arXiv:2412.01311v3 [quant-ph] 18 Aug 2025

Reduced Sampling Overhead for Probabilistic Error Cancellation by Pauli Error Propagation

Timon Scheiber[1,2], Paul Haubenwallner[1,2], and Matthias Heller[1,2]

[1]Fraunhofer Institute for Computer Graphics Research IGD

[2]Technical University of Darmstadt, Interactive Graphics Systems Group

Quantum error mitigation is regarded as a possible path to near-term quantum utility. The methods under the quantum error mitigation umbrella term, such as probabilistic error cancellation (PEC), zero-noise extrapolation (ZNE) or Clifford data regression (CDR) are able to significantly reduce the error for the estimation of expectation values, although at an exponentially scaling cost, i.e., in the sampling overhead. In this work, we present a method to reduce the sampling overhead of PEC through Pauli error propagation combined with classical preprocessing. Our findings indicate that this method significantly reduces sampling overheads for Clifford circuits, leveraging the well-defined interaction between the Clifford group and Pauli noise. Additionally, we show that the method is applicable to non-Clifford circuits, though with more limited effectiveness, primarily constrained by the number of non-Clifford gates present in the circuit. We further provide examples of Clifford sub-circuits commonly encountered in relevant calculations, such as resource state generation in measurement-based quantum computing.

1 Introduction

Quantum computing is expected to outperform classical computing in specific use cases within the near future [1, 2]. However, most of the existing algorithms showing a rigorously proven superior scaling compared to classical algorithms lie beyond the reach of current noisy intermediate scale quantum (NISQ) computers and will probably become relevant only after fault-tolerance is achieved [3, 4]. While recently tremendous progress in the realization of error corrected qubits has been made, both in terms of efficient encodings [5–7], and real hardware demonstrations [7–10], current quantum hardware is still far from being fault-tolerant. On the other side, evidence has been presented that the current generation of quantum hardware can access computational spaces, which might be out of reach even for advanced supercomputers [11–13]. Since these devices are still bound by noise, current NISQ-algorithms require aid by quantum error mitigation (QEM) schemes to be able to compete with classical solutions [14–16].

QEM methods mostly focus on quantum algorithms, that aim to estimate the expectation value $\langle A \rangle$ of some observable A, by reducing the noise induced bias at the cost of an increase in the variance of the estimate [16–18]. One of the earlier presented methods is the so-called probabilistic error cancellation (PEC) [15]. PEC aims to construct an ideal, noiseless circuit operation $\mathcal{U}(\rho) = U\rho U^\dagger$ by expanding it into an (overcomplete) basis of natively performable, noisy operations $\mathcal{O}_i$, which can be directly executed by the hardware. This expansion can be achieved in two ways: Either by *compensation*, where each gate operation $\mathcal{U}_i$ of a circuit $\mathcal{U} = \mathcal{U}_n \cdots \mathcal{U}_0$ is directly replaced by the superimposed operation $\mathcal{U}_i = \sum_j \eta_j \mathcal{O}_j$ for some real coefficients η_j, or alternatively by *inversion*, where for each noisy operation $\tilde{\mathcal{U}}_i = \Lambda_i \circ \mathcal{U}_i$ the mathematical inverse of the noise channel $\Lambda_i^{-1} = \sum_j \eta'_j \mathcal{O}_j$ is implemented directly before or after Λ_i to cancel the effect of the noise [19]. The implementation of this decomposition is performed probabilistically by sampling from a quasi-probability distribution defined by the linear combination with probabilities respecting the weights η_j. This implementation performs the ideal operation on average,

Timon Scheiber: timon.florian.scheiber@igd.fraunhofer.de

but it generally comes at the cost of an increase in the variance of the desired result described by the sampling overhead γ. The latter method has recently been demonstrated experimentally on a superconducting quantum chip [20]. In the reference, a sparse Pauli-Lindblad noise model was derived to efficiently characterize and learn the noise of sparsely connected quantum devices and estimate the inverse noise channels Λ_i^{-1} for different layers of noisy two-qubit gates. While the method delivers excellent results in terms of retrieving nearly bias-free estimates, the exponentially scaling increase in γ still limits the usefulness of the method to small circuits.

In this work, we alleviate this scaling issue by introducing a method to estimate and sample from a conjoint inverse noise channel (or fused noise channel) $\Lambda_{\text{global}}^{-1}$ in contrast to sampling from each inverse noise channel separately. To estimate this conjoint inverse noise channel we utilize Pauli error propagation through Clifford circuits. We show, that the proposed method can greatly reduce the required sampling overhead.

Related work and our main contributions

Probabilistic error cancellation, often denoted quasi-probability decomposition method, is a widely utilized technique for error mitigation. The primary challenge in its applicability lies in the sampling overhead. Numerous protocols have been developed to alleviate this issue [21–27]. For instance, Ref. [21] demonstrates that for highly local observables, the sampling overhead can be reduced by leveraging the fact that information and, hence, noise propagation, is constrained within a light cone originating from the observable. Other work [22] shows that the sampling overhead can be reduced by using the control variates method, commonly used in statistics for variance reduction of Monte-Carlo methods. Ref. [23] introduces a novel method with reduced sampling overhead for extracting non-linear features, e.g. the k-th moment of a density operator $\text{Tr}(\rho^k)$, by measuring a shifted observable on k copies of the original system.

In this work, we develop a method to reduce the sampling overhead for probabilistic error cancellation by leveraging the interference of local noise channels. The interference from sequential error channels can result in decreased sampling overheads [24], a principle that has been implicitly employed in recent tensor network based approaches [25, 26]. Although tensor networks are not constrained by circuit width, they require approximations to capture deep circuits leading to a bias in the mitigated expectation values. In contrast, our method only cancels those terms, which interfere destructively without any approximations such that no bias is introduced.

We calculate the interference based on Heisenberg propagation of noise channels within the context of Clifford circuits and Pauli rotation gates. By focusing our investigation on Clifford circuits and Pauli noise, we can effectively retrieve correlations from multiple layers of the quantum circuit. Given that the reduction in sampling overhead arises from the build-up of interference over various layers, our method achieves significantly greater reductions than those possible with tensor networks which are usually applied to low depth circuits. A further benefit of the Pauli error propagation method is that the inverse of the channels can be implemented by sampling from Pauli gates only and does not require any entangling gates, which might introduce more errors to the system.

A related work [27] employs a similar approach of error propagation for the case of phase errors to reduce the sampling overhead of PEC. However, in contrast to our approach, the authors consider cat qubits with dominant phase errors and a set of bias preserving gates and therefore only discusses the propagation of Pauli Z errors.

2 Probabilistic Error Cancellation for Pauli Errors

In this section we give a brief overview of PEC by inversion. We assume that the individual noise channels occurring during a gate operation can be described as an ideal Pauli channel, which can be assured by randomized compiling over the Pauli group (also known as Pauli twirling) [28]

$$\Lambda(\rho) = \sum_{i=1}^{N=4^n} c_i P_i \rho P_i^\dagger. \tag{1}$$

The summation runs over all elements P_i of the Pauli group of dimension n, where n denotes the number of qubits. The real, positive channel coefficients c_i sum up to one and can be interpreted as the probability of a Pauli error P_i occurring additionally to the effect of the ideal operation $\mathcal{U}$.

To perform PEC we assume that the exact coefficients of the correlated Pauli errors can be efficiently learned, for example by using techniques such as cycle benchmarking [20, 29].

An important property of Pauli noise channels is that their mathematical inverse also closely resembles a Pauli channel

$$\Lambda^{-1}(\rho) = \sum_{i=1}^{N} \tilde{c}_i P_i \rho P_i^\dagger, \tag{2}$$

however, with coefficients $\tilde{c}_i$ that are no longer guaranteed to be positive. While not being a channel in the mathematical sense, we will refer to the map defined by Eq. 2 as the inverse channel throughout this manuscript. The inverse is in general not a completely positive trace preserving (CPTP) map and can thus not be implemented by a single unitary operation. Its effect can however be realized *probabilistically*.

To implement the operation probabilistically, Eq. 2 can be restructured as follows:

$$\Lambda^{-1}(\rho) = \gamma \sum_{i=1}^{N} \mathrm{sgn}(\tilde{c}_i) \cdot p_i P_i \rho P_i^\dagger, \quad p_i = \frac{|\tilde{c}_i|}{\gamma}, \tag{3}$$

where the factor γ is given by

$$\gamma = \sum_{i=1}^{N} |\tilde{c}_i|. \tag{4}$$

The benefit of restructuring the inverse channel in this way stems from the fact that the coefficients p_i can now be interpreted as probabilities. Utilizing Eq. 3, the inverse is performed *on average* by sampling and applying the Pauli correction P_i directly before the noisy gate with corresponding probability p_i and multiplying the measured expectation value with the corresponding sign $\mathrm{sgn}(\tilde{c}_i) = s_i$ and γ-factor in post-processing.

To apply the method to circuits containing multiple noisy gate operations (or layers of parallel executed operations) $\tilde{\mathcal{U}} = \prod_{l=1}^{L} (\mathcal{U}_l \circ \Lambda_l)$, this process is repeated for each noisy layer $\tilde{\mathcal{U}}_l$ of the circuit, so that an individual correction is drawn and directly applied in front of the layer. Since each correction is attached to a corresponding sign and sampling overhead, the overall sign $s_{\mathrm{global}} = \prod_l s_l$ and γ-factor $\gamma_{\mathrm{total}} = \prod_l \gamma_l$ for the complete circuit are given by the product of the individual signs and sampling overheads respectively.

To retrieve the expectation value of the mitigated observable, several correction circuit instances are generated and executed. The individual results are multiplied by the total sampling overhead γ_{total}, and the respective global sign s_{global}. Finally, the mitigated expectation value is given by the average of the M executed correction circuits [15]

$$\langle A \rangle_{\mathrm{ideal}} = \frac{\gamma_{\mathrm{total}}}{M} \sum_{m=1}^{M} s_{\mathrm{global,m}} \langle A \rangle_{\mathrm{corr,m}} \tag{5}$$

where $\langle A \rangle_{\mathrm{corr,m}} = \mathrm{Tr}(A\, \tilde{\mathcal{U}}_{\mathrm{corr,m}}(\rho))$ refers to the expectation value of A evaluated on the m-th circuit.

As described by van den Berg et al. [20], the factor $\gamma \geq 1$ is directly related to the variance of the mitigated observable, which scales with γ^2. Since each noisy layer is associated with an individual noise channel, and thus an individual factor γ_l, the total γ-factor of the system increases exponentially. It follows that the variance of the mitigated expectation value and hence the number of shots required for a small sampling error, grows exponentially with the circuit depth.

3 Propagated Probabilistic Error Cancellation (pPEC) for Clifford circuits

One of the limiting factors of PEC is the large overhead to compensate for the rapid growth in variance. We approach this problem by introducing a method to decrease the sampling overhead of classical PEC, which we call (error)-propagated probabilistic error cancellation (pPEC). In pPEC the individual inverse noise channels of a given circuit are propagated to the start (or end) of the circuit and multiplied together, yielding a conjoint operation with smaller γ-factor. To estimate this fused inverse noise channel, we present two methods; a hardware-agnostic Monte-Carlo method which serves as a more educational example (see appendix A) as well as an analytic method based on a sparse Pauli-Lindblad noise model [20]. We motivate this approach by showing that it is favorable to sample from a conjoint inverse noise channel $\Lambda^{-1}_{\mathrm{global}}$ by comparing the sampling overhead $\gamma_{\mathrm{global}} = \gamma(\Lambda^{-1}_{\mathrm{global}})$ of applying the global inverse (fused noise channel for all layers) to the sampling overhead of correcting each layer individually $\gamma_{\mathrm{total}} = \prod_l \gamma(\Lambda_l^{-1})$.

In a naïve approach to estimate the global noise channel, one could try to directly apply a learning procedure to the full circuit which, in the general case, would require an exponentially scaling amount of measurements and is thus infeasible. However, the circuit can be decomposed into a product of individual noisy gate layers $\tilde{\mathcal{U}}_l$, each consisting of a (largely) learnable [30] noise channel Λ_l. It is customary to assume that single-qubit operations are effectively noiseless and only two-qubit operations contribute to the noise due to order of magnitude higher errors on most devices. The full circuit operation $\tilde{\mathcal{U}}$ can then be decomposed into a product of noisy circuit operations $\tilde{\mathcal{U}}_l$

$$\tilde{\mathcal{U}} = \prod_l \tilde{\mathcal{U}}_l = \prod_l (\mathcal{U}_l \circ \Lambda_l), \tag{6}$$

where we assume $\Lambda_l = \mathbb{1}$, if $\mathcal{U}_l$ consists solely of single-qubit gates. The effect of the noise can be canceled by probabilistically implementing the inverse of the noise channel Λ_l^{-1} directly before the application of the noisy gate $\tilde{\mathcal{U}}_l$. A graphical depiction of the method is presented in the upper part of Fig. 1.

Since we are only interested in calculating the global inverse noise channel and not the global noise channel itself, we consider the PEC-corrected circuit $\mathcal{U}(\rho) = \prod_{l=1}^{L} \mathcal{U}_l \Lambda_l \Lambda_l^{-1}$ as the point of departure. From this we can calculate the global inverse by propagating each inverse to the start of the circuit

$$\Lambda_{\text{global}}^{-1} = \prod_{l=L}^{1} \tilde{\Lambda}_l^{-1}, \tag{7}$$

where the individual, propagated operations are given by

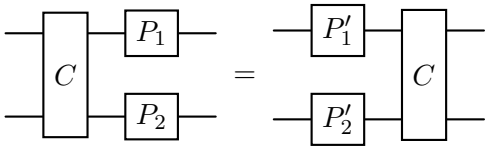
$$\tilde{\Lambda}_l^{-1} = \mathcal{U}_0 \cdots \mathcal{U}_{l-2} \mathcal{U}_{l-1} (\Lambda_l^{-1}), \tag{8}$$

with $\mathcal{U}_j(\Lambda_l^{-1}) = U_j \Lambda_l^{-1} U_j^\dagger$ describing the Heisenberg propagation of the inverse channels. Note that the product runs over the layers in reversed order, and we only conjugate the inverse noise channels with the ideal circuit operations and not the noise channels.

It is important to note, that for arbitrary gates the inverse after the propagation might no longer be described by a diagonal (inverse) Pauli channel. An important class of gates, for which the inverse noise channel will remain a diagonal Pauli channel are Clifford gates. This is the case, since elements of the Clifford group are defined by the property

$$CPC^\dagger = P', \tag{9}$$

where P and P' denote operators of the Pauli group and C an arbitrary operator of the Clifford group. Using Eq. (9) we are able to move the Pauli correction terms in front of a Clifford gate by exchanging it with the conjugated Pauli correction P'



. (10)

In general this conjugation can be performed efficiently by using a look-up table, which we provide in Appendix B for the convenience of the reader. This way the individual channels can be propagated to the start of the circuit and multiplied together yielding the global (inverse) noise channel.

The key advantage of fusing individual correction layers together into one global one stems from the fact, that different errors or corrections can interfere destructively. The identification of these corrections before applying the inverse probabilistically can thus reduce the required number of correction circuits by a sometimes large amount. For example, the PEC corrected expectation value with M independently sampled corrections can be expressed as

$$\langle A \rangle_{\text{PEC}} = \frac{\gamma_{\text{total}}}{M} \left[\sum_{m=1}^{M} s_m \langle A \rangle_{\text{corr,m}} \right], \tag{11}$$

where $\langle A \rangle_{\text{corr,m}}$ denotes the expectation value of A on the m-th circuit and s_m the corresponding global sign. By identifying the number of corrections $\mathcal{J}$ that interfere destructively, i.e., those for which the local corrections map to the same global correction with opposite sign, Eq. (11) can be restructured as

$$\begin{aligned} \langle A \rangle_{\text{PEC}} &= \frac{\gamma_{\text{total}}}{M} \left[\sum_{m=1}^{M-\mathcal{J}} s_m \langle A \rangle_{\text{corr,m}} + \underbrace{\sum_{j=1}^{\mathcal{J}} s_j \langle A \rangle_{\text{corr,j}}}_{0} \right] \\ &= \frac{\gamma_{\text{total}}}{M} \left[\sum_{m=1}^{M-\mathcal{J}} s_m \langle A \rangle_{\text{corr,m}} \right], \end{aligned} \tag{12}$$

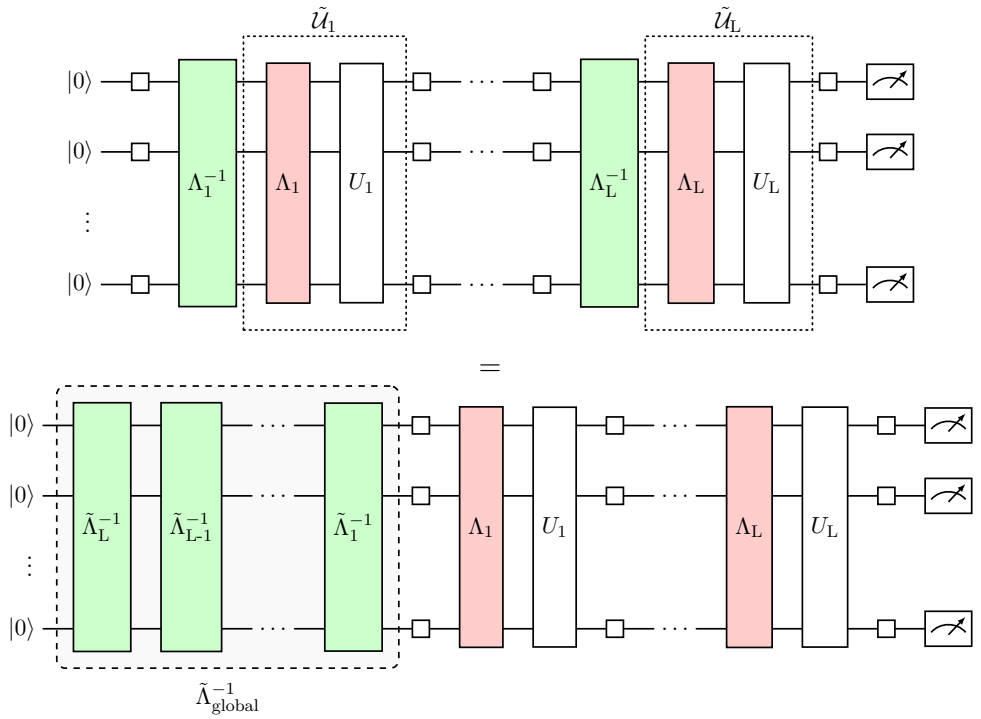


Figure 1: **Above:** graphical summary of the PEC method. The circuit consists of one qubit gates and layers of noisy two qubit gates $\tilde{\mathcal{U}}_l = \mathcal{U}_l \circ \Lambda_l$. The noise is probabilistically canceled by implementing the inverse of the noise channel and applying it in front of the noise. **Below:** the same circuit after propagation of the inverse noise channels to the start of the circuit. After multiplication the full noise of the circuit is canceled with the single layer correction $\Lambda_{\text{total}}^{-1}$.

where, by assumption, the last $\mathcal{J}$ correction circuits interfere destructively meaning that the individual expectation values cancel each other out. In contrast, by only sampling from the non-interfering corrections, the pPEC corrected expectation value with $M-\mathcal{J}$ independently drawn correction circuits is given by

$$\langle A \rangle_{\text{pPEC}} = \frac{\gamma_{\text{total}}}{M-\mathcal{J}} \left[\sum_{m=1}^{M-\mathcal{J}} s_m \langle A \rangle_{\text{corr,m}} \right], \quad (13)$$

which coincides with Eq. (12) up to a factor of $\frac{M-\mathcal{J}}{M}$. This factor can then be incorporated into the γ-factor giving the sampling overhead for pPEC

$$\gamma_{\text{pPEC}} = \gamma_{\text{total}} \cdot \frac{M-\mathcal{J}}{M}. \quad (14)$$

Hence, sampling from the pPEC distribution can be interpreted as sampling from a quasi-probability distribution with reduced overhead γ. The reduction stems from the fact that pPEC is able to identify the interfering corrections before the implementation. We provide a more illustrative example of how the interference of corrections follows from error propagation as well as a method to estimate a fused channel in Appendix A. Furthermore, we provide a more rigorous proof that the sampling from the global or fused inverse is always favorable in Appendix C.

3.1 XI-Reduction

To decrease the sampling overhead even further, we reduce the number of individual Pauli corrections that need to be applied at the start of the circuit, by leveraging the phase invariance of the computational basis states. Considering for example the Pauli Z operator it is easy to verify that $Z|0\rangle = |0\rangle$ and $Z|1\rangle = -|1\rangle \equiv |1\rangle$, up to a global phase.

This allows us to replace each Z operator with an identity operation if applied to a computational basis state without superposition, as for example at the start of the circuit. Furthermore, the Pauli Y operator can be expressed as $Y = -iX \cdot Z$ and thus $Y|0\rangle \equiv X|0\rangle$ and $Y|1\rangle \equiv X|1\rangle$, again omitting the global phase. These identities also hold for tensorized Paulis. For example, the Pauli strings $IXXZY \equiv IXXIX$ and $IYYIX \equiv IXXIX$ are both equivalent to the

same Pauli string after the reduction and can thus be corrected by the same operation. Using these reductions the total number of corrections and thus the maximum number of instances that need to be sampled individually reduces from 4^n to 2^n.

It is noteworthy that this reduction is inherently symmetric, in the sense that it can be applied either after propagation of the gates to the start or to the end of the circuit, directly before the measurement. This convenience directly follows from Born's rule, stating that only the squared magnitude of the amplitudes can be measured. For example in the computational basis the measurement of a quantum state $|\psi\rangle = \sum_{i\in\{0,1\}^n} c_i |i\rangle$ will result in the outcome i with probability $|c_i|^2$. Thus, any transformation $c_i \to e^{i\phi_i} c_i$ does not change the measurement outcome. Note, that the final measurement is always in the computational basis, when incorporating the final basis change into the circuit.

3.2 Readout-Error-Mitigation

Another common source of errors in quantum circuits are state preparation and measurement (SPAM) errors. Here, we utilize PEC for the mitigation of measurement errors and integrate them into the pPEC workflow. The readout error mitigation method presented in this work is based on assignment matrix inversion methods [31–33]. In these methods an assignment matrix A is calculated by preparing each basis state individually and measuring the noisy outcome. The matrix A correlates the ideal measurements with the noisy measurements via the relation $\vec{p}_{\text{noisy}} = A\vec{p}_{\text{ideal}}$. The idealized counts can then be retrieved by inverting the assignment matrix and multiplying the experimentally retrieved probability vector of the bit-strings with the inverse $\vec{p}_{\text{ideal}} = A^{-1}\vec{p}_{\text{noisy}}$. While this method almost fully eliminates readout errors, an exponential overhead for the preparation of the 2^n basis states is introduced, rendering the method infeasible for larger circuits.

To avoid this overhead we assume a tensor product noise model as presented in Ref. [31]

$$A = \bigotimes_{i=1}^{n} A^i = \bigotimes_{i=1}^{n} \begin{bmatrix} 1-\epsilon_i & \eta_i \\ \epsilon_i & 1-\eta_i \end{bmatrix}, \tag{15}$$

with ϵ_i and η_i denoting the $|0\rangle \to 1$ and $|1\rangle \to 0$ faulty readout probabilities on qubit i respectively. This imposes a significant simplification, but it can be considered reasonable for sparse measurement outcomes [34]. The tensor product of Eq. (15) runs over the individual 2×2 assignment matrices, which are defined for each individual qubit. These matrices are generally not equal among the qubits nor symmetrical under transposition, meaning that the individual bit-flip probabilities for $|0\rangle \to 1$ and $|1\rangle \to 0$ are not identical.

To apply the assignment matrix method in a fashion that integrates to the pPEC framework, we need to construct a measurement error channel that closely resembles a Pauli channel. This is achieved by symmetrizing the matrices A^i via randomized insertion of X gates [35], i.e., an X gate is applied with 50% probability directly before the measurement for each qubit and corrected in post-processing. This procedure symmetrizes the matrix $A^i_{\text{twirl}} = (A^i_{\text{twirl}})^T$ (see appendix D) which can then be incorporated into our framework using a Pauli X channel:

$$\Pi_{\text{meas}}(\rho) = \sum_m (1-p_x)\pi_m I\rho I\pi_m + p_x \pi_m X\rho X\pi_m, \tag{16}$$

where π_m are projection operators, fulfilling the relations $\pi_m\pi_{m'} = \delta_{mm'}\pi_m$, $\pi_m = \pi_m^\dagger$ and $\sum_m \pi_m = I$.

The bit-flip probabilities are explicitly given by the arithmetic average of the individual bit-flip probabilities, $p_{x,i} = \frac{\epsilon_i+\eta_i}{2}$.

3.3 Classical cost

We now discuss the classical cost of calculating the global or fused inverse noise channel. Our algorithm consists of two main steps, the channel propagation and the channel product. We first consider the cost associated with the propagation.

The commutation of a single Pauli with a Clifford gate can be performed by a look-up table in $\mathcal{O}(1)$. This propagation must be performed for all Paulis in a given channel, which can still be achieved with constant overhead, with the constant depending on the number of Paulis present in the channel. Since we only consider Pauli channels and Clifford gates, the total number of operators that need to be stored is conserved. The cost of propagating a single layer to the start of the circuit can be estimated as the overhead of commuting a noise channel of n Paulis by d layers. Hence, the propagation overhead for all Paulis scales polynomially as $\mathcal{O}(n \cdot d^2)$ due to each

layer's propagation to the circuit start with incrementally increasing depth d.

The bottleneck of our algorithm is the calculation of the product of k Pauli channels, where the number of terms, that one takes into account, can be chosen dynamically to restrict the computational time. This results in a general trade-off between classical and quantum resources. Taking the product of two Pauli channels $\sum_i^N c_i P_i \rho P_i$ with N respective elements, takes N^2 separate multiplications. The cost of fusing k channels can thus be upper bounded by $\mathcal{O}(N^k)$ which results in an exponential scaling in the classical cost. However, this scaling is a worst-case scaling without regarding interferences between channels.

For instance, the product of k single qubit noise channels acting on the same qubit seems to result in a $\mathcal{O}(4^k)$ scaling. Since the resulting operators of the multiplication are restricted to the 4 possible Pauli terms, in each step the corrections can be expressed as the 4 possible Pauli corrections. Hence, the scaling is given by taking $k \cdot 4^2$ multiplications, with overhead $\mathcal{O}(k)$ far fewer than the projected 4^k terms. This reduction in exponential overhead applies to any channel multiplication. The overhead is minimized when many corrections map to the same operator after an expansion step. As this effect is responsible for interference, multiplying channels with large shared support will lead to minimal classical overhead while reducing the quantum overhead significantly.

Up to this point the calculations presented here have been exact, preserving the bias-free property of PEC. If one allows for at least some small bias, the classical cost can be drastically reduced. The magnitude of the coefficients $|c_i|$ of the inverse channel $\Lambda^{-1}(\rho)$ can be regarded as the probability of applying a specific correction. Due to the multiplication of several corrections these values will be highly attenuated for some corrections while other coefficients will build up. By truncating all coefficients smaller than a specific cutoff ϵ, the truncated inverse channel can be expressed as

$$\Lambda^{-1}_{\text{trunc}}(\rho) = \sum_j \Theta(|c_j| - \epsilon) c_j P_j \rho P_j \tag{17}$$

where $\Theta(c_j)$ denotes the Heaviside step function and c_j is the coefficient of Pauli j.

Truncating the coefficients at each step introduces a slight bias but significantly reduces the overhead in the required multiplications so that smaller sampling overheads can be achieved.

4 Numerical simulations

Due to the large shot budget and time required to fully characterize the complete noise channels we restrict the experiments to numerical simulations. Since pPEC does not rely on approximations it is expected that the performance will be similar to regular PEC.

4.1 pPEC for gate-level noise

To demonstrate the efficiency of our proposed method, we compare PEC and pPEC for a random 10 qubit circuit consisting of Hadamard (H), Pauli (X,Y,Z), phase (S), controlled phase (C_Z) and measurement operations under gate level noise. For each two-qubit gate we apply a depolarizing channel

$$\Lambda_{\text{depol}}(\rho) = (1-p) I \rho I + \frac{p}{(N-1)} \sum_{i=2}^{N=4^2} P_i \rho P_i^\dagger, \tag{18}$$

where we chose the error probability $p = 2\%$ to resemble current hardware error rates [20]. It should be noted that processes such as hardware specific idle times of qubits which generally lead to dephasing and decay are not modeled in this description. Additionally, we integrate a tensor product, asymmetric readout error with a mean of $p_{\text{read}} = 2\%$, which we symmetrize using the method we introduced in Sec. 3.2.

Using a random Clifford circuit we benchmark the three different PEC protocols – PEC, pPEC and pPEC with XI-reduction. We perform the method by drawing corrections from each inverse noise channel directly (PEC) or by drawing from the pPEC distribution. For the considered example the pPEC circuits have been calculated by the direct product of all channels, which is generally possible for circuits with modest numbers of qubits. For larger circuits the same can be done, however the product needs to be truncated as the space of possible corrections grows exponentially. From each of these distributions a total of 40 correction circuits is drawn, which are simulated with 1024 shots per circuit instance.

To obtain a sufficient amount of statistics, a sample with a total of 1000 mitigated expectation values per method was generated. The resulting distribution is presented in Fig. 2. As expected, all PEC methods are able to retrieve the bias free expectation value on average, although at vastly

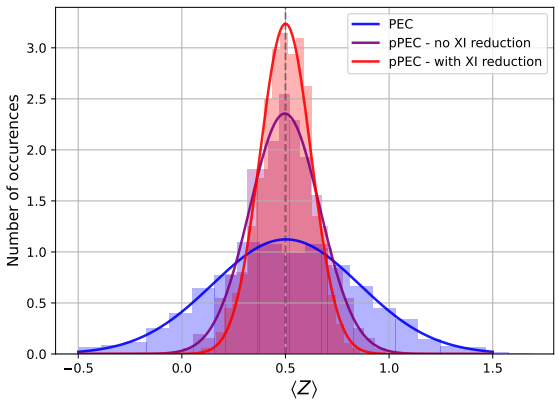


Figure 2: Mitigated circuits for the three tested method, PEC (blue), pPEC (purple), pPEC with XI-reduction (red). All methods are able to retrieve error free estimates of the noiseless observable. pPEC generally leads to smaller variances than the direct approach.

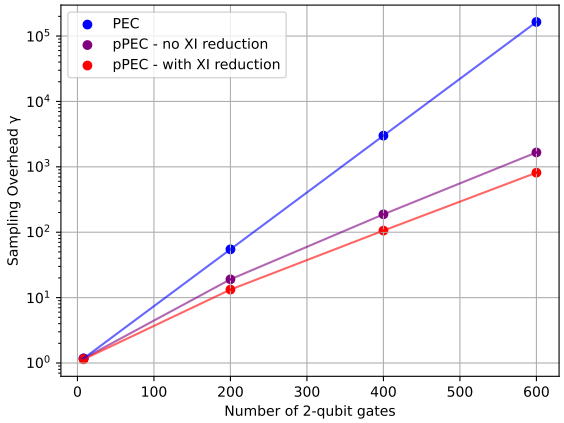


Figure 3: Calculated Sampling overhead averaged over 100 instances of random 5-qubit Clifford circuits at an increasing number of noisy operations.

different variances. Due to the promising results we now explore the scaling of the different approaches.

To investigate the expected scaling, we benchmark the different methods by calculating the global inverse for a sample of 100 random 5-qubit Clifford circuits for an increasing number of noisy two-qubit operations and calculate the expected γ-factors. The estimated total γ-factors for the direct PEC, propagated PEC and propagated PEC with XI-reduction are presented in Fig. 3. Since the sampling-overhead scales exponentially in the number of noisy operations, the data is plotted logarithmically. We achieve a far more favorable scaling for the propagated approach and a further decrease with the XI-reduction. However, it should be noted that the expected reduction can generally depend on the noise and structure of the circuit.

4.2 pPEC for circuit-level noise with SPL models

Sparse Pauli-Lindbladian (SPL) models are a recently introduced form of noise model to efficiently capture the noise of devices with limited physical qubit connectivity [20]. In the SPL model the noise is generated by a dissipative Lindbladian of the form

$$\mathcal{L}(\rho) = \sum_{k\in\mathcal{K}} \lambda_k (P_k \rho P_k^\dagger - \rho), \tag{19}$$

with model coefficients λ_k and Lindblad jump operators P_k given by a sparse subset $\mathcal{K}$ of the Pauli group. A noise channel, defined for a full layer of in-parallel executable gates, is described by the formula

$$\Lambda(\rho) = \prod_{k\in\mathcal{K}} (w_k \mathcal{I}(\cdot) + (1-w_k)\mathcal{P}_k(\cdot))\rho, \tag{20}$$

where the coefficients w_k are given by $w_k = (1 + e^{-2\lambda_k})/2$. The product runs over the indexed set $\mathcal{K}$ with usually far fewer terms than the dense noise model $|\mathcal{K}| \ll 4^n$.

$\mathcal{K}$ can be chosen in a way that only Pauli strings with no identity terms on at most two (physically connected) qubits are present. $\mathcal{P}_i(\cdot) \hat{=} P_i \cdot P_i^\dagger$ is a shorthand notation for a single Pauli channel in Kraus representation and the $(\cdot)$ symbol is a placeholder to illustrate that the whole product is applied to the system's density operator ρ.

In the following, we simulate a device with linear qubit topology, i.e., we choose the model Paulis in Eq. (20) by choosing weight-two (and one) Pauli strings with non-identity terms only on nearest neighbor qubits. As noise model we consider a quasi depolarizing channel by choosing the coefficients w_k to be homogeneous, allowing for easy tuning of the noise strength. We adjust the strength in accordance to the highest reported Pauli fidelity reported in Ref. [20], $f_{\max} \approx 0.996$, by utilizing $w_k = \frac{1+f_k}{2}$ where $f_k = \frac{1}{2^n}\mathrm{Tr}(P_k^\dagger \Lambda(P_k))$ is the Pauli fidelity for Pauli P_k with respect to the noise channel Λ.

We present our results of pPEC with XI reduction for random Clifford circuits of increasing

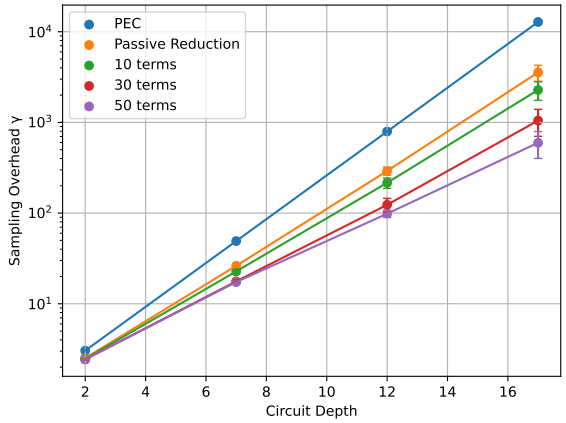


Figure 4: Calculated γ factors for quantum circuits of increasing circuit depth consisting of 10 qubits with an average Pauli fidelity of $f_{\text{avg}} = 0.996$. The results are averaged over 100 instances of random circuit instances, error bars indicate one standard deviation. The more channels are expanded, the higher the reduction in the sampling overhead γ is achieved, however at an increase in the classical preprocessing cost.

depth in Fig. 4. In contrast to the direct multiplication of individual channels, we first construct the global inverse channel in product form, which we expand term by term to reduce the sampling overhead. We find that for the SPL model the XI reduction leads to the vanishing of terms which contain only Z and I operations, which we denote as *passive reduction*. The details on how pPEC is applied to the SPL noise model are presented in Appendix E.

It is important to note that the ordering in which this expansion is performed is crucial to obtain meaningful reductions. Due to the large sampling overhead for the considered device noise we limit the investigation to 10 qubits. The data for higher qubit counts and lower error rates – as expected for future generations of hardware – is presented in Fig. 5.

Fig. 4 shows a clearly favorable scaling relative to direct PEC, reaching an improvement of an order of magnitude even for modest depths. On the other hand a direct comparison to Fig. 5 shows that the effect of pPEC is far less drastic for lower error rates. We attribute this to the fact, that the amount of interference is generally proportional to the magnitude of the channel coefficients $(1 - w_k)$. For lower error rates the magnitude will not be as large and thus deeper circuits are required to achieve comparable results. However, even for larger circuits and modest depths a noticeable reduction in the sampling overhead can be achieved. We therefore conjecture that pPEC is a practical approach to reduce the amount of quantum resources, as it is always favorable in comparison to direct PEC.

We now consider possible relevant applications in which pPEC might be applicable and help reduce the workload of the quantum computer.

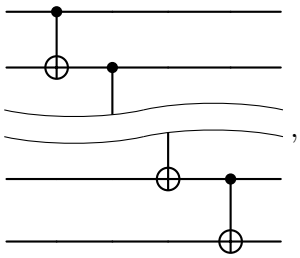
5 Applications

As described in the prior sections, the presented method is best suited for quantum circuits consisting of Clifford gates. While Clifford circuits themselves are not of any practical relevance due to their efficient simulability [36], larger Clifford structures do occur in many quantum circuits. In this section we give a brief overview of some possible applications of the proposed method.

5.1 Clifford structures in quantum circuits

5.1.1 Clifford building blocks

Clifford sub-circuits frequently occur in quantum circuits. In this section, we focus on two fundamental building blocks that often arise in quantum algorithms; the C_X-ladder as well as layers composed of SWAP gates. The C_X-ladder is commonly employed to implement multi-qubit rotations, which are integral to many circuits, as for example the Quantum Approximate Optimization Algorithm (QAOA) [37]. The general structure is given as



(21)

which consists of several C_X gates applied in series. Due to the sequential structure, the gates cannot be easily applied in parallel, rendering them ideal for a pPEC based approach.

Another often occurring Clifford building block is given by SWAP layers. For devices with sparse qubit connectivities, as for example superconducting quantum computers, interactions of

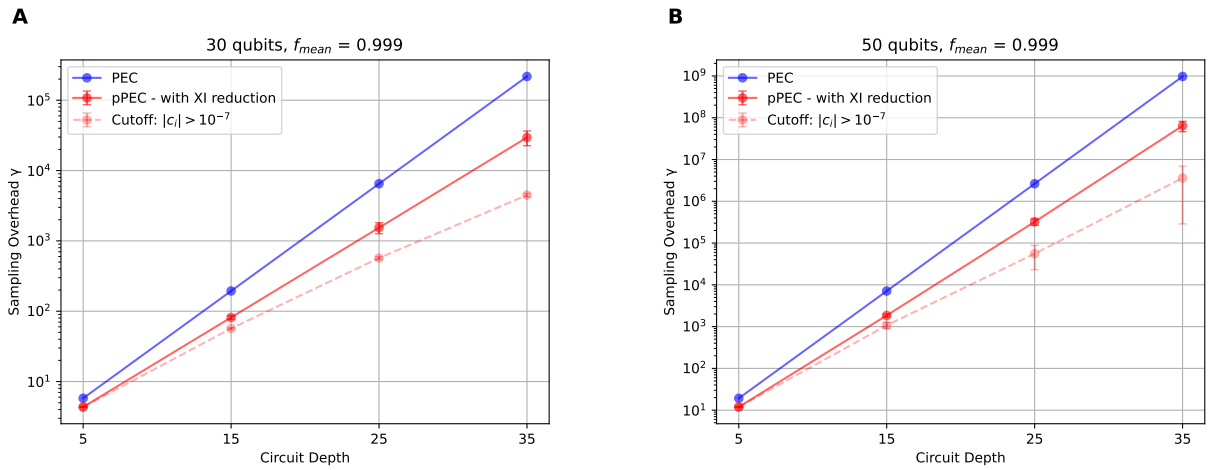
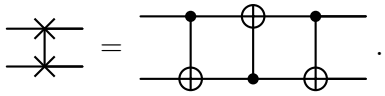


Figure 5: Performance of pPEC with XI reduction for future devices at 30 qubits (**A**) and 50 qubits (**B**). The results are averaged over 100 random circuit instances with error bars indicating one standard deviation. The amount of expanded channels was taken dynamically, ranging from 0 (lowest depth) to 200 (highest depth). The dashed curves show approximate versions of pPEC where channel coefficients with a magnitude smaller the 10^{-7} have been cut off. Since the reduction greatly reduces the number of channels, the number of expansion steps was increased by a factor of 4 for all calculations. While this method is more computationally efficient it generally comes at the cost of a bias.

physical qubits that are spatially separated need to be bridged by SWAP gates. SWAP gates are elements of the Clifford group as they can be decomposed to three consecutive C_X-gates:



(22)

Since these gates cannot be applied in parallel each C_X-gate would contribute with an individual noise channel Λ_i, which can be reduced with pPEC. Furthermore, SWAP gates are favorable for pPEC because the conjugation of a model Pauli does not change the weight of the Pauli string, i.e., the amount of non-identity terms in a Pauli string stays invariant (see Appendix B).

5.1.2 Reducing state preparations in VQE

A further interesting possible application for near-term quantum algorithms is the minimization of the number of state preparations in the Variational Quantum Eigensolver (VQE) [38]. The VQE algorithm aims to estimate the minimum eigenvalue of a given hermitian matrix H by preparing a parametrized state $|\psi(\theta_1, \theta_2, ..., \theta_n)\rangle = |\psi(\boldsymbol{\theta})\rangle$ and iteratively minimizing the measured energy

$$E = \min_{\boldsymbol{\theta}} \langle\psi(\boldsymbol{\theta})| H |\psi(\boldsymbol{\theta})\rangle \tag{23}$$

by optimizing the parameters of the trial wave function. The optimization is performed via a classical optimization loop, whereas the quantum computer is used to estimate the energy of each generation of parameters.

A limitation of this method is the required number of state preparations needed to measure a single instance of $\langle\psi(\boldsymbol{\theta})| H |\psi(\boldsymbol{\theta})\rangle$. Usually the matrix H is expressed as a weighted Pauli sum $\sum_j c_j P_j$ so that the total energy estimate reduces to estimating each individual Pauli expectation value separately. For a large number of Pauli terms the number of required measurements, and therefore state preparations, can rapidly become prohibitive.

To circumvent this problem, Ref. [39] introduces a method which aims to reduce the number of state preparations by grouping the Pauli operators into commuting groups, which can be measured in parallel. To measure the operators using a single state preparation,a basis change is performed, which is appended to the state-preparation circuit. The basis change circuit consists entirely of Clifford gates, as is expected for a mapping from Pauli operators to Pauli operators rendering these circuits ideal for pPEC. As an example, the mapping of the mutually commuting Pauli strings $XXXX$, $XXYY$, $XYXY$, $YXXY$ to the basis $ZIII$, $IZII$, $IIZI$, $IIIZ$ is depicted in Fig. 6. With the prior described noise model, with an average Pauli fidelity of $f_{\text{avg}} = 0.996$ on a linear device topology, we find an approximate $\gamma = 12.696$ for regular PEC.

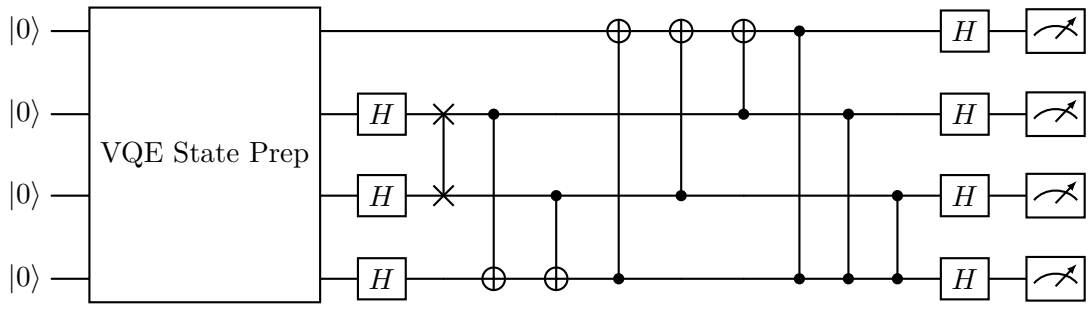


Figure 6: Example circuit consisting of a state preparation and change into the simultaneous eigenbasis of the Pauli operators $XXXX$, $XXYY$, $XYXY$ and $YXXY$ as presented in Ref. [39]. The state preparation prepares a possible ground state wave function according to some ansatz. The measurement basis is then rotated by a Clifford circuit to a basis where the observables of interest can be measured in parallel.

In contrast, our method results in a γ-factor of 7.335 for pPEC without the XI-reduction as well as $\gamma_{\mathrm{pPEC}_{XI}} = 5.449$ with XI reduction, which is about an order of magnitude improvement in the variance γ^2.

5.2 Measurement-based quantum computing

The application of pPEC within the model of measurement-based quantum computing (MBQC) [40] is another interesting possibility. In MBQC the computation is performed on a large, entangled resource state (in this example a graph state), realized only by Clifford operations

$$|G\rangle = \prod_{i,j \in E} C_{Zij} \, |+\rangle^{\otimes n} . \tag{24}$$

The state is constructed according to a graph $G = (V, E)$, where the vertices correspond to the individual qubits and the edges define which qubits are entangled via C_Z gates. To perform the computation, the individual qubits are measured in adaptive bases, performed by the transformation $M(\theta) = HR_z(\theta)$ which is equivalent to the implementation of a gate operation on the resource state. The randomness of a quantum measurement is compensated for by using classically controlled feed-forward operations.

We consider a two-dimensional lattice as shown in Fig. 7 as a resource state, which is universal for MBQC [41]. Such a graph state can be implemented in constant depth, with a minimum depth of 4 if the device is sufficiently connected. We again consider a device with a limited, linear connectivity and an average Pauli fidelity of $f_{\mathrm{avg}} =$

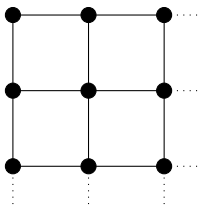


Figure 7: Depiction of the cluster state that can be used as a ressource for universal quantum computations in MBQC.

0.996. For the circuit to implement a graph state described by a 4×4 lattice (transpiled onto the hardware connectivity) we find for regular PEC a γ-factor of approximately $\gamma_{\mathrm{PEC}} \approx 4899.6$. Using pPEC we are able to reduce this factor to approximately $\gamma_{\mathrm{pPEC}_{XI}} \approx 839.1$. For improved fidelities $f_{\mathrm{avg}} = 0.9996$ and a 7×7 graph we find $\gamma_{\mathrm{PEC}} \approx 418.8$ and $\gamma_{\mathrm{pPEC}_{XI}} \approx 324.9$ which is a still considerable improvement of about 20%.

Our method is especially well suited for MBQC interesting since the resource state contains all noisy two-qubit operations while only consisting of Clifford gates. Thus, pPEC including the XI-reduction can be applied to the resource state preparation yielding favorable scaling in comparison to PEC. Further, pPEC can be applied to correct for the readout errors that occur during the measurement process. Note, that due to the feed-forward operation, assignment matrix methods can in general not be utilized, rendering a PEC approach more practical [42]. Considering the structure of the circuits and the fact that symmetrized measurement errors are defined by an X-channel, we can propagate these errors to the

start of the circuit. For example an X-correction before the measurement can be propagated behind the basis change operation $M(\theta)$ since

$$M(\theta)X = R_z(\theta)HX = ZM(\theta), \tag{25}$$

where we utilized that Pauli Z commutes with the $R_z(\theta)$ gate. Measurement errors can thus be freely propagated throughout the circuit, enabling their incorporation into the pPEC workflow.

5.3 Further possible applications

Apart from the circuits discussed, the method can be applied whenever Clifford circuits are present in circuits of practical interest. Potential algorithms include Clifford preconditioning of circuits [43] and applications of instantaneous quantum polynomial (IQP) circuits [44], utilizing partial propagation of Pauli errors that commute with the diagonal gates in the IQP circuit. While we do not explicitly investigate these approaches in this work, we consider them as promising avenues for further research.

6 Beyond Clifford Circuits

So far, we focused our approach to Clifford circuits. This choice is motivated by the fact that Clifford circuits propagate Pauli channels in a well-defined manner without introducing *branching*, allowing for the propagation of channels in deep (Clifford) circuits. Branching refers to the phenomenon where the commutation of a Pauli operator with a non-Clifford gate typically results in one Pauli term being mapped to a sum of Pauli terms. This effect arises from Heisenberg evolution and can be expressed as

$$G^\dagger P G = \sum_{P' \in \{X,Y,Z\}^{\otimes n}} \mathrm{Tr}(G^\dagger P' G) P' \tag{26}$$

where the summation is taken over the Pauli group of dimension n.

Generally, pPEC could thus be performed by tracking the evolution of the corrections as sums of Pauli terms. The problems with this approach are twofold. First, the number of Pauli operators, that need to be tracked, increases exponentially with the number of non-Clifford gates, rendering this method computationally expensive. Second, single correction operations can be mapped to large Pauli sums with highly non-local operators which cannot be implemented as simple single qubit Pauli gates. A probabilistic implementation of such operators using multi qubit gates might thus introduce significantly more errors into the system.

While both these issues can in principle be mitigated to some extent by either truncating small coefficients in Eq. (26) to keep the Pauli sums small [45] or by expressing the channel as a tensor network truncated to bond dimension one to only allow single qubit corrections [25], they still pose significant challenges.

An alternative method for handling non-Clifford gates is to restrict the non-Clifford gates to single-qubit Pauli rotation gates $R_P(\theta)$, which adhere to the general commutation rule

$$R_{P'}(\theta)P = PR_{P'}(-1^{\langle P,P'\rangle_{\mathrm{SP}}}\theta) \tag{27}$$

with $P, P' \in \{I, X, Y, Z\}$ and $\langle P, P'\rangle_{\mathrm{SP}}$ denoting the symplectic inner product of P and P', which is 0 when the Paulis commute and 1 otherwise. That is, the commutation will flip the sign of the rotation angle whenever P and P' do not commute.

Although this approach also introduces branching in the context of the rotation angles, a Pauli channel will still be mapped to a Pauli channel, enabling direct sampling of single-qubit Pauli corrections. Implementing this method requires to keep track of each sign flip for each correction separately and flipping the corresponding signs of the rotation angles before executing the circuit if the correction is sampled.

Since corrections with the same operator but different set of flips do not interfere, this will lead to generally lower reductions. The main hurdle introduced by keeping track of the signs stems from the fact that the number of possible corrections now scales exponentially in the circuit depth as well as the circuit width. However, this method might be feasible for low-magic circuits.

Simulation of the transverse-field Ising model

We conclude the section with considering a specific non-Clifford application, namely the simulation of the transverse-field Ising model using nearest neighbor interactions considered by Ref. [20], which describes magnetic spin-spin interactions

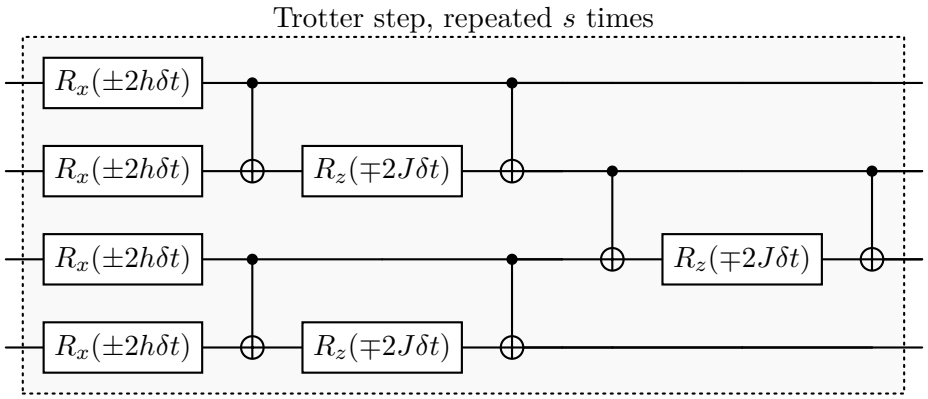


Figure 8: Circuit representation of the time evolution operator of the transverse field Ising model. The circuit shows a single Trotter step s for a model consisting of four qubits which is repeated several times until the circuit is measured. The $\pm$ signs in the circuit indicate that we flip the sign of the rotation angles relative to the Pauli correction that has been sampled.

with a local transverse field applied to each spin

$$H_{\text{Ising}} = -\sum_j J_{j,j+1} Z_j Z_{j+1} + \sum_j h_j X_j. \quad (28)$$

The time evolution induced by this Hamiltonian can be implemented on a gate-based quantum computer by utilizing the first order Trotter-Suzuki decomposition [46, 47] as

$$e^{-iHt} \approx (e^{iJ\sum_j Z_j Z_{j+1}\delta t} e^{-ih\sum_j X_j \delta t})^s, \quad (29)$$

where $\delta t = t/s$ is the total evolution time divided by the number of Trotter steps s and we assume equal coupling among all qubits, i.e., $h_i = h$ and $J_{ij} = J$. The implementation of the model on a gate-based quantum computer can then be performed by a set of single site rotation gates $R_X(2h\delta t)$ for the evolution under the transverse-field dynamics and controlled Z rotations, expressed as a rotation gate $R_Z(-2J\delta t)$ between two C_X gates accounting for the spin-spin interactions. We illustrate the circuit for a single Trotter step in Fig. 8. As demonstrated in Ref. [35], we implement the dynamics of a four qubit Hamiltonian with the parameters $h = 1, J = 0.15$ and $\delta t = 1/4$. To estimate the efficiency of our method we measure the magnetization $\mathbf{M} = \sum_i \langle X_i \rangle, \langle Y_i \rangle, \langle Z_i \rangle / N$, which represents the average polarization along a specified direction.

We simulate the dynamics over 16 Trotter steps while incorporating gate-level noise. Specifically, we apply a random 2-qubit Pauli error channel with an error probability $p = 0.01$ after each C_X gate. The experiment consists of a noiseless simulation, a noisy simulation, and corrections using PEC and pPEC. To calculate the pPEC inverse we perform a channel propagation for each inverse channel, tracking the sign of each passed rotation gate for each correction. In the expansion step we truncate the calculation of the channel products by only considering coefficients $|c_i| > 10^{-6}$ to keep the classical overhead small.

The results, presented in Fig. 9, illustrate the average magnetization of the four-qubit test circuit in the y-z plane. Both PEC and pPEC successfully retrieve the bias-free expectation values within the error bars for all data points. However, unlike the Clifford case, we observe a less significant reduction in the γ-factor, approximately 10% for the last Trotter step. We finally want to highlight that the circuit contains a higher number of non-Clifford gates compared to Clifford gates, making it one of the more challenging circuits for pPEC. While the results are far less drastic than the results for Clifford circuits, pPEC is still able to achieve a meaningful reduction.

7 Discussion and Conclusion

Quantum error mitigation is a promising avenue to achieve near-term quantum advantage. Even though tremendous progress in hardware error rates has been achieved, quantum error mitigation is still vital to obtain meaningful results on current hardware. However, one of the limiting

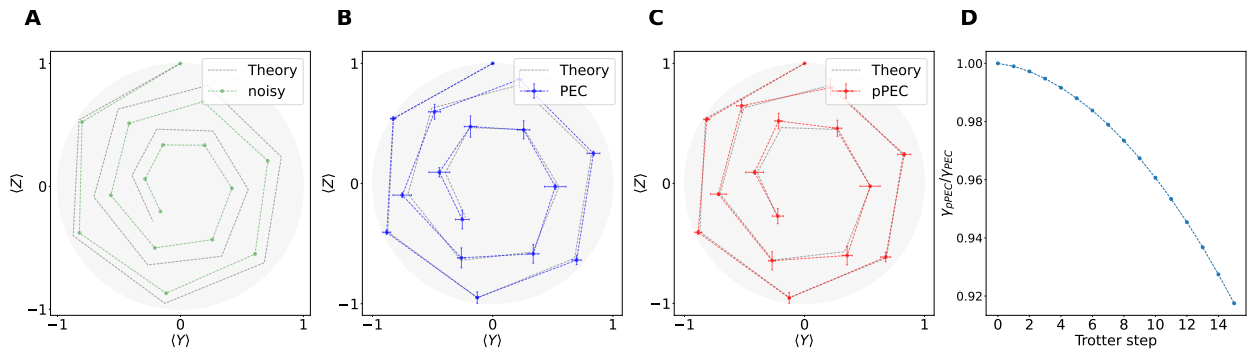


Figure 9: Dynamics of the average magnetization $\mathbf{M} = \sum_i \langle X_i \rangle, \langle Y_i \rangle, \langle Z_i \rangle / N$ for a linear chain of 4 qubits for an increasing number of Trotter steps. The results show the noisy simulation (**A**) the PEC corrected expectation value (**B**) and the pPEC corrected expectation value (**C**). In the experiment a total of 200 expectation values per data point has been taken. The error bars show one standard deviation of the estimated data. (**D**) sampling overhead reduction $\gamma_{\text{pPEC}}/\gamma_{\text{PEC}}$ relative to the number of Trotter steps.

factors for practical error mitigation still lies in the excessive amount of quantum resources required to mitigate large circuits.

In this work we presented a method to, in some cases drastically, reduce the sampling overhead of PEC for Clifford circuits. PEC in combination with Pauli error propagation is able to retrieve bias free expectation values of noisy circuits. The method itself delivers excellent results for current hardware error rates, reducing the sampling overhead by a few orders of magnitude for even moderately deep quantum circuits. We observed that the effect is far less drastic for reduced error rates, which are expected for future quantum device generations. However, even for future hardware error rates and quantum circuits consisting of 30-50 qubits a reduction of one to two orders of magnitude can be achieved, making pPEC an attractive choice whenever applicable.

We hope that the promising results of sampling from a fused inverse could spark further insights in the most effective form of canceling quantum noise. How to optimally estimate and sample from the global inverse is an interesting question for future research.

The clear bottleneck of our algorithm is the classical overhead to calculate the channel products. We showed that the expansion can be performed without tremendous overhead by truncation, yielding more limited results. A further interesting approach to efficiently expand the noise channel could be given by tensor network algorithms, which can approximate a large amount of data with few resources. By expressing the channels as matrix product operators (MPO) the channel products can be calculated as a contraction of several MPOs [48, 49]. The main problem lies in the fact that, to keep the bond dimension small, approximations need to be made which can lead to a bias in the estimated expectation values.

We presented basic examples of potential applications for our method, where Clifford circuits, despite being efficiently simulatable via the Gottesman-Knill theorem, are of practical interest. Finally, we tested our method on a non-Clifford example: the simulation of the transverse field Ising model. While the proposed method successfully retrieves the ideal value from noisy data, the reduction in sampling overhead is far lower than that observed for Clifford circuits. We find the exploration of circuits with low magic or measurement-based quantum computation (MBQC), particularly with respect to qubit recycling, to be especially promising avenues for further research.

Acknowledgements

This work was in part supported by the research project Zentrum für Angewandtes Quantencomputing (ZAQC), which is funded by the Hessian Ministry for Digital Strategy and Innovation and the Hessian Ministry of Higher Education, Research and the Arts. We acknowledge further support by the research project Leistungszentrum für innovative Therapeutika (TheraNova), which is funded by the Fraunhofer Gesellschaft and the Hessian Ministry for Higher Education, Research and the Arts.

References

[1] Andrew J Daley, Immanuel Bloch, Christian Kokail, Stuart Flannigan, Natalie Pearson, Matthias Troyer, and Peter Zoller. Practical quantum advantage in quantum simulation. *Nature*, 607(7920):667–676, 2022. DOI: 10.1038/s41586-022-04940-6.

[2] Alberto Di Meglio, Karl Jansen, Ivano Tavernelli, Constantia Alexandrou, Srinivasan Arunachalam, Christian W Bauer, Kerstin Borras, Stefano Carrazza, Arianna Crippa, Vincent Croft, et al. Quantum computing for high-energy physics: State of the art and challenges. *Prx quantum*, 5(3):037001, 2024. DOI: 10.1103/PRXQuantum.5.037001.

[3] Craig Gidney and Martin Ekerå. How to factor 2048 bit rsa integers in 8 hours using 20 million noisy qubits. *Quantum*, 5:433, 2021. DOI: 10.22331/q-2021-04-15-433.

[4] Markus Reiher, Nathan Wiebe, Krysta M Svore, Dave Wecker, and Matthias Troyer. Elucidating reaction mechanisms on quantum computers. *Proceedings of the national academy of sciences*, 114(29):7555–7560, 2017. DOI: 10.1073/pnas.1619152114.

[5] Sergey Bravyi, Andrew W Cross, Jay M Gambetta, Dmitri Maslov, Patrick Rall, and Theodore J Yoder. High-threshold and low-overhead fault-tolerant quantum memory. *Nature*, 627(8005):778–782, 2024. DOI: 10.1038/s41586-024-07107-7.

[6] Qian Xu, J Pablo Bonilla Ataides, Christopher A Pattison, Nithin Raveendran, Dolev Bluvstein, Jonathan Wurtz, Bane Vasić, Mikhail D Lukin, Liang Jiang, and Hengyun Zhou. Constant-overhead fault-tolerant quantum computation with reconfigurable atom arrays. *Nature Physics*, pages 1–7, 2024. DOI: 10.1038/s41567-024-02479-z.

[7] Dolev Bluvstein, Simon J Evered, Alexandra A Geim, Sophie H Li, Hengyun Zhou, Tom Manovitz, Sepehr Ebadi, Madelyn Cain, Marcin Kalinowski, Dominik Hangleiter, et al. Logical quantum processor based on reconfigurable atom arrays. *Nature*, 626(7997):58–65, 2024. DOI: 10.1038/s41586-023-06927-3.

[8] VV Sivak, Alec Eickbusch, Baptiste Royer, Shraddha Singh, Ioannis Tsioutsios, Suhas Ganjam, Alessandro Miano, BL Brock, AZ Ding, Luigi Frunzio, et al. Real-time quantum error correction beyond break-even. *Nature*, 616(7955):50–55, 2023. DOI: 10.1038/s41586-023-05782-6.

[9] Google Quantum AI. Suppressing quantum errors by scaling a surface code logical qubit. *Nature*, 614(7949):676–681, 2023. DOI: 10.1038/s41586-022-05434-1.

[10] Google Quantum AI. Quantum error correction below the surface code threshold. *Nature*, 638(8052):920–926, 2025. DOI: 10.1038/s41586-024-08449-y.

[11] Alexis Morvan, B Villalonga, X Mi, S Mandrà, A Bengtsson, PV Klimov, Z Chen, S Hong, C Erickson, IK Drozdov, et al. Phase transitions in random circuit sampling. *Nature*, 634(8033):328–333, 2024. DOI: 10.1038/s41586-024-07998-6.

[12] Youngseok Kim, Andrew Eddins, Sajant Anand, Ken Xuan Wei, Ewout Van Den Berg, Sami Rosenblatt, Hasan Nayfeh, Yantao Wu, Michael Zaletel, Kristan Temme, et al. Evidence for the utility of quantum computing before fault tolerance. *Nature*, 618(7965):500–505, 2023. DOI: 10.1038/s41586-023-06096-3.

[13] Kazuya Shinjo, Kazuhiro Seki, Tomonori Shirakawa, Rong-Yang Sun, and Seiji Yunoki. Unveiling clean two-dimensional discrete time quasicrystals on a digital quantum computer. *arXiv preprint arXiv:2403.16718*, 2024. DOI: 10.48550/arXiv.2403.16718.

[14] Piotr Czarnik, Andrew Arrasmith, Patrick J Coles, and Lukasz Cincio. Error mitigation with clifford quantum-circuit data. *Quantum*, 5:592, 2021. DOI: 10.22331/q-2021-11-26-592.

[15] Kristan Temme, Sergey Bravyi, and Jay M Gambetta. Error mitigation for short-depth quantum circuits. *Physical review letters*, 119(18):180509, 2017. DOI: 10.1103/physrevlett.119.180509.

[16] Zhenyu Cai, Ryan Babbush, Simon C Benjamin, Suguru Endo, William J Huggins, Ying Li, Jarrod R McClean, and Thomas E O'Brien. Quantum error mitigation. *Reviews of Modern Physics*, 95(4):045005, 2023. DOI: 10.1103/revmodphys.95.045005.

[17] Yihui Quek, Daniel Stilck França, Sumeet Khatri, Johannes Jakob Meyer, and Jens Eisert. Exponentially tighter bounds on lim-

itations of quantum error mitigation. *Nature Physics*, 20(10):1648–1658, 2024. DOI: 10.1038/s41567-024-02536-7.

[18] Ryuji Takagi, Suguru Endo, Shintaro Minagawa, and Mile Gu. Fundamental limits of quantum error mitigation. *npj Quantum Information*, 8(1):114, 2022. DOI: 10.1038/s41534-022-00618-z.

[19] Suguru Endo, Simon C Benjamin, and Ying Li. Practical quantum error mitigation for near-future applications. *Physical Review X*, 8(3):031027, 2018. DOI: 10.1103/physrevx.8.031027.

[20] 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*, pages 1–6, 2023. DOI: 10.1038/s41567-023-02042-2.

[21] Minh C Tran, Kunal Sharma, and Kristan Temme. Locality and error mitigation of quantum circuits. *arXiv preprint arXiv:2303.06496*, 2023. DOI: 10.48550/arXiv.2303.06496.

[22] Prasanth Shyamsundar and Wern Yeen Yeong. Cv4quantum: Reducing the sampling overhead in probabilistic error cancellation using control variates. *arXiv preprint arXiv:2502.08735*, 2025. DOI: 10.48550/arXiv.2502.08735.

[23] Benchi Zhao, Mingrui Jing, Lei Zhang, Xuanqiang Zhao, Yu-Ao Chen, Kun Wang, and Xin Wang. Retrieving nonlinear features from noisy quantum states. *PRX Quantum*, 5(2):020357, 2024. DOI: 10.1103/prxquantum.5.020357.

[24] Jiaqing Jiang, Kun Wang, and Xin Wang. Physical implementability of linear maps and its application in error mitigation. *Quantum*, 5:600, 2021. DOI: 10.22331/q-2021-12-07-600.

[25] Yuchen Guo and Shuo Yang. Quantum error mitigation via matrix product operators. *PRX Quantum*, 3(4):040313, 2022. DOI: 10.1103/prxquantum.3.040313.

[26] Sergei Filippov, Matea Leahy, Matteo AC Rossi, and Guillermo García-Pérez. Scalable tensor-network error mitigation for near-term quantum computing. *arXiv preprint arXiv:2307.11740*, 2023. DOI: 10.48550/arXiv.2307.11740.

[27] Mathys Rennela and Harold Ollivier. Low bit-flip rate probabilistic error cancellation. *arXiv preprint arXiv:2411.06422*, 2024. DOI: 10.48550/arXiv.2411.06422.

[28] Joel J Wallman and Joseph Emerson. Noise tailoring for scalable quantum computation via randomized compiling. *Physical Review A*, 94(5):052325, 2016. DOI: 10.1103/physreva.94.052325.

[29] 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):5347, 2019. DOI: 10.1038/s41467-019-13068-7.

[30] Senrui Chen, Yunchao Liu, Matthew Otten, Alireza Seif, Bill Fefferman, and Liang Jiang. The learnability of pauli noise. *Nature Communications*, 14(1):52, 2023. DOI: 10.1038/s41467-022-35759-4.

[31] Sergey Bravyi, Sarah Sheldon, Abhinav Kandala, David C Mckay, and Jay M Gambetta. Mitigating measurement errors in multiqubit experiments. *Physical Review A*, 103(4):042605, 2021. DOI: 10.1103/physreva.103.042605.

[32] Paul D Nation, Hwajung Kang, Neereja Sundaresan, and Jay M Gambetta. Scalable mitigation of measurement errors on quantum computers. *PRX Quantum*, 2(4):040326, 2021. DOI: 10.1103/prxquantum.2.040326.

[33] Lena Funcke, Tobias Hartung, Karl Jansen, Stefan Kühn, Paolo Stornati, and Xiaoyang Wang. Measurement error mitigation in quantum computers through classical bit-flip correction. *Physical Review A*, 105(6):062404, 2022. DOI: 10.1103/physreva.105.062404.

[34] Bo Yang, Rudy Raymond, and Shumpei Uno. Efficient quantum readout-error mitigation for sparse measurement outcomes of near-term quantum devices. *Physical Review A*, 106(1):012423, 2022. DOI: 10.1103/physreva.106.012423.

[35] Ewout Van Den Berg, Zlatko K Minev, and Kristan Temme. Model-free readout-error mitigation for quantum expectation values. *Physical Review A*, 105(3):032620, 2022. DOI: 10.1103/physreva.105.032620.

[36] Daniel Gottesman. The heisenberg representation of quantum computers. 1998. DOI: 10.48550/arXiv.quant-ph/9807006.

[37] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum approximate optimization algorithm. *arXiv preprint arXiv:1411.4028*, 2014. DOI: 10.48550/arXiv.1411.4028.

[38] 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(1):4213, 2014. DOI: 10.1038/ncomms5213.

[39] Pranav Gokhale, Olivia Angiuli, Yongshan Ding, Kaiwen Gui, Teague Tomesh, Martin Suchara, Margaret Martonosi, and Frederic T Chong. Minimizing state preparations in variational quantum eigensolver by partitioning into commuting families. *arXiv preprint arXiv:1907.13623*, 2019. DOI: 10.48550/arXiv.1907.13623.

[40] Hans J Briegel, David E Browne, Wolfgang Dür, Robert Raussendorf, and Maarten Van den Nest. Measurement-based quantum computation. *Nature Physics*, 5(1):19–26, 2009. DOI: 10.1038/nphys1157.

[41] Robert Raussendorf, Daniel E Browne, and Hans J Briegel. Measurement-based quantum computation on cluster states. *Physical review A*, 68(2):022312, 2003. DOI: 10.1103/physreva.68.022312.

[42] Riddhi S Gupta, Ewout Van Den Berg, Maika Takita, Diego Riste, Kristan Temme, and Abhinav Kandala. Probabilistic error cancellation for dynamic quantum circuits. *Physical Review A*, 109(6):062617, 2024. DOI: 10.1103/physreva.109.062617.

[43] Jiace Sun, Lixue Cheng, and Weitang Li. Toward chemical accuracy with shallow quantum circuits: A clifford-based hamiltonian engineering approach. *Journal of Chemical Theory and Computation*, 20(2):695–707, 2024. DOI: 10.1021/acs.jctc.3c00886.

[44] Dan Shepherd and Michael J Bremner. Temporally unstructured quantum computation. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 465(2105):1413–1439, 2009. DOI: 10.1098/rspa.2008.0443.

[45] Armando Angrisani, Alexander Schmidhuber, Manuel S Rudolph, M Cerezo, Zoë Holmes, and Hsin-Yuan Huang. Classically estimating observables of noiseless quantum circuits. *arXiv preprint arXiv:2409.01706*, 2024. DOI: 10.48550/arXiv.2409.01706.

[46] Hale F Trotter. On the product of semigroups of operators. *Proceedings of the American Mathematical Society*, 10(4):545–551, 1959. DOI: 10.2307/2033649.

[47] Masuo Suzuki. Generalized trotter's formula and systematic approximants of exponential operators and inner derivations with applications to many-body problems. *Communications in Mathematical Physics*, 51(2):183–190, 1976. DOI: 10.1007/bf01609348.

[48] Christopher J Wood, Jacob D Biamonte, and David G Cory. Tensor networks and graphical calculus for open quantum systems. *arXiv preprint arXiv:1111.6950*, 2011. DOI: 10.26421/qic15.9-10-3.

[49] Laurin E Fischer, Matea Leahy, Andrew Eddins, Nathan Keenan, Davide Ferracin, Matteo AC Rossi, Youngseok Kim, Andre He, Francesca Pietracaprina, Boris Sokolov, et al. Dynamical simulations of many-body quantum chaos on a quantum computer. *arXiv preprint arXiv:2411.00765*, 2024. DOI: 10.48550/arXiv.2411.00765.

A pPEC via MCMC method

Here, we give a description on how to estimate a global inverse noise channel using a Markov chain Monte Carlo (MCMC) simulation. While in general this approach is not scalable and applicable only to small system sizes it serves as an excellent educational example on how the interference of corrections reduces the sampling overhead.

The key insight to understand how the combination of channels reduces the sampling overhead lies in the interference of corrections. Considering for example a layer where a specific correction P_i has been sampled, it is possible that the effect of the correction is exactly canceled by a subsequent correction P_j, so that the same global correction could be achieved by performing no correction at all.

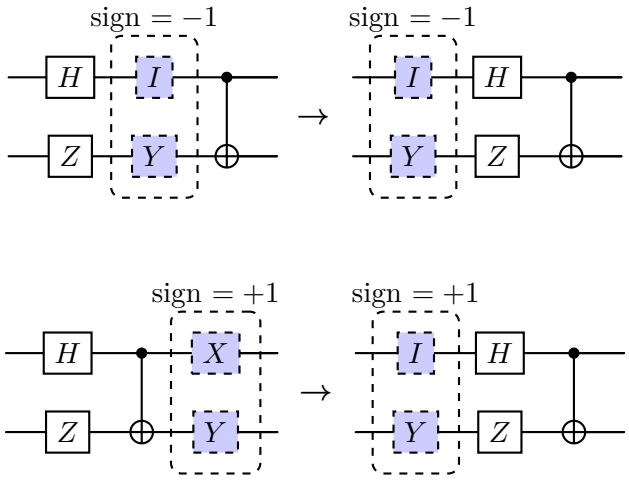


Figure 10: Example of interfering error paths. The drawn corrections (dashed blue gates) of the individual error paths amount, after the propagation, to the same correction although at a different sign. Calculating both paths is thus superfluous, since their contribution to the average can be easily accounted for in preprocessing.

A simple graphical illustration of interfering corrections is shown in Fig. 10. In this simplified model two corrections are drawn which lead to the same global correction, although at a different global sign. This pair of paths cancels each other in the average and can be omitted from sampling.

To quantify the amount of interference we utilize a MCMC algorithm that iterates reversely over each noisy layer in the circuit, drawing a specific Pauli correction P_l and corresponding sign s_l at each step. Consecutively, the drawn correction is propagated behind the next gate layer $P_l \to \tilde{P}_l$ using a look-up table (c.f. Appendix. B). At the next gate layer, a new Pauli correction P_{l-1} and sign s_{l-1} are drawn, which get multiplied with the propagated Pauli correction $P_{\text{prod}} = \tilde{P}_l P_{l-1}$ and the previous sign $s_{\text{prod}} = s_l s_{l-1}$. Then the product of the Pauli operators P_{prod} is propagated to the next layer, where the same process is repeated until the last noisy layer (or the initialization of the qubits) is reached.

As described in Sec. 3.1, at the initialization or the measurement layer, the corrections Y can be replaced by X and Z by I in each sampled final Pauli correction string via the XI-reduction.

Finally, each time a specific correction P_{final} was sampled, we save the number of occurrences and increment or decrement the index of draws depending on s_{final}. In this way, some corrections may interfere constructively, while others interfere destructively. Using this procedure, the reduction in the sampling overhead can be estimated by considering the amount of path interference

$$\frac{\gamma_{\text{ppec}}}{\gamma} = \frac{\#\text{ paths total} - \#\text{ interfering paths}}{\#\text{ paths total}}, \tag{30}$$

where a *path* refers to one iteration of the MCMC algorithm, in which the corrections have been propagated to the start or end of the circuit and a final global correction has been registered.

It is important to note that, due to the sign problem of the MCMC approach with quasi probability distributions, the number of Monte-Carlo samples needs to be scaled accordingly to address deep

circuits.

Algorithm 1 Mote-Carlo Simulation to estimate the global inverse

```
while n < samplesize do
    while layer index ≠ 0 do
        layer corr, sign ← sample correction and sign
        if layer index = N then
            global corr ← layer corr
            global sign ← sign
            Continue
        end if
        global corr ← propagate correction
        global corr ← global corr × layer corr
        global corr sign ← global sign × sign
        layer index ← layer index - 1
    end while
    Replace Y with X and Z with I in global correction
    add global sign × global correction to dictionary
end while
```

After the estimation of the global inverse, PEC can be performed by drawing from the estimated distribution obtained from the MCMC simulation. For a fixed number of circuit instances, a correction and corresponding sign are drawn, and the results are multiplied with the corresponding sign and averaged. Finally, the results need to be rescaled by the corresponding γ-factor. Due to the path interference, the total γ factor now amounts to

$$\gamma_{\text{ppec}} = \gamma \cdot \frac{\#\text{ samples} - \#\text{ interfering samples}}{\#\text{ samples}}, \tag{31}$$

showing the same structure as Eq. (14).

Note that due to the exponential scaling of the correction space, this approach requires an exponentially scaling amount of Monte-Carlo samples in order to converge. To circumvent prohibitive classical costs, the method can be truncated by for example only considering a select number of layers or performing the channel products analytically for a selected set of channels.

B Commutation Tables for Clifford Gates

In this section we showcase the basic commutation tables for the most used gates of the Clifford group of dimension $n = 1$ and $n = 2$. While it is generally true that each element of the Clifford group can be decomposed into H, S and C_X gates we provide commutation tables for a larger set of gates, since they are customary used in some decompositions or native gate sets of quantum hardware (such as the basis $\{X, S_x, R_z, C_X\}$).

C Conjoint implementation of inverse noise channels

We provide a simple prove that it is always favorable in terms of the sampling overhead to multiply two inverse Pauli channels together before probabilistically implementing them. We start the proof by considering a general inverse Pauli noise channel

$$\Lambda^{-1}(\rho) = \sum_{i=1}^{4^n} a_i P_i \rho P_i^\dagger = \sum_{i=1}^{4^n} a_i \mathcal{P}_i \quad a_i \in \mathbb{R} \tag{32}$$

C_X-Commutation table	
II	II
IX	XX
IY	XY
IZ	IZ
XI	XI
XX	IX
XY	IY
XZ	XZ
YI	YZ
YX	ZY
YY	ZX
YZ	YI
ZI	ZZ
ZX	YY
ZY	YX
ZZ	ZI

C_Z-Commutation table	
II	II
IX	ZX
IY	ZY
IZ	IZ
XI	XZ
XX	YY
XY	YX
XZ	XI
YI	YZ
YX	XY
YY	XX
YZ	YI
ZI	ZI
ZX	IX
ZY	IY
ZZ	ZZ

$SWAP$-Commutation table	
II	II
IX	XI
IY	YI
IZ	ZI
XI	IX
XX	XX
XY	YX
XZ	ZX
YI	IY
YX	XY
YY	YY
YZ	ZY
ZI	IZ
ZX	XZ
ZY	YZ
ZZ	ZZ

H-Commutation table	
I	I
X	Z
Y	Y
Z	X

S_X-Commutation table	
I	I
X	X
Y	Z
Z	Y

S_Y-Commutation table	
I	I
X	Z
Y	Y
Z	X

S_Z-Commutation table	
I	I
X	Y
Y	X
Z	Z

Table 2: Commutation tables for the most used gates of the Clifford group. The commutation tables where calculated via the relation $P' = CPC^\dagger$.

where $\mathcal{P}_i(\rho) \triangleq P_i \rho P_i^\dagger$ denotes the Kraus channel representation of a Pauli P_i. Considering that the sampling overhead γ is calculated as

$$\gamma(\Lambda^{-1}) = \gamma\left(\sum_{i=1}^{4^n} a_i \mathcal{P}_i\right) = \sum_{i=1}^{4^n} |a_i|, \tag{33}$$

it is easy to verify that the sampling overhead of implementing two inverse noise channels is given by

$$\gamma(\Lambda_1^{-1})\gamma(\Lambda_2^{-1}) = \sum_{i,j=1}^{4^n} |a_i||b_j|. \tag{34}$$

On the other hand, first multiplying both inverse channels and then calculating the sampling over-

head of the fused channel, yields

$$\begin{aligned}\gamma(\Lambda_1^{-1}\Lambda_2^{-1}) &= \gamma\left(\sum_{i=1}^{4^n} a_i\mathcal{P}_i \sum_{j=1}^{4^n} b_j\mathcal{P}_j\right)\\ &= \gamma\left(\sum_{k=1}^{4^n}\left(\sum_{i,j}^{4^n} a_i b_j \xi_{ijk}\right)\mathcal{P}_k\right) \quad \text{with} \quad \xi_{ijk} = \begin{cases}1 & \text{if } \mathcal{P}_i\mathcal{P}_j = \mathcal{P}_k\\ 0 & \text{otherwise}\end{cases}\\ &= \sum_{k}^{4^n}\left(|\sum_{i,j}^{4^n} a_i b_j \xi_{ijk}|\right)\\ &\leq \sum_{k}^{4^n}\left(\sum_{i,j}^{4^n} |a_i||b_j||\xi_{ijk}|\right)\\ &= \sum_{i,j=1}^{4^n} |a_i||b_j| = \gamma(\Lambda_1^{-1})\gamma(\Lambda_2^{-1}),\end{aligned} \tag{35}$$

where we made use of the triangle inequality as well as the fact that the total number of terms $a_i b_j$ needs to be conserved by the summation over k, i.e., ξ_{ijk} contains only one non-zero element for each product $\mathcal{P}_i\mathcal{P}_j$, $\sum_k |\xi_{ijk}| = 1\ \forall i,j$.

Thus, the cost of implementing two inverse channels separately is always greater or equal than the conjoint implementation. Note that in contrast to taking the product of two Pauli observables, the phase of the product of two Pauli channels can be omitted since $(e^{i\phi}P)\rho(e^{i\phi}P)^\dagger = P\rho P^\dagger$.

D Symmetrization of readout errors

We consider a readout error, which is described by a tensor product noise model

$$A = \bigotimes_{i=1}^{n} A^i = \bigotimes_{i=1}^{n}\begin{bmatrix}1-\epsilon_i & \eta_i\\ \epsilon_i & 1-\eta_i\end{bmatrix}, \tag{36}$$

such that the errors on each qubit occur independently. The measurement is described by a set of projectors $\pi_m = |m\rangle\langle m|$, which project onto the computational basis. The noisy projection operators are then given by

$$\tilde{\pi}_m = \sum_n A_{mn}|n\rangle\langle n| \tag{37}$$

where $A_{mn} = \langle m|A|n\rangle$ are the matrix elements of the error matrix. For the considered noise model the matrix A is sparse with only few non-vanishing matrix elements. The randomized insertion of X gates and subsequent correction by post-processing leads to a twirled readout error [35], which is given by

$$A_{\text{twirl}} = \frac{1}{2^n}\sum_{s\in\{0,1\}^n} X^{\otimes s} A X^{\otimes s,\dagger}, \tag{38}$$

where the $X^{\otimes s}$ operator corresponds to bit flips of the qubits indexed by s.

Since the noise model does not contain contributions of cross talk it suffices to consider each single

qubit error A^i twirled under a single X separately

$$\begin{aligned} A^i_{\text{twirl}} &= \frac{1}{2}(A^i + XA^iX^\dagger) \\ &= \frac{1}{2}\left(\begin{bmatrix} 1-\epsilon_i & \eta_i \\ \epsilon_i & 1-\eta_i \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1-\epsilon_i & \eta_i \\ \epsilon_i & 1-\eta_i \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\right) \\ &= \frac{1}{2}\left(\begin{bmatrix} 1-\epsilon_i & \eta_i \\ \epsilon_i & 1-\eta_i \end{bmatrix} + \begin{bmatrix} 1-\eta_i & \epsilon_i \\ \eta_i & 1-\epsilon_i \end{bmatrix}\right) \\ &= \begin{bmatrix} 1-p_x & p_x \\ p_x & 1-p_x \end{bmatrix} \\ &= (1-p_x)\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + p_x \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \end{aligned} \tag{39}$$

with $p_x = \frac{\epsilon_i+\eta_i}{2}$. Hence the randomized insertion of X gates transforms a general single qubit readout error channel into a Pauli X error channel.

We would like to note, that the method straight-forwardly generalizes to either the full readout error matrix or a model considering crosstalk of a subset of qubits. The integration of larger matrices comes at the cost of estimating each correlated measurement, but might generally reduce or even fully eliminate any bias.

E pPEC for SPL noise models

We now consider pPEC applied to sparse Pauli-Lindbladian (SPL) noise models. These models can be efficiently stored in a product representation of individual model Paulis which only model interactions between physically connected qubits. We depart from the product representation of an inverse Pauli channel as given in Ref. [20]:

$$\Lambda^{-1}(\rho) = \prod_{k\in\mathcal{K}} (w_k\mathcal{I}(\cdot) - (1-w_k)\mathcal{P}_k(\cdot))\rho. \tag{40}$$

Akin to the dense representation, the quasi probability implementation of the inverse channel is attached to a cost factor γ. For the SPL model the γ-factor is given by

$$\gamma = \prod_{k\in\mathcal{K}} (2w_k - 1)^{-1}, \tag{41}$$

which we aim to reduce using pPEC. Applying the pPEC method, each of the individually commuting Pauli channels are propagated to the start of the circuit by taking the conjugation with the circuit operation up to the select inverse channel,

$$\begin{aligned} \tilde{\Lambda}^{-1}(\rho) &= \prod_{k\in\mathcal{K}} (w_k\mathcal{C}(\mathcal{I})(\cdot) - (1-w_k)\mathcal{C}(\mathcal{P}_k)(\cdot))\rho \\ &= \prod_{k\in\mathcal{K}} (w_k\mathcal{I}(\cdot) - (1-w_k)\tilde{\mathcal{P}}_k(\cdot))\rho, \end{aligned} \tag{42}$$

where $\mathcal{C}$ denotes the Clifford circuit operation up to the respective inverse error channel and the operator $\tilde{\mathcal{P}}_k = CP_kC^\dagger \cdot C^\dagger P_k^\dagger C$ denotes the conjugated Pauli channel.

Applying this procedure to each correction layer, the global inverse is given by

$$\Lambda^{-1}_{\text{global}} = \prod_l \tilde{\Lambda}_l^{-1} = \prod_{k\in L\cdot\mathcal{K}} (w_k\mathcal{I}(\cdot) - (1-w_k)\tilde{\mathcal{P}}_k(\cdot))\rho \tag{43}$$

where the index k now runs over all model Paulis for the total number of layers L, yielding a total of $L\cdot|\mathcal{K}|$ terms. Again, as in the MCMC approach, a further reduction can be achieved by utilizing

the XI-reduction. In this model it suffices to replace the operators Y by X and Z by I in each of the individual Pauli channels in the product. As an immediate improvement all channels, which contain only Z and I Pauli operators can be directly omitted from the inverse noise channel and do not contribute to the sampling overhead, since

$$(w_k\mathcal{I} - (1-w_k)\mathcal{I}) = (2w_k - 1)\mathcal{I} \tag{44}$$

directly cancels with the corresponding term $(2w_k - 1)^{-1}$ in the γ factor. We denote this as *passive reduction*.

The global inverse noise channel will in general contain multiple Paulis that have been mapped to the same operator. The amount of terms in the product can then be directly reduced by multiplying terms with equal Paulis together

$$\begin{aligned}(w_1\mathcal{I} - (1-w_1)\mathcal{P}) \cdot (w_2\mathcal{I} - (1-w_2)\mathcal{P}) &= (w_1w_2 + (1-w_1)(1-w_2))\mathcal{I} - (w_1(1-w_2)) + (w_2(1-w_1))\mathcal{P} \\ &= w_3\mathcal{I} - (1-w_3)\mathcal{P}.\end{aligned}, \tag{45}$$

with $w_3 = (w_1w_2 + (1-w_1)(1-w_2))$. Since

$$(2w_1 - 1)^{-1}(2w_2 - 1)^{-1} = (2w_3 - 1)^{-1}, \tag{46}$$

this multiplication does not reduce the sampling overhead γ but can drastically reduce the number of total terms in the product.

As explained in Ref. [20] a further decrease in the sampling overhead can be achieved by explicitly expanding the product of Eq. (43)

$$\Lambda^{-1}(\rho) = \prod_{k\in\mathcal{K}} (w_k\mathcal{I}(\cdot) - (1-w_k)\mathcal{P}_k(\cdot))\rho \to \sum_i c_i\mathcal{P}_i(\rho). \tag{47}$$

Due to the multiplication of equal Pauli channels as well as the larger number of terms in this product, this expansion will yield far more drastic reductions than the expansion of an individual inverse channel given by Eq. (40).

In the worst case, the expansion will result in an exponential growth in the number of required parameters to characterize the noise channel. This leads to a general trade-off between required classical and quantum resources. However, the growth in the number of required coefficients can be reduced by utilizing specific term orderings before the expansion.

F Efficient expansion of SPL noise models

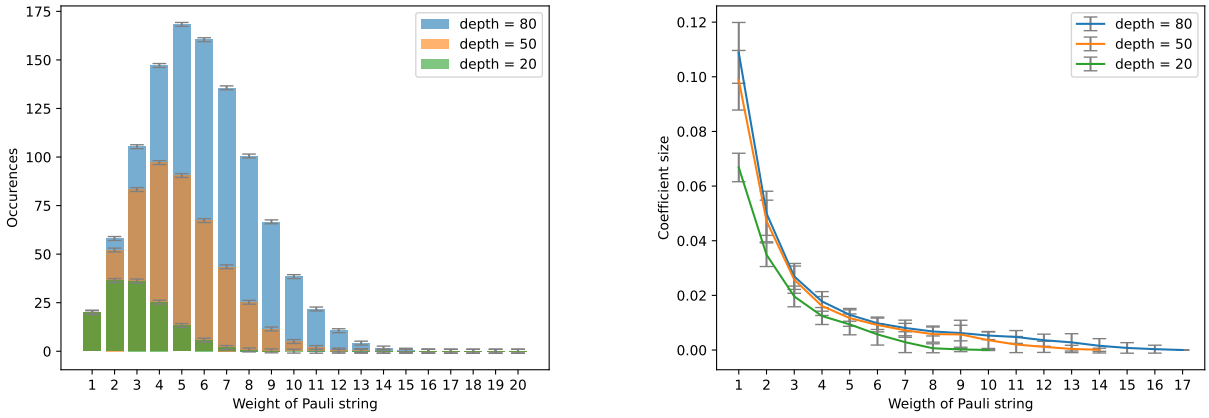
As stated in Ref. [20] the product structure of the noise model in SPL form can be explicitly expanded to reduce the sampling overhead γ, although at a computational cost in compute time as well as memory.

The number of required parameters and multiplications for the model scales quasi-exponentially, meaning that each term, that is absorbed into the sum, can in theory double the number of required coefficients. To evade this exponential increase, the expansion can be truncated cutting the multiplication off at a selected index

$$\Lambda^{-1}(\rho) = \prod_{k'\in\mathcal{K}'} (w_{k'}\mathcal{I}(\cdot) - (1-w_{k'})\mathcal{P}_{k'}(\cdot)) \left(\sum_i c_i\mathcal{P}_i(\cdot)\right)\rho. \tag{48}$$

The reduction in the sampling overhead can then be estimated by calculating the γ-factor of the expanded inverse channel

$$\gamma_{\text{reduction}} = \sum_i |c_i| \leq 1, \tag{49}$$



(a) (b)

Figure 11: Absolute value of coefficients and weights of the model Paulis for different circuit depths. The data is averaged over several instances of random Clifford circuits of 20 qubits. (a) Amount of Pauli strings relative to the number of non identity terms. For deeper circuits more Paulis with a higher weight are present. (b) Absolute value of the coefficient $|(1-w_k)|$ of the different Pauli strings after propagation relative to the amount of non identity terms in the Pauli string.

which is less or equal to 1 due to interference. While this approach is generally computationally more bearable, a direct expansion of the product, without exploiting structure, is still taxing and often inefficient. This inefficiency can be explained by considering a single expansion step

$$(w_i\mathcal{I}-(1-w_i)\mathcal{P}_i)\left(\sum_k c_k\mathcal{P}_k\right)=\left(w_i\sum_k c_k\mathcal{P}_k\mathcal{I}-(1-w_i)\sum_k c_k\mathcal{P}_k\mathcal{P}_i\right), \tag{50}$$

which leads to two separate sums, doubling the number of required coefficients. The amount of coefficients will however only double if no terms in the two sums are equal, that is no terms interfere. Luckily, this case is not of interest for pPEC, since the reduction in sampling overhead exactly stems from this interference.

One approach to maximize the amount of interference is the utilization of the subgroup structure of the Pauli operators. As a basic example, consider the expansion of the three single qubit Pauli channels

$$(w_1\mathcal{I}-(1-w_1)\mathcal{X})(w_2\mathcal{I}-(1-w_2)\mathcal{Y})(w_3\mathcal{I}-(1-w_3)\mathcal{Z})=(1-p_1-p_2-p_3)\mathcal{I}+p_1\mathcal{X}+p_2\mathcal{Y}+p_3\mathcal{Z} \tag{51}$$

which can be expanded without any increase in the number of coefficients. The cursive operators indicate that the operators act as channels in Kraus representation i.e. $\mathcal{X}=X(\cdot)X^\dagger$. The same holds trivially for any larger subgroup, as for example the 15 parameter group of $n=2$, the 63 parameter group of $n=3$ as well as any higher dimensional subgroup.

Further, the propagation of Paulis through deep circuits will lead to high weight Paulis as demonstrated in Fig. 11 (a). On the other hand, the size of the coefficients w_k is far smaller than those of low weight operators (Fig. 11 (b)) since it is highly unlikely that two corrections are mapped to the same high weight operator. Since the general reduction of the expansion is dependent on the magnitude of the channel coefficients, the absorption of low weight Paulis is more favorable.

Based on these heuristics we define a more efficient expansion of the product as follows. We start by ordering the terms in lexicographical ordering [33]

$$\begin{aligned} &III\preceq IIX\preceq IIY\preceq IIZ\\ \preceq\ &IXI\preceq IXX\preceq IXY\preceq IXZ\\ \preceq\ &IYI\preceq IYX\preceq IYY\preceq IYZ\preceq\dots \end{aligned} \tag{52}$$

and expanding the product of the first two terms. In each consecutive step we absorb a term only if its support is also in the continuously growing expanded inverse noise channel. If no Pauli with the same support is found, the next Pauli in the ordering is chosen.

Generally, after expanding the first truncated sum, the process can be iteratively applied to the remainder of the product, yielding several partially expanded sums

$$\Lambda^{-1}(\rho) = \prod_{k\in\mathcal{K}} (w_k\mathcal{I}(\cdot) - (1-w_k)\mathcal{P}_k(\cdot))\rho \to \left(\sum_{i_1} c_{i_1}\mathcal{P}_{i_1}(\cdot)\right)\left(\sum_{i_2} c_{i_2}\mathcal{P}_{i_2}(\cdot)\right)\ldots\left(\sum_{i_N} c_{i_N}\mathcal{P}_{i_N}(\cdot)\right)\rho. \quad (53)$$

Using this process the inverse channel can be expanded into multiple smaller sums, each with only a limited number of terms and total sampling overhead

$$\gamma_{\mathrm{pPEC}} = \gamma \prod_{j=1}^{N} \sum_{i} |c_{i_j}|. \quad (54)$$

While we do find that the lexicographical ordering leads to significant reductions it is possible that other orderings might lead to even more interference.