which all weight-one Pauli terms in the top box map to weight-two Paulis. In order for the resulting Paulis to satisfy property 2, they should all mutually anticommute and therefore lie in a single row or column of the 3×3 weight-two box. As a result, none of the remaining elements in the box commutes with all three of these Paulis. In order to satisfy property 2, the weight-one Pauli terms in the right box must therefore all map to Paulis within the same box. But this would imply case 1 and therefore leads to a contradiction. We conclude that the four classes discussed above capture all weight-two Hermitian Clifford operators.

4 Applications of the Pauli rotation twirl

In preparation for basis selection, discussed in Section 5, we now consider two applications of the Pauli rotation twirl. The first application shapes Pauli channels by averaging certain fidelities and relies on nested twirling with different twirl sets $\mathcal{J}$. Doing so reduces the number of unique Pauli fidelities that need to be learned and allows us to reduce the number of measurement bases used to learn the sparse Pauli-Lindblad noise model. The second application concerns the design of single-qubit correction sequences in learning circuits to enable extraction of individual channel fidelities. In both applications we fix $\theta = \pi/2$.

4.1 Rotation twirls for fidelity averaging

As seen from Table 1, weight-one Pauli X, Y, and Z rotations with angle $\theta = \pi/2$ result respectively in S_X, HZ and S operations, where S_X is the square root of X, H is the Hadamard gate, and S is the phase gate. These operations are easy to implement and have

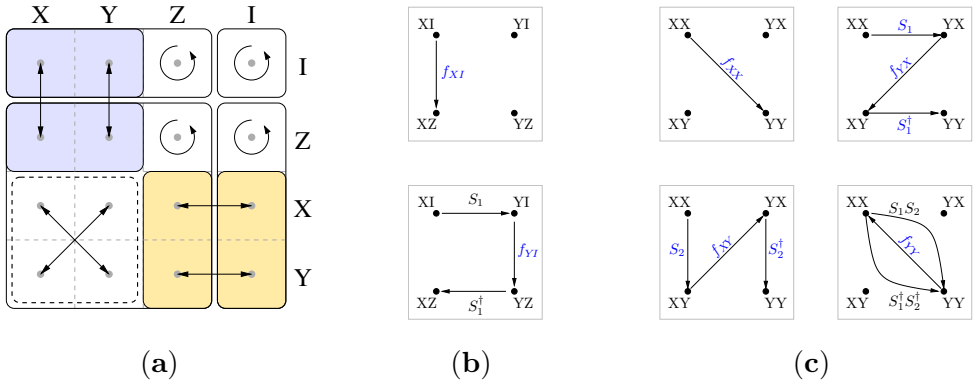

Figure 2: Illustration of (**a**) the conjugation of Pauli terms under class-3 operator CZ; (**b**) two trajectories of moving from XI to XZ when applying a phase twirl $\{S_1, I_1\}$ on qubit one. The effective transition fidelity resulting from the twirl is given by the average fidelity $(f_{XI}+f_{YI})/2$; (**c**) four trajectories of moving from XX to YY when applying phase twirls on both qubits. Here too the effective transition fidelity is given by the average of the path fidelities, assuming that single-qubit gates are noiseless.

the special property that their application permutes single-qubit Pauli terms. In particular, ignoring signs, conjugation with HZ exchanges Pauli X and Z terms, S exchanges X and Y, and S_X exchanges Y and Z terms.

We now explain how this property can be used to average certain fidelities in a noisy CZ gate, where it is assumed that the noise channel $\tilde{\Lambda}$ preceding the ideal CZ gate has already been shaped to a Pauli channel. For reference, we illustrate the Pauli transitions for the CZ gate in Figure 2(a).

Choosing a twirl set $\mathcal{J}_1 = \{II, ZI\}$ amounts to twirling the noisy CZ gate by Pauli rotations $\{I, R_Z(\frac{\pi}{2})\}$ on the first qubit. The operation resulting from the twirl is given by the expectation in Eq. (5), which is taken over the two values $j \in \mathcal{J}_1$. For $j = II$, all rotations reduce to the identity and the noise channel $\tilde{\Lambda}$ applies directly to XI, which results in fidelity of f_{XI}. Application of the ideal CZ gate then changes XI to XZ. On the other hand, for $j = ZI$, we start by applying a Pauli-Z rotation, which is equivalent to a phase gate and therefore changes XI to YI. The noise channel then incurs a fidelity term f_{YI}, but leaves the Pauli term itself unchanged. The ideal CZ gate and the second rotation gate that follows then successively map YI to YZ and XZ, which is the same final Pauli term as before (see the illustration in Figure 2(b)). To see why, observe that Eq. (6) can be rewritten as

$$R_{Q_j}(-\sigma_j\theta)OR_{P_j}(\theta) = O. \tag{8}$$

We conclude that twirling over $\mathcal{J}_1$ therefore results in a mapping from XI to XZ under a new noisy CZ gate with an expected fidelity of $(f_{XI} + f_{YI})/2$. The same applies for YI, and the additional twirling therefore has the effect of averaging fidelities f_{XI} and f_{YI}.

Applying twirl gates on both qubits, that is, adding a second twirl set $\mathcal{J}_2 = \{II, IZ\}$, does not change the above result, but does result in the averaging of the fidelities for Pauli terms IX and IY, as well as for XX, XY, YX, and YY, as shown in Figure 2(c). More generally, we see from the representation of class-3 gates in Figure 1 that twirling using Pauli CI and ID rotations effectively amounts to averaging the fidelities of corresponding terms in the first two columns as well as those in the last two rows. The same applies for Clifford operators in class 4.

The SWAP gate is an example of a class-2 operator where weight-one Paulis supported on the first qubit all map to weight-one terms supported on the second qubit, and vice versa due to Hermiticity. That means that there are now three possible rotations R_{P_j} we can apply on the first qubit, each resulting in the averaging of two of the Pauli terms. It may seem that nesting the twirls can help mix all terms, but doing so generates a total of $2^3 = 8$ paths through the noisy operator, which implies that the averaging cannot be uniform, since 3, the number of non-identity Pauli terms, does not divide 8, the number of elements in the twirl set. Instead, we can simply set $\mathcal{J} = \{X_1, Y_1, Z_1\}$ and twirl over $\{R_{X_1}, R_{Y_1}, R_{Z_1}\}$, which maps each single-qubit Pauli term to each of the three Pauli terms with equal probability. A similar twirl can be applied on the second qubit. The effect of applying this twirl on class-2 operators thus amounts to averaging the fidelities of corresponding elements in the first three columns and the last three rows in Figure 1, thus fully mixing the fidelities in the weight-two block. For class 1 it is possible to have one or three weight-one terms supported on the first qubit to map to weight-one terms. In case there is only one we average the fidelities in the other two columns, otherwise we can average the elements in the first three columns. The same averaging across rows applies for the weight-one terms in the right block, depending on the number of terms that remain weight one following conjugation. Since class-1 operators correspond to tensor products of two single-qubit Clifford operations, it immediately follows that the same techniques can be used for single-qubit gates.

In order to automatically determine the possible rotation twirls for a given Hermitian two-qubit Clifford, we can first classify the operator by conjugating all six weight-one Pauli terms. Depending on the resulting weights, and the terms for which the weight changes we can then easily determine the appropriate twirl.

Various twirl groups have been proposed in the literature, see for example [24, 27, 28], and it is natural to ask whether such twirls can further reduce the number of unique Pauli fidelities through averaging. The short answer is that this is indeed possible. However, implementing such twirls in the context of two-qubit gates may no longer be possible using only single-qubit gates. Inserting one or more two-qubit gates to twirl the noise associated with a single two-qubit gate clearly introduces disproportional levels of noise to the circuit and significantly complicates learning. Similar difficulties arise when considering twirling noise in the context of non-Clifford gates [29].

4.2 Extracting individual fidelities

As explained in Section 2, we can learn the fidelities of a Pauli noise channel associated with a Hermitian Clifford operator by applying the noisy operator an even number of times and taking measurements in different bases. Due to the presence of the Clifford gate, those learning sequences typically only allow us to learn pairwise products of certain fidelities. For instance, for a noisy CZ gate, starting with a Pauli XX term, we can learn the fidelity pair $f_{XX}f_{YY}$, as illustrated in Figure 3(a). In this case, application of the CZ gate changes XX to YY, which preserves the support and can be mapped back to the original Pauli by means of single-qubit $R_Z(\pi/2)$ operators or, equivalently, phase gates (see also [10, 25]). The insertion of these correction gates is shown in Figure 3(b) along with the changes of Pauli terms and fidelities. Although such a transformation is clearly always possible to correct terms that retain their support, it is worth asking if doing so affects the learnability of other terms in the same basis, for instance XI or IX. Care needs to be taken as well not to introduce inadvertent

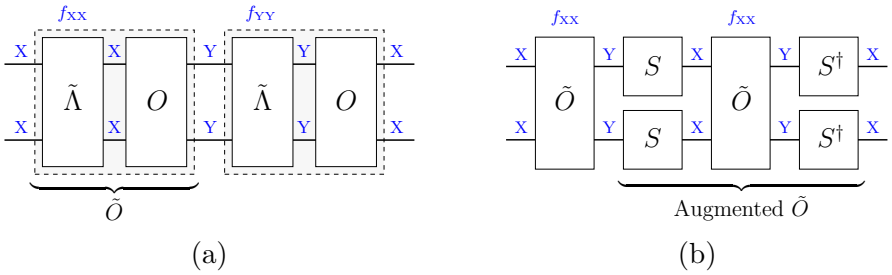

Figure 3: The noise channel of a Hermitian Clifford gate can be learned by repeated application of the gate. In (a) we trace the evolution a Pauli XX operator through two instances of a noisy gate $\tilde{O}$ implementing a CZ gate affected by a Pauli noise channel $\tilde{\Lambda}$. Applying the noise channel on a Pauli operator incurs a multiplicative fidelity term corresponding to that Pauli, in that case resulting in an overall fidelity of $f_{XX}f_{YY}$. To learn f_{XX} up to a sign, we can (b) insert a correction sequence of single-qubit gates, mapping YY back to XX after each application of $\tilde{O}$. This process can equivalently be interpreted as augmenting the second $\tilde{O}$ with a fixed twirl sequence.

sign changes in some of the fidelity terms. The learning sequence in Figure 3(b) can also be interpreted as applying two instances of the twirled noisy gate, one with a singleton twirl set $\mathcal{J}_1 = \{(I_1, I_2)\}$, and one $\mathcal{J}_2 = \{(Z_1, Z_2)\}$, where, with some abuse of notation, we use tuples to indicate a pair of single-qubit rotations. We see from Eq. (8) that this construction ensures that no spurious sign changes occur, but the question then changes to whether or not we can always implement the desired correction sequences using these singleton twirls.

Given an arbitrary basis ST with $S, T \in \{X, Y, Z\}$, our proposed approach is to insert single-qubit correction gates whenever ST maps to some other weight-two Pauli. This clearly works for Clifford operators in classes 1 and 2, since all weight-one terms retain their weight, allowing us to insert any single-qubit rotation operator. For classes 3 and 4, it can be seen from Figure 1 that we only ever need to exchange the terms in pairs (A, B) and (E, F). This can be done respectively using Pauli rotations CI and ID, both of which map to weight-one Paulis and are therefore feasible candidates for the (nested) twirl set. The lack of support changes for weight-one terms for operators in class 1 then allows us to conclude that the proposed construction can indeed generate all required correction sequences.

As seen from Figure 2(b), the insertion of correction sequences can affect the fidelity pairs that arise from an initial weight-one Pauli operator. Since the symmetry condition was assumed to hold only for Pauli pairs P_i and $Q_i = \sigma_i O P_i O^\dagger$, we can ask if the learning sequences arising from the proposed approach still allow us to learn the original fidelity pair $f_{P_i} f_{Q_i}$. In other words, can we still learn this fidelity pair in one of the nine two-qubit Pauli bases? To answer this question, we consider the four different classes in turn. The question does not apply for operators in class 1, since all Pauli terms maintain their original support. For class 2, we can assume that AI maps to IF, which implies that AF must map to itself. When working in the AF basis we therefore do not insert any correction gates, which means we can measure the fidelity pair for AI and IF in this basis. The same applies for all (weight-one) terms that change support as well as for Paulis CI and ID in class 4. What remains are the weight-two Pauli operators in classes 3 and 4 that map to weight-one Pauli operators, and vice versa. Clearly, since we only insert single-qubit correction gates for weight-two Paulis that retain their weight, we do not insert any such gates here and therefore always measure the original fidelity pairs in these bases. For class 4, we alternatively observe that for each row

and column in the weight-two box, there exists a weight-two Pauli that maps to itself. For any Pauli SI or IT we can therefore find a Pauli basis in the same row or column for which the circuit is run without correction gates, and for which we therefore measure the original fidelity pairs. We conclude that we can measure the original fidelity pairs in at least one basis for Paulis that change their support.

5 Basis selection using graph coloring and covering arrays

We now consider the selection of state preparation and measurement Pauli bases that allow us to measure all fidelities for our benchmark set $\mathcal{B}$, which then provides all the information needed to determine the sparse Pauli-Lindblad noise model parameters. To better explain the intuition behind our methods for basis selection, we begin with a simple setting. We consider a noisy layer of identity gates, for which fidelities can be measured individually rather than in pairs and equate $\mathcal{B}$ to the set of terms $\mathcal{K}$ for a two-local noise model. We then successively discuss the changes and considerations that apply for more general settings.

5.1 Basis selection for two-local models

In the most basic setting of a two-local noise model, Theorem 1 guarantees that it suffices to measure the fidelities of all weight-one Paulis on the model qubits, and all weight-two Paulis on neighboring qubits. For this it suffices to select a series of bases such that all nine $\{X, Y, Z\} \times \{X, Y, Z\}$ weight-two Pauli terms occur at least once for each connected pair of qubits.

We now leverage the combinatorial construction called *covering array* for basis selection. The covering array $\mathrm{CA}(N; t, k, v)$ is a N-by-k matrix with elements from an alphabet of size v, such that each N-by-t submatrix contains each of the v^t possible t-tuples of symbols in at least one row [30]. The covering array number $\mathrm{CAN}(t, k, v)$ is the smallest N for which $\mathrm{CA}(N; t, k, v)$ exists. In order to use these covering arrays, we start the basis-selection protocol by generating a graph with vertices corresponding to the qubits in the noise model and edges for the neighboring qubits, which can include virtual connections that indicate expected crosstalk terms between the given pairs of qubits. Once the graph has been generated we apply graph coloring and find a covering array $\mathrm{CA}(N; 2, k, 3)$ of strength $t = 2$, where k matches the number of colors and the symbols of the alphabet are set to $\{X, Y, Z\}$ (for more information on covering arrays of order three with strength two, see [31]). By mapping each of the colors to a column in the covering array we can then construct N bases. Each basis is generated by a single row of the array, which prescribes the mapping of a color to a Pauli term. The Pauli basis string can then be formed by looking up the terms associated with the color assigned to qubits in the graph coloring. We summarize the different steps in Figure 4.

An equivalent approach for basis selection for two-local models was proposed independently in [32], but without making the connection to covering arrays. Perfect hash families, which are related to covering arrays [33], were used for basis selection in [34] to measure all k-local Pauli observables. A randomized approach to basis selection in this context, aimed at reducing the number of measurements rather than the number of unique bases, was proposed in [35] and derandomized in [36].

5.2 Reduction of the basis elements

When noise occurs in the context of general layers of Hermitian Clifford gates, noise channel fidelities typically occur as the product $f_i f_j$ of the fidelities corresponding to Paulis P_i and P_j. For convenience we refer to P_i and P_j as the Pauli pair or Pauli terms that appear in the fidelity, and $f_i f_j$ the fidelity or fidelity pair for the Pauli pair. Although it is not always possible to uniquely determine the individual fidelities, there are some cases, as discussed in Section 4.2, where it is nevertheless possible to do so by inserting correction sequences in the learning circuits. These sequences are aimed at correcting Pauli terms that maintain their support following conjugation by the gates in the layer, but one side effect of these sequences is that they also affect the terms that appear in other fidelity pairs. If these terms change support we may require strengthening the symmetry condition to apply to all such pairs as well. We partially addressed this in Section 4.2, where we showed that for one-local terms P there is always at least one basis in which we can learn the original fidelity pair, which would then allow resolution of the individual fidelities as a consequence of the symmetry condition. This result does not generally extend to Pauli terms that overlap with multiple two-qubit gates. For instance, consider a weight-two Pauli string P_1P_2 such that each term P_i is supported on a different two-qubit gate. If application of the gates does not change the support, and therefore maps the Pauli to some weight-two Pauli Q_1Q_2, then for each $i \in \{1, 2\}$ there is at least one basis for gate i for which Pauli P_i maps to the original term Q_i. However, since the bases for the gates may be chosen independently, there is no guarantee that we select a pair of bases for which the original Pauli pair appears simultaneously, that is, for which P_1P_2 maps to Q_1Q_2. Those cases either require additional learning bases or stronger symmetry assumptions.

Combining Pauli twirls with the rotational twirls, introduced in Section 4.1, allows us to do away with correction sequences altogether: rather than correcting the Pauli terms that appear in each fidelity pair, we shape the noise channel such that certain fidelities are averaged and become equivalent. Whenever several fidelities are equivalent it suffices to measure any one of them to know them all, which allows us to reduce the number of measurement bases. Consider for example the classes of two-qubit gates in Figure 1. For classes 1 and 2, the fidelities for all Paulis with the same support are averaged. It therefore suffices to take measurements in, say, the CD basis (where C and D are mapped to appropriate Pauli terms), which includes terms for all three support CI, ID, and CD. The rotation twirl set for classes 3 and 4 is smaller and therefore leads to averaging of smaller subsets of fidelities. Nevertheless, even under the weak symmetry assumption, we can still measure all desired fidelities by measuring for instance only in the CD and AF bases. In this case, measuring the fidelities for Pauli terms that are supported on multiple gates does not require any special treatment. The implication of these simplifications is that we can merge any pair of vertices in the graph whose qubits are subject to the same two-qubit gate. The number of basis elements or symbols required for the resulting vertex is one for gates from classes 1 or 2, and two for classes 3 and 4. Vertices corresponding to idle qubits or qubits on which single-qubit operations apply also require only a single symbol. After merging the vertices, we can delete all vertices with a single symbol because the corresponding term in the learning basis can be arbitrarily set to any Pauli. Following these transformations we can apply graph coloring and determine the bases using a (binary) covering array $\mathrm{CA}(2, k', 2)$, where the number of colors k' can differ from that of the original graph, and where the alphabet size v is now equal to 2 (explicit

k	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$\mathrm{CAN}(2,k,2)$	4	4	5	6	6	6	6	6	6	7	7	7	7	7	8	8	8	8	8
$\mathrm{CAN}(3,k,2)$	–	8	8	10	12	12	12	12	12	12	15	16	16	17	17	18	18	18	18
$\mathrm{CAN}(4,k,2)$	–	–	16	16	21	24	24	24	24	24	24	32	35	35	35	35	36	39	39
$\mathrm{CAN}(2,k,3)$	9	9	9	11	12	12	13	13	14	15	15	15	15	15	15	15	15	15	15
$\mathrm{CAN}(3,k,3)$	–	27	27	33	33	39	42	45	45	45	45	45	45	51	51	58	59	59	59

Table 2: Number of rows (learning bases) for the smallest covering arrays of different sizes known at the time of writing [38, 39]. The values for $\mathrm{CAN}(2,k,2)$ match the Rényi lower bound [40] and are therefore optimal.

constructions for binary covering arrays can be found in [37]). It can be seen from Table 2, that the smaller alphabet size leads to a substantial reduction in the number of learning bases. When extracting the learning bases from the graph coloring and the binary covering array, we expand each symbol for merged vertices to two gate-dependent Pauli terms, one for each qubit associated with the original vertices. We illustrate the basis-selection protocol in the case of Pauli twirls, with and without additional rotation twirls in Figure 4

5.3 Basis selection for ℓ-local noise models

We now show how the learning protocol can be modified to support the generalization from the two-local Pauli-Lindblad noise model to any weight ℓ-local noise models. We can define an ℓ-local noise model by selecting support sets $\{\mathcal{S}_i\}_i$ defined in Theorem 1 with $\max_i |\mathcal{S}_i| = \ell$ and setting $\mathcal{K} = \bigcup_i \mathcal{P}_i$, where $\mathcal{P}_i$ is the set of all Pauli terms supported on $\mathcal{S}_i$. In order to learn the model parameters is then suffices to learn the fidelities for Pauli terms in $\mathcal{B} := \mathcal{K}$. For this we again set up a graph with vertices corresponding to the model qubits. The edges are generated based on each set $\mathcal{S}_i$ such that the graph has a clique on the vertices associated with the qubits in $\mathcal{S}_i$. After this, depending on the twirl group, we can merge the vertices of the qubits in each of the gates and color the graph. Observe that addition of the local cliques ensures that, for each i, all vertices in $\mathcal{S}_i$ are assigned different colors. The basis selection is then done using a covering array $\mathrm{CA}(\ell, k, v)$, where k is the number of colors in the (reduced) graph, and v is 2 or 3 depending on whether or not we apply rotational twirling. Mapping the symbols back to Pauli terms proceeds as before.

5.4 Further extensions

Finally, we briefly consider basis selection for general sets of Pauli terms $\mathcal{B}$ and an isolated noise channel. It is well known that sets of mutually commuting Paulis observables can be measured simultaneously by means of a basis change effected by an appropriate Clifford circuit. In this case, we can determine the minimum number of learning bases by forming a graph with vertices corresponding to the elements in $\mathcal{B}$ and edges between vertices whose Pauli terms commute and then finding the maximum clique cover. Alternatively, we can connect Pauli terms that do not commute and find a graph coloring. Each of the cliques in the cover, or terms with equal cover, can then be measured simultaneously under the appropriate basis transformation. However, this approach is generally not practically feasible on noisy quantum processors due to the excessive noise generated by the Clifford circuit implementing the basis change. This problem can be overcome, possibly at the expense of more bases, by allowing Pauli operators to be measured simultaneously only if the Pauli terms on each of the

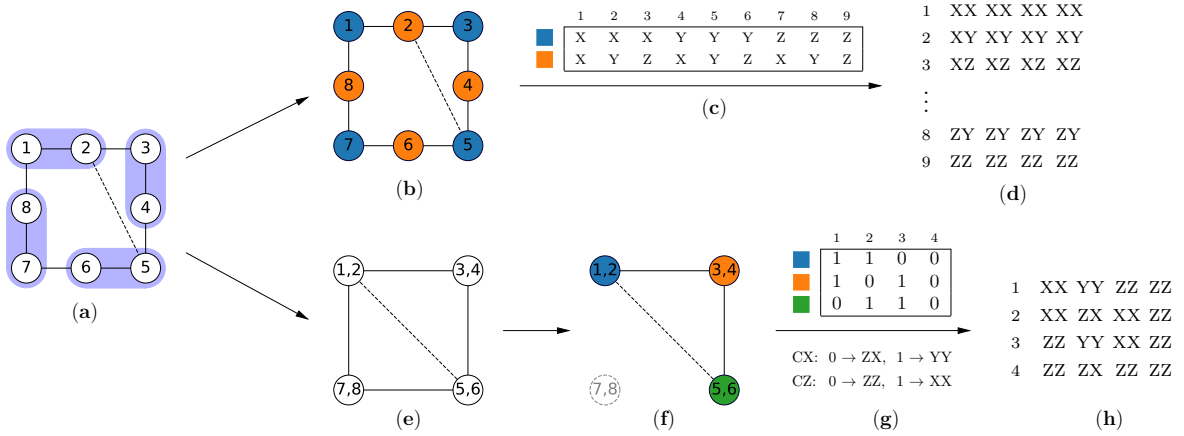

Figure 4: Application of the learning protocol to (a) a layer of gates CZ$(1,2)$, CX$(3,4)$, CZ$(5,6)$, and Swap$(7,8)$ on an example topology with presumed crosstalk between qubits 2 and 5. The top row shows basis selection with regular Pauli twirling. Here, we start by (b) finding a graph coloring for the qubit topology. We then find (c) a covering array CA(2,2,3) with symbols $\{X, Y, Z\}$, here shown in transposed form. Each column identifies the mapping of colors to symbols for a single basis. Applying the mapping gives (d) the desired learning bases. The bottom row illustrates basis selection when using a Pauli rotation twirl in addition to the regular Pauli twirl. In this case, we first (e) merge the vertices of the qubits of each gate, and, depending on the type of gate, associate with each vertex one or two symbols. We can then (f) delete all vertices that require only a single symbol, for instance for the swap gate, since any basis terms on those qubits allow us to learn all desired fidelities. We then color the resulting graph and (g) find a CA(2,3,2) covering array and define a gate-dependent mapping of symbols to basis terms. Finally, we (h) apply the mapping to obtain the different bases. Note that the chromatic number of the graphs corresponding to basic and rotation twirling can differ; in the present example the chromatic numbers are two and three, respectively. However, if the additional crosstalk occurred on qubits 2–4, rather than 2–5, the chromatic numbers would change to three and two, respectively.

qubits commute. When considering learning of noise in the context of gates a more elaborate approach is needed that takes advantage of the fact that, under the symmetry assumption, we can learn the fidelities for Pauli pair P_i and P_j by measuring either of the two Pauli operators. The equivalence of fidelities due to rotational twirling further increases the number of Pauli observables that can be measured to obtain the desired fidelities. We will not discuss any of these extensions here.

6 Conclusions

We have shown how the combined use of graph coloring and covering arrays gives rise to an elegant protocol for basis selection in the learning of sparse Pauli-Lindblad noise models. Complementary to this we have introduced Pauli-rotation twirls and shown how they can be used to reduce the number of unique fidelities in Pauli noise channels. Together, these techniques enable the practical learning of noise models for highly-connected qubit topologies with added connections to model crosstalk between known qubit pairs. Based on the the covering array sizes listed in Table 2, we can learn two-local noise model for Pauli-twirled channels even when the qubit topology has a chromatic number up to twenty by leveraging the covering array in [41]. Adding a rotation twirl allows us to reduce the alphabet size

of the covering arrays, which drastically reduces the required number of learning bases and opens up the avenue to learning three-local noise models. Reduction of the number of unique Pauli fidelities by means of the rotation twirl is generally applicable and could be used, for instance, to reduce the number of measurements required for averaged circuit eigenvalue sampling [42, 43].

Acknowledgements

EvdB would like to thank Dr. Jose Torres-Jimenez for providing access to the uniform covering arrays on [39] and for pointing out useful references. We would also like to thank the referees for providing additional pointers to relevant literature.

References

[1] 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).

[2] Jarrod R. McClean, Jonathan Romero, Ryan Babbush, and Alán Aspuru-Guzik. "The theory of variational hybrid quantum-classical algorithms". New Journal of Physics **18**, 023023 (2016).

[3] Jules Tilly, Hongxiang Chen, Shuxiang Cao, Dario Picozzi, Kanav Setia, Ying Li, Edward Grant, Leonard Wossnig, Ivan Rungger, George H Booth, et al. "The variational quantum eigensolver: a review of methods and best practices". Physics Reports **986**, 1–128 (2022).

[4] Ewout van den Berg, Zlatko K. Minev, and Kristan Temme. "Model-free readout-error mitigation for quantum expectation values". Physical Review A **105**, 032620 (2022).

[5] Sergey Bravyi, Sarah Sheldon, Abhinav Kandala, David McKay, and Jay M. Gambetta. "Mitigating measurement errors in multiqubit experiments". Physical Review A **103**, 042605 (2021).

[6] Rebecca Hicks, Christian W. Bauer, and Benjamin Nachman. "Readout rebalancing for near-term quantum computers". Phys. Rev. A **103**, 022407 (2021).

[7] Kristan Temme, Sergey Bravyi, and Jay M. Gambetta. "Error mitigation for short-depth quantum circuits". Physical Review Letters **119**, 180509 (2017).

[8] Ying Li and Simon C. Benjamin. "Efficient variational quantum simulator incorporating active error minimization". Physical Review X **7**, 021050 (2017).

[9] Abhinav Kandala, Kristan Temme, Antonio D. Córcoles, Antonio Mezzacapo, Jerry M. Chow, and Jay M. Gambetta. "Error mitigation extends the computational reach of a noisy quantum processor". Nature **567**, 491–495 (2019).

[10] Ewout van den Berg, Zlatko K. Minev, Abhinav Kandala, and Kristan Temme. "Probabilistic error cancellation with sparse Pauli-Lindblad models on noisy quantum processors". Nature Physics **19**, 1116–1121 (2023).

[11] Suguru Endo, Simon C. Benjamin, and Ying Li. "Practical quantum error mitigation for near-future applications". Physical Review X **8**, 031027 (2018).

[12] Andrea Mari, Nathan Shammah, and William J Zeng. "Extending quantum probabilistic error cancellation by noise scaling". Physical Review A **104**, 052607 (2021).

[13] Samuele Ferracin, Akel Hashim, Jean-Loup Ville, Ravi Naik, Arnaud Carignan-Dugas, Hammam Qassim, Alexis Morvan, David I. Santiago, Irfan Siddiqi, and Joel J. Wallman. "Efficiently improving the performance of noisy quantum computers". Quantum **8**, 1410 (2024).

[14] Benjamin McDonough, Andrea Mari, Nathan Shammah, Nathaniel T. Stemen, Misty Wahl, William J. Zeng, and Peter P. Orth. "Automated quantum error mitigation based on probabilistic error reduction". In 2022 IEEE/ACM Third International Workshop on Quantum Computing Software (QCS). Pages 83–93. IEEE (2022).

[15] Youngseok Kim, Andrew Eddins, Sajant Anand, Ken Xuan Wei, Ewout van den Berg, Sami Rosenblatt, Hasan Nayfeh, Yantao Wu, Michael Zaletel, Kristan Temme, and Abhinav Kandala. "Evidence for the utility of quantum computing before fault tolerance". Nature **618**, 500–505 (2023).

[16] Charles H. Bennett, Gilles Brassard, Sandu Popescu, Benjamin Schumacher, John A. Smolin, and William K. Wootters. "Purification of noisy entanglement and faithful teleportation via noisy channels". Phys. Rev. Lett. **76**, 722–725 (1996).

[17] Emanuel Knill. "Fault-tolerant postselected quantum computation: Threshold analysis" (2004) arXiv:quant-ph/0404104.

[18] Oliver Kern, Gernot Alber, and Dima L. Shepelyansky. "Quantum error correction of coherent errors by randomization". The European Physical Journal D-Atomic, Molecular, Optical and Plasma Physics **32**, 153–156 (2005).

[19] Michael R Geller and Zhongyuan Zhou. "Efficient error models for fault-tolerant architectures and the Pauli twirling approximation". Physical Review A **88**, 012314 (2013).

[20] Joel J. Wallman and Joseph Emerson. "Noise tailoring for scalable quantum computation via randomized compiling". Physical Review A **94**, 052325 (2016).

[21] Steven T. Flammia and Joel J. Wallman. "Efficient estimation of Pauli channels". ACM Transactions on Quantum Computing **1**, 1–32 (2020).

[22] Alexander Erhard, Joel J. Wallman, Lukas Postler, Michael Meth, Roman Stricker, Esteban A. Martinez, Philipp Schindler, Thomas Monz, Joseph Emerson, and Rainer Blatt. "Characterizing large-scale quantum computers via cycle benchmarking". Nature Communications **10**, 1–7 (2019).

[23] Shelby Kimmel, Marcus P. da Silva, Colm A. Ryan, Blake R. Johnson, and Thomas Ohki. "Robust extraction of tomographic information via randomized benchmarking". Phys. Rev. X **4**, 011050 (2014).

[24] Jonas Helsen, Xiao Xue, Lieven M. K. Vandersypen, and Stephanie Wehner. "A new class of efficient randomized benchmarking protocols". npj Quantum Information **5**, 1–9 (2019).

[25] Senrui Chen, Yunchao Liu, Matthew Otten, Alireza Seif, Bill Fefferman, and Liang Jiang. "The learnability of Pauli noise". Nature Communications **14**, 52 (2023).

[26] Zhenyu Cai and Simon C Benjamin. "Constructing smaller Pauli twirling sets for arbitrary error channels". Scientific reports **9**, 1–11 (2019).

[27] Arnaud Carignan-Dugas, Joel J. Wallman, and Joseph Emerson. "Characterizing universal gate sets via dihedral benchmarking". Phys. Rev. A **92**, 060302 (2015).

[28] A. K. Hashagen, S. T. Flammia, D. Gross, and J. J. Wallman. "Real randomized benchmarking". Quantum **2**, 85 (2018).

[29] David Layden, Bradley Mitchell, and Karthik Siva. "Theory of quantum error mitigation for non-Clifford gates" (2024). arXiv:2403.18793.

[30] Charles J. Colbourn. "Combinatorial aspects of covering arrays". Le Matematiche **59**, 125–172 (2004). url: `https://lematematiche.dmi.unict.it/index.php/lematematiche/article/view/166`.

[31] Jose Torres-Jimenez, Brenda Acevedo-Juárez, and Himer Avila-George. "Covering array EXtender". Applied Mathematics and Computation **402**, 126122 (2021).

[32] Jose Este Jaloveckas, Minh Tham Pham Nguyen, Lilly Palackal, Jeanette Miriam Lorenz, and Hans Ehm. "Efficient learning of sparse Pauli Lindblad models for fully connected qubit topology" (2023). arXiv:2311.11639.

[33] Charles Colbourn. "Covering perfect hash families and covering arrays of higher index". International Journal of Group Theory **13**, 293–305 (2024).

[34] Jordan Cotler and Frank Wilczek. "Quantum overlapping tomography". Phys. Rev. Lett. **124**, 100401 (2020).

[35] Tim J. Evans, Robin Harper, and Steven T. Flammia. "Scalable Bayesian Hamiltonian learning" (2019). arXiv:1912.07636.

[36] Hsin-Yuan Huang, Richard Kueng, and John Preskill. "Efficient estimation of Pauli observables by derandomization". Phys. Rev. Lett. **127**, 030503 (2021).

[37] Jose Torres-Jimenez and Eduardo Rodriguez-Tello. "New bounds for binary covering arrays using simulated annealing". Information Sciences **185**, 137–152 (2012).

[38] Charles Colbourn. `https://www.public.asu.edu/~ccolbou/src/tabby/catable.html`.

[39] Jose Torres-Jimenez. `https://www.tamps.cinvestav.mx/~oc/`.

[40] Alfréd Rényi. "Foundations of probability". Wiley. New York, USA (1971).

[41] Kari J. Nurmela. "Upper bounds for covering arrays by tabu search". Discrete applied mathematics **138**, 143–152 (2004).

[42] Steven T. Flammia. "Averaged Circuit Eigenvalue Sampling". In François Le Gall and Tomoyuki Morimae, editors, 17th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2022). Volume 232 of Leibniz International Proceedings in Informatics (LIPIcs), pages 4:1–4:10. Dagstuhl, Germany (2022). Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[43] Evan T. Hockings, Andrew C. Doherty, and Robin Harper. "Scalable noise characterisation of syndrome extraction circuits with averaged circuit eigenvalue sampling" (2024). arXiv:2404.06545.