

Certification and Classification of Linear Quantum Error Mitigation Methods

Zach Blunden-Codd¹ and Mohamed Tamaazousti²

¹Université Grenoble Alpes, CEA, List, F-38000 Grenoble, France

²Université Paris-Saclay, CEA, List, F-91120, Palaiseau, France

Numerous mitigation methods exist for quantum noise suppression, making it challenging to identify the optimum approach for a specific application – especially as ongoing advances in hardware tuning and error correction are expected to reduce logical error rates. In order to facilitate the ‘future proof’ application-dependent comparison of mitigation methods, we develop a set of quantitative metrics that account for continual improvements in logical gate quality. We use these metrics to define qualitative criteria (e.g. scalability, efficiency, and robustness to characterised imperfections in the mitigation implementation), which we combine into application-specific certifications. We then provide a taxonomy of linear mitigation methods, characterising them by their features and requirements. Finally, we use our framework to produce and evaluate a mitigation strategy. A mitigation strategy is a collections of mitigation methods and compilation procedures designed to mitigate all relevant errors for a given piece of characterised hardware. Our example mitigation strategy is targeted at mitigating the outputs of hardware suffering from stochastic noise and/or rotational errors. We find the most significant determinant of efficient mitigation is accurate and precise characterisation.

Contents

1	Introduction	2
1.1	Context and Motivation	2
1.2	Contributions	3
1.3	Assumptions	4
1.4	Overview	5
2	Certification Scheme	5
2.1	Linear Mitigation	6
2.2	Circuit-Independent Noise Modelling	7
2.3	Quantitative Mitigation Method Metrics	9
2.4	Regimes	13
2.5	Qualitative Certification Criteria	14

Zach Blunden-Codd: zblundencodd@gmail.com

3	Characterisation of Noise	17
3.1	Characterisation	17
3.2	Circuit-Independent Noise Models	18
3.3	Stochastic Noise	19
3.4	Rotational Errors	21
3.5	(Stochastic) Over-Rotation Errors	22
4	Taxonomical Classification of Mitigation Methods	23
4.1	Nomenclature	23
4.2	Variant Generation Procedures	24
4.3	Pre-tailoring: Local Cancellation	29
4.4	Global Mitigation	30
4.5	Local Mitigation	33
4.6	Mitigation Method Catalogue	35
5	Case Study: Method Certification	38
5.1	Certification for Stochastic Noise	38
5.2	Certification for Rotational Errors	39
5.3	Numerical Analysis	42
6	Case Study: Strategy Certification	50
6.1	The USEM:ORE strategy	50
6.2	Method Certification	54
6.3	Strategy Certification	59
7	Concluding Remarks	59
7.1	Further Work	59
	References	60
	Appendices	63

1 Introduction

In this section we introduce and motivate our study (Sec. 1.1), highlight our contributions (Sec. 1.2), discuss the assumptions inherent to our study (Sec. 1.3), and provide an overview of the remainder of this document (Sec. 1.4). This document also has a companion piece: Ref. [1], which follows the same basic layout but has more extensive cross-referenced appendices that contain all the derivations required for this work. In App. B we provide a complete worked example for the mitigation of uniform dephasing errors, to aid comprehension.

1.1 Context and Motivation

Quantum devices inherently suffer from noise [2]. While unavoidable at present, one effective way to reduce its impact on computations is to employ error mitigation techniques (see [3, 4]). Error mitigation permits one to increase the accuracy of quantum calculations at the cost of a longer runtime (and/or increased numbers of qubits).

There are a great many mitigation methods available, all with different parameters, assumptions, requirements, trade-offs, etc. Our aim is to provide a more comprehensive

framework to facilitate the selection of mitigation methods for particular applications implemented on well-characterised hardware. We build on the existing literature to simplify the assessment of mitigation methods.

In Ref. [5] the authors discuss a trade-off between the sampling cost (due to the increased variance of mitigated circuits) and the increase in the runtime of an individual circuit (due to additional gates being used to effect the mitigation) for methods using identity insertion. We combine both these factors and the increases in the number of qubits required for the mitigation into a single metric: the runtime scaling, which measures the effective increase in total runtime due to the mitigation.

In Ref. [6] a lower bound is provided for the sampling cost of a variety of classes of mitigation methods as a function of the maximum bias for any observable measurement. We provide a strategy to calculate an upper bound for the post-mitigation bias without requiring quantum measurements of the observable, or simulation of the circuit.

In Ref. [7] they provide order of magnitude estimates for the sampling costs and fidelities of a variety of classes of mitigation methods. We provide finer grained measures of bias and sampling cost to permit the comparison of mitigation methods of the same class, such as different variants of zero-noise extrapolation or probabilistic error cancellation.

Error mitigation is only one of three main types of noise suppression:

1. **Hardware Tuning** [8]: Improving the quality of physical gates and qubits through improved manufacturing techniques, pulse shaping, lowering temperatures, feedback control, etc.
2. **Error Correction** [9]: Improving the quality of logical gates and qubits (and therefore outputs) for each run of the circuit through redundancy, encoding, real time intervention, etc.
3. **Error Mitigation** [3, 4]: Improving the quality of observable expectation values on average for an ensemble of runs through the use of symmetry verification, effective quasi-stochastic superpositions of gates, etc.

Although we focus on error mitigation, our framework is designed to be generalisable. It accounts for future reductions in the gate-wise noise amplitude of quantum circuits, due to hardware tuning and/or error correction, by giving results in terms of the overall noise level. The cost of error mitigation has been shown to grow exponentially with the number of gates (for stochastic noise with fixed gate-wise error probability, see Ref. [6]). This means that the only way to mitigate larger circuits efficiently is to reduce the gate-wise error probability, which means we can have the same noise level with a greater number of gates. The error probability could be reduced using error correction but in this case we should mitigate the logical qubits rather than physical qubits, in order to profit from the noise reduction.

It should be straightforward to extend this framework to incorporate some aspects of error correction, such as: increased numbers of qubits and increased execution time. In order to generalise the framework further, to incorporate the costs associated with hardware tuning, and the classical computing resources required for error correction, we might have to use energy or resource consumption as our metric of choice to characterise the cost of noise suppression (akin to Ref. [10]).

1.2 Contributions

1. **Taxonomy (Sec. 4)**: We introduce a taxonomy dedicated to linear mitigation methods for quantum noise suppression, categorising them according to key features such

as the use of hidden inverses, custom noise amplification channels, and identity insertion techniques. This taxonomy enables a systematic assessment of the utility and limitations of individual mitigation features, and provides a foundation for generalising and optimising existing methods.

2. **Certification (Sec. 2.3):** There are myriad mitigation methods in the literature and it can be difficult to determine the most appropriate method for a given application. We provide a selection of novel (proxy bias, noise boundary, runtime scaling, and area scale factor) and existing (sampling cost and length scale factor) quantitative metrics for the circuit-independent evaluation of linear mitigation methods (i.e. these metrics do not require numerical simulation of circuits or their implementation on a quantum computer). We then use these metrics to determine the compliance of a method with a set of qualitative criteria (robust, scalable, efficient, precise, etc.) that can be used to certify the method for a particular task.
3. **Case Studies:** We carry out the certification procedure for a selection of mitigation methods (Sec. 5) for stochastic noise (Sec. 3.3) and rotational errors (Sec. 3.4). We also introduce the idea of a mitigation strategy (combination of mitigation method, gate set, and compilation scheme; see Sec. 2.5.2) and certify it to be USEful (Sec. 2.5.2) if it is a scalable strategy that is efficacious for the production of the mitigated circuit for any quantum algorithm. Finally, we study the USEM:ORE strategy (Sec. 6.1) for the universal and scalable mitigation of hardware suffering from (stochastic) over-rotation errors (Sec. 3.5) and demonstrate that it complies with the USEful certification.

1.3 Assumptions

- **Linear Methods:** Our discussion is limited to linear mitigation methods (Sec. 2.1) since they are particularly amenable to circuit-independent analysis¹. There are many non-linear methods (such as virtual distillation [11–13] and symmetry verification [7]) that would also be worthy of study but are beyond the scope of this paper.
- **Noise Models:** Though we provide some relatively general analyses (see Sec. 2.3), our main focus is on stochastic noise (Sec. 3.3) and rotational errors (Sec. 3.4). So our conclusions should not be applied to other noise channels without further verification. We assume perfect characterisation of our noise models but in practice there will always be limitations to the precision of characterisation, which will affect the final bias.
- **Large Circuit Limit:** One of our core focuses is scalability. So a number of our criteria are only appropriate in the asymptotic regime of large circuits (Sec. 2.4.3). Care should be taken when applying our analysis to shorter circuits.
- **Small Noise Levels:** Our analysis of bias assumes the small noise level regime (Sec. 2.4.2). Additional analysis would be required to generalise to larger noise levels.

¹In this work an analysis is circuit independent if it depends only on coarse properties of the circuit such as the size or number of gates and not on the precise nature of the algorithm/gates being implemented.

- **Pauli Observables:** We assume that our measured observables are single qubit Pauli operators, this assumption can be imposed without loss of generality, but would require extra gates and post-processing for the measurement of arbitrary observables.
- **Gate Errors:** We focus on noise related to the application of gates but one could also apply some of these methods to SPAM (state preparation and measurement) errors, if one assigned the effective noise channels to virtual identity gates.
- **Uniform Noise Models:** For simplicity we assume that our noisy circuits have uniform noise. That is, every gate suffers from the same noise (see Sec. 2.2.2). However, our analysis can be generalised to non-uniform noise models (e.g. see Sec. 4.4.2).

1.4 Overview

The structure of this paper is the following:

1. **Certification Scheme (Sec. 2):** We provide quantitative metrics for the evaluation of mitigation methods: the proxy bias, sampling cost, area scale factor, runtime scaling, and noise boundary. Then we use these metrics to define qualitative criteria for certifying mitigation methods (Sec. 2.5.1) and strategies (Sec. 2.5.2).
2. **Characterisation of noise models (Sec. 3):** We ‘characterise’ the three types of noise we analyse in this work: stochastic noise (Sec. 3.3), rotational errors (Sec. 3.4), and (stochastic) over-rotation errors (Sec. 3.5).
3. **Taxonomical Classification Scheme (Sec. 4):** We classify the resources we have available to modify error channels: customised channels (Sec. 4.2.3), identity insertions (Sec. 4.2.4), hidden inverses (Sec. 4.2.2); the scope of noise amplification (Sec. 4.1): synchronous, asynchronous, and local; and the type of pre-tailoring we apply: none or local cancellation (Sec. 4.3). Then we produce a taxonomy of mitigation methods derived from various combinations of these aspects (Fig. 1).
4. **Case Study for Methods (Sec. 5):** We consider the mitigation methods in our taxonomy (Table 1) and evaluate them with respect to the metrics and certification criteria developed in Sec. 2 for stochastic noise (Sec. 3.3) and rotational errors (Sec. 3.4).
5. **Case Study for Strategy (Sec. 6):** We use the results of Sec. 5 to design USEM:ORE, a mitigation strategy for hardware suffering from (stochastic) over-rotation errors and certify that this strategy is USEful (see Sec. 2.5.2).
6. **Conclusions:** We conclude by summarising our contributions and discussing possible applications of our work (see Sec. 7).

2 Certification Scheme

Here we will work through the quantitative metrics (Sec. 2.3) and qualitative criteria (Sec. 2.5) that we develop for the certification of mitigation methods. But first, we shall briefly discuss linear mitigation methods and their advantages.

2.1 Linear Mitigation

The purpose of quantum error mitigation is to suppress the bias (ε_o) of an observable ($\hat{O}$) estimate generated using a quantum computer. We define the bias by:

$$\varepsilon_o = \left| \langle \tilde{O} \rangle_M - \langle \hat{O} \rangle \right|, \quad (1)$$

where $\langle \hat{O} \rangle$ is the target observable expectation in the absence of noise and $\langle \tilde{O} \rangle_M$ is the mitigated estimate.

Linear² mitigation is a way of mitigating an observable using a linear combination of (noisy) observable estimates from noisy circuits:

$$\langle \tilde{O} \rangle_M = \sum_{i=0}^{N_C} c_{\text{GL};i} \langle \tilde{O} \rangle_i \approx \langle \hat{O} \rangle, \quad (2)$$

where $c_{\text{GL};i}$ is the (global) mitigation coefficient for the i th observable; $\langle \tilde{O} \rangle_i$ is the i th noisy observable estimate, e.g. the result of the measurement of observable $\hat{O}$ using the i th noisy circuit variant ($\tilde{\mathcal{C}}_i$); N_C is the number of circuits used for the mitigation; and $\langle \hat{O} \rangle$ is the noise-free observable expectation. We can also define our mitigation in terms of noisy states ($\tilde{\rho}_i$) or noisy circuit variants ($\tilde{\mathcal{C}}_i$):

$$\tilde{\rho}_M = \sum_{i=0}^{N_C} c_{\text{GL};i} \tilde{\rho}_i \approx \rho, \quad \tilde{\mathcal{C}}_M = \sum_{i=0}^{N_C} c_{\text{GL};i} \tilde{\mathcal{C}}_i \approx \mathcal{C}, \quad (3)$$

where we use the same coefficients ($c_{\text{GL};i}$) for observables (Eq. (2)), states, and circuits. Here $\tilde{\rho}_M$ is the mitigated state, $\tilde{\rho}_i$ is the noisy states output by the i th noisy circuit, and ρ is the (target) noise-free state. $\tilde{\mathcal{C}}_M$ is the mitigated circuit, $\tilde{\mathcal{C}}_i$ is the (noisy) channel of the i th circuit, and $\mathcal{C}$ is the (noise-free) channel corresponding to the circuit we wish to implement. To move between the representations we use:

$$\langle \tilde{O} \rangle_M = \text{Tr} [\hat{O} \tilde{\rho}_M] = \text{Tr} [\hat{O} \tilde{\mathcal{C}}_M(\rho_0)], \quad \langle \tilde{O} \rangle_i = \text{Tr} [\hat{O} \tilde{\rho}_i] = \text{Tr} [\hat{O} \tilde{\mathcal{C}}_i(\rho_0)], \quad (4)$$

where (ρ_0) is the initial state input into the circuit. Some common examples of linear mitigation are probabilistic error cancellation [14], subspace expansion [15], and Richardson extrapolation [16].

Linear mitigation methods are interesting because it is possible to develop analytic bounds for their bias and costs (see Sec. 2.3). An additional advantage of linear mitigation methods is that they can be composed (see Secs 4.4.2 and 4.5.2). So we can combine multiple mitigation methods in order to mitigate different sections of our circuit or different types of noise. In the next section we will discuss how we can model the effect of noise in a circuit-independent way (Sec. 2.2) and then use these models to analyse mitigation methods (Sec. 2.3).

²An example of a non-linear mitigation method is virtual distillation (see Refs [11–13]), where the effective state after mitigation is the original noisy state raised to some integer power and is not a linear combination of physical states, in general.

2.2 Circuit-Independent Noise Modelling

Here we model the effect of noise in a circuit-independent way. The idea is to find a model for error channels that denotes them as linear combinations of (c-number) explicit functions of some noise amplitude (x) and (q-number) quantum channels, which are independent of the noise amplitude. This allows one to model the effect of multiple noise channels within a circuit without requiring precise knowledge of the circuit itself or the precise nature of the quantum channels within our error channels.

2.2.1 Channels

In this work we focus on noise associated to gates so we model each noisy gate (e.g. $\tilde{U}$) as a noise-free gate (e.g. U) followed³ by an error channel ($\mathcal{E}$). Each error channel has a noise amplitude⁴ (x) where the larger the noise parameter the larger the effect of the noise (in general⁵). We separate the full error channel ($\mathcal{E}$) into N_N sub-components, each with a scalar valued coefficient ($f_k(x)$) that depends on the noise amplitude and a ‘channel-valued’ term ($\mathcal{N}_k$) that is independent of the amplitude (x):

$$\mathcal{E} = \sum_{k=0}^{N_N} f_k(x) \mathcal{N}_k, \quad (5)$$

where we assume:

$$\mathcal{E} \xrightarrow{x \rightarrow 0} \mathcal{I}, \quad (6)$$

where $\mathcal{I}$ is the identity channel, which has no effect on the quantum state. For the rest of this work we also assume that:

$$f_k(0) = \delta_{0k}, \quad \mathcal{N}_0 = \mathcal{I}, \quad (7)$$

where δ_{0k} is the Kronecker delta (so zero unless $k = 0$, when it is unity) and that $\mathcal{N}_k$ is a completely-positive and trace-preserving (CPTP) channel.

To illustrate this notation we consider the example of a local dephasing (DPh) noise channel:

$$\mathcal{E}_{\text{DPh}}(\rho) = (1 - p) \rho + p Z \rho Z, \quad (8)$$

where ρ is the density operator representing a quantum state, p is the noise probability, and Z is the Pauli Z operator. For this channel we can define:

$$\mathcal{E}_{\text{DPh}} = f_0(x) \mathcal{N}_0 + f_1(x) \mathcal{N}_1, \quad x = p, \quad f_0(x) = 1 - x, \quad \mathcal{N}_0 = \mathcal{I}, \quad (9)$$

$$N_N = 2, \quad f_1(x) = x, \quad \mathcal{N}_1(\rho) = Z \rho Z, \quad (10)$$

which satisfies our requirements since $f_0(0) = 1$, $f_1(0) = 0$, and $\mathcal{N}_0$ and $\mathcal{N}_1$ are CPTP channels.

³The analysis is analogous when the noise channel occurs before the noise-free gate so we will not explicitly consider that case in this work.

⁴In this work we mainly focus on noise models with a single noise parameter but the analysis can be readily generalised to multi-noise parameter noise models.

⁵When we have coherent errors sometimes larger noise amplitudes can result in smaller errors due to periodicity in the noise parameter but we will focus on the small noise amplitude regime, $x \ll 1$, in this work so that shouldn't be an issue.

2.2.2 Circuits

We assume a uniform noise model, meaning each error channel in the circuit shares the same type and noise amplitude. This means all error channels in the circuit have the same functional form for the coefficients (f_k); and the same noise amplitude (x); but they could have different associated channels ($\mathcal{N}_k$). The uniformity assumption is limiting, but we can account for non-uniform noise models by sequentially applying this procedure for every error channel that appears in the circuit. We assume there are N_G (uniform) noisy gates in the circuit and therefore N_G error channels. Between the noisy gates we permit any number of noise-free gates and we place no restriction on the form of the noise-free components of the noisy gates. The result is that our noisy circuit ($\tilde{\mathcal{C}}$) can be expressed by the channel:

$$\tilde{\mathcal{C}} = \sum_{\{l_k\}: \sum_{j=0}^{N_N} l_j = N_G} g_{\{l_k\}}(x) \check{\chi}_{\{l_k\}}, \quad \text{where } g_{\{l_k\}}(x) = \left(\prod_{k=0}^{N_N} (f_k(x))^{l_k} \right), \quad (11)$$

but this expression needs a bit of unpacking. We use the short-hand notation:

$$\{l_k\} = \{l_k\}_{k=0}^{N_N} = l_0, l_1, \dots, l_{N_N}. \quad (12)$$

The sum includes every list of l_k 's ($\{l_k\}$ where $0 \leq l_k \leq N_G$) for which the sum of the list elements (l_k 's) is equal to the number of noise channels in the circuit (N_G):

$$\sum_{j=0}^{N_N} l_j = N_G. \quad (13)$$

$\check{\chi}_{\{l_k\}}$ is a quasi-channel⁶ and is the linear combination of all the channels that can be obtained by replacing l_k of the error channels in the circuit with $\mathcal{N}_k$ for each k ($0 \leq k \leq N_N$). $\check{\chi}_{\{l_k\}}$ is independent of the noise amplitude (x) since all noise amplitude dependence is encoded in $g_{\{l_k\}}(x)$.

As an example we will consider the case of a circuit with N_G gates, each of which are followed by a local dephasing noise channel ($\mathcal{E}_{\text{DPh}}$), so the noisy circuit is given by:

$$\tilde{\mathcal{C}} = \bigcirc_{i=1}^{N_G} (\mathcal{E}_{\text{DPh}} \circ \mathcal{U}_i) = \sum_{l_0, l_1: l_0 + l_1 = N_G} g_{l_0, l_1}(x) \check{\chi}_{l_0, l_1}, \quad (14)$$

where:

$$\mathcal{U}_i(\rho) = U_i \rho U_i^\dagger, \quad (15)$$

is the quantum channel associated with the quantum gate U_i , and (see Eqs (9) and (11)):

$$g_{l_0, N_G - l_0} = (1 - x)^{l_0} x^{N_G - l_0}, \quad \check{\chi}_{l_0, N_G - l_0} = \sum_{j=1}^{\frac{N_G!}{l_0!(N_G - l_0)!}} \bigcirc_{i=1}^{N_G} \left(\mathcal{N}_{S_{i; l_0, j}} \circ \mathcal{U}_i \right), \quad (16)$$

where $S_{i; l_0, j}$ is either zero or one, i.e. $\mathcal{N}_{S_{i; l_0, j}}$ is either $\mathcal{N}_0$ or $\mathcal{N}_1$. More concretely $S_{i; l_0, j}$ is the i th element of the j th permutation of a set of l_0 zeros and $l_1 = N_G - l_0$ ones, we sum

⁶Quasi-channels are linear sums of completely-positive and trace-preserving (CPTP) channels but may not be CPTP themselves. $\check{\chi}_{\{l_k\}}$ is a special type of quasi-channel where all the linear coefficients are equal to unity.

over all possible distinct permutations. Since $l_1 = N_G - l_0$ these are the only terms that need to be considered.

Fortunately, we do not need to understand the structure of $\check{\chi}_{\{l_k\}}$ since it is independent of the noise amplitude (x). For the following analysis all we need to know is that a decomposition of the form in Eq. (11) exists and the definition of $g_{\{l_k\}}(x)$. The nature of $\check{\chi}_{\{l_k\}}$ is irrelevant for our calculations as long as it is independent of the noise amplitude and we know its multiplicity (how many CPTP channels it comprises, see Eq. (17)).

2.3 Quantitative Mitigation Method Metrics

Mitigation suppresses the bias (Eq. (1)) at the cost of an increase in the variance, thus we require longer runtimes to obtain mitigated results to the same precision as unmitigated results. In the following sections, we derive a novel, circuit-independent, upper bound for the bias – termed the proxy bias (Secs 2.3.1 and 2.3.3) – as well as a new characterisation of runtime scaling, which quantifies the relative increase in runtime resulting from mitigation (Sec. 2.3.4). We combine the two to determine the noise boundary (Sec. 2.3.5), a new measure of the maximum noise level (see Eq. (24)) that can be mitigated for a given target bias and tolerated runtime scaling. We then use these metrics to help us define unambiguous qualitative criteria for certification (Sec. 2.5).

2.3.1 Proxy Bias

In order to determine the true bias (Eq. (1)) we would have to implement or simulate the full quantum circuit both with and without noise. This is not feasible for (novel) large-scale circuits, where no ground truth is available. As an alternative we provide the proxy bias (Eq. (21)), which is an upper bound for the true bias that can be determined without any explicit simulation, once we have a model such as that derived in Sec. 2.2.2. This facilitates the comparison of mitigation methods.

It is useful to define the multiplicity⁷ of (i.e. the number of CPTP channels that make up) the quasi-channel $\check{\chi}_{\{l_k\}}$:

$$\# [\check{\chi}_{\{l_k\}}] = \frac{N_G!}{\prod_{j=0}^{N_N} (l_j!)}.$$
 (17)

This means any physical (i.e. properly normalised) state (ρ_0) acted on by the quasi-channel $\check{\chi}_{\{l_k\}}$ will have an expectation value that satisfies:

$$0 \leq \left| \text{Tr} \left[\hat{O} \check{\chi}_{\{l_k\}} (\rho_0) \right] \right| \leq \# [\check{\chi}_{\{l_k\}}],$$
 (18)

if $\hat{O} \in (\text{I}, \text{X}, \text{Y}, \text{Z})^{N_Q}$ is a Pauli word operator and N_Q is the number of qubits. We have this upper bound because we know that the maximum expectation magnitude for any Pauli observable with respect to a physical state is unity and CPTP channels (e.g. $\mathcal{N}_k$) acting on a physical state return physical states. Since there are $\# [\check{\chi}_{\{l_k\}}]$ physical channels that make up $\check{\chi}_{\{l_k\}}$ that is the maximum possible observable expectation after passing through the quasi-channel: $\check{\chi}_{\{l_k\}}$.

⁷In general the multiplicity of a quasi-channel is the sum of the magnitudes of the linear coefficients pre-multiplying the CPTP channels that are summed together to give the quasi-channel. Since all the linear coefficients of $\check{\chi}_{\{l_k\}}$ are unity the multiplicity $\check{\chi}_{\{l_k\}}$ is just the number of constituent channels.

So the noisy state will have a bias given by (Eqs (1), (4), and (11)):

$$\begin{aligned}\varepsilon_o &= \left| \text{Tr} [\hat{O}\rho] - \text{Tr} [\hat{O}\tilde{\mathcal{C}}(\rho_0)] \right| \\ &= \left| \left(1 - (f_0(x))^{N_G}\right) \text{Tr} [\hat{O}\rho] - \sum_{\{l_k\}: \sum_{j=0}^{N_N} l_j = N_G; l_0 \neq N_G} g_{\{l_k\}}(x) \text{Tr} [\hat{O}\check{\chi}_{\{l_k\}}(\rho_0)] \right| \leq \varepsilon_\heartsuit,\end{aligned}\tag{19}$$

where we have separated out the noise-free component (ρ), i.e. that corresponding to the case where we act with the identity component ($\mathcal{N}_0$) of every noise channel (see Eq. (7)). In this case $l_0 = N_G$ so (see Eq. (12)):

$$\{l_k\} = N_G, 0, \dots, 0,\tag{20}$$

and $\check{\chi}_{N_G,0,\dots,0}(\rho_0) = \rho$ is the noise-free state. $\hat{O}$ is any Pauli operator, and we define:

$$\begin{aligned}\varepsilon_\heartsuit &= \left| 1 - (f_0(x))^{N_G} \right| + \sum_{\{l_k\}_k: \sum_{k=0}^{N_N} l_k = N_G; l_0 \neq N_G} |g_{\{l_k\}}(x)| \# [\check{\chi}_{\{l_k\}}] \\ &= \left| 1 - (f_0(x))^{N_G} \right| + \left(\sum_{k=0}^{N_N} |f_k(x)| \right)^{N_G} - |f_0(x)|^{N_G},\end{aligned}\tag{21}$$

as the (unmitigated) proxy bias.

If we consider our local dephasing noise example (see Eqs (9) and (14)) we find the proxy bias is given by (see Eqs (9) and (21)):

$$\varepsilon_\heartsuit = 2 \left(1 - (1-x)^{N_G} \right),\tag{22}$$

since $0 \leq x \leq 1$. This is upper bounded by 2 (which occurs in the limit $N_G \rightarrow \infty$ or $x = 1$). This bound is tight since it is indeed the maximum possible error for a Pauli expectation, which must be in the range -1 to 1 (inclusive). However, in general we do not expect this method to yield a tight upper bound on the bias. However, by choosing the decomposition components carefully one can make the bound tighter.

2.3.2 Noise Level

The noise level (e_G) is the leading order term of the (unmitigated) proxy bias (Eq. (21)) in the small noise regime ($N_G |x| \ll 1$). So if we define:

$$f_k(x) = f_k(0) + f'_k(x) + \mathcal{O}_S(x^2), \quad f_0(0) = 1, \quad f_{k>0}(0) = 0, \quad f'_k(x) = \mathcal{O}_S(x),\tag{23}$$

where $\mathcal{O}_S(x)$ is of linear order in x (see Sec. 2.4.2), then the noise level is given by:

$$e_G = N_G \sum_{k=0}^{N_N} |f'_k(x)|.\tag{24}$$

We find this to be a useful parameter to quantify the strength of the noise and it is proportional to the noise level defined in Ref. [7] for stochastic noise (see Sec. 3.3). It is independent of the precise structure of the circuit, only depending on the number of noise channels the circuit contains. If we have multiple different types of error channels then

the noise level of the full circuit is just the sum of the noise levels due to the constituent channels. For lower noise levels we expect smaller biases, for the unmitigated circuits, but owing to the complex interaction between gates and noise channels this will not always hold in practice. The noise level is always an upper bound for the bias but is not guaranteed to be tight, one should choose the channel decomposition (see Eq. (5)) that results in the smallest noise level to achieve the tightest bound.

If we consider our local dephasing noise example (see Eqs (9) and (14)) we find the noise level is given by (see Eq. (22)):

$$e_G = 2N_G x, \quad (25)$$

since (Eq. (9)):

$$f_0(0) = 1, \quad f'_0(x) = -x, \quad f_1(0) = 0, \quad f'_1(x) = x. \quad (26)$$

2.3.3 Proxy Bias after Mitigation

The proxy bias after mitigation is simply the proxy bias (see Sec. 2.3.1) of the mitigated channel, which serves as an upper bound for the bias of the mitigated state (for a Pauli observable). So the lower the proxy bias the better the mitigation, in general (the true bias after mitigation will depend on the specific structure of the circuit and could be significantly less than the proxy bias). In this work we consider linear mitigation and assume that all the noisy circuits used for our mitigation have the same noise model but different noise amplitudes. In this case the mitigated circuit has the channel (Eqs (3) and (11)):

$$\tilde{\mathcal{C}}_{;M} = \sum_{i=0}^{N_C} c_{G; i} \tilde{\mathcal{C}}_{; i} = \sum_{\{l_k\}_k: \sum_{k=0}^{N_N} l_k = N_G} g_{M, \{l_k\}}(x) \check{\chi}_{\{l_k\}}, \quad \text{where} \quad g_{M, \{l_k\}}(x) = \sum_{i=0}^{N_C} c_{G; i} g_{\{l_k\}}(x_i), \quad (27)$$

and x_i is the noise amplitude of the i th noisy circuit used for the mitigation (generally we assume $x_0 = x$ is the original noise level). Notice the $\check{\chi}_{\{l_k\}}$ terms are not altered by the mitigation only their coefficients ($g_{M, \{l_k\}}(x)$). The proxy bias is therefore given by (see Eq. (21)):

$$\varepsilon_\heartsuit = |1 - g_{M, \{N_G, 0, \dots, 0\}}(x)| + \sum_{\{l_k\}_k: \sum_{k=0}^{N_N} l_k = N_G; l_0 \neq N_G} |g_{M, \{l_k\}}(x)| \# [\check{\chi}_{\{l_k\}}]. \quad (28)$$

In this work we will only make use of the most significant order approximation to the proxy bias (i.e. the noise level after mitigation). We provide example derivations of the proxy bias in Apps B.4.4 and B.5.3.

2.3.4 Runtime Scaling

Although mitigation reduces the bias, it will increase the variance (σ^2) and may increase the number of gates required within a circuit, and therefore each circuit's runtime. In order to account for both the increase in variance and the increase in the runtime we introduce the runtime scaling (S_{GI}), which is the ratio of the total runtime required in the mitigated case to the total runtime in the unmitigated case to obtain the same precision. Naturally, the smaller the runtime scaling the better (for a fixed bias). For the equations below,

we assume that the observables measured are Pauli word operators⁸ $\hat{O} \in \{I, X, Y, Z\}^{\otimes N_Q}$, where N_Q is the number of qubits.

We define the unmitigated total runtime by:

$$t_0 = N_R (\tau_E + \tau_0), \quad (29)$$

where N_R is the number of repeated runs (or shots) of a circuit we must make (on average) to obtain our target variance (σ^2). For the worst case observable expectation we have:

$$N_R = \frac{1}{\sigma^2}. \quad (30)$$

τ_E is the average execution time of our unmitigated circuit and τ_0 is the time taken for initialisation and readout (treated as independent of the size of the circuit). For mitigated circuits the total runtime is given by:

$$t_M = C_{GI}^2 N_R (F_{GI,L} \tau_E + \tau_0), \quad (31)$$

where $F_{GI,L}$ is the length scale factor, i.e. the average increase in runtime owing to the mitigation; and C_{GI} is the global sampling cost for the mitigation, i.e. the scale factor for the increase in the number of samples required for the mitigation to obtain the same variance as the unmitigated circuit [14]:

$$C_{GI} = \sum_{i=0}^{N_C} |c_{GI;i}|, \quad F_{GI,L} = \frac{\sum_{i=0}^{N_C} |c_{GI;i}| \tau_{E,i}}{C_{GI} \tau_E}, \quad (32)$$

where $\tau_{E,i}$ and $c_{GI;i}$ are the average execution time and the mitigation coefficient, respectively, of the i th circuit variant (see Eq. (3) and Ref. [14]).

The runtime scaling (S_{GI}) is given by the ratio of the mitigated runtime to the unmitigated runtime:

$$S_{GI} = \frac{t_M}{t_0} \approx \begin{cases} C_{GI}^2 & \text{if } \tau_E \ll \tau_0, \\ C_{GI}^2 F_{GI,L} & \text{if } \tau_E \gg \tau_0. \end{cases} \quad (33)$$

Since we will mainly focus on the large circuit regime (see Sec. 2.4.3) we will use the latter limit.

We can also define the width scale factor:

$$F_{GI,W} = \frac{N_{Q,M}}{N_Q}, \quad (34)$$

which is the ratio of qubits required for the mitigation ($N_{Q,M}$) to those required for the original noisy circuit (N_Q). This provides a natural definition for the area scale factor:

$$F_{GI,A} = F_{GI,L} [F_{GI,W}], \quad (35)$$

which expresses the number of times we could run the original circuit using the time and resources required to run the mitigated circuit, by running the original circuit multiple times in parallel on the circuit used for the mitigated results. We take the largest integer smaller than $F_{GI,W}$ because we cannot split up the qubits required to implement the original circuit between runs, in general. In this more general case the runtime scaling is given by:

$$S_{GI} = \frac{t_M [F_{GI,W}]}{t_0} \approx \begin{cases} C_{GI}^2 [F_{GI,W}] & \text{if } \tau_E \ll \tau_0, \\ C_{GI}^2 F_{GI,A} & \text{if } \tau_E \gg \tau_0. \end{cases} \quad (36)$$

However, for the methods we consider in this work $F_{GI,W} = 1$.

⁸The application of the derivation in Sec. 2.3.4 to non-Pauli observables is straightforward since any observable can be decomposed as a function of Pauli observables.

2.3.5 Noise Boundaries

It is often possible to reduce the bias at the cost of increasing the runtime scaling, for a given mitigation method (such as zero-noise extrapolation). We call this ‘increasing the order of the mitigation’ and define an order parameter N_M , where larger N_M means, in general, lower bias but longer runtime. For most applications one must solve a Pareto optimisation problem where the weighting given to the (proxy) bias and the runtime scaling will depend on the application.

In order to permit the comparison of mitigation methods that manage the bias-runtime-scaling (or equivalently bias-variance) trade-off relation differently, we introduce the noise boundary (e_B). The noise boundary is the maximum noise level (Sec. 2.3.2) that can be mitigated, given that we require a bias smaller than or equal to $\varepsilon_\star$ and a runtime scaling less than or equal to $S_\star$ (where the exact values of $\varepsilon_\star$ and $S_\star$ will be set by the problem we wish to solve and the equipment we have available). Naturally, the larger the noise boundary, the better the mitigation. A general expression for the noise boundary can be difficult to formulate analytically so we will focus on the large circuit regime (see Sec. 2.4.3), considering only the most significant contributions to the proxy bias and runtime scaling⁹.

In order to calculate the noise boundary we first characterise the proxy bias and the runtime scaling as functions of the number of noisy gates (N_G), noise level (e_G), and the order of mitigation (N_M) for a given mitigation method and noise model:

$$\varepsilon_{\heartsuit;M} = f_{\heartsuit}(N_G, e_G, N_M), \quad S_{GI} = f_{RS}(N_G, e_G, N_M). \quad (37)$$

Then we solve these equations simultaneously with $\varepsilon_{\heartsuit;M} = \varepsilon_\star$ and $S_{GI} = S_\star$ to obtain e_G as a function of $\varepsilon_\star$ and $S_\star$ (and possibly N_G). In this case $e_G = e_B$ (i.e. the noise boundary):

$$e_B = f_{NB}(N_G, \varepsilon_\star, S_\star). \quad (38)$$

The larger the boundary the higher the noise level we can mitigate, so larger boundaries imply better mitigation methods. However, it is the rate of change with $S_\star$ that is more interesting when we are comparing methods in the large circuit regime (see Sec. 2.4.3).

For an examples of a noise boundary calculation see App. B.5.5.

2.4 Regimes

In order to permit the comparison of different mitigation methods in a tractable way we will focus on two main regimes: large circuits and small noise levels. To facilitate our analysis we define our order parameter:

$$\mathcal{O}(f(y_0, y_1, \dots)), \quad (39)$$

as a function that tends to zero in the limit $f \rightarrow 0$. f is an arbitrary function that will indicate the most significant term of the Laurent series of $\mathcal{O}$, up to constant scale factors. We add a subscript (or two) to $\mathcal{O}$ in order to specify the regime that we are in.

⁹Where approximations to the runtime scaling and/or bias must be made one should err on the side of caution (i.e. round up the bias and runtime scaling) so that the noise boundary is guaranteed to be an upper bound (to the specified order).

2.4.1 Small Amplitude Regime

In all regimes we assume that the noise amplitude of each individual gate is small, e.g. $x, x^2 N_G \ll 1$, where N_G is the number of noisy gates in the circuit. So we can neglect all but the smallest powers of x in our calculations, remaining terms can be replaced by:

$$\mathcal{O}_0(f(x)), \quad (40)$$

where $\mathcal{O}_0 \rightarrow 0$ as $x \rightarrow 0$.

2.4.2 Small Noise Level Regime

In the small noise level regime, the noise level (a function of the noise amplitude: x and the number of gates: N_G , see Sec. 2.3.2) in the circuit is small (e.g. $e_G \ll 1$) so that we consider only the highest orders in e_G . In this regime we denote the neglected component:

$$\mathcal{O}_S(f(e_G)), \quad (41)$$

this tends to zero in the limit $e_G \rightarrow 0$, which is the small noise limit. It is in this regime that we calculate the proxy bias but it may be possible to generalise to other regimes.

2.4.3 Large Circuit Regime

In the large circuit regime we assume that the number of gates in the circuit is large ($N_G \gg 1$) so that we consider only the highest orders in N_G . In this regime we denote the neglected component:

$$\mathcal{O}_L(f(N_G)), \quad (42)$$

this tends to zero in the limit $N_G \rightarrow \infty$ (so N_G will generally appear in the denominator). The large circuit limit is when the number of gates tends to infinity but the (full circuit) noise level (e_G) stays fixed, so this requires that the (gate-wise) noise level tends to zero, i.e. $x \rightarrow 0$ as $N_G \rightarrow \infty$ such that e_G is independent of N_G (at least to leading order). If we are in the regime of large circuits and low noise levels then we use $\mathcal{O}_{LS}$ or, if the behaviour is the same for either large circuits or small noise levels, we use $\mathcal{O}_{L/S}$.

2.5 Qualitative Certification Criteria

We can use our quantitative metrics (Sec. 2.3) to derive some qualitative criteria for certifying our mitigation methods (Sec. 2.5.1) and strategies (Sec. 2.5.2). One can pick and choose the relevant certification criteria depending on the target application.

2.5.1 Mitigation Methods

We certify mitigation methods with respect to a particular noise model (such as stochastic noise (Sec. 3.3) or rotational errors (Sec. 3.4)) as their behaviour can depend strongly on the type of noise they are mitigating. We propose the following certification criteria:

1. **Scalable (S):** A scalable mitigation method is one for which the runtime scaling (S_{GI}) is finite and independent of circuit size (N_G) to leading order, in the large circuit regime ($N_G \gg 1$, see Sec. 2.4.3) when the noise level (e_G , Sec. 2.3.2) is independent of N_G . i.e. $S_{GI} < \infty$ for $N_G \rightarrow \infty$ if $e_G < \infty$. This means we can mitigate arbitrarily large circuits without a change in the runtime scaling provided the noise amplitude

(x) is reduced by a proportionate amount, so that the noise level (e_G) remains fixed. This will generally mean the length of circuit we can mitigate using our method will be inversely proportional to the noise amplitude.

2. **Unbounded (U):** An unbounded (in principle) mitigation method can mitigate arbitrarily large noise levels, assuming we permit arbitrarily large runtime scaling, in the large circuit limit, i.e. $e_B \rightarrow \infty$ (Sec. 2.3.5) as $S_\Phi \rightarrow \infty$ for fixed ε_Φ and $N_G = \infty$.
3. **Precise (P):** A precise mitigation method is either unbiased or has a bias after mitigation that can be made arbitrarily small for suitably large runtime scaling, in the small noise regime, i.e. $\varepsilon_{\mathcal{Q};M} \rightarrow 0$ for $S_\Phi \rightarrow \infty$, where $e_G \ll 1$.
4. **Efficient (E):** An efficient mitigation method has a runtime scaling (S_{GI} , Sec. 2.3.4) that tends to unity (i.e. cost free) as the noise level tends to zero, for fixed target bias (ε_Φ). i.e. $S_{GI} \rightarrow 1$ as $e_G \rightarrow 0$ for ε_Φ fixed.
5. **Robust (R):** A robust mitigation method has metrics (i.e. runtime scaling, Sec. 2.3.4, and proxy bias, Sec. 2.3.3) that are independent of deviations from the target noise due to small, characterised, imperfections in variant generation (see Sec. 4.2), to leading order. In general, we expect the process of generating noisy gate variants, i.e. noise amplification or transformation (e.g. see Secs 4.2.2 and 4.2.3) to be imperfect and to result in noise differing slightly from the target noise. A method is robust if these imperfections, provided they are characterised and of the same order as the original noise amplitude (x), do not change the value of the metrics to leading order in the noise level for the small noise regime (Sec. 2.4.2) and to leading order in the number of gates for the large circuit regime (Sec. 2.4.3). If the coefficients, but not the order, of the leading order terms change then it is quasi-robust (qR). A (quasi-)hardware-robust method is (q)R for the imperfections that occur on a particular hardware, not necessarily for all possible imperfections. Imperfections must be well characterised and taken into account in the coefficient and metric determination.

Some of the implications of these certifications are give below:

- **Scalable** mitigation methods can take advantage of reductions in noise levels (e.g. through hardware tuning [8] or error correction [9]) to mitigate larger circuits without an increase in relative cost.
- **Unbounded** mitigation methods can mitigate, arbitrarily large, arbitrarily noisy circuits with a runtime that depends only on their noise level and (possibly) their target bias (in the large circuit limit).
- **Precise** mitigation methods make high precision applications, e.g. precise characterisation, possible. Since they can make the bias after mitigation arbitrarily small.
- **Efficient** mitigation methods make high precision applications practical (in the small noise regime). Since the cost of mitigation disappears as the (pre-mitigation) noise level decreases.
- **Robust** methods maintain their performance in the face of inevitable hardware imperfections.

A mitigation method that is scalable, unbounded, precise, efficient, and robust is certified as a SUPER mitigation method. If the method is only quasi-robust, we certify it SUPEqR. This latter designation is sufficient for inclusion in an efficacious mitigation strategy (see Sec. 2.5.2).

2.5.2 Mitigation Strategies

A mitigation strategy (e.g. see Sec. 6.1) is a complete prescription for producing mitigated observable estimates from a quantum algorithm using a particular piece of noisy quantum hardware. They include:

- **Physical Operations:** The (characterised) physical operations¹⁰ the quantum hardware can implement.
- **Mitigation Methods:** A (set of) mitigation method(s) designed to obtain mitigated observable estimates using the available physical operations.
- **Compilation Scheme:** A scheme for converting target algorithms into mitigated quantum circuits.
- **Additional Implementation-Dependent Requirements:** Depending on the hardware available and certification required (Sec. 2.5) there may be additional considerations. For example, for a scalable strategy we require that the noise level be independent of the circuit length in order to ensure the runtime scaling is independent of the circuit length (see Sec. 2.5.1).

The idea is that a particular hardware will have a basis set of physical operations that it can implement, each of which will have an associated noise model. Mitigation methods will generally¹¹ require additional gates or physical operations to implement the mitigation, beyond those required to implement the unmitigated target circuit. The mitigation methods used within the mitigation strategy should be tailored to the hardware they are designed to mitigate and the gate set(s) they require must be implementable (and, for noise-aware (Sec. 4.1) strategies, characterisable) on that hardware. A mitigation method is **compatible** with a piece of hardware if the hardware can generate all the gates or physical operations required to implement the mitigation method. We introduce three strategy certification criteria:

1. **Universal:** A universal mitigation strategy can mitigate any quantum algorithm¹². It is designed to be implemented using a particular universally-mitigatable piece of hardware, which is a piece of hardware that can generate all the gates or physical operations required to produce a mitigated circuit for any quantum algorithm. The strategy is hardware specific so the mitigation methods it contains may not be able to mitigate any possible noise model (unlike, for example, the PEC strategy of [14]). Universality requires only that the mitigation methods can mitigate the noise (or some of the noise) that occurs on the specified hardware. The advantage of this restriction is that we can target specific noise models and focus on scalability and efficacy rather than requiring our methods to be completely general.

¹⁰Physical operations are a generalisation of gates and consists of all gates implementable on a piece of quantum hardware as well as all other transformations it can induce. Examples of physical operations are the process of changing a single (or small subset) of parameters of the effective Hamiltonian that generates a particular gate, measurements, the turning on or off of an external field, a period of the free evolution of a qubit, and a period of evolution under a static or dynamic external field.

¹¹Identity insertion based methods (Sec. 4.2.4) can often perform mitigation without requiring any gates beyond those required for a particular algorithm.

¹²If we have a ‘quantum accelerator’, i.e. a piece of hardware that is not universal but targeted at a particular task, we can relax the universality requirement and require only that the mitigation strategy can mitigate any algorithm the ‘quantum accelerator’ is designed to implement.

2. **Scalable:** A scalable mitigation strategy consists of only scalable mitigation methods (see Sec. 2.5.1). This means we can mitigate arbitrarily large circuits without a diverging runtime scaling (Sec. 2.3.4). There is quite convincing evidence that, for stochastic noise¹³, the sampling cost of any mitigation method scales exponentially with the number of gates (N_G) for constant (gate-wise) noise amplitude (x), e.g. see Ref. [6]. Thus, for fixed x no mitigation method is scalable (for stochastic noise). Therefore, we are forced to consider reductions in the noise amplitude (e.g. through hardware tuning [8] or error correction [9]) as part of our mitigation strategy when we scale to larger circuits. It is why we consider a method scalable if the runtime scaling is finite in the limit $N_G \rightarrow \infty$ when the limit is taken such that the noise level (e_G , Sec. 2.3.2) is kept constant. One must ensure that the methods are still scalable when implemented using the quantum hardware specified in the strategy.
3. **Efficacious:** A mitigation strategy is efficacious if all the mitigation methods it contains are precise, efficient, unbounded, and robust (or at least quasi-hardware-robust), see Sec. 2.5.1. This means, given a sufficiently high tolerated runtime scaling (S_Φ) it can reduce the bias (of the circuits it can mitigate) to any desired level no matter what the initial noise level, i.e. we can ensure:

$$\varepsilon_{\heartsuit;M} \leq \varepsilon_\Phi, \quad (43)$$

for S_Φ sufficiently large, regardless of the initial value of e_G . It also means that any improvements to our hardware that lead to reductions in the circuit noise level will reduce the runtime scaling required for mitigation. Lastly, quasi-hardware-robustness means that (for the hardware in our strategy) the runtime scaling and proxy bias are the same, to leading order, when actually implemented on our hardware as in the ideal case (where the noise tailoring required for mitigation does not introduce any new errors). Without quasi-hardware-robustness there is no guarantee that the other certifications will be meaningful for our strategy in practice.

Universal strategies are required for universal quantum computation, scalable strategies are worth investing in long term, and efficacious strategies are more likely to be practical. If a strategy is universal, scalable, and efficacious then we certify it as a USEful mitigation strategy. In the next section we investigate and characterise our target noise models.

3 Characterisation of Noise

Here we introduce some of our notation for noise channels (Sec. 3.2) as well as Pauli transfer matrices (Sec. 3.2.1), a useful representation for analysing noise models. Then we present characterisation data for three types of noise: stochastic noise (SN, Sec. 3.3), rotational errors (RE, Sec. 3.4), and over-rotation errors (ORE, Sec. 3.5).

3.1 Characterisation

We consider a noise channel to be characterised when we can express it as a Kraus map [2] or transfer matrix (Sec. 3.2.1). With access to quantum hardware one can characterise (i.e. determine the structure of) noise channels using gate-set tomography [17], classical

¹³Pure rotational errors are generally cheaper to mitigate than stochastic noise and can be scalably mitigated even if the noise level (Sec. 2.3.2) grows with the square root of the number of gates (see Table 4).

simulation, etc. But, for illustrative purposes, we will consider three explicit classes of noise: stochastic noise (SN, Sec. 3.3), rotational errors (RE, Sec. 3.4), and (stochastic) over-rotation errors (ORE, Sec. 3.5). It can be useful to model noise channels in different ways for different applications.

3.2 Circuit-Independent Noise Models

We consider noise to be associated with a gate. A noisy gate ($\tilde{U}$) is equivalent to a unitary (noise-free) gate (U) followed by an error channel¹⁴ ($\mathcal{E}$), which may or may not be unitary. We augment the standard circuit notation to account for noisy gates by representing noisy gates (which may not be unitaries) as gates (e.g. U) with a tilde atop:

$$\boxed{\tilde{U}} = \boxed{U} \boxed{\mathcal{E}}. \quad (44)$$

We represent general channels (e.g. $\mathcal{C}$) using a calligraphic script while for noise-free unitaries: we use upright text for named unitaries (e.g. Z and P_i , for the Pauli Z operator and the i th Pauli word operator, respectively) and italics for generic unitaries (e.g. U and V). Error channels are represented by $\mathcal{E}$ (see Sec. 2.2.1). We reserve $\mathcal{C}$ (see Sec. 2.2.2) to represent the channel associated to a full circuit. We use curly brackets to denote a stochastic or quasi-stochastic ensemble, e.g. the channel of Eq. (5) of Sec. 2.2.1 can be represented by:

$$\boxed{\mathcal{E}} = -\left\{f_k(x) : \boxed{\mathcal{N}_k}\right\}_{k=0}^{N_N}, \quad (45)$$

where

$$\sum_{k=0}^{N_N} |f_k| = 1 \implies \text{stochastic}, \quad \sum_{k=0}^{N_N} f_k = 1 \implies \text{quasi-stochastic}. \quad (46)$$

For example a (stochastic) Pauli (P) noise channel can be defined by:

$$\boxed{\mathcal{E}_P} = -\left\{p_{P,k} : \boxed{P_k}\right\}_{k=0}^{4^{N_Q}-1}, \quad p_{P,0} = 1 - p, \quad p = \sum_{k=1}^{4^{N_Q}-1} p_{P,k}, \quad (47)$$

where $p_{P,k} \geq 0$ is the probability of the operator P_k occurring on a given run and p is the probability of any error occurring. $P_k \in \{I, X, Y, Z\}^{\otimes N_Q}$ is the k th Pauli word (tensor products of Pauli operators) for N_Q qubits (see App. B.1.1); P_0 is the global identity. An alternative way to represent an error channel is with a (Pauli) transfer matrix.

3.2.1 Transfer Matrices

The (Pauli) transfer matrix ($\hat{T}_{\mathcal{E}}$) associated with a channel ($\mathcal{E}$) is defined by [2, 17]:

$$\left(\hat{T}_{\mathcal{E}}\right)_{i,j} = \frac{1}{2^{N_Q}} \text{Tr} [P_i \mathcal{E}(P_j)], \quad (48)$$

¹⁴Our analysis will be based on modelling error channels as following their gate but all the analysis is analogous if the channel precedes the gate.

where N_Q is the number of qubits in the channel and $P_k \in \{I, X, Y, Z\}^{\otimes N_Q}$ is the k th Pauli word operator. Transfer matrices have the nice property that the composition of two channels is just the product of their transfer matrices:

$$\hat{T}_{\mathcal{E}_2 \circ \mathcal{E}_1} = \hat{T}_{\mathcal{E}_2} \hat{T}_{\mathcal{E}_1}. \quad (49)$$

If a transfer matrix is diagonalisable then we can write it as:

$$\hat{T}_{\mathcal{E}} = \hat{W} \hat{D} \hat{W}^{-1}. \quad (50)$$

If, in addition, $\hat{W}$ is independent of the noise amplitude (x) of the channel $\mathcal{E}$; then we have a model similar to that of Sec. 2.2.1 and all the x dependence is included in the diagonal matrix ($\hat{D}$). This property is particularly useful for local mitigation strategies (see Sec. 4.5) as it massively simplifies the problem. In this case the noise channel can be fully characterised by its eigenvalues ($\lambda_m(x)$) and the matrix $\hat{W}$. There are some useful properties of error channels that we shall discuss in the following sections.

3.2.2 Compatibility

Noise channels are compatible if they commute (i.e. $\mathcal{N}_k$ and $\mathcal{N}_l$ are compatible if they satisfy the relation $[\mathcal{N}_k, \mathcal{N}_l]_- = 0$) and an error channel decomposition is self-compatible if all its noise channels are compatible. A self-compatible error channel is one that has a diagonalisation matrix ($\hat{W}$, see Sec. 3.2.1) that is independent of the noise amplitude, in which case it boasts a self-compatible decomposition.

3.2.3 Closure

An error channel decomposition (see Eq. (45) and Eq. (5) of Sec. 2.2.1) is closed if the span of the constituent noise channels (i.e. $\{\mathcal{N}_k\}_{k=0}^{N_N}$) is closed under composition. Composing two closed channels (with the same constituent noise channels) next to each other will result in a composite channel with the same constituent channels, but potentially different parameters (e.g. noise amplitude).

3.3 Stochastic Noise

(Single parameter) stochastic noise (SN) can be expressed by:

$$\boxed{\mathcal{E}_{\text{SN}}} = \begin{cases} \text{---} & \text{with probability } 1 - p \\ \boxed{\mathcal{N}} & \text{with probability } p. \end{cases} \quad (51)$$

where $\mathcal{N}$ is some noise channel (that may or may not cause an actual error to occur depending on the current state the qubits are in), which must be a physical (i.e. CPTP) channel for the error channel ($\mathcal{E}$) to be physical. p is the probability of such a noise channel occurring and this is the noise amplitude (i.e. $x = p$). So this channel can be characterised by (see Eq. (5) of Sec. 2.2.1):

$$\begin{aligned} f_0 &= 1 - p, & f_1 &= p, \\ \mathcal{N}_0 &= \mathcal{I}, & \mathcal{N}_1 &= \mathcal{N}, \end{aligned} \quad (52)$$

where $\mathcal{I}$ is the identity channel and does nothing. We can derive the corresponding proxy bias (upper bound on deviation) as follows (see Eq. (21) of Sec. 2.3.1):

$$\varepsilon_{\heartsuit} = 2 \left(1 - (1 - p)^{N_G} \right). \quad (53)$$

So, in the limit of infinitely many applications of the channel, we have a proxy bias of two, which is indeed the maximum possible bias for a Pauli operator, so the proxy bias may well be a tight bound. The noise level (e_G) is the leading order component of the proxy bias (see Sec. 2.3.2), for stochastic noise:

$$e_G = e_{\text{SN}} = \varepsilon_{\heartsuit} + \mathcal{O}_S(e_{\text{SN}}^2), \quad (54)$$

where:

$$e_{\text{SN}} = 2pN_G. \quad (55)$$

This is the same order as the noise level of Ref. [3] but has an additional factor of two in order that the noise level be an upper bound (to first order).

If we have uniform stochastic noise affecting N_G gates in our circuit then the channel associated to the circuit can be represented by (see Eq. (11) of Sec. 2.2.2):

$$\tilde{\mathcal{C}}_{N_G, p} = \sum_{k=0}^{N_G} (1 - p)^{N_G - k} p^k \check{\chi}_{\text{SN}, k}, \quad (56)$$

where $\check{\chi}_{\text{SN}, k}$ is independent of p (see Sec. 2.2.2):

$$\check{\chi}_{\text{SN}, k} = \check{\chi}_{N_G - k, k}, \quad (57)$$

i.e. $\check{\chi}_{\text{SN}, k}$ is a linear combination of all the channels that can be obtained from the original noisy circuit by replacing $N_G - k$ of the error channels with the identity channel ($\mathcal{I}$) and k with $\mathcal{N}$. The multiplicity (see Eq. (17)) is given by:

$$\# [\check{\chi}_{\text{SN}, k}] = \frac{N_G!}{k! (N_G - k)!}. \quad (58)$$

All stochastic noise channels are self-compatible since the identity channel commutes with any other channel but they are only closed if:

$$\mathcal{N}^2 = a\mathcal{I} + (1 - a)\mathcal{N}, \quad (59)$$

for some $0 \leq a \leq 1$ (i.e. the product of two constituent channels is a linear combination of the existing constituent channels). The channel has eigenvalues given by:

$$\lambda_m(p) = (1 - p) + p\lambda_{\mathcal{N}, m}, \quad (60)$$

where $\lambda_{\mathcal{N}, m}$ is the m th eigenvalue of $\mathcal{N}$. If the error channel is closed then $\lambda_{\mathcal{N}, 0} = 1$ and $\lambda_{\mathcal{N}, 1} = -a$:

$$\lambda_0(p) = 1, \quad \lambda_1(p) = 1 - (1 + a)p. \quad (61)$$

3.4 Rotational Errors

Rotational errors (RE) are unitary (coherent) errors generated by a Hermitian unitary operator (P):

$$\boxed{\mathcal{E}_{\text{RE}}} = \boxed{e^{i\frac{\phi}{2}P}}, \quad (62)$$

where the noise amplitude is the rotation angle (i.e. $x = \phi$) and P is some self-inverse operator ($P^2 = \mathbb{I}$). This channel can be characterised by (see Eq. (5) of Sec. 2.2.1):

$$\begin{aligned} f_0 &= \cos^2 \frac{\phi}{2}, & f_1 &= \sin \phi, & f_2 &= \sin^2 \frac{\phi}{2}, \\ \mathcal{N}_0 &= \mathcal{I}, & \check{\mathcal{N}}_1(\rho) &= i \frac{P\rho - \rho P}{2}, & \mathcal{N}_2(\rho) &= P\rho P. \end{aligned} \quad (63)$$

In this case $\mathcal{N}_0$ and $\mathcal{N}_2$ are CPTP channels, as Sec. 2.2.1 requires, but $\check{\mathcal{N}}_1$ is not. However $\check{\mathcal{N}}_1$ is normalised and can be decomposed in terms of CPTP channels:

$$\check{\mathcal{N}}_1 = \frac{1}{2}\mathcal{N}_1 - \frac{1}{2}\mathcal{N}_3, \quad (64)$$

where $\mathcal{N}_1$ and $\mathcal{N}_3$ are the CPTP channels:

$$\mathcal{N}_1(\rho) = S_P \rho S_P^\dagger, \quad \mathcal{N}_3(\rho) = S_P^\dagger \rho S_P, \quad S_P = e^{i\frac{\pi}{4}P} = \frac{\mathbb{I} + iP}{\sqrt{2}}, \quad (65)$$

so we can perform all the analyses that we need to. The proxy bias is given by (see Eq. (21) of Sec. 2.3.1):

$$\varepsilon_\heartsuit = 1 - 2 \cos^{2N_G} \frac{\phi}{2} + (1 + \sin |\phi|)^{N_G}. \quad (66)$$

Clearly the bound is not tight in this case since the noise level tends to infinity as $N_G \rightarrow \infty$. The noise level (e_G) is the leading order component of the proxy bias (see Sec. 2.3.2), for rotational errors:

$$e_G = e_{\text{RE}} = \varepsilon_\heartsuit + \mathcal{O}_S(e_{\text{RE}}^2), \quad (67)$$

where:

$$e_{\text{RE}} = N_G |\phi|. \quad (68)$$

If we have uniform rotational errors affecting N_G gates in our circuit then the channel associated to the circuit can be represented by (see Eq. (11) of Sec. 2.2.2):

$$\tilde{\mathcal{C}}_{N_G, \phi} = \sum_{k=0}^{2N_G} \cos^{2N_G-k} \frac{\phi}{2} \sin^k \frac{\phi}{2} \check{\chi}_{\text{RE}, k}, \quad (69)$$

where $\check{\chi}_{\text{RE}, k}$ is independent of ϕ , and the multiplicity (see Eq. (17)) is given by:

$$\# [\check{\chi}_{\text{RE}, k}] = \sum_{l=0}^{\lfloor \frac{k}{2} \rfloor} \frac{2^{k-2l} (N_G!)}{l! (k-2l)! (N_G + l - k)!}. \quad (70)$$

The factors of two and the summation are due to the fact there are different combinations of $\mathcal{N}_0$, $\check{\mathcal{N}}_1$, and $\mathcal{N}_2$ that lead to the same prefactor in Eq. (69).

All rotational noise channels are self-compatible since $\mathcal{N}_0$, $\check{\mathcal{N}}_1$, and $\mathcal{N}_2$ commute with each other and they are also closed since:

$$\mathcal{N}_0 \circ \mathcal{N}_k = \mathcal{N}_k, \quad \check{\mathcal{N}}_1^2 = \frac{1}{2}\mathcal{N}_2 - \frac{1}{2}\mathcal{N}_0, \quad \mathcal{N}_1 \circ \mathcal{N}_2 = -\mathcal{N}_1, \quad \mathcal{N}_2^2 = \mathcal{N}_0. \quad (71)$$

The channel has eigenvalues given by:

$$\lambda_{\mathcal{E},0}(\phi) = 1, \quad \lambda_{\mathcal{E},1}(\phi) = e^{i\phi}, \quad \lambda_{\mathcal{E},2}(\phi) = e^{-i\phi}. \quad (72)$$

3.5 (Stochastic) Over-Rotation Errors

Now (stochastic) over-rotation errors (ORE) are rotational errors that occur when a gate is produced by an operation equivalent to a rotation generated by some Hermitian unitary operator (P) and there is an overshoot¹⁵. So we rotate by more than the target angle. Thus over-rotation errors are generated by the same operator that generates the gate. In order to have a more realistic, and general, scenario we consider stochastic over-rotation errors. Here we have a random distribution of rotational errors, which is equivalent to a rotational error channel followed by a stochastic error channel, with the same generator:

$$\boxed{\mathcal{E}_{\text{ORE}}} = \boxed{e^{i\frac{\phi}{2}P}} \left\{ \begin{array}{l} \text{---} \quad \text{with probability } 1-p \\ \boxed{P} \quad \text{with probability } p. \end{array} \right. \quad (73)$$

This channel has a channel decomposition given by:

$$\begin{aligned} f_0 &= \cos^2 \frac{\phi}{2} - p \cos \phi, & f_1 &= (1-2p) \sin \phi, & f_2 &= \sin^2 \frac{\phi}{2} + p \cos \phi, \\ \mathcal{N}_0 &= \mathcal{I}, & \check{\mathcal{N}}_1(\rho) &= i \frac{P\rho - \rho P}{2}, & \mathcal{N}_2(\rho) &= P\rho P. \end{aligned} \quad (74)$$

So we have the same constituent noise channels as for rotational errors (see Sec. 3.4). The proxy bias is given by (see Eq. (21) of Sec. 2.3.1):

$$\varepsilon_{\heartsuit} = 1 - 2 \left(\cos^2 \frac{\phi}{2} - p \cos \phi \right)^{N_G} + (1 + |1-2p| \sin |\phi|)^{N_G}, \quad (75)$$

and we define the noise level (e_{ORE}) as the leading order component of the proxy bias for over-rotation errors:

$$e_{\text{ORE}} = 2N_G p + N_G |\phi| = e_{\text{SN}} + e_{\text{RE}} = \varepsilon_{\heartsuit} + \mathcal{O}_S(e_{\text{RE}}^2, e_{\text{SN}}^2). \quad (76)$$

Over-rotation error channels are self-compatible and closed since the noise channels are the same as for rotational error channels (c.f. Eqs (63) and (74)). The channel has eigenvalues given by:

$$\lambda_{\mathcal{E},0}(p, \phi) = 1, \quad \lambda_{\mathcal{E},1}(p, \phi) = (1-2p) e^{i\phi}, \quad \lambda_{\mathcal{E},2}(p, \phi) = (1-2p) e^{-i\phi}. \quad (77)$$

In the next section we will look in more detail at some mitigation methods and their components.

¹⁵We also have under-rotation errors where there is an undershoot and we rotate less than we aim for. For ease of notation we shall consider under-rotation as a negative over-rotation.

4 Taxonomical Classification of Mitigation Methods

Here we classify our mitigation methods based on their component parts and the variant generation procedures they make use of. That is, the procedures used to produce the collection of noisy (variant) circuits that are used in the mitigation (i.e. $\{\tilde{\mathcal{C}}_{;i}\}_{i=0}^{N_C}$, see Eq. (3)). First, we discuss the naming convention we adopt for our mitigation methods (Sec. 4.1); then we present the components and variant generation procedures (Secs 4.2, 4.3, 4.4, and 4.5) that we consider; and, finally, we present the methods that we study in this work, in the context of a taxonomical scheme (Sec. 4.6).

4.1 Nomenclature

We give our mitigation methods a four component name (see Table 1 for examples):

$$[\text{Pre-tailoring}][\text{Variant generation}][\text{Scope}]\text{M}:[\text{Noise Awareness}], \quad (78)$$

where each component uses an acronym.

The pre-tailoring strategies are used to convert the noise models of the noisy gates to different forms before the mitigation proper. In this work we introduce and use only one: local cancellation (LC), which we use to convert non-Pauli to Pauli noise (see Sec. 4.3).

In order to effect the mitigation we need to combine a number of different variants of our original circuit. Each circuit is produced by replacing (some of) our noisy gates with other gates, which we denote noisy gate variants. These have the same noise-free components as the gates they replace but different noise channels. We consider four variant generation strategies, which can also be used in combination: custom channels (C), these are tailored (i.e. custom made) error channels inserted after the noisy gate in order to produce new gate variants with amplified errors (see Sec. 4.2.3); hidden inverses (HI), this is a strategy for reversing the sign (direction) of coherent error channels (see Sec. 4.2.2); identity insertion (II), this is a noise amplification strategy that involves alternating a noisy gate and its noisy inverse (see Sec. 4.2.4); and tuned identity insertion (TII), this is a variant of identity insertion where we tune the number of identity insertions in order to reduce the runtime scaling (see Sec. 4.2.4).

There are three possible scopes for noise amplification: synchronous (S), where we use the same variant generation procedure for each gate being mitigated within a particular circuit; local (L), where we have a mitigation strategy that is equivalent to mitigating each gate¹⁶ independently; and asynchronous (A), the most general scheme, where we can have different variant generation procedures for different gates within the same circuit (this implicitly includes local methods).

Noise awareness denotes whether the method uses quantitative knowledge about the noise models to choose the mitigation coefficients (see Sec. 2.1), which we denote noise aware (NA); or not, which we denote knowledge free (KF). In order to specify the type of noise our method is optimised for, we can replace NA with an acronym for the noise model, e.g. SN for stochastic noise, RE for rotational errors, or ORE for (stochastic) over-rotation errors.

In the following subsections we will explore some of the components of our taxonomy in more detail.

¹⁶The definition of a gate is used quite flexibly for local mitigation, each ‘gate’ could be a single recognised gate, a collection of adjacent gates, or a physical operation that is part of what makes up a gate.

4.2 Variant Generation Procedures

In order to effect our mitigation we make use of a number of noisy gate variants ($\tilde{U}_i$):

$$\tilde{U}_i = U \mathcal{E}_i, \quad (79)$$

which have the same noise-free component (U) but different noise channels ($\mathcal{E}_i$).

4.2.1 Type Preserving Procedures

For global mitigation (Sec. 4.4) we focus on type preserving variant generation procedures (TPP, akin to Ref. [18] but extended to more general noise models). These procedures result in noisy gate variants with channels of the same type (see Sec. 3) as the original channel. This means we can use the same form of model (see Sec. 2.2.1) for the amplified channel as for the original channel, which simplifies the mitigation procedure (see Sec. 4.4.1). This means we require:

$$\mathcal{E}_i = \begin{cases} \begin{cases} 1 - p_i & : \text{---} \\ p_i & : \mathcal{N} \end{cases} & \text{for stochastic noise,} \\ e^{i \frac{\phi_i}{2} P} & \text{for rotational errors.} \end{cases}, \quad (80)$$

with $p_i \neq p$ or $\phi_i \neq \phi$ for stochastic noise and rotational errors, respectively. The restrictions on the channels imposed by the requirement for type preservation will vary depending on the variant generation procedures one has access to and the noise channels that need mitigating. The hidden inverse procedure should be a TPP if it is implementable (i.e. if it is possible to reverse the sign of coherent error channels, see Sec. 4.2.2). The use of custom channels (i.e. placing a single tailoring noise channel after the noisy gate of the same type¹⁷ as the original noise channel) is a TPP for rotational errors (Sec. 3.4); and also for stochastic noise (Sec. 3.3), provided that the channel is closed (Sec. 3.2.3). For identity insertion to be a TPP, we have the same requirements as for custom channels but we also require that the noise channels of the noisy gates commute with the noise-free components of the noisy gates. There are other variant production techniques that we do not explicitly consider here, e.g. the tuning of Hamiltonian parameters (or pulse stretching, see Refs [16, 19, 20]), and these will have different requirements to be TPPs.

For local mitigation (Sec. 4.5) we do not require TPPs and it is sufficient for the error channels being mitigated to be self-compatible (see Sec. 3.2.2) and implementable. This means for identity insertion the noise component must commute with the noise-free component of the noisy gate. For the hidden inverse procedure it must be possible to reverse the sign of coherent error channels.

If we use the completely general form of probabilistic error cancellation (see Ref. [16]) then we do not explicitly require TPPs or self-compatible error channels. We only need access to a set of (noisy) gates or physical operations with our target gate within their span (e.g. see Ref. [14]).

¹⁷If we permit custom (tailoring) channels of more general types, i.e. not necessarily of the same type as the noise channel they are tailoring, or we permit the use of multiple sequential custom channels, then the use of custom channels is a TPP for more general noise models.

4.2.2 Hidden Inverses

Hidden inverses (HIs) are used to invert coherent error channels. The idea comes from Refs [21, 22]. Their method is designed for Hermitian (self-inverse) gates with coherent error channels.

Gates are implemented by a sequence of physical operations, if the order and direction of these physical operations can be reversed then we should be able to produce the inverse of our original gate, e.g. if our noisy gate ($\tilde{U}$) is produced using the physical operations $\tilde{A}$, $\tilde{B}$, and $\tilde{D}$ ([21]):

$$\tilde{U} = \tilde{A} \circ \tilde{B} \circ \tilde{D}, \quad (81)$$

then the inverse gate is given by:

$$\widetilde{U^\dagger} = \widetilde{D^{-1}} \circ \widetilde{B^{-1}} \circ \widetilde{A^{-1}}. \quad (82)$$

If our gate (without noise) is Hermitian then the noise-free components of both implementations ($\tilde{U}$ and $\widetilde{U^\dagger}$) are the same (i.e. $U = U$). This means the only difference between $\tilde{U}$ and $\widetilde{U^\dagger}$ should be the noise channel. If the noise is completely coherent then $\widetilde{U^\dagger}$ should have an error channel that is the inverse of the error channel of $\tilde{U}$. So, if we have a fully coherent error channel then the effective error channel of a hidden inverse is given by:

$$\boxed{\mathcal{E}}_{\text{HI}} = \left(\boxed{\mathcal{E}} \right)^\dagger. \quad (83)$$

We can expand the idea of Refs [21, 22] to encompass gates suffering from over-rotation errors (ORE, see Sec. 3.5), Hermitian or not. We assume our target rotation is ω ($0 \leq \omega \leq 2\pi$) and that we consistently over-rotate by an angle ϕ . So the noisy gate is given by:

$$\boxed{\widetilde{U(\omega)}} = \boxed{e^{i\frac{\omega+\phi}{2}P}} = \boxed{U(\omega)} \boxed{\mathcal{E}_{\text{RE}}}, \quad (84)$$

where:

$$\boxed{U(\omega)} = \boxed{e^{i\frac{\omega}{2}P}}, \quad \boxed{\mathcal{E}_{\text{RE}}} = \boxed{e^{i\frac{\phi}{2}P}}, \quad (85)$$

are the noise-free gate and noise channel respectively. We assume that it is a true over-rotation so our target (ω) and erroneous (ϕ) rotations always have the same sign, i.e. $\omega\phi \geq 0$ regardless of the value of ω (this approach also works if we have a true under-rotation error, i.e. we always have $\omega\phi \leq 0$). In this case we define the hidden inverse (target) angle:

$$\omega_{\text{HI}} = \omega - 2\frac{\omega}{|\omega|}\pi, \quad (86)$$

so $\omega_{\text{HI}}\omega < 0$ and the hidden inverse is a gate with a target angle of ω_{HI} instead of ω :

$$\boxed{\widetilde{U(\omega)}}_{\text{HI}} = \boxed{\widetilde{U(\omega_{\text{HI}})}} = \boxed{e^{i\frac{\omega_{\text{HI}}+\phi_{\text{HI}}}{2}P}} = \boxed{U(\omega - 2\pi\omega/|\omega|)} \boxed{e^{i\frac{\phi_{\text{HI}}}{2}P}}. \quad (87)$$

$$= \boxed{-1} \boxed{U(\omega)} \boxed{\mathcal{E}_{\text{RE}}}_{\text{HI}} = \boxed{U(\omega)} \boxed{\mathcal{E}_{\text{RE}}}_{\text{HI}}. \quad (88)$$

Since a global sign for a gate is irrelevant (for a quantum channel) this is equivalent to our original noisy gate but with a rotational error ϕ_{HI} instead of ϕ . Since we assume the sign

of the rotational error is always the same as that of the target rotation (and ω_{HI} has the opposite sign to ω) this means that ϕ_{HI} will have the opposite sign to ϕ . Thus $\tilde{U}_{\text{HI}}$ is a gate variant with the opposite rotational error to our original gate, but the same noise-free component. If the magnitude of the over-rotational error is fixed, then we have $\phi = -\phi_{\text{HI}}$ and therefore $\mathcal{E}_{\text{RE;HI}} = \mathcal{E}_{\text{RE}}^\dagger$, as with the standard hidden inverse procedure. However, as we will see in Sec. 6.2.2, we can tolerate some difference in the magnitudes of ϕ_{HI} and ϕ as long as their signs are opposed and they are the same order of magnitude (i.e. $\phi_{\text{HI}}^2 \ll \phi$, $\phi^2 \ll \phi_{\text{HI}}$).

So we should be able to apply the hidden inverse procedure to Hermitian operators/gates and operators/gates suffering from correlated over-rotation errors (i.e. when the relative sign between the rotational error and the target rotation is fixed and independent of the rotation angle).

4.2.3 Custom channels

We can also produce gate variants by adding a custom (C) tailoring channel ($\mathcal{E}_{\text{C},i}$) after the original noisy gate:

$$\boxed{\tilde{U}}_i = \boxed{\tilde{U}} \boxed{\mathcal{E}}_{\text{C},i} = \boxed{U} \boxed{\mathcal{E}} \boxed{\mathcal{E}}_{\text{C},i}, \quad (89)$$

where $\tilde{U}_i$ is the i th tailored gate (i.e. noisy gate variant), $\tilde{U}$ is the original noisy gate, $\mathcal{E}$ is the original noise channel, and the effective noise channel for the i th variant is given by:

$$\boxed{\mathcal{E}}_i = \boxed{\mathcal{E}} \boxed{\mathcal{E}}_{\text{C},i}. \quad (90)$$

This idea is implemented for Pauli noise channels in Refs [23–25].

For this work we use a tailoring error channel that has the same type as the original error channel, so (Secs 3.3 and 3.4):

$$\boxed{\mathcal{E}}_{\text{C},i} = \begin{cases} \begin{cases} 1 - r_i & : \text{---} \\ r_i & : \boxed{\mathcal{N}} \end{cases} & \text{for stochastic noise,} \\ \boxed{e^{i\frac{\theta_i}{2}P}} & \text{for rotational errors.} \end{cases} \quad (91)$$

The custom stochastic noise channel involves doing nothing with probability $1 - r_i$ and applying $\mathcal{N}$ with probability r_i , while the rotational channel involves deterministically applying a rotational error of θ_i generated by the same P that generates the original error channel.

If we apply the tailoring channel (Eq. (91)) we end up with the effective channel in Eq. (80) with:

$$p_i = p + r_i(p_{\text{Max}} - p), \quad \phi_i = \phi + \theta_i, \quad (92)$$

where p and ϕ are the original unamplified errors and we have assumed that the stochastic noise channel is closed (see Eq. (59) of Sec. 3.3, and Secs 3.2.3 and 4.2.1). Since $0 \leq r_i \leq 1$ we have (see Eq. (59)):

$$p_{\text{Max}} = 1 - ap, \quad (93)$$

as the largest stochastic error we can realise using a custom channel, assuming $p \ll 1$.

We have found the optimal distributions for minimising the sampling cost of the mitigation are given by:

$$r_i = \frac{1 - \cos \frac{i\pi}{N_M}}{2}, \quad \text{and} \quad \theta_0 = 0, \quad \theta_i = \frac{(2i-1)\pi}{N_M}, \quad (94)$$

for stochastic and rotational errors, respectively, where $N_M + 1$ is the number of variants in the set of noisy gates ($0 \leq i \leq N_M$, see Ref. [1] for more details). We choose r_i and θ_i independently of the original errors (p and ϕ), since otherwise we are in the realm of feedback control rather than mitigation.

The custom channels we use are optimised to reduce the sampling cost using quantitative knowledge of the noise model. In Ref. [26] they investigate optimising (effectively) customised channels to reduce the bias, for a fixed sampling cost. They do without explicit knowledge of the noise model since they use Hamiltonian rescaling. Their work complements our own and, for stochastic noise, our custom channels could be considered as special cases of their extremal Chebyshev nodes (not their tilted Chebyshev nodes).

Customised channels and Hidden Inverses If we have access to both hidden inverses (see Sec. 4.2.2) and custom channels (i.e. CHI) then we can use:

$$\phi_i = \phi + \theta_i, \quad \theta_i = \frac{2i\pi}{N_C}, \quad \phi_{N_C} = \phi_{\text{HI}} = -\phi, \quad (95)$$

for mitigating rotational errors, where for $i = N_C$ we use a hidden inverse and for the other (non-zero) i s (i.e. $0 < i < N_C$) we use custom channels.

4.2.4 Identity Insertion

Identity insertion (II, see Ref. [27]) is a knowledge-free noise-amplification technique. We amplify the noise by inserting resolutions of the identity, in terms of our noisy gate and its noisy inverse (which may be the same gate if the noise-free gate is Hermitian), before the original noisy gate, in the hope that this will amplify the noise channel without changing its structure. II requires the noisy gate ($\tilde{U}$) and its (noisy) inverse¹⁸ ($U^\dagger$) to have the same noise channel ($\mathcal{E}$) and for this noise channel to commute with both the noisy and noisy inverse gates. For standard II the gate $\tilde{U}_{;i}$ is produced using i identity insertions (see Eq. (44)):

$$-\boxed{\tilde{U}}_i = -(\boxed{\tilde{U}} \boxed{U^\dagger})^i \boxed{\tilde{U}} = -(\boxed{U} \boxed{\mathcal{E}} \boxed{U^\dagger} \boxed{\mathcal{E}})^i \boxed{U} \boxed{\mathcal{E}} = -\boxed{U} (\boxed{\mathcal{E}})^{2i+1}. \quad (97)$$

So the effective error channel of the i th noisy variant ($\tilde{U}_{;i}$) is:

$$-\boxed{\mathcal{E}}_i = (\boxed{\mathcal{E}})^{2i+1}. \quad (98)$$

¹⁸The noisy inverse gate ($\tilde{U}^\dagger$) is given by:

$$-\boxed{\tilde{U}^\dagger} = -\boxed{U^\dagger} \boxed{\mathcal{E}} \quad (96)$$

and is distinct from the inverse noisy gate ($(\tilde{U})^{-1}$) so $\tilde{U}\tilde{U}^\dagger \neq \mathbb{I}$, in general.

To first order, in the original noise amplitude (e.g. p or ϕ), the noise amplitude of the amplified gate ($\tilde{U}_{;i}$) is $2i + 1$ times the noise amplitude of the unamplified gate ($\tilde{U}$) and we have:

$$p_i = \frac{1 - (1 - p(1 + a))^{2i+1}}{1 + a}, \quad \phi_i = (2i + 1) \phi. \quad (99)$$

for stochastic noise and rotational errors, respectively, where we have assumed the stochastic noise channel is closed ($0 \leq a \leq 1$, see Secs 3.2.3 and 3.3).

Tuned Identity Insertion Tuned identity insertion (TII) is a variant of the identity insertion method where, rather than using i identity insertions for the i th noisy gate variant ($\tilde{U}_{;i}$), we make the number of identity insertions, used for each variant, a free parameter (m_i):

$$\boxed{\tilde{U}}_i = -(\boxed{\tilde{U}} - \boxed{\tilde{U}^\dagger})^{m_i} \boxed{\tilde{U}} = -\boxed{U}(-\boxed{\mathcal{E}})^{2m_i+1}. \quad (100)$$

We choose m_i to minimise the runtime scaling (see Sec. 2.3.4). We have not yet found an analytic expression for m_i so we optimise it numerically (see Table 7 and Figs (2)-(7)). To first order the noise amplitude of the amplified gate is $2m_i + 1$ times the noise amplitude of the unamplified gate ($\tilde{U}$) as we have:

$$p_i = \frac{1 - (1 - p(1 + a))^{2m_i+1}}{1 + a}, \quad \phi_i = (2m_i + 1) \phi. \quad (101)$$

for stochastic noise and rotational errors, respectively, where we have assumed the stochastic noise channel is closed (c.f. Eqs (99) and (101)).

Custom channel Identity Insertion Custom channel identity insertion (CII) is a variant of identity insertion where we produce a single noisy variant (and its noisy inverse) of our noisy gate (we do not need to produce the original noisy gate at all) with amplified noise using custom error channels. For rotational errors we have:

$$\begin{aligned} \boxed{\tilde{U}}_T &= \boxed{\tilde{U}} - \boxed{\mathcal{E}}_C = \boxed{U} - \boxed{\mathcal{E}} - \boxed{\mathcal{E}}_C = \boxed{U} - \boxed{\mathcal{E}}_T, & \boxed{\mathcal{E}}_C &= \boxed{e^{i\frac{\theta}{2}P}}, \\ \boxed{\tilde{U}^\dagger}_T &= \boxed{\tilde{U}^\dagger} - \boxed{\mathcal{E}}_C = \boxed{U^\dagger} - \boxed{\mathcal{E}} - \boxed{\mathcal{E}}_C = \boxed{U^\dagger} - \boxed{\mathcal{E}}_T, & \boxed{\mathcal{E}}_T &= \boxed{e^{i\frac{\phi+\theta}{2}P}}. \end{aligned} \quad (102)$$

We then use identity insertion as normal but with our amplified gate (and its amplified inverse), so the noisy gate variants we use for the mitigation are:

$$\boxed{\tilde{U}}_i = -(\boxed{\tilde{U}}_T - \boxed{\tilde{U}^\dagger}_T)^i \boxed{\tilde{U}}_T = -\boxed{U}(-\boxed{\mathcal{E}}_T)^{2i+1}. \quad (103)$$

Here we have assumed the noisy inverse gate is amplified by the same amount as the original noisy gate and that they have the same error channel after amplification. For rotational errors we have found that the optimal amplification is given by:

$$\theta = \frac{2\pi}{3}, \quad \implies \quad \phi_i = (2i + 1) \left(\frac{2\pi}{3} + \phi \right), \quad (104)$$

for unbiased local mitigation. Here ϕ_i is just a variant of Eq. (99) with the original noise level replaced by the amplified noise (i.e. ϕ replaced by $\phi + \theta$). If we permit different custom channels for the noisy gate and its inverse we can achieve lower runtime scalings (Sec. 2.3.4) at the cost of additional complexity. There is no equivalent approach for stochastic noise; as stochastic noise is not periodic.

4.3 Pre-tailoring: Local Cancellation

Prior to performing the mitigation proper we can apply some tailoring to our gates to make them more amenable to mitigation.

A common strategy is Pauli twirling [28–31] (or randomised compiling [32]), which is a knowledge-free tailoring strategy to convert non-Pauli error channels into Pauli error channels (Eq. (47) of Sec. 3.2). It involves stochastically sandwiching noisy gates between other (twirling) operators. In order for it to function correctly the twirling procedure should not introduce any non-Pauli errors, which, in general, means the twirling operators should be noise free.

A similar strategy is Clifford twirling where one requires an enlarged twirling set, but the resulting channel is depolarising. In Ref. [33] they propose Clifford twirling the entire circuit, to obtain global depolarising noise, which is particularly easy to mitigate.

We introduce local cancellation (LC) as an alternative to Pauli twirling, which achieves similar ends but has different requirements. The idea is that one finds, for each noisy gate, a stochastic combination of noisy gate variants that is equivalent to the noise-free gate followed by a stochastic noise channel. The full procedure is given in App. C.11 and requires knowledge of the noise model. However, here we shall consider the special case of applying local cancellation to a gate suffering from an over-rotation error for which we can apply hidden inverses perfectly (see Sec. 4.2.2), to invert the over-rotation error channel.

If we assume we have a noisy gate with an over-rotation error ϕ then our locally cancelled gate is given by (see Eqs (51), (62), and (83)):

$$\boxed{\tilde{U}}_{\text{LC}} = \begin{cases} \frac{1}{2} : \boxed{\tilde{U}} \\ \frac{1}{2} : \boxed{\tilde{U}}_{\text{HI}} \end{cases} = \boxed{U} \begin{cases} \frac{1}{2} : \boxed{\mathcal{E}_{\text{RE}}} \\ \frac{1}{2} : \boxed{\mathcal{E}_{\text{RE}}}_{\text{HI}} \end{cases} = \boxed{U} \boxed{\mathcal{E}_{\text{SN}}}, \quad (105)$$

where $\tilde{U}$ is the original noisy gate, $\tilde{U}_{\text{HI}}$ is its hidden inverse, U is its noise-free component, $\mathcal{E}_{\text{RE}}$ is its original rotational error noise channel, $\mathcal{E}_{\text{RE}}_{\text{HI}}$ is the noise channel of its hidden inverse, and $\mathcal{E}_{\text{SN}}$ is its stochastic noise channel after local cancellation, where:

$$\boxed{\mathcal{E}_{\text{RE}}} = \boxed{e^{i\frac{\phi}{2}P}}, \quad \boxed{\mathcal{E}_{\text{RE}}}_{\text{HI}} = \boxed{e^{-i\frac{\phi}{2}P}}, \quad \boxed{\mathcal{E}_{\text{SN}}} = \begin{cases} \text{---} & \text{with prob. } 1 - p_{\text{LC}}, \\ \boxed{P} & \text{with prob. } p_{\text{LC}}. \end{cases} \quad (106)$$

So, instead of implementing our original noisy gate for every run of the circuit, we implement the original gate half the time and its hidden inverse the other half of the time. Since this is stochastic (the coefficients are true probabilities) this tailoring is free (i.e. it has a runtime scaling, Eq. (33), of unity), and since it is a linear ‘mitigation’ (Sec. 2.1) it can be composed with other linear mitigation procedures as a pre-tailoring step (see Sec. 4.4.2). The effective error probability (p_{LC}) of the effective stochastic noise channel is given by:

$$p_{\text{LC}} = \frac{1 - \cos \phi}{2} = \frac{\phi^2}{4} + \mathcal{O}_0(\phi^4), \quad (107)$$

(see App. C.11 for details). This means, if the original noise level is given by e_{RE} , after local cancellation the noise level is given by (see Sec. 2.3.2):

$$e_{\text{SN}} = 2N_{\text{G}}p_{\text{LC}} + \mathcal{O}_{\text{L}}\left(\frac{e_{\text{RE}}^4}{N_{\text{G}}^3}\right) = \frac{e_{\text{RE}}^2}{2N_{\text{G}}}, \quad e_{\text{RE}} = N_{\text{G}}\phi. \quad (108)$$

So, for fixed noise level (e_{RE}), the effective noise level after local cancellation will tend to zero in the large circuit limit (Sec. 2.4.3). Therefore local cancellation will not only tailor but will also mitigate the noise in the small noise level or large circuit regime.

4.4 Global Mitigation

When we apply global mitigation methods we mitigate uniform noise (i.e. a collection of gates with the same type of error channel, e.g. all stochastic or rotational but not necessarily the same noise channel: $\mathcal{N}$ or Hermitian operator: P) with the same noise amplitude (e.g. the same ϕ or p). If we wish to mitigate multiple different types of error channels with global mitigation methods then we must compose multiple separate methods (see Sec. 4.4.2). We will start by providing an example of the derivation of linear mitigation coefficients for the mitigation of uniform stochastic noise or rotational errors.

4.4.1 Mitigation Coefficients

We consider noisy circuit variants (e.g. $\tilde{\mathcal{C}}_{;i}$) with gates suffering from stochastic noise (Sec. 3.3) or rotational errors (Sec. 3.4) with noise amplitude p_i or ϕ_i , respectively, for circuit variant i . In these cases we have a simple model of the output of the full noisy circuit (Eqs (56) and (69)):

$$\tilde{\mathcal{C}}_{;i} = \begin{cases} \sum_{k=0}^{N_{\text{G}}} (1-p)^{N_{\text{G}}-k} p^k \check{\chi}_{\text{SN},k} & \text{for stochastic noise,} \\ \sum_{k=0}^{2N_{\text{G}}} \cos^{2N_{\text{G}}-k} \frac{\phi}{2} \sin^k \frac{\phi}{2} \check{\chi}_{\text{RE},k} & \text{for rotational errors,} \end{cases}$$

where $\check{\chi}_{\text{RE},0} = \mathcal{C} = \check{\chi}_{\text{SN},0}$ are the noise-free circuits. Now the mitigated circuit is a linear superposition of these noisy circuits (Eq. (3) of Sec. 2.1):

$$\tilde{\mathcal{C}}_{;\text{M}} = \sum_{i=0}^{N_{\text{C}}} c_{\text{GL};i} \tilde{\mathcal{C}}_{;i},$$

so the noise-free circuit can be obtained if we require:

$$\begin{cases} \sum_{i=0}^{N_{\text{C}}} c_{\text{GL};i} (1-p_i)^{N_{\text{G}}} = 1, & \sum_{i=0}^{N_{\text{C}}} c_{\text{GL};i} (1-p_i)^{N_{\text{G}}-k} p_i^k = 0 & \text{for stochastic noise,} \\ \sum_{i=0}^{N_{\text{C}}} c_{\text{GL};i} \cos^{2N_{\text{G}}} \frac{\phi_i}{2} = 1, & \sum_{i=0}^{N_{\text{C}}} c_{\text{GL};i} \cos^{2N_{\text{G}}-k} \frac{\phi_i}{2} \sin^k \frac{\phi_i}{2} = 0, & \text{for rotational errors,} \end{cases} \quad (109)$$

for $0 < k \leq N_{\text{M}}$ where:

$$N_{\text{M}} = \begin{cases} N_{\text{G}}, & \text{for stochastic noise,} \\ 2N_{\text{G}}, & \text{for rotational errors,} \end{cases} \quad (110)$$

which will result in unbiased mitigation.

If we can accept a biased mitigation we could use:

$$\sum_{i=0}^{N_{\text{M}}} c_{\text{GL};i} = 1, \quad \sum_{i=0}^{N_{\text{M}}} c_{\text{GL};i} p_i^k = 0, \quad (111)$$

for $0 < k \leq N_M \leq N_G$ for stochastic noise and:

$$\sum_{i=0}^{N_M} c_{\text{GL};i} = 1, \quad \sum_{i=0}^{N_M} c_{\text{GL};i} \sin \phi_i \sin^{2(m-1)} \frac{\phi_i}{2} = 0, \quad \sum_{i=0}^{N_M} c_{\text{GL};i} \sin^{2m} \frac{\phi_i}{2} = 0, \quad (112)$$

for rotational errors. In the latter case we use the first $N_M \leq 2N_G$ equations (iterating from left to right of Eq. (112) for sequential values of m until we reach $N_M + 1$ equations). This should neglect the mitigation of higher order terms¹⁹ (e.g. terms of the same order as $\mathcal{O}_0(p^{N_M+1})$ or $\mathcal{O}_0(\phi^{N_M+1})$), which will have a smaller contribution to the overall error.

The mitigation coefficients that satisfy Eqs (111) and (112) are given by:

$$c_{\text{GL};i} = \begin{cases} \prod_{m=0:m \neq i}^{N_M} \frac{p_m}{p_m - p_i}, & \text{for stochastic noise,} \\ \prod_{j=0,j \neq i}^{N_M} \frac{\sin \frac{\phi_j}{2}}{\sin \frac{\phi_j - \phi_i}{2}}, & \text{for rotational errors,} \end{cases} \quad (113)$$

(where the rotational expression only holds for even N_M). The strategy outlined above is for synchronous mitigation strategies, where all the noisy gates are amplified by the same amount. So all the noisy circuits ($\tilde{\mathcal{C}}_{;i}$) suffer from uniform noise. For asynchronous strategies we amplify each noisy gate by a different amount so even if the original circuit suffered from uniform noise the variant circuits ($\tilde{\mathcal{C}}_{;i}$) may not. All concatenated local methods (see Sec. 4.5.2) are examples of asynchronous methods but the only global non-synchronous method we consider is identity-insertion-assisted asynchronous mitigation (IIAM) otherwise known as RIIM [5, 34]. In App. C.8 we provide mitigation coefficients for IIAM that can be used for mitigating both stochastic noise and rotational errors (and perhaps other error types), generalising Ref. [34] (which targets stochastic noise).

4.4.2 Mitigation Composition

Mitigation composition involves (effectively) applying mitigation strategies sequentially to the same circuit to strip away one uniform noise layer at a time. The full implementation is discussed in Ref. [1] but we will present a brief example. We assume that we have a circuit ($\tilde{\mathcal{C}}_{\text{SN,RE}}$) suffering from independent stochastic noise and rotational errors. First, we derive two independent mitigation strategies. The first one mitigates stochastic noise (see Eq. (3)):

$$\tilde{\mathcal{C}}_{\text{SN};M} = \sum_{i=0}^{N_{M,\text{SN}}} c_{\text{SN};i} \tilde{\mathcal{C}}_{\text{SN};i} = \mathcal{C} + \mathcal{O}_S \left(e_{\text{SN}}^{N_{M,\text{SN}}+1} \right), \quad (114)$$

where $\tilde{\mathcal{C}}_{\text{SN};M}$ is the mitigated version of a circuit suffering from stochastic noise, $N_{M,\text{SN}}$ is the order of the stochastic mitigation, $\tilde{\mathcal{C}}_{\text{SN};i}$ is the i th noisy circuit variant (suffering from stochastic noise) that we use in our mitigation, and $\mathcal{C}$ is the noise-free (target) circuit. The second strategy is to mitigate rotational errors (see Eq. (3)):

$$\tilde{\mathcal{C}}_{\text{RE};M} = \sum_{j=0}^{N_{M,\text{RE}}} c_{\text{RE};j} \tilde{\mathcal{C}}_{\text{RE};j} = \mathcal{C} + \mathcal{O}_S \left(e_{\text{RE}}^{N_{M,\text{RE}}+1} \right), \quad (115)$$

where $\tilde{\mathcal{C}}_{\text{RE};M}$ is the mitigated version of a circuit suffering from rotational errors, $N_{M,\text{RE}}$ is the order of the rotational mitigation, and $\tilde{\mathcal{C}}_{\text{RE};j}$ is the j th noisy circuit variant (suffering

¹⁹In order for us to achieve exponential suppression of errors we require $p_i^2 \ll p$ and/or $\phi_i^2 \ll \phi$ for all i .

from rotational errors) that we use in our mitigation. Now, we define variant generation procedures $\mathfrak{G}_{\text{SN};i}$ and $\mathfrak{G}_{\text{RE};j}$ where:

$$\tilde{\mathcal{C}}_{\text{SN};i} = \mathfrak{G}_{\text{SN};i} \circ \tilde{\mathcal{C}}_{\text{SN}}, \quad \tilde{\mathcal{C}}_{\text{RE};j} = \mathfrak{G}_{\text{RE};j} \circ \tilde{\mathcal{C}}_{\text{RE}}, \quad (116)$$

where $\tilde{\mathcal{C}}_{\text{SN}}$ and $\tilde{\mathcal{C}}_{\text{RE}}$ are the original noisy circuits suffering from stochastic noise or rotational errors, respectively, and $\tilde{\mathcal{C}}_{\text{SN};j}$ and $\tilde{\mathcal{C}}_{\text{RE};j}$ are the j th amplified (or otherwise modified) circuits used for the mitigations (see Secs 2.1 and 4.2). So, for example, $\mathfrak{G}_{\text{RE};j}$ could be a procedure that amplifies all the rotational error amplitude from ϕ to ϕ_j (see Sec. 4.2.3), replaces all the gates with rotational errors with their hidden inverses (see Sec. 4.2.2), or applies identity insertion j times before each gate with a rotational error (see Sec. 4.2.4).

Now, if we have a circuit that suffers from both rotational errors and stochastic noise ($\tilde{\mathcal{C}}_{\text{SN,RE}}$) then we simply apply the mitigation method for rotational errors followed by²⁰ the one for stochastic noise. The composed mitigation strategy is given by:

$$\tilde{\mathcal{C}}_{\text{SN,RE};M} = \sum_{i=0}^{N_C} c_{\text{Gl};i} \tilde{\mathcal{C}}_{\text{SN,RE};i} = \mathcal{C} + \mathcal{O}_S \left(e_{\text{SN}}^{N_{\text{M,SN}}+1}, e_{\text{RE}}^{N_{\text{M,RE}}+1} \right), \quad (117)$$

where:

$$\tilde{\mathcal{C}}_{\text{SN,RE};i} = \mathfrak{G}_{\text{SN};i_2} \circ \mathfrak{G}_{\text{RE};i_1} \circ \tilde{\mathcal{C}}_{\text{SN,RE}}, \quad c_{\text{Gl};i} = c_{\text{SN};i_2} c_{\text{RE};i_1}, \quad (118)$$

$$N_C = (N_{\text{M,SN}} + 1)(N_{\text{M,RE}} + 1) - 1 \quad i_1 = i \bmod N_{\text{M,RE}}, \quad i_2 = \left\lfloor \frac{i}{N_{\text{M,RE}}} \right\rfloor, \quad (119)$$

where $\left\lfloor \frac{i}{N_{\text{M,RE}}} \right\rfloor$ is the integer part of the division of i by $N_{\text{M,RE}}$ and $i \bmod N_{\text{M,RE}}$ is the remainder.

In general, our sampling cost is the product of the sampling costs of the constituent mitigation methods. So, for our example, we have (Eq. (32)):

$$C_{\text{Gl}} = \sum_{i=0}^{N_C} |c_{\text{Gl};i}| = C_{\text{SN}} C_{\text{RE}}, \quad C_{\text{SN}} = \sum_{j=0}^{N_{\text{M,SN}}} |c_{\text{SN};j}|, \quad C_{\text{RE}} = \sum_{j=0}^{N_{\text{M,RE}}} |c_{\text{RE};j}|, \quad (120)$$

and the length scale factor is given by the sum of the length scale factors of the constituent mitigation methods minus one less than the number of constituent mitigation methods, so for our example we have (Eq. (32)):

$$F_{\text{Gl,L}} = \frac{\sum_{i=0}^{N_C} |c_{\text{Gl};i}|}{C_{\text{Gl}}} \frac{\tau_{\text{E},i}}{\tau_{\text{E}}} = F_{\text{SN,L}} + F_{\text{RE,L}} - 1, \quad (121)$$

$$F_{\text{SN,L}} = \frac{\sum_{j=0}^{N_{\text{M,SN}}} |c_{\text{SN};j}|}{C_{\text{SN}}} \frac{\tau_{\text{E,SN},j}}{\tau_{\text{E}}}, \quad F_{\text{RE,L}} = \frac{\sum_{j=0}^{N_{\text{M,RE}}} |c_{\text{RE};j}|}{C_{\text{RE}}} \frac{\tau_{\text{E,RE},j}}{\tau_{\text{E}}}, \quad (122)$$

where τ_{E} is the runtime of the unmodified circuit, $\tau_{\text{E},i}$ is the runtime of the i th amplified circuit ($\tilde{\mathcal{C}}_{\text{SN,RE};i}$), $\tau_{\text{E,SN},j}$ is the runtime of the j th stochastically amplified circuit ($\tilde{\mathcal{C}}_{\text{SN};j}$), $\tau_{\text{E,RE},j}$ is the runtime of the j th rotationally amplified circuit ($\tilde{\mathcal{C}}_{\text{RE};j}$).

To the most significant order the mitigated noise level is the sum of the noise levels of the constituent mitigation strategies, so, for our example:

$$\varepsilon_{\text{Q};M} = \mathcal{O}_S \left(e_{\text{SN}}^{N_{\text{M,SN}}+1}, e_{\text{RE}}^{N_{\text{M,RE}}+1} \right), \quad (123)$$

assuming e_{SN} and e_{RE} are of the same order.

²⁰We assume in Eqs (117) and (118) that the two constituent mitigation methods commute, i.e. $\mathfrak{G}_{\text{SN};i_2}$ and $\mathfrak{G}_{\text{RE};i_1}$ commute for all i_1, i_2 . This may not be the case if we use identity insertion to amplify the errors. If the mitigation of one type of error changes the amplitude or type of another we must take this into account when applying our subsequent mitigation strategies.

4.5 Local Mitigation

When we apply local mitigation methods we (effectively) mitigate each gate (which could be a composite gate) independently, defining:

$$\boxed{\tilde{U}}_{\text{M}} = -\left\{c_{\text{L};i} : \boxed{\tilde{U}}_i\right\}_{i=0}^{N_{\text{L},\text{C}}} = -\boxed{U} \left\{c_{\text{L};i} : \boxed{\mathcal{E}}_i\right\}_{i=0}^{N_{\text{L},\text{C}}} \approx -\boxed{U}, \quad (124)$$

where the definition of the noisy gate variant $\tilde{U}_i$ and the values of $c_{\text{L};i}$ will depend on the mitigation method we use (see Table 1). U is the noise-free component of the noisy gate and $\mathcal{E}_i$ is the error channel associated to the i th noisy gate variant. We use $N_{\text{L},\text{C}}$ extra noisy gate variants (in addition to the original noisy gate). The curly brackets denote a quasi-stochastic combination, so the channel $\tilde{U}_{\text{M}}$ (equivalent to the effective gate we denote $\tilde{U}_{\text{M}}$) is given by (Eq. (79)):

$$\tilde{U}_{\text{M}} = \sum_{i=0}^{N_{\text{L},\text{C}}} c_{\text{L};i} \tilde{U}_i, \quad \text{where} \quad \tilde{U}_i(\rho) = \mathcal{E}_i(U\rho U^\dagger). \quad (125)$$

All unbiased local mitigation methods are examples of probabilistic error cancellation (PEC, see Ref. [14, 16, 35]) but rather than using the universal mitigation gate set of Ref. [14] or allowing complete generality (see Ref. [6]), we focus on finding optimal gate sets for particular noise models, building on the work of Ref. [18].

4.5.1 Eigenvalue Mitigation

If the transfer matrices (e.g. $\hat{T}_{\mathcal{E}_i}$, see Sec. 3.2.1) of all our noisy gate variant error channels (e.g. $\mathcal{E}_i$) can be diagonalised in the same basis, i.e. we have (Eq. (50)):

$$\hat{T}_{\mathcal{E}_i} = \hat{W} \hat{D}_i \hat{W}^{-1},$$

where $\hat{W}$ is independent of i , then we can perform eigenvalue mitigation. Eq. (124) (with equality instead of approximate equality) is equivalent to the requirement that the quasi-stochastic sum of our error channels is equivalent to the identity channel, which means:

$$\hat{T}_{\text{I}} = \sum_{i=0}^{N_{\text{L},\text{C}}} c_{\text{L};i} \hat{T}_{\mathcal{E}_i} = \hat{W} \left(\sum_{i=0}^{N_{\text{L},\text{C}}} c_{\text{L};i} \hat{D}_i \right) \hat{W}^{-1} \implies \sum_{i=0}^{N_{\text{L},\text{C}}} c_{\text{L};i} \hat{D}_i = \hat{T}_{\text{I}} \implies \sum_{i=0}^{N_{\text{L},\text{C}}} c_{\text{L};i} \lambda_{i,k} = 1, \quad (126)$$

where $\hat{T}_{\text{I}}$ is the identity channel transfer matrix, $c_{\text{L};i}$ is the i th mitigation coefficient, and $\lambda_{i,k}$ is the k th eigenvalue of the i th noisy gate variant used in our mitigation. So we require (at most) one (distinct) noisy gate variant for each eigenvalue of our noisy transfer matrix, for unbiased mitigation.

4.5.2 Mitigation Concatenation

To mitigate an entire circuit we must concatenate the mitigation strategies for the individual gates. This can be considered as a special case of mitigation composition (see Sec. 4.4.2) where each of the constituent strategies applies only to a single gate. Concatenated local mitigation (see [36]) involves determining mitigation coefficients ($c_{\text{L};j}$) and producing noisy gate variants ($\tilde{U}_{j;}$) for each gate independently and then concatenating

these (locally) mitigated gates in order to produce a complete mitigated circuit. As an example we consider the case of a noisy circuit ($\tilde{\mathcal{C}}$) consisting of two noisy gates:

$$\boxed{\tilde{\mathcal{C}}} = \boxed{\tilde{V}_1} \boxed{\tilde{V}_2}, \quad \boxed{\tilde{V}_1} = \boxed{V_1} \boxed{\mathcal{E}_1}, \quad \boxed{\tilde{V}_2} = \boxed{V_2} \boxed{\mathcal{E}_2}. \quad (127)$$

We replace each noisy gate with a mitigated version:

$$\boxed{\tilde{V}_1}_M = - \left\{ c_{L;1,j} : \boxed{\tilde{V}_1}_j \right\}_{j=0}^{N_{L,C,1}}, \quad \boxed{\tilde{V}_2}_M = - \left\{ c_{L;2,j} : \boxed{\tilde{V}_2}_j \right\}_{j=0}^{N_{L,C,2}}, \quad (128)$$

where $N_{L,C,m}$ is the number of extra gate variants required to mitigate the m th gate and $\tilde{V}_{m;j}$ is the j th noisy variant of the m th gate in our circuit. We then construct the mitigated circuit as the concatenation of the two mitigated gates:

$$\begin{aligned} \boxed{\tilde{\mathcal{C}}}_M &= \boxed{\tilde{V}_1}_M \boxed{\tilde{V}_2}_M = - \left\{ c_{L;1,j} : \boxed{\tilde{V}_1}_j \right\}_{j=0}^{N_{L,C,1}} \left\{ c_{L;2,j} : \boxed{\tilde{V}_2}_j \right\}_{j=0}^{N_{L,C,2}} \\ &= \left\{ c_{\text{Gl};i} : \boxed{\tilde{\mathcal{C}}}_i \right\}_{i=0}^{N_C}, \end{aligned} \quad (129)$$

where:

$$\begin{aligned} N_C &= (N_{L,C,1} + 1)(N_{L,C,2} + 1) - 1, \quad c_{\text{Gl};i} = c_{L;1,i_1} c_{L;2,i_2}, \quad \boxed{\tilde{\mathcal{C}}}_i = \boxed{\tilde{V}_1}_{i_1} \boxed{\tilde{V}_2}_{i_2}, \\ i_1 &= i \bmod N_{L,C,1}, \quad i_2 = \left\lfloor \frac{i}{N_{L,C,1}} \right\rfloor, \end{aligned} \quad (130)$$

where $\left\lfloor \frac{i}{N_{L,C,1}} \right\rfloor$ is the integer part of the division of i by $N_{L,C,1}$ and $i \bmod N_{L,C,1}$ is the remainder. N_C is the total number of extra circuits required for the global mitigation.

The sampling cost (see Sec. 2.3.4) for a concatenated local mitigation with N_G gates is given by:

$$C_{\text{Gl}} = \prod_{m=1}^{N_G} C_{L,m}, \quad C_{L,m} = \sum_{j=0}^{N_{L,C,m}} |c_{L;m,j}|, \quad (131)$$

where $C_{L,m}$ is the sampling cost for the m th gate. The length scale factor (see Sec. 2.3.4) is given by:

$$F_{\text{Gl},L} = \sum_{m=1}^{N_G} F_{L,L,m} \frac{\tau_{E,L,m}}{\tau_E}, \quad F_{L,L,m} = \frac{\sum_{j=0}^{N_{L,C,m}} |c_{L;m,j}| \tau_{E,L,m,j}}{C_{L,m} \tau_{E,L,m}}, \quad \tau_E = \sum_{m=1}^{N_G} \tau_{E,L,m}, \quad (132)$$

where $F_{L,L,m}$ is the length scale factor of the m th gate, $\tau_{E,L,m}$ is the unmitigated execution time of the m th gate, $\tau_{E,L,m,j}$ is the execution time of the j th noisy variant of the m th gate, and τ_E is the unmitigated execution time of the entire circuit. If we have a uniform noise model and local execution time ($\tau_E = \tau_{E,L} N_G$ and $\tau_{E,L,m} = \tau_{E,L}$, for all m); such that the sampling costs and length scale factors are identical for every gate ($F_{L,L,m} = F_{L,L}$ and $C_{L,m} = C_L$ (we could use the geometric mean for rigour)), we find:

$$F_{\text{Gl},L} = F_{L,L}, \quad C_{\text{Gl}} = C_L^{N_G}. \quad (133)$$

So if the local cost is linear in the noise amplitude (x : p or ϕ for stochastic noise or rotational errors, respectively), we find:

$$C_L = 1 + bx + \mathcal{O}_0(x^2) = e^{bx} (1 + \mathcal{O}_0(x^2)) \implies C_{\text{Gl}} = e^{bN_G x} (1 + \mathcal{O}_0(N_G x^2)), \quad (134)$$

that the global cost grows exponentially with the noise level. So a linear local cost is sufficient to ensure a finite cost in the large gate limit ($N_G \rightarrow \infty$) provided we have a constant noise level. At least for stochastic noise and rotational errors where the noise level is proportional to $N_G x$ (see Secs 2.3.2, 3.3, and 3.4).

In order to determine the most significant contributions to the proxy bias for concatenated mitigation methods we separate our circuit into collections of gates with uniform noise models, determine the most significant term in the proxy bias of each and then add them up to obtain the total noise level (see Sec. 4.4.2). This requires modelling the global effect of our mitigated channels on the circuit (see Secs 2.2.1, 2.2.2, and 2.3.2), which will, in general, be qualitatively different to the original unmitigated channels. If, however, the local mitigation strategy preserves the structure of the channel and only modifies the noise amplitude (x is reduced to x_M) then the leading contribution to the proxy bias is simply the noise level (Eq. (21)) with x replaced by x_M . So we have (see Secs 3.3 and 3.4):

$$\varepsilon_{\heartsuit;M} = \begin{cases} 2p_M N_G + \mathcal{O}_{LS} \left((p_M N_G)^2 \right), & \text{for stochastic noise,} \\ N_G |\phi_M| + \mathcal{O}_{LS} \left((N_G \phi_M)^2 \right), & \text{for rotational errors.} \end{cases} \quad (135)$$

where p_M is the stochastic noise amplitude after mitigation and ϕ_M is the rotational noise amplitude after mitigation.

4.6 Mitigation Method Catalogue

We now summarise the collection of mitigation methods derived from combinations of the variant generation procedures discussed above, which form the foundation of our taxonomy. Table 1 summarises the linear mitigation methods examined in this work, detailing their characteristics and their locations in the appendix. Each method can be combined with local cancellation (see Sec. 4.3) when addressing non-stochastic noise. Fig. 1 contextualises these methods within the overall taxonomy, illustrating that synchronised and concatenated local methods are subset of asynchronous methods, as the latter encompasses all possible approaches for generating circuit variants. Local mitigation methods can also (sometimes) be considered as synchronised mitigation applied to a single gate, this equivalence holds if we use a type preserving procedure (Sec. 4.2.1). In the following section we will certify our mitigation methods with respect to the criteria in Sec. 2.5.1.

Table 1: **Table of mitigation methods.** The naming convention is explained in Sec. 4.1. Key: C-Custom channel (Sec. 4.2.3), HI-Hidden Inverses (Sec. 4.2.2), T-Tuned (Sec. 4.2.4), II-Identity Insertion (Sec. 4.2.4); AM-Asynchronous Mitigation (Sec. 4.1), SM-Synchronous Mitigation (Sec. 4.1), LM-Local Mitigation (Sec. 4.5); KF-Knowledge Free (Sec. 4.1), NA-Noise Aware (Sec. 4.1). Where neither NA nor KF are specified the method is noise aware; except for IIAM, which is knowledge free. LC- can be prefixed to the other methods, which indicates we perform local cancellation (Sec. 4.3) before we carry out the other method (LC is knowledge free if we have rotational errors and perfect hidden inverses). An unbiased method has no (intrinsic) bias (Sec. 2.3.1) after mitigation so the only bias will be due to uncertainty in the characterisation or implementation of the mitigation.

Name	Alternative Acronym	Custom Channel	Hidden Inverse	Identity Insertion	Local	Synchronous	Noise Aware	Biased	Coefficients
CHILM		✓	✓	×	✓	×	✓	×	App. C.3
CHISM		✓	✓	×	×	✓	✓	×	App. C.4
CIILM		✓	×	✓	✓	×	✓	×	App. C.5
CLM		✓	×	×	✓	×	✓	×	App. C.6
CSM		✓	×	×	×	✓	✓	×	App. C.7
IIAM	RIIM [34]	×	×	✓	×	×	×	✓	App. C.8
IILM:KF		×	×	✓	✓	×	×	✓	App. C.9
IILM:NA		×	×	✓	✓	×	✓	×	App. C.9
IISM:KF	FIIM [34]	×	×	✓	×	✓	×	✓	App. C.10
IISM:NA		×	×	✓	×	✓	✓	✓	App. C.10
LC-		×	✓	×	✓	×	✓	✓	App. C.11
TIILM		×	×	✓	✓	×	✓	×	App. C.12

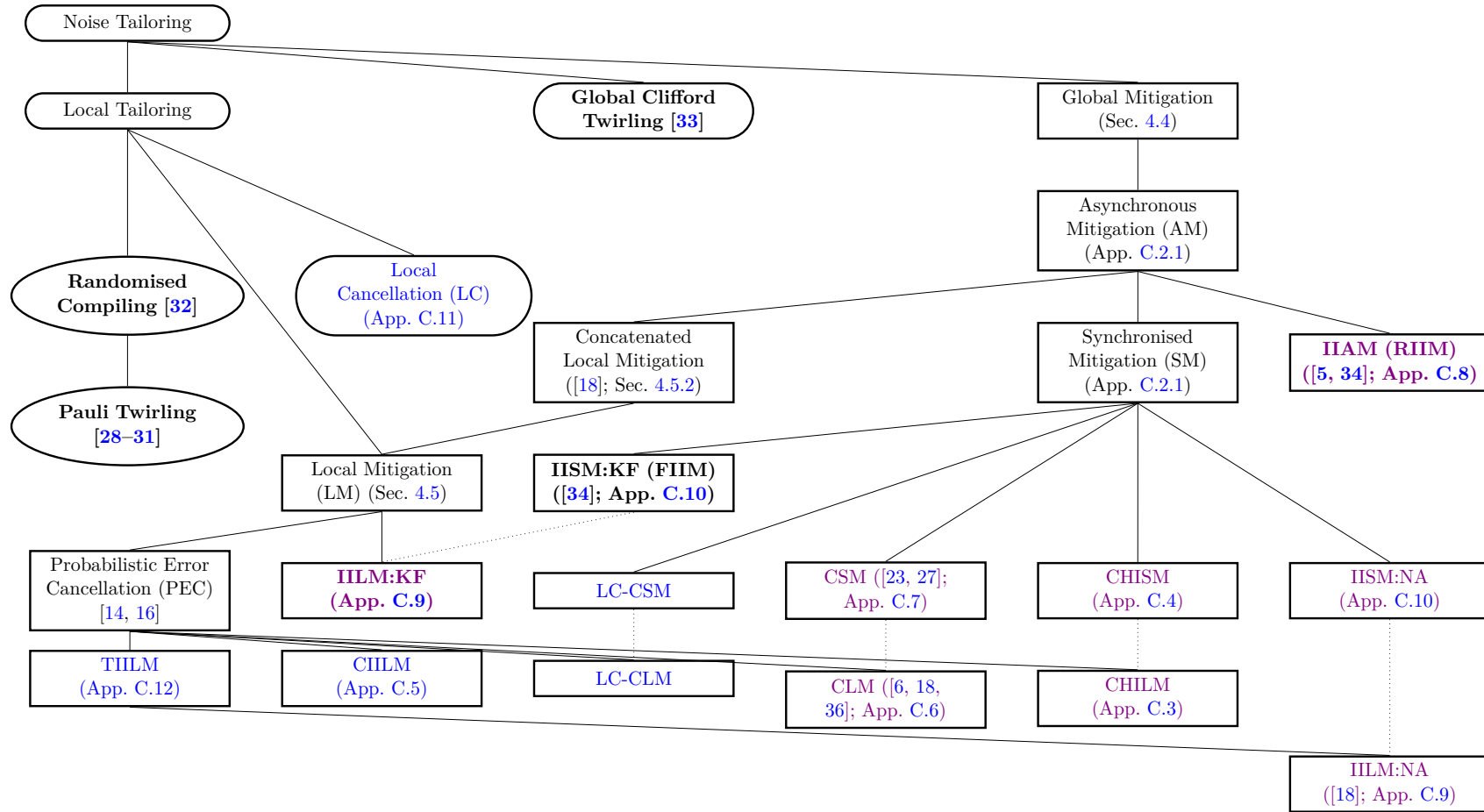


Figure 1: Taxonomy of the mitigation methods in Table 1 (this is far from a complete taxonomy, e.g. see [3, 4]): Rectangles are for mitigation methods, ellipses for tailoring methods designed to produce Pauli error channels from non-Pauli error channels, and rounded-vertex rectangles are for more general tailoring methods. Text in blue indicates methods developed in this work and text in violet indicates methods adapted or optimised within this work. Methods in bold font are knowledge free. A method that is below and linked to another method with a solid line is a special case of that method, if it is linked with a dotted line it is a special case only when the noise type is preserved by the variant generation procedure (see Sec. 4.2.1).

5 Case Study: Method Certification

Here certify the mitigation methods presented in Table 1 for stochastic noise (Sec. 5.1) and rotational errors (Sec. 5.2). We first present the quantitative metrics (Tables 2 and 4) from Sec. 2.3 in appropriate regimes (Sec. 2.4) then use these metrics to certify the methods (Tables 3 and 5) with respect to the criteria from Sec. 2.5.1.

We conclude by providing a numerical analysis (Sec. 5.3) of the methods for a simple simulated circuit (Sec. 5.3.1) suffering from stochastic over-rotation errors to elucidate our analytical discussion.

5.1 Certification for Stochastic Noise

We consider uniform closed stochastic noise (see Secs 3.3 and 3.2.3). We gauge the strength of the noise by the size of e_{SN} (Eq. (55)):

$$e_{\text{SN}} = 2pN_G.$$

The noise level (Sec. 2.3.2) of the circuit being mitigated (in the small noise level regime, Sec. 2.4.2). Here p is the (gate-wise) noise amplitude, i.e. the error probability for each gate, and N_G is the number of noisy gates.

In Table 2 we provide the leading order expressions for the proxy bias (i.e. the (mitigated) noise level, see Sec. 2.3.3), runtime scaling (Sec. 2.3.4), and noise boundary (Sec. 2.3.5) in the appropriate regimes (see Sec. 2.4). We then use these quantitative metrics in order to certify our mitigation methods in Table 3.

Table 2: **Quantitative Mitigation Metrics for Stochastic Noise.** The noise level (i.e. first order term of the proxy bias, see Sec. 2.3.1), runtime scaling (Sec. 2.3.4), and noise boundary (Sec. 2.3.5) are recorded for a number of mitigation methods. We consider both the small noise (see Sec. 2.4.2) and large circuit (see Sec. 2.4.3) regimes. For the runtime (Sec. 2.3.4) calculations we assume that all gates and custom error channels (Sec. 4.2.3) take the same time to implement. N_M , where given, is the order of the mitigation (see Sec. 2.3.5). All results are for the mitigation of uniform closed (see Sec. 3.2.3) stochastic noise (see Sec. 3.3). * Unmitigated results relate to the original noisy circuit. ** IIAM results are given for the combined large circuit and small noise regime, for more information see App. C.8. *** IISM results apply to both the knowledge-free (KF) and noise-aware (NA) variants and are given in the small noise regime. The runtime scaling is given to leading order for $N_M \gg 1$, for more information see Ref. [1]. **** $0 \leq a \leq 1$ is a property of the original noise model (see Eq. (59) of Sec. 3.3). Further details of each method can be found in Table 1 and Fig. 1 of Sec. 4. The naming convention is explained in Sec. 4.1.

Metric	Noise Level ($e_{\text{G:M}}$)	Excess Runtime Scaling ($S_{\text{GI}} - 1$)	Runtime Scaling (S_{GI})	Noise Boundary (e_{B})
Regime	Small Noise ($e_{\text{SN}} \ll 1$)	Small Noise ($e_{\text{SN}} \ll 1$)	Large Circuit ($N_G \gg 1$)	Large Circuit ($N_G \gg 1$)
Unmitigated*	e_{SN}	0	1	ε_+
CLM	0	$\left(2 + \frac{1}{2N_G}\right) e_{\text{SN}}$	$e^{2e_{\text{SN}}}$	$\frac{\ln(S_+)}{2}$
CSM	0	$\left(\frac{8}{3} - \frac{1}{6N_G^2}\right) N_G e_{\text{SN}}$	$2 \cosh^2 \sqrt{2N_G} e_{\text{SN}}$	$\frac{\ln^2(2S_+)}{8N_G}$
IIAM **	$\frac{e_{\text{SN}}^{N_M+1}}{(N_M+1)!}$	$\frac{N_G^{2N_M}}{(N_M!)^2} - 1$	$\frac{N_G^{2N_M}}{(N_M!)^2}$	ε_+
IILM:KF	$\frac{e_{\text{SN}}^{N_M+1}}{N_G^{N_M}} \left(\frac{1+a}{2}\right)^{N_M} \frac{\prod_{i=0}^{N_M} (2i+1)}{(N_M+1)!}$	$\geq \frac{3 \times 4^{N_G}}{2} - 1$	$\geq \frac{3 \times 4^{N_G}}{2}$	ε_+
IILM:NA	0	$\frac{3 \times 4^{N_G}}{2} - 1$	$\frac{3 \times 4^{N_G}}{2} e^{\frac{3}{4}(1+a)e_{\text{SN}}}$	ε_+
IISM***	$\frac{e_{\text{SN}}^{N_M+1} N_G! \prod_{i=0}^{N_M} (2i+1)}{N_G^{N_M+1} (N_G - N_M - 1)! (N_M+1)!}$	$\approx \frac{4^{N_M+1}}{\pi} - 1$	$\approx \frac{4^{N_M+1}}{\pi}$	$\approx \frac{\ln(2\varepsilon_+)}{4 \frac{\ln(\pi S_+)}{2}}$
TIILM****	0	1	$e^{4(1+a)e_{\text{SN}}} (1 + 2 \ln 2)$	$\frac{\ln \frac{S_+}{1+2 \ln 2}}{4(1+a)}$

Table 3: **Qualitative Certification Criteria for Stochastic Noise.** We assess each mitigation method against the criteria in Sec. 2.5.1. $\checkmark_q$ indicates a method is quasi-robust. $\checkmark_S$ and $\checkmark_{qS}$ indicate a method is robust or quasi-robust, respectively, in the small noise regime (see Sec. 2.4.2). Further details of each method can be found in Table 1 and Fig. 1 of Sec. 4. Notice we have one extra method compared to Table 2 because IISM:KF and IISM:NA have qualitatively different robustness but identical quantitative metrics. The naming convention is explained in Sec. 4.1.

Criteria	Scalable	Unbounded	Precise	Efficient	Robust
Unmitigated	$\checkmark$	$\times$	$\times$	$\checkmark$	$\checkmark$
CLM	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
CSM	$\times$	$\times$	$\checkmark$	$\checkmark$	$\checkmark_S$
IAM	$\times$	$\times$	$\checkmark$	$\times$	$\checkmark$
IILM:KF	$\times$	$\times$	$\checkmark$	$\times$	$\checkmark$
IILM:NA	$\times$	$\times$	$\checkmark$	$\times$	$\checkmark_q$
IISM:KF	$\checkmark$	$\times$	$\checkmark$	$\times$	$\checkmark$
IISM:NA	$\checkmark$	$\times$	$\checkmark$	$\times$	$\checkmark_{qS}$
TIILM	$\checkmark$	$\checkmark$	$\checkmark$	$\times$	$\checkmark_q$

As we see in Table 3 the only method that meets all the criteria, and therefore can be certified SUPER (see Sec. 2.5.1), is CLM (local mitigation with custom channels, see App. C.6). So this is the optimal method to use, if it is available. CLM and TIILM are the only methods that could be expected to mitigate large noise levels (i.e. of order unity or above, see noise boundary column of Table 2). If we lack knowledge about the noise model then we should use IISM:KF but this will only be effective if the noise level is reasonably small (i.e. $e_{SN} < \frac{1}{2}$) and IISM:KF is much less efficient than CLM (see runtime scaling, small noise column of Table 2).

5.2 Certification for Rotational Errors

For rotational errors (see Sec. 3.4), we gauge the strength of the noise by the size of e_{RE} (Eq. (68)):

$$e_{RE} = N_G |\phi|.$$

The noise level (Sec. 2.3.2) of the circuit being mitigated (in the small noise regime, Sec. 2.4.2). Here ϕ is the (gate-wise) noise amplitude, i.e. the rotational error for each gate, and N_G is the number of noisy gates.

In Table 4 we provide the leading order expressions for the proxy bias (i.e. the (mitigated) noise level, see Sec. 2.3.1), runtime scaling (Sec. 2.3.4), and noise boundary (Sec. 2.3.5) in the appropriate regimes (see Sec. 2.4). We then use these quantitative metrics in order to certify our mitigation methods in Table 5.

As we see in Table 5 there are a number of methods that meet all criteria, and therefore can be certified SUPER (see Sec. 2.5.1). As for stochastic noise, CLM is featured but CHILM and LC-CLM are much more efficient and have noise boundaries with much better scaling (see runtime scaling and noise boundary column of Table 4). In fact, for a fixed noise level the sampling costs of both of these methods tends to unity in the large circuit limit ($N_G \rightarrow \infty$), so they would be effectively free in this limit. If we lack knowledge about the noise model then we could use IISM:KF but this will only be effective if the noise level is reasonably small (i.e. $e_{RE} < \frac{1}{2}$). A better choice would be LC-IISM:KF (if it is available) as this has a noise boundary that scales with the square root of the number of gates (see noise boundary column of Table 4). So, though not efficient, it satisfies all

Table 4: Quantitative Mitigation Metrics: Rotational Errors. The noise level (i.e. first order term of the proxy bias, see Sec. 2.3.1), runtime scaling (Sec. 2.3.4), and noise boundary (Sec. 2.3.5) are recorded for a number of mitigation methods. We consider both the small noise (see Sec. 2.4.2) and large circuit (see Sec. 2.4.3) regimes. For the runtime (Sec. 2.3.4) calculations we assume that all gates and custom error channels (Sec. 4.2.3) take the same time to implement. N_M , where given, is the order of the mitigation (see Sec. 2.3.5). All results are for the mitigation of uniform rotational errors (see Sec. 3.4). * Unmitigated results relate to the original noisy circuit. ** IIAM results are given for the combined large circuit and small noise regime, for more information see App. C.8. *** IISM results apply to both the knowledge-free (KF) and noise-aware (NA) variants and are given in the small noise regime. The runtime scaling is given to leading order for $N_M \gg 1$, for more information see Ref. [1]. **** CHISM and CSM results are shown for $e_{RE} \leq \frac{\pi}{2}$. We do not have analytic results for larger noise levels for these methods but the best strategy would be to concatenate (Secs 4.4.2 and 4.5.2) multiple instances together, splitting the noisy gates into sets with lower noise levels; so that, for each set, $e_{RE} \leq \frac{\pi}{2}$. For CSM to reach the boundary ($e_B = \frac{\pi}{2}$) we require $S_\Phi \geq \frac{3+\sqrt{10}}{2}$ (see Ref. [1]). ***** The LC-CSM noise boundary is only valid for large noise levels ($e_{RE} \gg 1$, see Ref. [1]). Further details of each method can be found in Table 1 and Fig. 1 of Sec. 4. The naming convention is explained in Sec. 4.1.

Metric	Noise Level ($e_{G:M}$)	Excess Runtime Scaling ($S_{GI} - 1$)	Runtime Scaling (S_{GI})	Noise Boundary (e_B)
Regime	Small Noise ($e_{SN} \ll 1$)	Small Noise ($e_{SN} \ll 1$)	Large Circuit ($N_G \gg 1$)	Large Circuit ($N_G \gg 1$)
Unmitigated*	e_{RE}	0	1	ε_+
CHILM	0	$\left(1 + \frac{1}{4N_G}\right) \frac{e_{RE}^2}{N_G}$	$1 + \frac{e_{RE}^2}{N_G}$	$\sqrt{N_G(S_\Phi - 1)}$
CHISM****	0	$\left(\frac{4}{3} - \frac{1}{12N_G^2}\right) e_{RE}^2$	$\frac{3}{2\cos^2 e_{RE}} - \frac{\tan e_{RE}}{2e_{RE}}$	$\frac{\pi}{2} - \sqrt{\frac{3}{2S_\Phi}}$
CIILM	0	5	$6e^{4\sqrt{3}e_{RE}}$	$\frac{\ln\left(\frac{S_\Phi}{6}\right)}{\frac{4\sqrt{3}}{2}}$
CLM	0	$\left(2 + \frac{1}{N_G}\right) e_{RE}$	$e^{2e_{RE}}$	$\frac{\ln S_\Phi}{2}$
CSM*****	0	$3e_{RE}$	$1 + \sin e_{RE} (3 \cos e_{RE} + \sin e_{RE})$	$\frac{\pi}{2}$
IIAM**	$\frac{e_{RE}^{N_M+1}}{(N_M+1)!}$	$\frac{N_G^{2N_M}}{(N_M!)^2} - 1$	$\frac{N_G^{2N_M}}{(N_M!)^2}$	ε_+
IILM:KF	$\frac{e_{RE}^{N_M+1}}{N_G^{N_M}} \prod_{i=0}^{N_M} \frac{(2i+1)}{(N_M+1)!}$	$\geq \frac{3 \times 4^{N_G}}{2} - 1$	$\geq \frac{3 \times 4^{N_G}}{2}$	ε_+
IILM:NA	0	$\frac{15}{7} \times \left(\frac{7}{2}\right)^{2N_G} - 1$	$\frac{15}{7} \times \left(\frac{7}{2}\right)^{2N_G}$	ε_+
IISM***	$\frac{e_{RE}^{N_M+1} N_G! \prod_{i=0}^{N_M} (2i+1)}{N_G^{N_M+1} (N_G - N_M - 1)! (N_M+1)!}$	$\approx \frac{4^{N_M+1}}{\pi} - 1$	$\approx \frac{4^{N_M+1}}{\pi}$	$\approx \frac{\ln(\varepsilon_+)}{4 \frac{\ln(\pi S_\Phi)}{2}}$
LC-	$\frac{e_{RE}}{2N_G}$	0	1	$\sqrt{2N_G \varepsilon_+}$
LC-CLM	0	$\left(1 + \frac{1}{4N_G}\right) \frac{e_{RE}^2}{N_G}$	$1 + \frac{e_{RE}^2}{N_G}$	$\sqrt{N_G(S_\Phi - 1)}$
LC-CSM*****	0	$\left(\frac{4}{3} - \frac{1}{12N_G^2}\right) e_{RE}^2$	$\left(2 - \frac{\tanh e_{RE}}{e_{RE}}\right) \cosh^2 e_{RE}$	$\frac{\ln(2S_\Phi)}{2}$
LC-IIAM	$\frac{e_{RE}^{2(N_M+1)}}{(2N_G)^{N_M+1} (N_M+1)!}$	$\frac{N_G^{2N_M}}{(N_M!)^2} - 1$	$\frac{N_G^{2N_M}}{(N_M!)^2}$	ε_+
LC-IISM***	$\frac{e_{RE}^{2(N_M+1)} N_G! \prod_{i=0}^{N_M} (2i+1)}{2^{N_M+1} N_G^{2(N_M+1)} (N_G - N_M - 1)! (N_M+1)!}$	$\approx \frac{4^{N_M+1}}{\pi} - 1$	$\approx \frac{4^{N_M+1}}{\pi}$	$\approx \sqrt{N_G 4 \frac{\ln(\varepsilon_+)}{\ln(\pi S_\Phi)}}$
LC-TIILM	0	1	$2 \left(1 + \sqrt{\frac{2}{N_G}} e_{RE}\right)$	$\sqrt{\frac{N_G}{2}} \left(\frac{S_\Phi}{2} - 1\right)$
TIILM	0	1	$e^{2e_{RE}} (1 + \pi)$	$\frac{\ln \frac{S_\Phi}{1+\pi}}{2}$

Table 5: Qualitative Certification Criteria: Rotational Errors. We assess each mitigation method against the criteria in Sec. 2.5.1. $\checkmark_q$ indicates a method is quasi-robust. $\checkmark_S$ and $\checkmark_{qS}$ indicate a method is robust or quasi-robust, respectively, in the small noise regime (see Sec. 2.4.2). * Unmitigated results relate to the original noisy circuit. **** CSM and CHISM are unbounded provided we are willing to concatenate (Secs 4.4.2 and 4.5.2) multiple instances together to deal with large noise levels. ***** results for CIILM are shown using the same custom channel for the noisy gate and its inverse, if we use separate channels for each we obtain lower sampling costs and an efficient mitigation, see Ref. [1]. Further details of each method can be found in Table 1 and Fig. 1 of Sec. 4. Notice we have one extra method compared to Table 4 because IISM:KF and IISM:NA have qualitatively different robustness but identical quantitative metrics. The naming convention is explained in Sec. 4.1.

Criteria	Scalable	Unbounded	Precise	Efficient	Robust
Unmitigated*	$\checkmark$	$\times$	$\times$	$\checkmark$	$\checkmark$
CHILM	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark_q$
CHISM****	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark_q$
CIILM*****	$\checkmark$	$\checkmark$	$\checkmark$	$\times$	$\checkmark_q$
CLM	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
CSM****	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark_S$
IIAM	$\times$	$\times$	$\checkmark$	$\times$	$\checkmark$
IILM:KF	$\times$	$\times$	$\checkmark$	$\times$	$\checkmark$
IILM:NA	$\times$	$\times$	$\checkmark$	$\times$	$\checkmark_q$
IISM:KF	$\checkmark$	$\times$	$\checkmark$	$\times$	$\checkmark$
IISM:NA	$\checkmark$	$\times$	$\checkmark$	$\times$	$\checkmark_{qS}$
LC-	$\checkmark$	$\checkmark$	$\times$	$\checkmark$	$\checkmark_q$
LC-CLM	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark_q$
LC-CSM	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark_{qS}$
LC-IIAM	$\times$	$\times$	$\checkmark$	$\times$	$\checkmark_q$
LC-IISM	$\checkmark$	$\checkmark$	$\checkmark$	$\times$	$\checkmark_{qS}$
LC-TIILM	$\checkmark$	$\checkmark$	$\checkmark$	$\times$	$\checkmark_q$
TIILM	$\checkmark$	$\checkmark$	$\checkmark$	$\times$	$\checkmark_q$

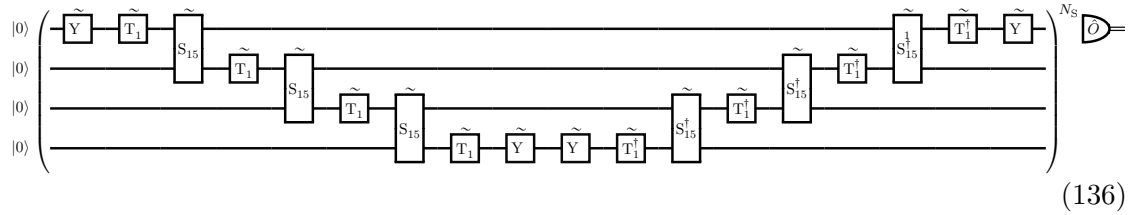
the other criteria. It is worth noting that LC used on its own (see LC- row of Table 4) gives unbiased results in the large circuit limit ($N_G \rightarrow \infty$, for fixed e_{RE}), with no excess sampling cost.

5.3 Numerical Analysis

For our numerical analysis we simulate the mitigation of the circuit in Sec. 5.3.1, suffering from (stochastic) over-rotation errors (see Sec. 3.5), for a range of mitigation methods. Details about the simulation methods used can be found in App. D. Our numerical results support our analytical analysis.

5.3.1 Benchmarking Circuit

The circuit we use for benchmarking is:



where $|0\rangle$ is the $+1$ eigenstate of the Pauli Z operator. N_S is the number of sets of gates we use in our circuit, X, Y, and Z are the standard Pauli operators and:

$$T_1 = e^{i\frac{\pi}{8}X}, \quad S_{15} = e^{i\frac{\pi}{4}Z \otimes Z}, \quad (137)$$

where the S_{15} acts on two qubits, in this case always nearest neighbours. We define our benchmarking bias as:

$$\varepsilon = \sqrt{\frac{\varepsilon_X^2 + \varepsilon_Y^2 + \varepsilon_Z^2}{3}}, \quad (138)$$

where (see Eqs (1) and (4)):

$$\varepsilon_{\hat{O}} = \text{Tr} \left[\hat{O} (\mathcal{C}(\rho_0) - \tilde{\mathcal{C}}_M(\rho_0)) \right], \quad (139)$$

where $\rho_0 = \bigotimes_{n=1}^4 |0\rangle$ is the initial state, $\mathcal{C}$ is the channel corresponding to the noise-free circuit, $\tilde{\mathcal{C}}_M$ is the (effective) channel corresponding to the mitigated circuit (see Sec. 2.1), and $\hat{O}$ corresponds to the Pauli X, Y, or Z operator measured on the first qubit. We use Eq. (138) rather than the fidelity (as suggested in Ref. [7]) since our noise-free state ($\mathcal{C}(\rho_0) = \rho_0$) is a product state of $+1$ Z eigenstates, so the fidelity would not adequately account for errors in other degrees of freedom. The runtime scaling is determined in the negligible initialisation and readout time limit (see Eq. (33)):

$$S_{GI} = C_{GI}^2 F_{GI,L}. \quad (140)$$

For our numerical analysis we consider two different length circuits, with $N_S = 1$ (18 gates) and $N_S = 10$ (180 gates), to demonstrate the scalability (or otherwise) of our methods. Scalable methods are those where the runtime scaling is independent of the number of gates (N_G) for a fixed noise level (e_{SN} or e_{RE}).

We choose these circuit and observable definitions because it ensures that all the error channels contribute to the final bias and thus all must be mitigated to mitigate the circuit completely.

5.3.2 Noise Model

For our numerical analysis we use a uniform (stochastic) over-rotation error (see Sec. 3.5) noise model. So every gate is followed by a stochastic noise channel (Sec. 3.3), with noise amplitude p , and a rotational error channel (Sec. 3.4), with noise amplitude ϕ . These error channels are both generated by the same operator as the gate itself; i.e. an error channel generated by X follows each T_1 (acting on the same qubit), an error channel generated by Y follows each Y (acting on the same qubit), and an error channel generated by $P_{15} = Z \otimes Z$ follows S_{15} (acting on the same qubits).

To further simplify the analysis we consider the two extremes: pure stochastic noise, i.e. $\phi = 0$ (see Sec. 5.3.3), and pure rotational errors, i.e. $p = 0$ (see Sec. 5.3.4). All rotational errors act in the same direction²¹ (i.e. have the same sign).

5.3.3 Marking the Bench: Stochastic Noise

In Figs 2 and 3 we plot the runtime scaling (Eq. (140)) against the bias (Eq. (138)) for various mitigation methods, demonstrating the superiority of CLM for almost all regimes.

The vertical black lines in Figs 2 and 3 show the bias before mitigation, horizontal lines are the runtime scalings of unbiased mitigation methods, and each point represents a different order of mitigation (N_M) for the biased methods (some biased methods become unbiased for sufficiently large N_M). Increasing N_M results in lower bias but higher sampling cost. Some methods have both unbiased and biased variants. Methods not featured have runtime scalings outside the range of the plot. For example, IILM has a very large runtime scaling (see Table 2). For scalable methods (see Sec. 2.5.1) the runtime scaling for a mitigation with a given bias should not increase with the number of gates, for a constant noise level, so the top and bottom plots of Figs 2 and 3 (with the same noise level: e_{SN}) should look the same for scalable methods.

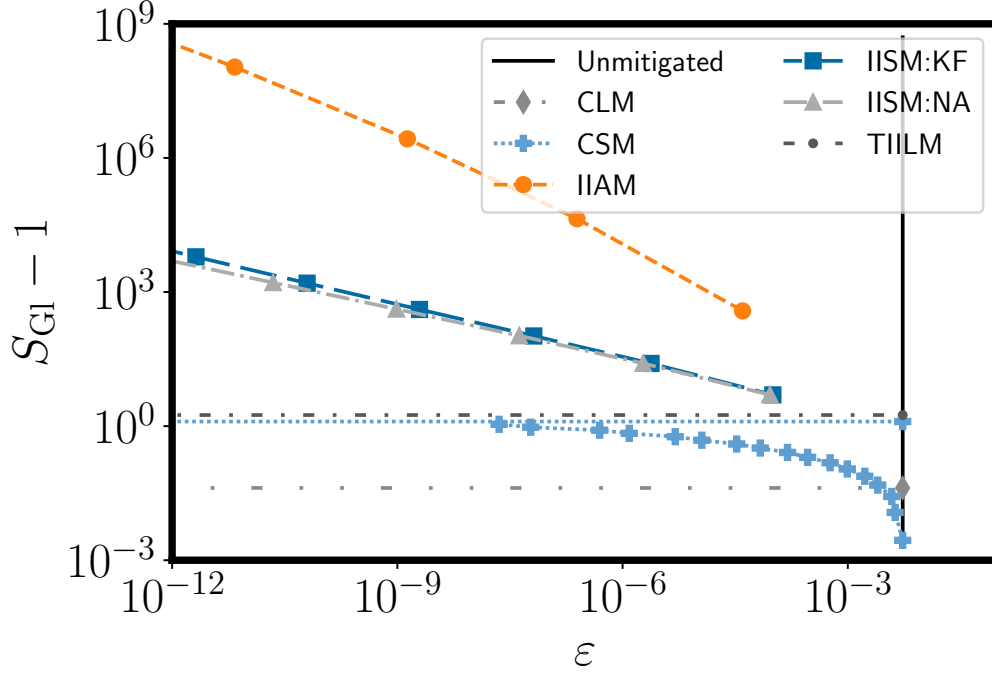
This numerical analysis also enables us to study mitigation methods that we have not yet derived analytic formula for. We consider a biased version of CSM, where we use $N_M < N_G$ (see Secs 4.2.3 and 4.4.1 and Eqs (111) and (112)), to compliment the unbiased version that features in Tables 2 and 3. At present there is not an analytic expression for the proxy bias of this method but the numerical results look promising. Though the unbiased method is not scalable for stochastic noise (see Tables 2 and 3), the biased version may well be, since the runtime scaling required for a given residual bias (bias after mitigation) seems to depend only on the noise level ($e_G = e_{SN}$) and not the number of gates ($N_G = 18N_S$), see Figs 2 and 3.

5.3.4 Marking the Bench: Rotational Errors

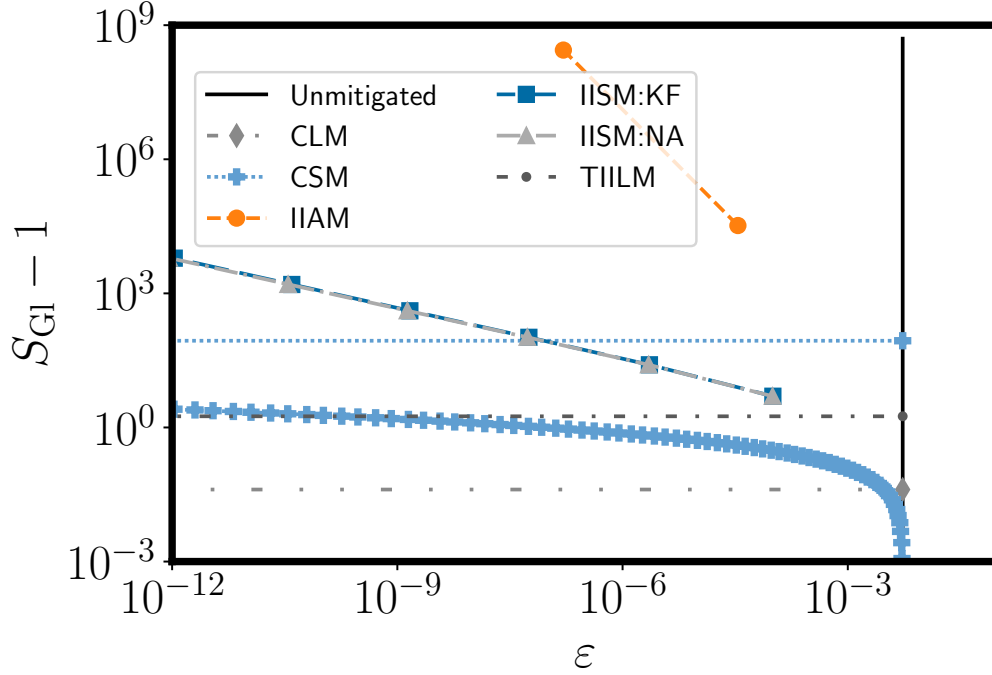
In Figs 4 and 5 we plot the runtime scaling (Eq. (140)) against the bias (Eq. (138)) for various mitigation methods, demonstrating the superiority of CHILM for almost all regimes.

What is more, for CHILM, the larger the circuit we wish to mitigate the smaller the runtime scaling (for fixed noise level: e_{RE}). This means CHILM is not just scalable it improves as we scale up the circuit size (for constant noise level). So we can classify CHILM as super-scalable. The vertical black lines in Figs 4 and 5 show the bias before

²¹If the inverse gates had rotational error channels that were the inverse of those of the forward acting gates then these error channels would cancel and we would need only to mitigate the $\tilde{Y}$ gates (which are self-inverse) to mitigate the entire circuit. This is a known disadvantage of using a circuit that is equivalent to the identity in the absence of noise.

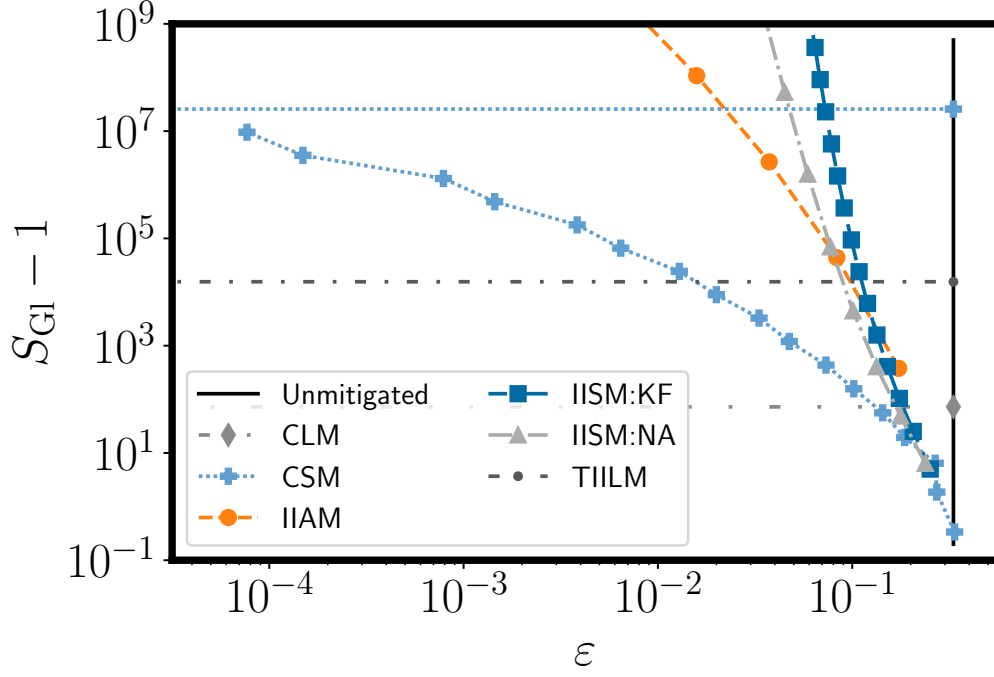


(a) $e_{\text{SN}} = 0.02$, $e_{\text{RE}} = 0$, $N_S = 1$ (for TIILM $m_0 = 0$, $m_1 = 246$).

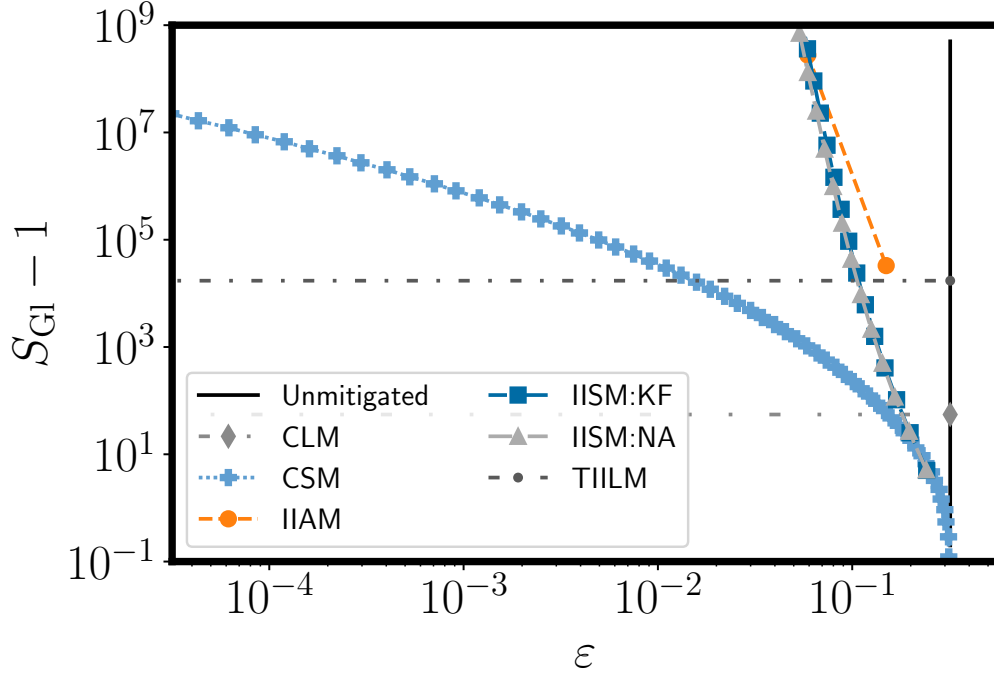


(b) $e_{\text{SN}} = 0.02$, $e_{\text{RE}} = 0$, $N_S = 10$ (for TIILM $m_0 = 0$, $m_1 = 2478$).

Figure 2: Runtime scaling (Eq. (140)) versus bias (Eq. (138)) for a circuit (Sec. 5.3.1) with stochastic noise (see Secs 3.3 and 5.3.2) with noise amplitude $p = \frac{e_{\text{SN}}}{2N_G}$, where $N_G = 18N_S$ is the number of gates. The vertical black lines show the bias before mitigation, horizontal lines are the sampling costs of unbiased methods, and each point represents a different order of mitigation (N_M). The methods from Table 2 that are not featured have runtime scalings higher than 10^9 , so they do not fit in the plot. m_0 and m_1 are the parameters used for TIILM (see Sec. 4.2.4).

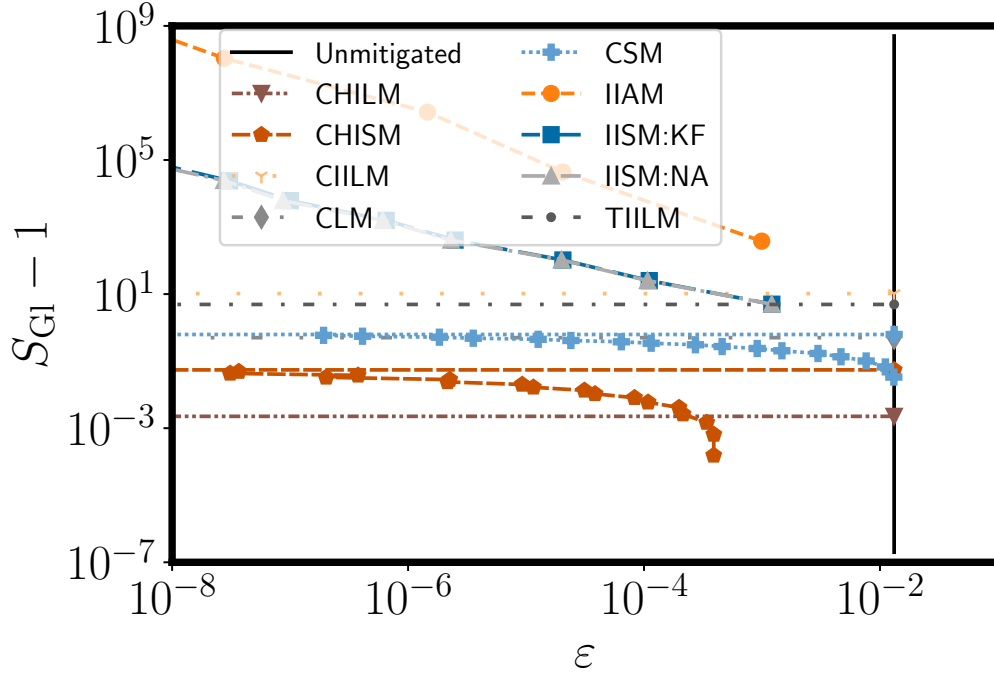


(a) $e_{SN} = 2$, $e_{RE} = 0$, $N_S = 1$ (for TIILM $m_0 = 0$, $m_1 = 16$).

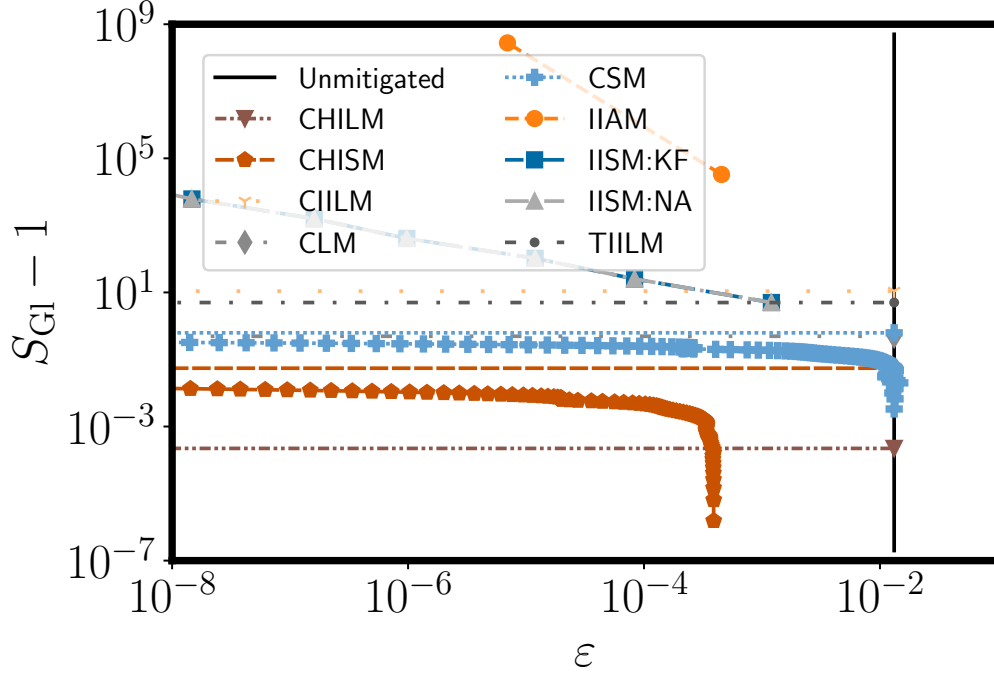


(b) $e_{SN} = 2$, $e_{RE} = 0$, $N_S = 10$ (for TIILM $m_0 = 0$, $m_1 = 169$).

Figure 3: Runtime scaling (Eq. (140)) versus bias (Eq. (138)) for a circuit (Sec. 5.3.1) with stochastic noise (see Secs 3.3 and 5.3.2) with noise amplitude $p = \frac{e_{SN}}{2N_G}$, where $N_G = 18N_S$ is the number of gates. The vertical black lines show the bias before mitigation, horizontal lines are the sampling costs of unbiased methods, and each point represents a different order of mitigation (N_M). The methods from Table 2 that are not featured have runtime scalings higher than 10^9 , so they do not fit in the plot. m_0 and m_1 are the parameters used for TIILM (see Sec. 4.2.4).

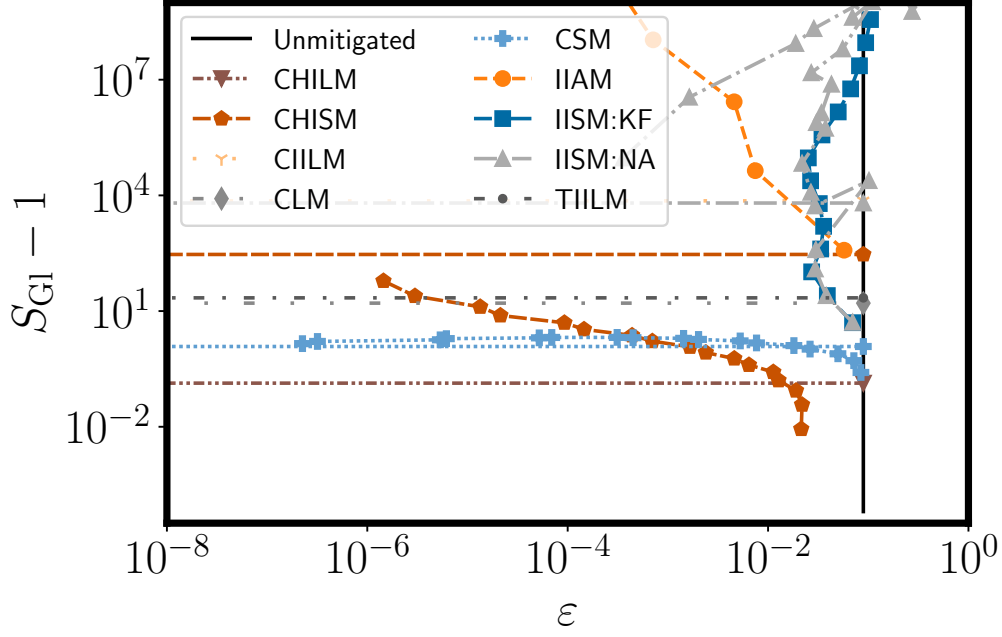


(a) $e_{SN} = 0$, $e_{RE} = 0.2$, $N_S = 1$ (for TIILM $m_0 = 0$, $m_1 = 58$, $m_2 = 200$).

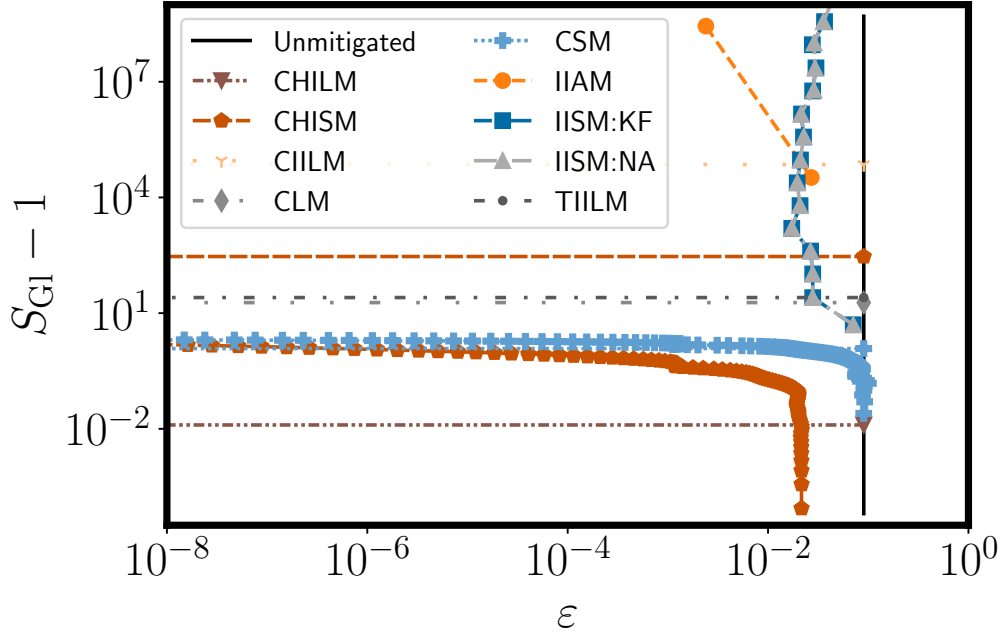


(b) $e_{SN} = 0$, $e_{RE} = 0.2$, $N_S = 10$ (for TIILM $m_0 = 0$, $m_1 = 589$, $m_2 = 2002$).

Figure 4: Runtime scaling (Eq. (140)) versus bias (Eq. (138)) for a circuit (Sec. 5.3.1) with rotational errors (see Secs 3.4 and 5.3.2) with noise amplitude $\phi = \frac{e_{RE}}{N_G}$, where $N_G = 18N_S$ is the number of gates. The vertical black lines show the bias before mitigation, horizontal lines are the sampling costs of unbiased methods, and each point represents a different order of mitigation (N_M). The methods from Table 4 that are not featured have runtime scalings higher than 10^9 , so they do not fit in the plot; or use LC pre-tailoring and are featured in Fig. 6. m_0 , m_1 , and m_2 are the parameters used for TIILM (see Sec. 4.2.4).

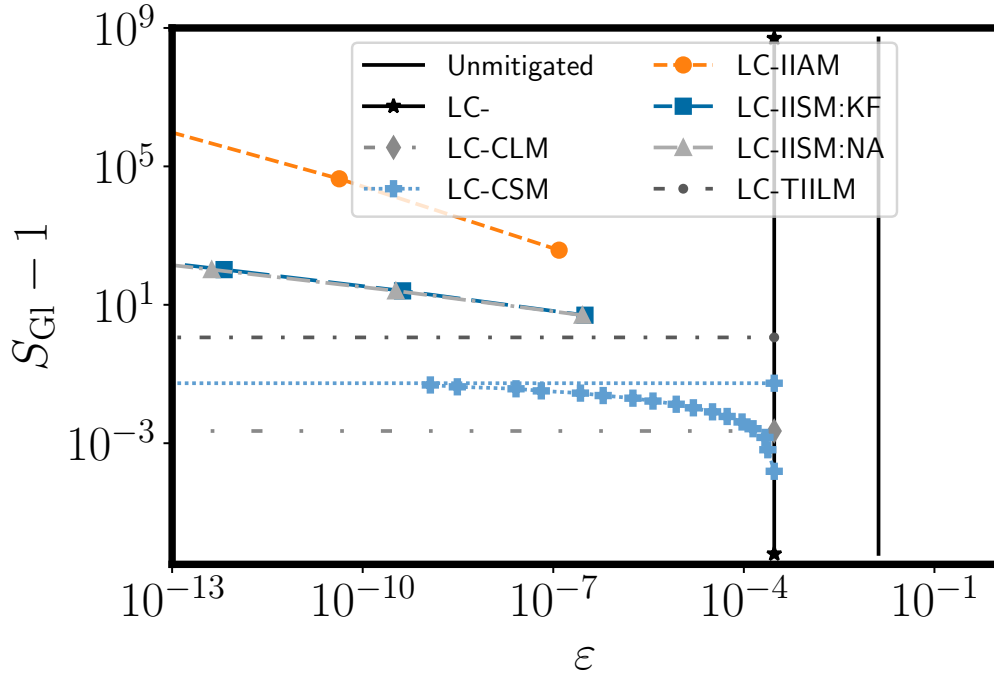


(a) $e_{\text{SN}} = 0$, $e_{\text{RE}} = 1.5$, $N_S = 1$ (for TIILM $m_0 = 0$, $m_1 = 15$, $m_2 = 34$).

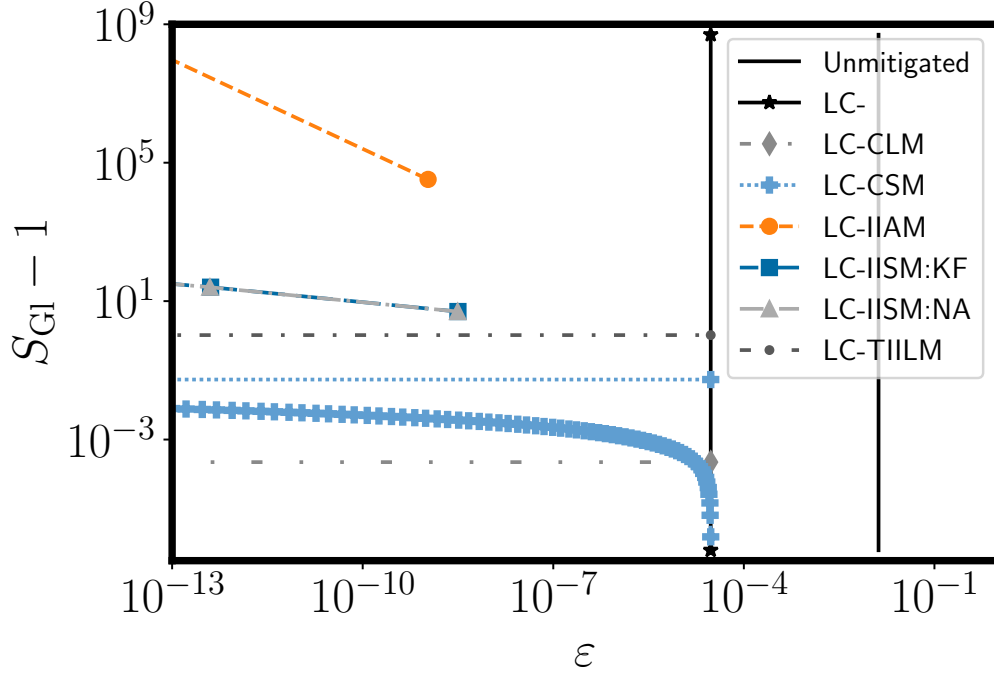


(b) $e_{\text{SN}} = 0$, $e_{\text{RE}} = 1.5$, $N_S = 10$ (for TIILM $m_0 = 0$, $m_1 = 152$, $m_2 = 341$).

Figure 5: Runtime scaling (Eq. (140)) versus bias (Eq. (138)) for a circuit (Sec. 5.3.1) with rotational errors (see Secs 3.4 and 5.3.2) with noise amplitude $\phi = \frac{e_{\text{RE}}}{N_G}$, where $N_G = 18N_S$ is the number of gates. The vertical black lines show the bias before mitigation, horizontal lines are the sampling costs of unbiased methods, and each point represents a different order of mitigation (N_M). The methods from Table 4 that are not featured have runtime scalings higher than 10^9 , so they do not fit in the plot; or use LC pre-tailoring and are featured in Fig. 7. m_0 , m_1 , and m_2 are the parameters used for TIILM (see Sec. 4.2.4).

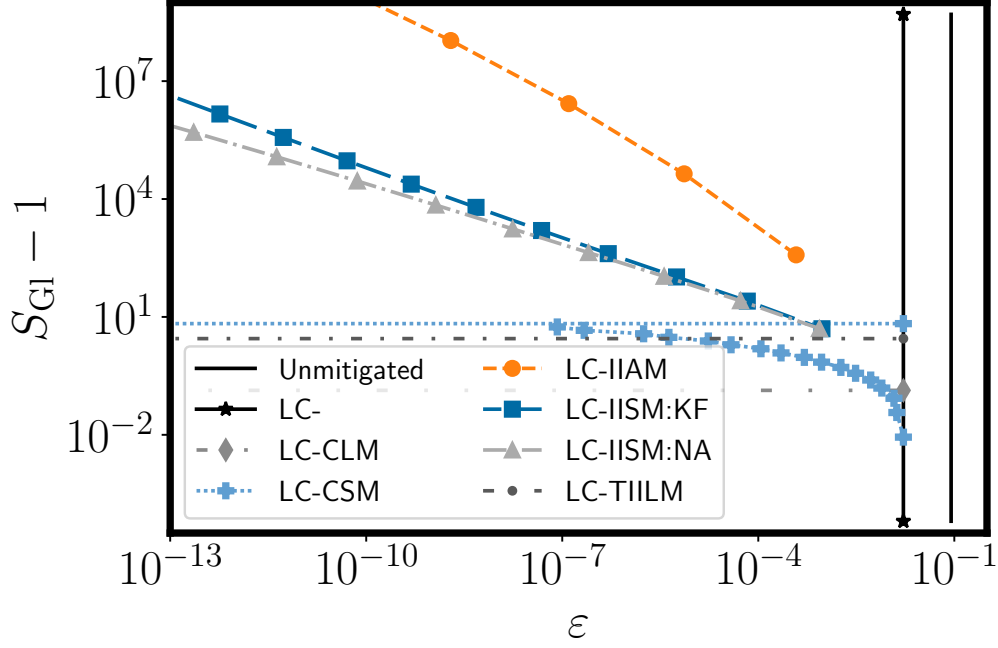


(a) $e_{SN} = 0$, $e_{RE} = 0.2$, $N_S = 1$ (for TIILM $m_0 = 0$, $m_1 = 1066$).

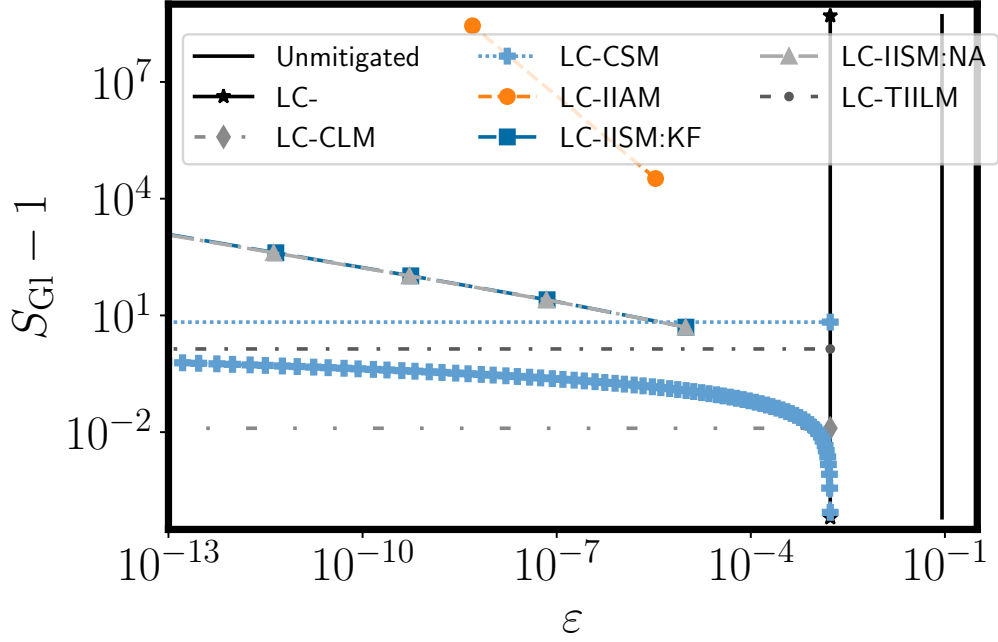


(b) $e_{SN} = 0$, $e_{RE} = 0.2$, $N_S = 10$ (for TIILM $m_0 = 0$, $m_1 = 34068$).

Figure 6: Runtime scaling (Eq. (140)) versus bias (Eq. (138)) after local cancellation (LC) for a circuit (Sec. 5.3.1) with rotational errors (see Secs 3.4 and 5.3.2) with noise amplitude $\phi = \frac{e_{RE}}{N_G}$, where $N_G = 18N_S$ is the number of gates. The vertical black lines show the bias before mitigation (the one with stars is after we apply LC but before subsequent mitigation methods), horizontal lines are the sampling costs of unbiased methods, and each point represents a different order of mitigation (N_M). The methods from Table 4 that are not featured have runtime scalings higher than 10^9 , so they do not fit in the plot; or do not use LC pre-tailoring and are featured in Fig. 4. m_0 and m_1 are the parameters used for TIILM (see Sec. 4.2.4).



(a) $e_{\text{SN}} = 0$, $e_{\text{RE}} = 1.5$, $N_S = 1$, (for TIILM $m_0 = 0$, $m_1 = 135$).



(b) $e_{\text{SN}} = 0$, $e_{\text{RE}} = 1.5$, $N_S = 10$ (for TIILM $m_0 = 0$, $m_1 = 4487$).

Figure 7: Runtime scaling (Eq. (140)) versus bias (Eq. (138)) after local cancellation (LC) for a circuit (Sec. 5.3.1) with rotational errors (see Secs 3.4 and 5.3.2) with noise amplitude $\phi = \frac{e_{\text{RE}}}{N_G}$, where $N_G = 18N_S$ is the number of gates. The vertical black lines show the bias before mitigation (the one with stars is after we apply LC), horizontal lines are the sampling costs of unbiased methods, and each point represents a different order of mitigation (N_M). The methods from Table 4 that are not featured have runtime scalings higher than 10^9 , so they do not fit in the plot; or do not use LC pre-tailoring and are featured in Fig. 5. m_0 and m_1 are the parameters used for TIILM (see Sec. 4.2.4).

mitigation, horizontal lines are the runtime scalings of unbiased mitigation methods, and each point represents a different order of mitigation (N_M) for the biased methods (some biased methods become unbiased for sufficiently large N_M). Increasing N_M results in lower bias but higher sampling cost.

Methods not featured have runtime scalings outside the range of the plot. For example, IILM has a very large runtime scaling (see Table 4). It is worth noting that, for large enough noise levels, increasing the order of mitigation can result in a lower quality mitigation, as is the case for IISM (see Fig. 5). For scalable methods (see Sec. 2.5.1) the runtime scaling for a mitigation with a given bias should not increase with the number of gates, for a constant noise level, so the top and bottom plots of Figs 4 and 5 (with the same noise level: e_{RE}) should look the same; or, if we have a super-scalable method like CHILM, the sampling cost will be lower in the bottom plots.

In Figs 6 and 7 we provide a similar analysis but make use of local cancellation (LC) to tailor the noise model before applying the mitigation methods (for stochastic noise). We find LC-CLM performs identically to CHILM, which is not surprising as they make use of the same noisy gate variants, and all other methods perform much better with local cancellation. The black line with a star on either end is the bias after applying local cancellation, without additional mitigation.

In the next section we will use these results to develop a mitigation strategy for a piece of hardware suffering from (stochastic) over-rotation errors (Sec. 3.5) and certify that it is USEful (see Sec. 6.3).

6 Case Study: Strategy Certification

Here we present (Sec. 6.1) and certify (Secs 6.2 and 6.3) USEM:ORE (Universal, Scalable, and Efficacious Mitigation for Over-Rotation Errors), a mitigation strategy for a piece of hardware suffering from (stochastic) over-rotation errors. We certify it as USEful (universal, scalable, and efficacious) using the certification criteria of Sec. 2.5.2.

6.1 The USEM:ORE strategy

In order to construct a mitigation strategy (Sec. 2.5.2) we need to carry out the following steps:

1. **Choose Application:** specify our target application and determine appropriate metrics (e.g. Sec. 2.3) and certifications (e.g. Sec. 2.5) for our strategy (Sec. 6.1.1).
2. **Characterise Hardware:** determine the elementary physical operations our hardware can implement and their associated error channels (noise) (Sec. 6.1.2).
3. **Compare Mitigation Methods:** select the mitigation methods for our mitigation strategy; using information from the hardware characterisation, metrics, and certification (Sec. 6.1.3).
4. **Collate Gate Set:** designate a set of gates (or physical operations) that spans the space of gates needed for our calculations and mitigations (Sec. 6.1.4).
5. **Compilation Scheme:** design a compilation scheme that permits the production of mitigated versions of the algorithms we wish to implement (Sec. 6.1.5).

6.1.1 Application Specification

Our application is implementing mitigated versions of any quantum algorithm, taking maximal advantage of improvements in hardware. Therefore we want USEM:ORE to be universal, scalable, and efficacious (see Sec. 2.5.2) for (stochastic) over-rotation errors. This means we will be interested in finding SUPER (see Sec. 2.5.1) mitigation methods. This requires the analysis of the proxy bias and runtime scaling of mitigation methods (see Sec. 2.3) in the large circuit and small noise regimes (see Sec. 2.4).

6.1.2 Hardware Characterisation

For illustrative purposes we consider a piece of quantum hardware for which the native physical operations are rotations generated by any Pauli word operator (P_i) of any angle (ω), e.g. we can produce:

$$e^{i\frac{\omega}{2}P_i}, \quad \text{for any} \quad P_i = \bigotimes_{n=1}^{N_Q} \vec{\sigma}_{[n], \lfloor \frac{i}{4^{n-1}} \rfloor \bmod 4}, \quad (141)$$

where ω is the target rotation angle and N_Q is the number of qubits the physical operation acts on. $\vec{\sigma}_{[n],j}$ is the j th single qubit Pauli operator for qubit n ($\vec{\sigma}_{[n],0} = I$, $\vec{\sigma}_{[n],1} = X$, $\vec{\sigma}_{[n],2} = Y$, and $\vec{\sigma}_{[n],3} = Z$), $\lfloor \frac{i}{4^{n-1}} \rfloor$ is the integer part of the division of i by 4^{n-1} , and $k \bmod 4$ is the remainder after dividing k by 4. This means we have²² $P_0 = I$, $P_1 = X$, $P_2 = Y$, $P_3 = Z$, $P_{14} = Z \otimes Y$, and $P_{15} = Z \otimes Z$ (for $N_Q \geq 2$, where we suppress leading identities in the tensor product, see Eq. (141)). To simplify our notation we define:

$$P_i = e^{i\frac{\pi}{2}P_i}, \quad S_i = e^{i\frac{\pi}{4}P_i} = \sqrt{P_i}, \quad T_i = e^{i\frac{\pi}{8}P_i} = (P_i)^{\frac{1}{4}}. \quad (142)$$

For this piece of hardware we assume each physical operation is noisy and suffers from (stochastic) over-rotation errors. This means each time we try to implement one of the native physical operations we rotate slightly too far or not far enough and the amount of over-rotation for each implementation is random (see Sec. 3.5). We assume that the over-rotations follow a Gaussian distribution (with mean μ_{ORE} and standard deviation σ_{ORE}) and therefore a noisy gate can be modelled by:

$$\boxed{\widetilde{e^{i\frac{\omega}{2}P_i}}} = \boxed{e^{i\frac{\omega}{2}P_i}} \boxed{e^{i\frac{\phi}{2}P_i}} \begin{cases} \text{---} & \text{with probability } 1 - p \\ \boxed{P_i} & \text{with probability } p, \end{cases} \quad (143)$$

where ϕ is the noise amplitude of the rotational (component of the) error channel and p is the noise amplitude of the stochastic (component of the) error channel:

$$\phi = \mu_{\text{ORE}}, \quad p = \frac{1 - e^{-\frac{\sigma_{\text{ORE}}^2}{2}}}{2} = \frac{\sigma_{\text{ORE}}^2}{4} + \mathcal{O}_0(\sigma_{\text{ORE}}^4). \quad (144)$$

The noise channels have eigenvalues of the form (Eq. (77) of Sec. 3.5):

$$\lambda_{\mathcal{E},0}(p, \phi) = 1, \quad \lambda_{\mathcal{E},1}(p, \phi) = (1 - 2p)e^{i\phi}, \quad \lambda_{\mathcal{E},2}(p, \phi) = (1 - 2p)e^{-i\phi}.$$

We will assume there can be different noise amplitudes (ϕ and p) for each physical operation that we mitigate so we benefit from using a local mitigation method.

²²Operators ($\vec{\sigma}_{[n]}$) acting on higher index (larger n) qubits are denoted to the left of those acting on lower index qubits, in our notation for Pauli words, so that the i index is independent of the number of qubits (N_Q) for sufficiently large N_Q (i.e. $N_Q > \log_4 i$, otherwise P_i is not well defined). We neglect to denote leading order identities for notational convenience.

6.1.3 Mitigation Methods

There are a variety of local mitigation methods that we could choose from (see Table 1 of Sec. 4.6). Since our noise has both stochastic and rotational components (see Eq. (143)) it seems reasonable to choose a method that performs well for both types of noise (see Tables 3 and 5 of Secs 5.1 and 5.2, respectively). The obvious choice is CLM but CHILM is much more efficient for rotational errors. So we propose C(HI)LM, that is, if we have access to hidden inverse (for a particular physical operation) we will use CHILM but if not we will settle for CLM. We will use the version designed to mitigate rotation errors (but modify the mitigation coefficients to account for the stochastic contribution). Since the noise channels for stochastic over-rotation errors and rotational errors are the same (c.f. Secs 3.4 and 3.5).

This means we find a linear combination of noisy gate variants that is equivalent to the noise-free target gate (Eq. (124) of Sec. 4.5):

$$\boxed{\widetilde{U}}_{\text{M}} = -\left\{c_{L;j} : \boxed{\widetilde{U}}_j\right\}_{j=0}^2 = \boxed{U}, \quad (145)$$

where we use 3 gate variants in our mitigation (i.e. $N_{L,C} = 2$), with indices 0, 1, and 2. The noisy gates to be mitigated have the form (Eq. (143)):

$$\boxed{\widetilde{U}} = \boxed{U} \boxed{\mathcal{E}}, \quad \text{where} \quad \boxed{U} = \boxed{e^{i\frac{\phi}{2}P_i}}, \quad (146)$$

is the noise-free component of the gate and $\mathcal{E}$ is the error channel:

$$\boxed{\mathcal{E}} = \boxed{e^{i\frac{\phi}{2}P_i}} \begin{cases} \text{---} & \text{with probability } 1-p \\ \boxed{P_i} & \text{with probability } p. \end{cases}, \quad (147)$$

The noisy variants ($\widetilde{U}_j$) that we use for the mitigation are generated using custom channels (for CLM) or custom channels and a hidden inverse (for CHILM) and are given by²³:

$$\begin{cases} \boxed{\widetilde{U}}_j = \boxed{U} \boxed{e^{i\frac{\theta_j}{2}P_i}}, & \text{for CLM (where } 0 \leq j \leq 2) \\ \boxed{\widetilde{U}}_j = \boxed{U} \boxed{e^{i\frac{\theta_j}{2}P_i}}, \quad \boxed{\widetilde{U}}_2 = \boxed{U}_{\text{HI}}, & \text{for CHILM (where } 0 \leq j < 2), \end{cases} \quad (148)$$

where the controlled rotations (θ_j) we apply after our noisy gate are (Eqs (94) and (95)):

$$\begin{cases} \theta_0 = 0, & \theta_1 = \frac{\pi}{2}, & \theta_2 = \frac{3\pi}{2}, & \text{for CLM} \\ \theta_0 = 0, & \theta_1 = \pi, & \theta_2 = 0, & \text{for CHILM} \end{cases} \quad (149)$$

and for CHILM we use a hidden inverse for the last (unamplified) gate variant, so the total rotational errors²⁴ of the gate variants are given by:

$$\begin{cases} \phi_0 = \phi, & \phi_1 = \phi + \frac{\pi}{2}, & \phi_2 = \phi + \frac{3\pi}{2}, & \text{for CLM} \\ \phi_0 = \phi, & \phi_1 = \phi + \pi, & \phi_2 = -\phi, & \text{for CHILM,} \end{cases} \quad (150)$$

²³Here, for simplicity, we assume our custom amplifying channels can be implemented perfectly without introducing additional errors, for a more robust analysis see Sec. 6.2.2.

²⁴The total rotational errors of the tailored noisy gates are given by $\phi_j = \phi + \theta_j$ for gate variants without hidden inverses and $\phi_j = -\phi + \theta_j$ when we use the hidden inverse. ϕ is the rotational error of the original noisy gate and θ_j is the rotational angle of the custom channel.

where:

$$\begin{aligned} \boxed{\widetilde{U}}_j &= \boxed{U} \boxed{\mathcal{E}}_j, & \boxed{\mathcal{E}}_j &= \boxed{e^{i\frac{\phi_j}{2} P_i}} \left\{ \begin{array}{ll} \text{---} & \text{with probability } 1 - p \\ \boxed{\widetilde{P_i}} & \text{with probability } p. \end{array} \right. \end{aligned} \quad (151)$$

We assume that the stochastic noise level remain unchanged, i.e. $p_j = p$ for both methods.

Now the eigenvalues associated with the transfer matrices ($\hat{T}_{\mathcal{E};j}$) that characterise the effect of the (amplified) error channels ($\mathcal{E};j$) are given by (see Eq. (77)):

$$\lambda_{j,0} = \lambda_0(p_j, \phi_j) = 1, \quad \lambda_{j,1} = \lambda_1(p_j, \phi_j) = (1 - 2p_j) e^{i\phi_j}, \quad (152)$$

$$\lambda_{j,2} = \lambda_2(p_j, \phi_j) = (1 - 2p_j) e^{-i\phi_j}. \quad (153)$$

In order to perform an unbiased mitigation we must determine mitigation coefficients that satisfy (see Eq. (126) of Sec. 4.5.1):

$$\sum_{j=0}^{N_{L,C}} c_{L;j} \lambda_{j,k} = 1,$$

for (integer) $0 \leq k \leq 2$, i.e. we require:

$$\begin{cases} c_{L;0} = \frac{\cos \phi}{1-2p}, & c_{L;1} = \frac{1-\cos \phi - 2p \sin \phi}{2(1-2p)}, & c_{L;2} = \frac{1-\cos \phi - 2p \sin \phi}{2(1-2p)}, & \text{for CLM} \\ c_{L;0} = \frac{1}{2}, & c_{L;1} = \frac{1-(\cos \phi)^{-1} - 2p}{2(1-2p)}, & c_{L;2} = \frac{1}{2(1-2p) \cos \phi}, & \text{for CHILM,} \end{cases} \quad (154)$$

as the coefficients of our mitigation methods.

6.1.4 Gate Set

We desire a universal (see Sec. 2.5.2) mitigation strategy so we require a gate set that includes: a collection of gates (or physical operations) that is sufficient to implement any algorithm we wish to apply and, also, all the gates (or physical operations) needed to mitigate that collection of gates. There are multiple different gate sets we could choose that would satisfy our criteria. We choose the following (Eq. (141)):

$$T = T_3 = e^{i\frac{\pi}{8}Z}, \quad S = S_3 = e^{i\frac{\pi}{4}Z}, \quad Z = P_3 = e^{i\frac{\pi}{2}Z}, \quad (155)$$

$$S_1 = e^{i\frac{\pi}{4}X}, \quad X = P_1 = e^{i\frac{\pi}{2}X}, \quad (156)$$

$$S_{15} = e^{i\frac{\pi}{4}Z \otimes Z}, \quad (157)$$

i.e. we choose the standard T, S, and Pauli Z gate (Eq. (155)); the Pauli X gate and its square root ($S_1 = \sqrt{X}$, Eq. (156)); and one two-qubit gate: S_{15} . We require a copy of each of the single qubit gates for each of the qubits that we wish to act on and a copy of the two qubit gate for each pair of qubits we wish to act on. We could do without X, Z, and/or S if we wished to, since $T^2 = S$, $S^2 = Z$, and $S_1^2 = X$.

This gate set is also sufficient for our mitigation since the amplification channels (see Eqs (148) and (149)) we need for C(HI)LM can be decomposed as:

$$e^{i\frac{\pi}{4}P_3} = S_3 = S, \quad e^{i\frac{\pi}{2}P_3} = P_3 = Z, \quad e^{i\frac{3\pi}{4}P_3} = SZ, \quad (158)$$

$$e^{i\frac{\pi}{4}P_1} = S_1, \quad e^{i\frac{\pi}{2}P_1} = P_1 = X, \quad e^{i\frac{3\pi}{4}P_1} = S_1X \quad (159)$$

$$e^{i\frac{\pi}{4}P_{15}} = S_{15}, \quad e^{i\frac{\pi}{2}P_{15}} = P_{15} = Z \otimes Z, \quad e^{i\frac{3\pi}{4}P_{15}} = S_{15}(Z \otimes Z), \quad (160)$$

where each row of operators can be used to mitigate (all the gates in) the corresponding row of our gate set (Eqs (155), (156), and (157), see Sec. 6.1.3).

If we wish to mitigate any gate using CHILM we do not require the S_i gates for the mitigation but we do require the hidden inverse implementation of the gate we wish to mitigate, which is the same as the original gate except that the rotational error of the noise channel has its sign reversed (see Secs 4.2.2 and 6.1.2), e.g. if we have the noisy gate (Eq. (146)):

$$\boxed{\widetilde{U}} = \boxed{\widetilde{e^{i\frac{\omega}{2}P_i}}} = \boxed{e^{i\frac{\omega}{2}P_i}} \boxed{e^{i\frac{\phi}{2}P_i}} \left\{ \begin{array}{l} \text{---} \quad \text{with probability } 1 - p \\ \boxed{P_i} \quad \text{with probability } p, \end{array} \right. \quad (161)$$

then the hidden inverse of this noisy gate is given by:

$$\boxed{\widetilde{U}}_{\text{HI}} = \boxed{\widetilde{e^{i\frac{\omega}{2}P_i}}}_{\text{HI}} = \boxed{e^{i\frac{\omega}{2}P_i}} \boxed{e^{-i\frac{\phi}{2}P_i}} \left\{ \begin{array}{l} \text{---} \quad \text{with probability } 1 - p \\ \boxed{P_i} \quad \text{with probability } p. \end{array} \right. \quad (162)$$

So we need to add the hidden inverse implementation, of any gate we wish to mitigate using CHILM, to our gate set (Eqs (155), (156), and (157)), which could double the number of gates in our gate set.

6.1.5 Compilation Scheme

The compilation scheme is simple. We produce a quantum circuit that is equivalent to the algorithm we wish to implement (how we go about doing this is beyond the scope of this work). Then we decompose every gate in this circuit in terms of the gates within our gate set (Eqs (155), (156), and (157)). If we wish to be more efficient we can combine the previous two steps and optimise the compilation scheme in order to minimise the circuit length and/or runtime scaling. Once we have a circuit that implements our algorithm, written in terms of gates from our gate set, we simply replace every gate with its mitigated version (Eq. (145)). This procedure means we end up with a mitigated circuit that only requires us to make use of gates that we can implement on our hardware (see Secs 6.1.3 and 6.1.4), and this mitigated circuit is the circuit we will run on our quantum hardware to implement our algorithm.

6.1.6 Additional Requirements

Since we desire a scalable mitigation strategy we require the noise level to be independent of circuit length (see Sec. 2.5.1). This means some level of hardware tuning, i.e. improving gate implementation to reduce errors; or error correction, i.e. actively correcting the errors that occur during the application of logical gates, might be required in order to permit the mitigation of larger circuits.

6.2 Method Certification

We can certify our mitigation methods (CLM and CHILM, see Sec. 6.1.3) with regards to the criteria introduced in Sec. 2.5.1, but first we need to calculate the quantitative metrics (Sec. 2.3).

6.2.1 Quantitative Mitigation Metrics

The proxy bias (Sec. 2.3.1) is zero for both our methods ($\varepsilon_{\odot, \text{M}} = 0$) since they are (intrinsically) unbiased and if our characterisation is perfect the mitigated gates will be error free.

To calculate the runtime scaling (Sec. 2.3.4) we must first calculate the (local) sampling cost²⁵ (see Eqs (32), (131), and (154)):

$$C_L = \sum_{j=0}^2 |c_{L,j}| = \begin{cases} \begin{cases} \frac{\cos \phi + |\sin \phi|}{1-2p} & \text{for } |\phi| > 2p, \\ \frac{2 \cos \phi - 1 + 2p}{1-2p} & \text{for } |\phi| < 2p, \end{cases} & \text{for CLM} \\ \frac{1}{(1-2p) \cos \phi}, & \text{for CHILM,} \end{cases} \quad (163)$$

and the (local) length scale factor (Eqs (32), (132), and (154)):

$$F_{L,L} = \frac{\sum_{j=0}^2 |c_{L,j}| \tau_{E,L,j}}{C_L \tau_{E,L}} = \begin{cases} \begin{cases} \frac{3 \cos \phi - 1 + 2p + 5 |\sin \phi|}{2(\cos \phi + |\sin \phi|)} & \text{for } \phi < -2p, \\ \frac{7 \cos \phi - 5 + 10p - \sin \phi}{2(2 \cos \phi - 2 + 4p)} & \text{for } |\phi| < 2p, \\ \frac{\cos \phi + 1 - 2p + 5 |\sin \phi|}{2(\cos \phi + |\sin \phi|)} & \text{for } \phi > 2p, \end{cases} & \text{for CLM} \\ \frac{3 - (1-2p) \cos \phi}{2}, & \text{for CHILM,} \end{cases} \quad (164)$$

where we assume every gate takes the same amount of time²⁶ to implement so we use (see Eqs (158), (159), and (160)):

$$\begin{cases} \frac{\tau_{E,L,0}}{\tau_{E,L}} = 1, & \frac{\tau_{E,L,1}}{\tau_{E,L}} = 2, & \frac{\tau_{E,L,2}}{\tau_{E,L}} = 3, & \text{for CLM} \\ \frac{\tau_{E,L,0}}{\tau_{E,L}} = 1, & \frac{\tau_{E,L,1}}{\tau_{E,L}} = 2, & \frac{\tau_{E,L,2}}{\tau_{E,L}} = 1, & \text{for CHILM.} \end{cases} \quad (165)$$

The full runtime scaling (for negligible initialisation time, see Eqs (33), (131), and (132)): $S_{\text{GI}} = C_L^{2N_G} F_{L,L}$, is not particularly informative so we will instead expand it in the small noise regime ($e_{\text{SN}}, e_{\text{RE}} \ll 1$):

$$S_{\text{GI}} = \begin{cases} \begin{cases} 1 + 2 \left(1 + \frac{1}{4N_G}\right) e_{\text{SN}} + 2 \left(1 + \frac{3}{4N_G}\right) e_{\text{RE}} & \text{for } \phi < -2p, \\ 1 + 4 \left(1 + \frac{3}{8N_G}\right) e_{\text{SN}} + \frac{e_{\text{RE}}}{2N_G} & \text{for } -2p < \phi < 0, \\ 1 + 4 \left(1 + \frac{3}{8N_G}\right) e_{\text{SN}} - \frac{e_{\text{RE}}}{2N_G} & \text{for } 0 < \phi < 2p, \\ 1 + 2 \left(1 - \frac{1}{4N_G}\right) e_{\text{SN}} + 2 \left(1 + \frac{3}{4N_G}\right) e_{\text{RE}} & \text{for } \phi > 2p, \end{cases} & \text{for CLM} \\ 1 + 2 \left(1 + \frac{1}{4N_G}\right) e_{\text{SN}} + \left(1 + \frac{1}{4N_G}\right) \frac{e_{\text{RE}}^2}{N_G}, & \text{for CHILM,} \end{cases} \quad (166)$$

(where we have neglected terms of order $\mathcal{O}_{\text{S}}(e_{\text{SN}}^2, e_{\text{SN}} e_{\text{RE}}, e_{\text{RE}}^2)$ for CLM and those of order $\mathcal{O}_{\text{S}}(e_{\text{SN}}^2, e_{\text{SN}} e_{\text{RE}}^2, e_{\text{RE}}^4)$ for CHILM). In the large circuit regime ($N_G \gg 1$) the runtime scaling

²⁵For simplicity we assume every gate suffers from the same noise amplitude (i.e. we have a uniform noise model). This simplification can be easily relaxed by replacing the local sampling cost with the geometric mean of the sampling costs for each gate, but, as long as the noise amplitudes are all of similar orders of magnitude, our derivation is relevant.

²⁶In this section we assume every gate takes the same amount of time to implement, this is a slightly different approach to that used in deriving Table 4 where we assumed that every custom channel takes the same amount of time to implement as the gate it amplifies. So for CLM we used $\frac{\tau_{E,L,2}}{\tau_{E,L}} = 2$ in Table 4 rather than $\frac{\tau_{E,L,2}}{\tau_{E,L}} = 3$ as used in Eqs (164) and (165), see Sec. 2.3.4 for more information on the derivation.

is given by:

$$S_{\text{GI}} = \begin{cases} \begin{cases} e^{2(e_{\text{SN}} + e_{\text{RE}})} \left(1 + \mathcal{O}_L \left(\frac{e_{\text{SN}}}{N_G}, \frac{e_{\text{SN}}^2}{N_G}, \frac{e_{\text{RE}}}{N_G}, \frac{e_{\text{RE}}^2}{N_G} \right) \right) & \text{for } e_{\text{RE}} > e_{\text{SN}}, \\ e^{4e_{\text{SN}}} \left(1 + \mathcal{O}_L \left(\frac{e_{\text{SN}}}{N_G}, \frac{e_{\text{RE}}}{N_G}, \frac{e_{\text{RE}}^2}{N_G} \right) \right) & \text{for } e_{\text{RE}} < e_{\text{SN}}, \end{cases} & \text{for CLM} \\ e^{2e_{\text{SN}}} \left(1 + \mathcal{O}_L \left(\frac{e_{\text{SN}}}{N_G}, \frac{e_{\text{SN}}^2}{N_G}, \frac{e_{\text{RE}}^2}{N_G} \right) \right), & \text{for CHILM.} \end{cases} \quad (167)$$

Here N_G is the number of mitigated gates and we have used (Eqs (55) and (68)):

$$e_{\text{SN}} = 2pN_G, \quad e_{\text{RE}} = N_G |\phi|. \quad (168)$$

Interestingly, Eq. (167) means, if we have access to hidden inverses, the rotational noise level is largely irrelevant in the large circuit regime (provided the rotational noise level is independent of N_G).

Determining the noise boundary (see Sec. 2.3.5) is now just a case of inverting the large-circuit-regime runtime equation (Eq. (167)) in the large circuit limit ($N_G \rightarrow \infty$). We find:

$$\begin{cases} \begin{cases} e_B = \frac{\ln S_\star}{2} & \text{for } e_{\text{RE}} > e_{\text{SN}}, \\ e_{\text{SN},B} = \frac{\ln S_\star}{4} & \text{for } e_{\text{RE}} < e_{\text{SN}}, \end{cases} & \text{for CLM} \\ e_{\text{SN},B} = \frac{\ln S_\star}{2}, & \text{for CHILM,} \end{cases} \quad (169)$$

where e_B is the total noise boundary and $e_{\text{SN},B}$ is the noise boundary for stochastic noise. So, if we have a tolerated runtime scaling of $S_\star$ we can mitigate circuits with any (finite) noise levels (e_{SN} and e_{RE}) such that:

$$\begin{cases} e_{\text{RE}} < e_{\text{SN}} < e_{\text{SN},B} \quad \text{or} \quad (e_{\text{RE}} + e_{\text{SN}} < e_B \quad \text{and} \quad e_{\text{RE}} > e_{\text{SN}}), & \text{with CLM} \\ e_{\text{SN}} < e_{\text{SN},B} \quad \text{and} \quad e_{\text{RE}} \text{ is finite,} & \text{with CHILM,} \end{cases} \quad (170)$$

in the large circuit regime. So CHILM offers a distinct advantage, particularly if e_{RE} is the most significant contribution to the noise.

6.2.2 Hardware Robustness

The test for hardware robustness is relatively simple (see Sec. 2.5.1). We assume that every gate has independent (i.e. non uniform) noise, that we assume has been well-characterised. Then we calculate the proxy bias (which is still zero) and runtime scaling (Sec. 2.3.4) of mitigation using these noisy gates for variant generation rather than the noise-free amplification gates (Eqs (150) and (151)) and hidden inverse procedures (Eq. (161)) that we considered in Secs 6.2, 6.1.4, and 6.1.3, which did not add any uncontrolled errors into our circuit beyond those present in the original noisy gate. If the runtime scaling does not change (to leading order) then the method is robust for our hardware.

The two mitigation protocols we implement are given by (see Eqs (142), (145), (148), and (150)):

$$\begin{aligned} \text{---} \tilde{U} \text{---}_M &= \text{---} \tilde{U} \text{---} \left\{ \begin{array}{l} \tilde{C}_{C;0} : \text{---} \\ \tilde{C}_{C;1} : \text{---} \tilde{S}_i \text{---} \\ \tilde{C}_{C;2} : \text{---} \tilde{P}_i \text{---} \tilde{S}_i \text{---} \end{array} \right\} = \left\{ \begin{array}{l} \tilde{C}_{\text{CHI};0} : \text{---} \tilde{U} \text{---} \\ \tilde{C}_{\text{CHI};1} : \text{---} \tilde{U} \text{---} \tilde{P}_i \text{---} \\ \tilde{C}_{\text{CHI};2} : \text{---} \tilde{U}_{\text{HI}} \text{---} \end{array} \right\} = \text{---} \tilde{U} \text{---}, \quad (171) \end{aligned}$$

where the first decomposition is CLM and the second CHILM. Now, we assume our noisy gate $\tilde{U}$ has rotational and stochastic noise amplitudes ϕ and p , respectively (see Eq. (143)), as before. However, we also assume that our mitigation gates and the hidden inverse procedure are noisy, so we have (see Eq. (73)):

$$\begin{array}{|c|} \hline \tilde{U} \\ \hline \end{array} = \begin{array}{|c|} \hline U \\ \hline \end{array} \begin{array}{|c|} \hline \mathcal{E}_{\text{ORE}; p, \phi} \\ \hline \end{array}, \quad \begin{array}{|c|} \hline \tilde{U} \\ \hline \end{array} = \begin{array}{|c|} \hline U \\ \hline \end{array} \begin{array}{|c|} \hline \mathcal{E}_{\text{ORE}; y_{\text{HI}} p, z_{\text{HI}} \phi} \\ \hline \end{array}, \quad (172)$$

$$\begin{array}{|c|} \hline \tilde{P}_i \\ \hline \end{array} = \begin{array}{|c|} \hline P_i \\ \hline \end{array} \begin{array}{|c|} \hline \mathcal{E}_{\text{ORE}; z_P p, z_P \phi} \\ \hline \end{array}, \quad \begin{array}{|c|} \hline \tilde{S}_i \\ \hline \end{array} = \begin{array}{|c|} \hline S_i \\ \hline \end{array} \begin{array}{|c|} \hline \mathcal{E}_{\text{ORE}; z_S p, z_S \phi} \\ \hline \end{array}. \quad (173)$$

$\tilde{P}_i$; $\tilde{S}_i$; and $\tilde{U}_{\text{HI}}$ are noisy, with noise amplitudes $z_P \phi$ and $y_P p$; $z_S \phi$ and $y_S p$; and $-z_{\text{HI}} \phi$ (where $z_{\text{HI}} > 0$ otherwise the hidden inverse failed) and $y_{\text{HI}} p$, respectively. This means the rotational noise amplitudes of our noisy variants are given by:

$$\begin{cases} \phi_0 = \phi, & \phi_1 = (1 + z_S) \phi + \frac{\pi}{2}, & \phi_2 = (1 + z_S + z_P) \phi + \frac{3\pi}{2}, & \text{for CLM} \\ \phi_0 = \phi, & \phi_1 = (1 + z_P) \phi + \pi, & \phi_2 = -z_{\text{HI}} \phi, & \text{for CHILM,} \end{cases} \quad (174)$$

(where the gates are numbered from top to bottom in Eq. (171)) and their stochastic noise amplitudes are:

$$\begin{cases} p_0 = p, & p_1 = (1 + y_S) p, & p_2 = (1 + y_S + y_P) p, & \text{for CLM} \\ p_0 = p, & p_1 = (1 + y_P) p, & p_2 = y_{\text{HI}} p, & \text{for CHILM.} \end{cases} \quad (175)$$

So:

$$\begin{array}{|c|} \hline \tilde{U}_j \\ \hline \end{array} = \begin{array}{|c|} \hline U \\ \hline \end{array} \begin{array}{|c|} \hline \mathcal{E}_j \\ \hline \end{array} = \begin{array}{|c|} \hline U \\ \hline \end{array} \begin{array}{|c|} \hline e^{i \frac{\phi_j}{2} P_i} \\ \hline \end{array} \begin{cases} \text{---} & \text{with probability } 1 - p_j \\ \begin{array}{|c|} \hline P_i \\ \hline \end{array} & \text{with probability } p_j, \end{cases} \quad (176)$$

is the j th gate variant used to mitigate the noisy gate $\tilde{U}$.

Now the eigenvalues associated with the transfer matrices ($\hat{T}_{\mathcal{E}_j}$) that characterise the effect of the (amplified) error channels ($\mathcal{E}_j$) are given by Eq. (152) as before and we must solve Eq. (126) to find our coefficients (where $c_{L;j} = \tilde{c}_{C;j}$ for CLM and $c_{L;j} = \tilde{c}_{\text{CHI};j}$ for CHILM). The resulting coefficients are too cumbersome to write in full here but we can examine their first order expansions (in the noise amplitudes), i.e. we neglect terms of order $\mathcal{O}_0(p^2, p\phi, \phi^2)$. We also make the assumption that all the noise amplitudes have the same order of magnitude, i.e. $(1 + z_S + z_P + z_{\text{HI}}) \phi, (1 + y_S + y_P + y_{\text{HI}}) p \ll 1$.

The coefficients for CLM (to leading order) are:

$$\tilde{c}_{C;0} = 1 + 2p, \quad \tilde{c}_{C;1} = -p - \frac{\phi}{2}, \quad \tilde{c}_{C;2} = -p + \frac{\phi}{2}, \quad (177)$$

these are independent of y_S, z_S, y_P, z_P so CLM is robust to using noisy amplification gates since the runtime scaling is a function of the coefficients in Eq. (177) (to leading order).

The coefficients for CHILM (to leading order) are:

$$\begin{aligned} \tilde{c}_{\text{CHI};0} &= \frac{z_{\text{HI}}}{1 + z_{\text{HI}}} + \frac{(1 + z_{\text{HI}})(z_{\text{HI}} - y_{\text{HI}}) - (z_{\text{HI}} + y_{\text{HI}}) z_P}{(1 + z_{\text{HI}})^2} p + \frac{z_{\text{HI}}(z_{\text{HI}} - 3z_P - 1)}{12(1 + z_{\text{HI}})} \phi^2, \\ \tilde{c}_{\text{CHI};1} &= -\frac{z_{\text{HI}} + y_{\text{HI}}}{1 + z_{\text{HI}}} p - \frac{z_{\text{HI}}}{4} \phi^2, \\ \tilde{c}_{\text{CHI};2} &= \frac{1}{1 + z_{\text{HI}}} + \frac{2y_{\text{HI}}(1 + z_{\text{HI}}) + (z_{\text{HI}} + y_{\text{HI}}) z_P}{(1 + z_{\text{HI}})^2} p + \frac{z_{\text{HI}}(2z_{\text{HI}} + 3z_P + 4)}{12(1 + z_{\text{HI}})} \phi^2, \end{aligned} \quad (178)$$

so these do depend on $y_{\text{HI}}, z_{\text{HI}}$ and z_{P} . Thus we must calculate the runtime scaling explicitly to determine if CHILM is robust. CHILM's runtime scaling is given by:

$$\begin{aligned} S_{\text{GI}} &= 1 + 2 \left(1 + \frac{1}{4N_{\text{G}}} \right) \frac{y_{\text{HI}} + z_{\text{HI}}}{1 + z_{\text{HI}}} e_{\text{SN}} + \left(1 + \frac{1}{4N_{\text{G}}} \right) z_{\text{HI}} \frac{e_{\text{RE}}^2}{N_{\text{G}}} + \mathcal{O}_{\text{S}}(e_{\text{SN}}^2, e_{\text{SN}} e_{\text{RE}}^2, e_{\text{RE}}^4) \\ &= e^{2 \frac{y_{\text{HI}} + z_{\text{HI}}}{1 + z_{\text{HI}}} e_{\text{SN}}} \left(1 + \mathcal{O}_{\text{L}} \left(\frac{e_{\text{SN}}}{N_{\text{G}}}, \frac{e_{\text{SN}}^2}{N_{\text{G}}}, \frac{e_{\text{RE}}^2}{N_{\text{G}}} \right) \right), \end{aligned} \quad (179)$$

and the noise boundary is given by:

$$e_{\text{SN,B}} = \frac{1 + z_{\text{HI}}}{y_{\text{HI}} + z_{\text{HI}}} \frac{\ln S_{\dagger}}{2}. \quad (180)$$

So, though CHILM is not robust, it is quasi-robust (see Sec. 2.5.1) since the orders of the most significant order terms for its noise boundary and runtime scaling are independent of the $y_{\text{HI}}, z_{\text{HI}}$ and z_{P} , even though the values of the coefficients are not. The runtime scaling and noise boundary are robust to noise in the custom channel (i.e. the values of z_{P} and y_{P}) and, if $y_{\text{HI}} < 1$, using imperfect hidden inverses could actually perform better than using noise-free hidden inverses.

6.2.3 Method Certification

We can use our metrics to certify our mitigation methods (see Sec. 2.5.1).

1. **Scalable:** Both methods are scalable since their runtime scaling in the large circuit regime depends only on the noise level (e_{SN} and/or e_{RE}) and not the number of gates (N_{G} , see Eq. (167)). However, CHILM scales better than CLM as CHILM's runtime is independent of the rotational noise level (e_{RE}) in the large circuit limit.
2. **Unbounded** (in principle): The noise boundary (see Eq. (169)) grows as the logarithm of the tolerated runtime scaling ($S_{\dagger}$) for both methods. So we can mitigate arbitrary noise levels with either methods, if we are willing to wait long enough. Thus both methods are unbounded but CHILM is independent of the rotational noise level, so, again, it has a slight edge.
3. **Precise:** Since both methods are unbiased they are both precise and can be used to reduce the bias to arbitrarily small levels, if we are willing to wait long enough.
4. **Efficient:** Both methods are also efficient since the runtime scaling in the small noise limit ($e_{\text{SN}}, e_{\text{RE}} \rightarrow 0$) is unity (see Eq. (166)). So they can make good use of any reduction in noise level. However, CHILM is more efficient than CLM since it is independent of e_{RE} to first order.
5. **Robust** (see Sec. 6.2.2): CLM is robust for our hardware since noise in the mitigation gates does not change the runtime cost to leading order in the large circuit or small noise regime provided the noise is characterised and the same order of magnitude as that of the gate being mitigated. CHILM is quasi-robust (see Sec. 2.5.1) since noise in the hidden inverse procedure can change the effective size of the stochastic noise level but it is robust to noise in the amplification channel.

So CLM can be certified SUPER (scalable, unbounded, precise, efficient, and robust) and CHILM is a SUPEqR (scalable, unbounded, precise, efficient, and quasi-robust) mitigation method. Though CHILM is not robust it does scale better than CLM in general. In the next section we shall certify the mitigation strategy as a whole.

6.3 Strategy Certification

We can now certify our full mitigation strategy according to the criteria in Sec. 2.5.2.

1. **Universal:** We have demonstrated (see Sec. 6.1.3) that we can mitigate any gate within our gate set (Eqs (155), (156), and (157)) using only gates that are already within the gate set to perform the mitigation (see Eqs (158), (159), and (160) of Sec. 6.1.4, the exception being for CHILM where we also require hidden inverse implementations of the gate to be mitigated), if all the gates suffer from stochastic over-rotation errors (see Sec. 6.2.2). Now, all that remains is to show that our gate set is universal. To do this we will provide decompositions of the gates within a universally recognised universal gate set (T-Z45, H-Hadamard, CX-Controlled Not, e.g. see Refs [37, 38]) in terms of our (universally mitigatable) gate set (Eqs (155), (156), and (157)):

$$\begin{aligned} T = T_3, \quad CZ = e^{-3i\frac{\pi}{4}} (S_3 Z \otimes S_3 Z) S_{15}, \quad H = e^{i\frac{\pi}{2}} S_1 S_3 S_1, \quad CX = (I \otimes H) CZ (I \otimes H). \end{aligned} \quad (181)$$

2. **Scalable:** Since C(HI)LM is scalable (see Sec. 6.2.3) even when using imperfect custom channels and hidden inverses to effect the mitigation (see Sec. 6.2.2) the full mitigation strategy is scalable.
3. **Efficacious:** C(HI)LM is SUPE(q)R (see Sec. 6.2.3), so it is unbounded, precise, efficient and quasi-robust for hardware suffering from over-rotation errors (see Sec. 6.2.2) therefore it is efficacious (the quasi-robustness ensures that the method certifications are still relevant for the noisy hardware we consider).

Thus our mitigation strategy (USEM:ORE) can be certified USEful for hardware suffering from stochastic over-rotation errors and we can successfully conclude our certification.

7 Concluding Remarks

In this work, we introduced a unified framework for the classification and certification of quantum error mitigation (QEM) methods. Our approach combines quantitative metrics – such as proxy bias, runtime scaling, and noise boundaries – with qualitative criteria including scalability, robustness, and efficiency. These tools enable a systematic evaluation of linear mitigation methods and their applicability across different noise models.

Through our taxonomy and case studies, we demonstrated how the proposed certification scheme can be applied to compare and design mitigation strategies tailored to specific quantum hardware and noise profiles.

7.1 Further Work

Future efforts should aim to:

- (i) Extend the certification framework to non-linear and hybrid error mitigation methods as well as additional noise models (e.g. amplitude damping [2]). In Ref. [1] we consider virtual distillation – a non-linear mitigation method [11–13] – applied to circuits suffering from global depolarising noise and calculate the relevant quantitative metrics (runtime scaling, proxy bias, and noise boundary).

- (ii) Incorporate realistic experimental imperfections into the certification process and make use of real noise characterisation data obtained by tomography.
- (iii) Explore integration with other approaches to noise suppression, such as error correction and/or hardware tuning (e.g. incorporate metrics such as energy consumption [10]).

A promising direction is the development of automated standardised certification pipelines combining theoretical metrics and experimental data to enable adaptive mitigation strategies in near-term quantum devices.

Acknowledgments

Plots and simulations were produced using Python (using matplotlib, mpmath, numpy, scipy, and sympy) and the analytic analysis made use of Mathematica. This publication was made possible by the use of the FactoryIA supercomputer, financially supported by the Ile-de-France Regional Council. Thanks to David Herrera-Martí for securing the funding that made this work possible and for his contributions throughout. Many thanks to Vasily Sazonov for his part in developing the ideas that motivated and formed the basis of this work, his advice, and his comments on the manuscript.

References

- [1] Z J. Blunden-Codd. “Classification and certification of quantum error mitigation methods: Extended edition”. unpublished (2025).
- [2] John Preskill. “Lecture notes for ph219/cs219: Quantum information chapter 3”. http://theory.caltech.edu/~preskill/ph219/chap3_15.pdf (2018). [Online; accessed 14-November-2022].
- [3] 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” (2023). [arXiv:2210.00921](https://arxiv.org/abs/2210.00921).
- [4] Suguru Endo, Zhenyu Cai, Simon C. Benjamin, and Xiao Yuan. “Hybrid quantum-classical algorithms and quantum error mitigation”. *Journal of the Physical Society of Japan* **90**, 032001 (2021).
- [5] Vincent R. Pascuzzi, Andre He, Christian W. Bauer, Wibe A. de Jong, and Benjamin Nachman. “Computationally efficient zero-noise extrapolation for quantum-gate-error mitigation”. *Phys. Rev. A* **105**, 042406 (2022).
- [6] Ryuji Takagi, Suguru Endo, Shintaro Minagawa, and Mile Gu. “Fundamental limits of quantum error mitigation”. *npj Quantum Information* **8**, 114 (2022).
- [7] Zhenyu Cai. “A practical framework for quantum error mitigation” (2023). [arXiv:2110.05389](https://arxiv.org/abs/2110.05389).
- [8] C. J. Ballance, T. P. Harty, N. M. Linke, M. A. Sepiol, and D. M. Lucas. “High-fidelity quantum logic gates using trapped-ion hyperfine qubits”. *Phys. Rev. Lett.* **117** (2016).
- [9] Google Quantum AI and Collaborators. “Quantum error correction below the surface code threshold”. *Nature* **638**, 920–926 (2024).
- [10] Marco Fellous-Asiani, Jing Chai, Yvain Thonnart, Hui Ng, Robert Whitney, and A. Auffèves. “Optimizing resource efficiencies for scalable full-stack quantum computers”. *PRX Quantum* **4** (2023).

- [11] Bálint Koczor. “Exponential error suppression for near-term quantum devices”. *Phys. Rev. X* **11**, 031057 (2021).
- [12] William J. Huggins, Sam McArdle, Thomas E. O’Brien, Joonho Lee, Nicholas C. Rubin, Sergio Boixo, K. Birgitta Whaley, Ryan Babbush, and Jarrod R. McClean. “Virtual distillation for quantum error mitigation”. *Phys. Rev. X* **11**, 041036 (2021).
- [13] Piotr Czarnecki, Andrew Arrasmith, Lukasz Cincio, and Patrick J. Coles. “Qubit-efficient exponential suppression of errors” (2021). [arXiv:2102.06056](#).
- [14] Suguru Endo, Simon C. Benjamin, and Ying Li. “Practical quantum error mitigation for near-future applications”. *Phys. Rev. X* **8**, 031027 (2018).
- [15] Nobuyuki Yoshioka, Hideaki Hakoshima, Yuichiro Matsuzaki, Yuuki Tokunaga, Yasunari Suzuki, and Suguru Endo. “Generalized quantum subspace expansion”. *Phys. Rev. Lett.* **129**, 020502 (2022).
- [16] Kristan Temme, Sergey Bravyi, and Jay M. Gambetta. “Error mitigation for short-depth quantum circuits”. *Phys. Rev. Lett.* **119**, 180509 (2017).
- [17] Daniel Greenbaum. “Introduction to quantum gate set tomography” (2015).
- [18] Andrea Mari, Nathan Shammah, and William J. Zeng. “Extending quantum probabilistic error cancellation by noise scaling”. *Phys. Rev. A* **104**, 052607 (2021).
- [19] 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).
- [20] Youngseok Kim, Christopher J. Wood, Theodore J. Yoder, Seth T. Merkel, Jay M. Gambetta, Kristan Temme, and Abhinav Kandala. “Scalable error mitigation for noisy quantum circuits produces competitive expectation values”. *Nature Physics* **19**, 752–759 (2023).
- [21] Bichen Zhang, Swarnadeep Majumder, Pak Hong Leung, Stephen Crain, Ye Wang, Chao Fang, Dripto M. Debroy, Jungsang Kim, and Kenneth R. Brown. “Hidden inverses: Coherent error cancellation at the circuit level”. *Phys. Rev. Appl.* **17**, 034074 (2022).
- [22] V Leyton-Ortega, S Majumder, and R C Pooser. “Quantum error mitigation by hidden inverses protocol in superconducting quantum devices”. *Quantum Science and Technology* **8**, 014008 (2022).
- [23] Ying Li and Simon C. Benjamin. “Efficient variational quantum simulator incorporating active error minimization”. *Phys. Rev. X* **7**, 021050 (2017).
- [24] 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” (2022). [arXiv:2201.10672](#).
- [25] Ewout van den Berg, Zlatko K. Mineev, Abhinav Kandala, and Kristan Temme. “Probabilistic error cancellation with sparse pauli–lindblad models on noisy quantum processors”. *Nature Physics* **19**, 1116–1121 (2023).
- [26] Michael Krebsbach, Björn Trauzettel, and Alessio Calzona. “Optimization of richardson extrapolation for quantum error mitigation”. *Phys. Rev. A* **106**, 062436 (2022).
- [27] Tudor Giurgica-Tiron, Yousef Hindy, Ryan LaRose, Andrea Mari, and William J. Zeng. “Digital zero noise extrapolation for quantum error mitigation”. In 2020 IEEE International Conference on Quantum Computing and Engineering (QCE). *Pages* 306–316. (2020).
- [28] Joe O’Gorman, Naomi H. Nickerson, Philipp Ross, John JL Morton, and Simon C.

- Benjamin. “A silicon-based surface code quantum computer”. [npj Quantum Information](#) **2**, 15019 (2016).
- [29] 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).
- [30] Charles H. Bennett, David P. DiVincenzo, John A. Smolin, and William K. Wootters. “Mixed-state entanglement and quantum error correction”. [Phys. Rev. A](#) **54**, 3824–3851 (1996).
- [31] E. Knill. “Fault-tolerant postselected quantum computation: Threshold analysis” (2004). [arXiv:quant-ph/0404104](#).
- [32] Joel J. Wallman and Joseph Emerson. “Noise tailoring for scalable quantum computation via randomized compiling”. [Phys. Rev. A](#) **94**, 052325 (2016).
- [33] Jarrod Ryan McClean, Sergio Boixo Castrillo, Craig Gidney, and Vadim Smelyanskiy. “Reducing errors with circuit gauge selection” (2022). US Patent App. 17/777,708.
- [34] Andre He, Benjamin Nachman, Wibe A. de Jong, and Christian W. Bauer. “Zero-noise extrapolation for quantum-gate error mitigation with identity insertions”. [Phys. Rev. A](#) **102**, 012426 (2020).
- [35] Christophe Piveteau, David Sutter, and Stefan Woerner. “Quasiprobability decompositions with reduced sampling overhead”. [npj Quantum Information](#) **8**, 12 (2022).
- [36] Zhenyu Cai. “Multi-exponential error extrapolation and combining error mitigation techniques for nisq applications”. [npj Quantum Information](#) **7**, 80 (2021).
- [37] P. B. M. Sousa and R. V. Ramos. “Universal quantum circuit for n-qubit quantum gate: A programmable quantum gate”. [Quantum Info. Comput.](#) **7**, 228–242 (2007).
- [38] Lars Steffen. “Quantum teleportation and efficient process verification with superconducting circuits”. PhD thesis. ETH Zurich. (2013).
- [39] H.P. Breuer and F. Petruccione. “The theory of open quantum systems”. Oxford University Press. (2002).
url: <https://books.google.fr/books?id=0Yx5VzaMYm8C>.
- [40] John Preskill. “Quantum computing in the nisq era and beyond”. [Quantum](#) **2**, 79 (2018).
- [41] Seth T. Merkel, Jay M. Gambetta, John A. Smolin, Stefano Poletto, Antonio D. Córcoles, Blake R. Johnson, Colm A. Ryan, and Matthias Steffen. “Self-consistent quantum process tomography”. [Phys. Rev. A](#) **87**, 062119 (2013).
- [42] Sergey Bravyi, Sarah Sheldon, Abhinav Kandala, David C. McKay, and Jay M. Gambetta. “Mitigating measurement errors in multiqubit experiments”. [Physical Review A](#) **103** (2021).
- [43] Wolfram Research, Inc. “Mathematica, Version 13.0”. Champaign, IL, 2023.
- [44] Andre He, Benjamin Nachman, Wibe A. de Jong, and Christian W. Bauer. “Resource efficient zero noise extrapolation with identity insertions” (2020). [arXiv:2003.04941v1](#).

Appendix

Appendices

A	Glossary of Terms	64
B	Worked Example: IISM:NA	66
B.1	Notation	66
B.2	Circuit	67
B.3	Assumptions and Requirements	68
B.4	Noise Model Characterisation	68
B.5	Mitigation: IISM:NA	71
C	Mitigation Methods	78
C.1	Noise Models	79
C.2	Linear Mitigation	79
C.3	CHILM	81
C.4	CHISM	83
C.5	CIILM	84
C.6	CLM	85
C.7	CSM	87
C.8	IHAM	88
C.9	IILM	92
C.10	IISM	93
C.11	LC	94
C.12	TIILM	96
D	Numerical	98
D.1	Simulation	98
D.2	Precision and Code	103
D.3	Parameters	104

Appendix

A Glossary of Terms

Here we present a glossary of terms that we use in this work in alphabetical order, with references to where they are used in the main text.

Term	Description	Reference
Area Scale Factor ($F_{\text{GL,A}}$)	The product of the length and width scale factors, representing the resource overhead of mitigation per run of the circuit.	Sec. 2.3.4
Asynchronous Mitigation (AM)	Mitigation where different gates can be amplified by different amounts in each circuit variant.	Sec. 4.1
Bias (ε)	Deviation of the measured observable expectation value from the true noiseless value.	Sec. 2.1
Certification	Process of assessing a mitigation method using predefined quantitative and qualitative metrics.	Sec. 2.5
Circuit Variant	One of a selection of noisy implementations of a noise-free target circuit.	Sec. 4.2
CPTP	Completely Positive and Trace Preserving	Ref. [39]
Custom Channel (C)	A tailored noise channel inserted after a gate to amplify or modify its noise characteristics.	Sec. 4.2.3
Efficacious	Mitigation strategy certification criteria for when all mitigation methods within the strategy are SUPER or SUPEqR.	Sec. 2.5.2
Efficient	Property of a mitigation method with unit runtime scaling in the limit of vanishing noise.	Sec. 2.5.1
Error Channel ($\mathcal{E}$)	Mathematical model representing the effect of errors in a quantum gate.	Sec. 2.2.1
Error Mitigation	Technique to reduce the impact of noise on quantum computations without reducing the noise level of any individual run.	Sec. 1.1
Gate Variant	One of a selection of noisy implementations of a noise-free target gate.	Sec. 4.2
Hidden Inverse (HI)	A gate variant constructed to have the opposite coherent error to the original gate, used for local cancellation.	Sec. 4.2.2
Identity Insertion (II)	A technique that amplifies noise by inserting sequences of gates and their inverses.	Sec. 4.2.4
Knowledge Free (KF)	Mitigation methods that do not require quantitative knowledge of the noise model.	Sec. 4.1
Length Scale Factor ($F_{\text{GL,L}}$)	The average increase in circuit runtime due to mitigation.	Sec. 2.3.4
Linear Mitigation	Mitigation methods that use linear combinations of noisy circuit outputs to estimate noise-free results.	Sec. 2.1
Local Cancellation (LC)	A tailoring method that converts coherent noise to incoherent/stochastic noise.	Sec. 4.3
Local Mitigation (LM)	Mitigation applied independently to each gate or small group of nearby gates.	Sec. 4.1
Mitigation Method	A procedure for mitigating the noise of a quantum circuit.	Sec. 4
Mitigation Strategy	A procedure for producing mitigated circuits for quantum algorithms implemented on a selected and characterised piece of quantum hardware. May consist of multiple mitigation methods.	Sec. 2.5.2
NISQ	Noisy Intermediate-Scale Quantum: The current era of quantum devices with limited qubit counts and significant noise.	Ref. [40]
Noise Amplitude (x , p or ϕ)	Parameter quantifying the strength of noise in a quantum channel.	Sec. 2.2.1
Noise Aware (NA)	Mitigation methods that use quantitative knowledge of the noise model to optimise performance.	Sec. 4.1
Noise Boundary (e_B)	The maximum noise level that can be mitigated for a given tolerated runtime scaling and target bias.	Sec. 2.3.5
Noise Channel ($\mathcal{N}$)	A CPTP map that is part of a linear decomposition of an error channel.	Sec. 2.2.1
Noise Level (e_G)	Parameter quantifying the strength of noise in a quantum circuit to leading order.	Sec. 2.3.2
ORE	Over-Rotation Errors: An error channel equivalent to that produced by a rotation in Hilbert space rotating too far.	Sec. 3.5

Appendix

Term	Description	Reference
PEC	Probabilistic Error Cancellation: A linear mitigation method using quasi-probability decompositions.	Ref. [16]
Precise	Property of a mitigation method that can reduce the effective noise level to an arbitrarily small amount.	Sec. 2.5.1
Proxy Bias ($\varepsilon_\odot$)	An upper bound to the true bias, calculable without explicit circuit simulation.	Sec. 2.3.1
QEM	Quantum Error Mitigation: See Error Mitigation	
Robust	Property of a mitigation method with runtime scaling independent of small characterised imperfections in method implementation.	Sec. 2.5.1
Rotational Errors (RE)	A coherent error channel equivalent to a (unitary) rotation generated by a Hermitian unitary (self-inverse) operator.	Sec. 3.4
Runtime Scaling (S)	The ratio of the total runtime required for a mitigated circuit to that of the unmitigated circuit to achieve the same precision.	Sec. 2.3.4
Sampling Cost (C)	The increase in the number of samples (shots) required for a mitigated circuit to preserve the statistical precision scaling of the original circuit.	Sec. 2.3.4
Scalable	Property of a mitigation method that allows it to maintain performance as circuit size increases for fixed noise level.	Secs 2.5.1 and 2.5.2
SPAM Errors	State Preparation and Measurement Errors: Errors associated with initialising and measuring quantum states.	Ref. [41]
Stochastic Noise (SN)	Random noise model where errors occur probabilistically.	Sec. 3.3
SUPEqR	Scalable, Unbounded, Precise, Efficient, quasi-Robust: Certification for mitigation methods that are robust to leading order.	Sec. 2.5
SUPER	Scalable, Unbounded, Precise, Efficient, Robust: Certification for the highest standard of mitigation methods.	Sec. 2.5
Synchronous Mitigation (SM)	A mitigation method where all gates, within a particular circuit variant, are amplified by the same amount.	Sec. 4.1
Taxonomy (of mitigation methods)	Structured classification of error mitigation techniques based on their characteristics and requirements.	Sec. 4
Tuned Identity Insertion (TII)	An optimised version of identity insertion where the number of identity insertions is chosen to minimise the runtime scaling.	Sec. 4.2.4
Unbounded	Property of a mitigation method that can mitigate arbitrarily high noise levels.	Sec. 2.5.1
Uniform Noise Model	Noise model where all gates have the same type of noise channel and the same noise amplitude.	Sec. 2.5.2
Universal	Property of a mitigation strategy that allows it to provide a mitigated circuits for any quantum algorithm.	Sec. 2.5.2
USEful	Certification that a mitigation strategy is Universal, Scalable, and Efficacious	Sec. 2.5.2
USEM:ORE	Universal, Scalable, Efficacious, Mitigation for Over-Rotation Errors: A mitigation strategy strategy certified USEful for hardware with over-rotation errors.	Sec. 6.1
Width Scale Factor ($F_{\text{GI},W}$)	The average increase in qubits required per circuit due to mitigation.	Sec. 2.3.4

Appendix

B Worked Example: IISM:NA

Here we provide calculations of proxy bias, before and after mitigation, and the sampling cost; for hardware suffering from uniform dephasing errors, mitigated using IISM:NA (noise-aware identity-insertion-assisted synchronous mitigation).

B.1 Notation

We use Pauli words (App. B.1.1) and transfer matrices (App. B.1.2).

B.1.1 Pauli Words and Bloch Decomposition

$\vec{P}$ is the Pauli word vector and:

$$P_i = \left(\vec{P} \right)_i, \quad P_i \in \{I, X, Y, Z\}^{\otimes N_Q}, \quad (\text{B.182})$$

where P_i is the i th Pauli word. For concreteness we define:

$$\left(\vec{P} \right)_i = P_i = \bigotimes_{n=1}^{N_Q} \left(\vec{\sigma}_{[n]} \right)_{\left\lfloor \frac{i}{4^{n-1}} \right\rfloor \bmod 4}, \quad (\text{B.183})$$

where a subscript $[n]$ denotes that an operator acts on the n th qubit and $\vec{\sigma}_{[n]}$ is the Pauli vector for the n th qubit: $(\vec{\sigma}_{[n]})_0 = I$, the two by two identity matrix; $(\vec{\sigma}_{[n]})_1 = X$, the two by two Pauli X matrix; $(\vec{\sigma}_{[n]})_2 = Y$, the two by two Pauli Y matrix; and $(\vec{\sigma}_{[n]})_3 = Z$, the two by two Pauli Z matrix. If we wish to denote the n th qubit operator of the i th component of $\vec{P}$ we use the notation:

$$P_{i,n} = \left(\vec{\sigma}_{[n]} \right)_{\left\lfloor \frac{i}{4^{n-1}} \right\rfloor \bmod 4}. \quad (\text{B.184})$$

Some examples of Pauli words are:

$$P_0 = I = I \otimes \dots I, \quad P_1 = I \otimes \dots X, \quad P_2 = I \otimes \dots Y, \quad P_3 = I \otimes \dots Z, \quad (\text{B.185})$$

$$P_{15} = I \otimes \dots Z \otimes Z, \quad P_{27} = I \otimes \dots X \otimes Y \otimes Z, \quad (\text{B.186})$$

though generally we neglect leading identities.

The (generalised) Bloch decomposition is a vector representation of a density matrix in the basis of Pauli words:

$$\rho = \frac{\vec{b} \cdot \vec{P}}{2^{N_Q}}, \quad (\text{B.187})$$

where N_Q is the number of qubits and $\vec{b}$, the generalised Bloch vector, is a vector of expectations for each of the Pauli words, so:

$$\left(\vec{b} \right)_i = \langle P_i \rangle = \text{Tr}[\rho P_i], \quad (\text{B.188})$$

and therefore $-1 \leq \left(\vec{b} \right)_i \leq 1$.

Appendix

B.1.2 Transfer Matrices

The output of a CPTP (completely-positive and trace-preserving) error channel (with input state ρ) is a density matrix:

$$\tilde{\rho} = \mathcal{E}(\rho). \quad (\text{B.189})$$

If $\vec{b}$ is the Bloch vector of ρ and $\vec{\tilde{b}}$ is the Bloch vector of $\tilde{\rho}$ we have:

$$\frac{1}{2^{N_Q}} \sum_{j=0}^{4^{N_Q}-1} (\vec{b})_j P_j = \frac{\vec{b} \cdot \vec{P}}{2^{N_Q}} = \tilde{\rho} = \mathcal{E}(\rho) = \frac{\mathcal{E}(\vec{b} \cdot \vec{P})}{2^{N_Q}} = \frac{1}{2^{N_Q}} \sum_{j=0}^{4^{N_Q}-1} (\vec{b})_j \mathcal{E}(P_j), \quad (\text{B.190})$$

where we have used the linearity of the channel and the Bloch decomposition of the density operators (Eq. (B.187)). Thus we can express the channel as a map between Bloch vectors:

$$(\vec{\tilde{b}})_i = \frac{1}{2^{N_Q}} \text{Tr} \left[\sum_{j=0}^{4^{N_Q}-1} (\vec{b})_j P_j P_i \right] = \text{Tr} [P_i \tilde{\rho}] = \text{Tr} [P_i \mathcal{E}(\rho)] \quad (\text{B.191})$$

$$= \frac{1}{2^{N_Q}} \sum_{j=0}^{4^{N_Q}-1} \text{Tr} [P_i \mathcal{E}(P_j)] (\vec{b})_j, \quad (\text{B.192})$$

so we have:

$$\vec{\tilde{b}} = \hat{T}_{\mathcal{E}} \vec{b}, \quad (\text{B.193})$$

where we define the Pauli transfer matrix as:

$$(\hat{T}_{\mathcal{E}})_{i,j} = \frac{1}{2^{N_Q}} \text{Tr} [P_i \mathcal{E}(P_j)]. \quad (\text{B.194})$$

Since $\vec{\tilde{b}}$ must have all elements between -1 and 1 (see Eq. (B.188)) $\hat{T}_{\mathcal{E}}$ must also have all elements between -1 and 1 , otherwise there would be some choices of $\vec{b}$ that would lead to unphysical $\vec{\tilde{b}}$.

B.2 Circuit

We consider a situation where we wish to implement the circuit:

$$\boxed{C} = \boxed{U_1} - \boxed{U_2} - \dots - \boxed{U_{N_G}}, \quad (\text{B.195})$$

of N_G gates where U_n is the n th noise-free gate. However, owing to noise affecting our gates, we actually implement the circuit:

$$\boxed{\tilde{C}} = \boxed{\tilde{U}_1} - \boxed{\tilde{U}_2} - \dots - \boxed{\tilde{U}_{N_G}}, \quad (\text{B.196})$$

where $\tilde{U}_n$ is the noisy gate given by:

$$\boxed{\tilde{U}_n} = \boxed{U_n} - \boxed{\mathcal{E}}, \quad (\text{B.197})$$

and $\mathcal{E}$ is its error channel.

Appendix

B.3 Assumptions and Requirements

1. We assume a uniform noise model, which means every gate has an identical noise channel.
2. We assume that the noise-free components of all gates commute with their error channels, a requirement for most methods using identity insertion.
3. We assume that the noisy inverse gate has the same error channel as the noisy gate it is the (near) inverse of, a requirement for most methods using identity insertion.
4. We use a single qubit dephasing error as our noise channel for illustrative purposes, but clearly this would not be appropriate for two qubit gates.
5. We assume there are no SPAM (state preparation and measurement) errors, as our mitigation methods are designed to mitigate gates, there are other procedures for mitigating measurements [42].

B.4 Noise Model Characterisation

We assume a dephasing noise channel with noise amplitude (error probability) $0 \leq x \leq 1$:

$$\mathcal{E}(\rho) = (1 - x)\rho + xZ\rho Z, \quad (\text{B.198})$$

where Z is a Pauli Z operator. Using circuit notation we have:

$$\boxed{\mathcal{E}} = \begin{cases} \text{---} & \text{with probability } 1 - x, \\ \boxed{Z} & \text{with probability } x. \end{cases} \quad (\text{B.199})$$

B.4.1 Transfer Matrix

The transfer matrix for dephasing noise is given by (Eq. (B.194)):

$$\hat{T}_{\mathcal{E}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 - 2x & 0 & 0 \\ 0 & 0 & 1 - 2x & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (\text{B.200})$$

this is a diagonal matrix with the eigenvalues:

$$\lambda_0 = 1, \quad \lambda_1 = 1 - 2x, \quad (\text{B.201})$$

both of which are repeated.

B.4.2 Canonical Decomposition

We can decompose the channel into its canonical form by defining:

$$\mathcal{E} = f_0(x)\mathcal{N}_0 + f_1(x)\mathcal{N}_1, \quad f_0(x) = 1 - x, \quad \mathcal{N}_0 = \mathcal{I}, \quad (\text{B.202})$$

$$f_1(x) = x, \quad \mathcal{N}_1(\rho) = Z\rho Z, \quad (\text{B.203})$$

where $\mathcal{I}$ is the identity channel (returns the input unchanged). It is worth noting here that $\mathcal{N}_n$ is independent of the noise amplitude: x ; and that both $\mathcal{I}$ and $\mathcal{N}_n$ are CPTP

Appendix

(completely-positive and trace-preserving) channels. This means that if they act on a physical state they will return a physical state.

Since the identity channel commutes with all other channels this decomposition is self-compatible (i.e. its constituent noise channels commute). It is also closed under composition since the composition of two consecutive dephasing noise channels is equivalent to a single dephasing noise channel with a modified noise amplitude. That is, if we define:

$$\mathcal{E}_x = (1-x)\mathcal{I} + x\mathcal{N}_1, \quad \mathcal{E}_y = (1-y)\mathcal{I} + y\mathcal{N}_1, \quad (\text{B.204})$$

then:

$$\mathcal{E}_x \circ \mathcal{E}_y = \mathcal{E}_y \circ \mathcal{E}_x = ((1-x)(1-y) + xy)\mathcal{I} + (x(1-y) + y(1-x))\mathcal{N}_1, \quad (\text{B.205})$$

(since $\mathcal{N}_1^2 = \mathcal{I}$) which is just the dephasing noise channel $\mathcal{E}_z$ where:

$$z = x + y(1-2x). \quad (\text{B.206})$$

B.4.3 Circuit Noise Model

The channel that corresponds to the full circuit (see Eq. (B.196)) is:

$$\begin{aligned} \tilde{\mathcal{C}} &= \tilde{\mathcal{U}}_{N_G} \circ \tilde{\mathcal{U}}_{N_G-1} \circ \dots \tilde{\mathcal{U}}_1 = \mathcal{U}_{N_G} \circ \mathcal{E} \circ \mathcal{U}_{N_G-1} \circ \mathcal{E} \circ \dots \mathcal{U}_1 \circ \mathcal{E} \\ &= (1-x)^{N_G} \mathcal{U}_{N_G} \circ \mathcal{U}_{N_G-1} \circ \dots \mathcal{U}_1 \\ &\quad + x(1-x)^{N_G-1} (\mathcal{U}_{N_G} \circ \mathcal{N}_1 \circ \mathcal{U}_{N_G-1} \circ \dots \mathcal{U}_1 + \dots + \mathcal{U}_{N_G} \circ \mathcal{U}_{N_G-1} \circ \dots \mathcal{U}_1 \circ \mathcal{N}_1) \\ &\quad + \dots + x^{N_G} \mathcal{U}_{N_G} \circ \mathcal{N}_1 \circ \mathcal{U}_{N_G-1} \circ \mathcal{N}_1 \circ \dots \mathcal{U}_1 \circ \mathcal{N}_1, \end{aligned} \quad (\text{B.207})$$

where we have used the canonical error channel decomposition of Eq. (B.202); $\tilde{\mathcal{C}}$ is the channel corresponding to the whole circuit; and $\tilde{\mathcal{U}}_n$, $\mathcal{U}_n$, and $\mathcal{E}$ are the channels corresponding to the n th noisy-gate, noise-free component of the noisy gate, and error channel of the noisy gate, respectively.

We can write the expansion in Eq. (B.207) more succinctly in the form:

$$\tilde{\mathcal{C}} = \sum_{k=0}^{N_G} (1-x)^{N_G-k} x^k \check{\chi}_{\text{SN},k}, \quad (\text{B.208})$$

where:

$$\check{\chi}_{\text{SN},0} = \mathcal{U}_{N_G} \circ \mathcal{U}_{N_G-1} \circ \dots \mathcal{U}_1, \quad (\text{B.209})$$

$$\check{\chi}_{\text{SN},1} = \mathcal{U}_{N_G} \circ \mathcal{N}_1 \circ \mathcal{U}_{N_G-1} \circ \dots \mathcal{U}_1 + \dots + \mathcal{U}_{N_G} \circ \mathcal{U}_{N_G-1} \circ \dots \mathcal{U}_1 \circ \mathcal{N}_1, \quad (\text{B.210})$$

...

$$\check{\chi}_{\text{SN},N_G} = \mathcal{U}_{N_G} \circ \mathcal{N}_1 \circ \mathcal{U}_{N_G-1} \circ \mathcal{N}_1 \circ \dots \mathcal{U}_1 \circ \mathcal{N}_1, \quad (\text{B.211})$$

i.e. $\check{\chi}_{\text{SN},k}$ is the term containing all the constituent maps of $\tilde{\mathcal{C}}$ that include exactly k occurrences of the channel $\mathcal{N}_1$. Clearly, $\check{\chi}_{\text{SN},0}$ is equivalent to the noise-free target circuit ($\mathcal{C}$, see Eq. (B.195)):

$$\check{\chi}_{\text{SN},0} = \mathcal{U}_{N_G} \circ \mathcal{U}_{N_G-1} \circ \dots \mathcal{U}_1 = \mathcal{C}. \quad (\text{B.212})$$

We christen the number of constituent channels of $\check{\chi}_{\text{SN},k}$ its multiplicity:

$$\# [\check{\chi}_{\text{SN},k}] = \frac{N_G!}{k!(N_G-k)!}. \quad (\text{B.213})$$

Appendix

B.4.4 Proxy Bias and Noise Level

The bias (ε_O) due to noise in a circuit for an observable ($\hat{O}$, here assumed to be a Pauli word operator²⁷: $\hat{O} = P_i$) is given by:

$$\varepsilon_O = \left| \langle \tilde{O} \rangle - \langle \hat{O} \rangle \right|, \quad (\text{B.214})$$

where:

$$\langle \hat{O} \rangle = \text{Tr} [\hat{O} \mathcal{C}(\rho_0)], \quad \langle \tilde{O} \rangle = \text{Tr} [\hat{O} \tilde{\mathcal{C}}(\rho_0)], \quad (\text{B.215})$$

where ρ_0 is the initial state of the qubits, $\langle \hat{O} \rangle$ is the noise-free observable expectation and $\langle \tilde{O} \rangle$ is the noisy observable expectation. Using our notation in Eq. (B.208) we can re-express the bias in terms of our chi-terms (see Eq. (B.213)):

$$\begin{aligned} \varepsilon_O &= \left| \text{Tr} [\hat{O} (\tilde{\mathcal{C}}(\rho_0) - \mathcal{C}(\rho_0))] \right| = \left| \text{Tr} \left[\hat{O} \left(\sum_{k=0}^{N_G} (1-x)^{N_G-k} x^k \check{\chi}_{\text{SN},k}(\rho_0) - \check{\chi}_{\text{SN},0}(\rho_0) \right) \right] \right| \\ &= \left| \left((1-x)^{N_G} - 1 \right) \text{Tr} [\hat{O} \check{\chi}_{\text{SN},0}(\rho_0)] + \sum_{k=1}^{N_G} (1-x)^{N_G-k} x^k \text{Tr} [\hat{O} \check{\chi}_{\text{SN},k}(\rho_0)] \right| \\ &\leq \left(1 - (1-x)^{N_G} \right) + \sum_{k=1}^{N_G} (1-x)^{N_G-k} x^k \frac{N_G!}{k! (N_G-k)!} = 2 \left(1 - (1-x)^{N_G} \right), \quad (\text{B.216}) \end{aligned}$$

where we have inverted the binomial expansion in the final step and in the penultimate step used the triangle inequality and the fact that (see Eq. (B.213)):

$$-\# [\check{\chi}_{\text{SN},k}] \leq \text{Tr} [\hat{O} \check{\chi}_{\text{SN},k}(\rho_0)] \leq \# [\check{\chi}_{\text{SN},k}]. \quad (\text{B.217})$$

This holds because each of the sub-maps in the sum that makes up $\check{\chi}_{\text{SN},k}(\rho_0)$ is a CPTP map (see Eq. (B.209)), which means when it acts on ρ_0 (a physical state) it returns a physical state. Since the modulus of the expectation of a Pauli operator with respect to a physical state (ρ) is bounded by unity:

$$-1 \leq \text{Tr} [\rho P_i] \leq 1, \quad (\text{B.218})$$

the maximum magnitude of the expectation with respect to $\check{\chi}_{\text{SN},k}(\rho_0)$ occurs when every sub-map expectation has the same sign. In this case we simply count the sub-maps (i.e. use the multiplicity) to find the bounds for the expectation with respect to $\check{\chi}_{\text{SN},k}(\rho_0)$. We define the upper bound to the bias in Eq. (B.216) as the proxy bias:

$$\varepsilon_{\heartsuit} = 2 \left(1 - (1-x)^{N_G} \right), \quad (\text{B.219})$$

which is independent of the structure of the gates (U_n) within the circuit, depending only on how many error channels there are and what their noise amplitude is.

We define the noise level (e_G) of the circuit as the first order expansion of the proxy bias in the small noise ($xN_G \ll 1$) regime:

$$e_G = 2xN_G. \quad (\text{B.220})$$

²⁷We choose a Pauli word operator for our proxy bias observable to simplify the calculation but this is not really necessary. In fact, all we need is an operator with an expectation (on a physical state) bounded between -1 and 1 . Since rescaling and shifting an observable does not change its fractional error, we can obtain a bound for any observable simply by rescaling the proxy bias given here.

Appendix

B.5 Mitigation: IISM:NA

We will now work through the mitigation method evaluation procedure for IISM:NA (noise-aware identity-insertion-assisted synchronous mitigation). This method uses knowledge about the noise model to effect the mitigation. It makes use of identity insertions to amplify the errors for each gate in the circuit. It is a synchronous mitigation strategy so it amplifies the errors of each gate by the same amount for each circuit variant.

B.5.1 Identity Insertion

Identity insertion involves resolving the (noisy) identity before every noisy gate, in terms of the noisy gate and its inverse. So:

$$\begin{aligned} \boxed{\tilde{U}_n}_i &= \left(\boxed{\tilde{U}_n} \boxed{U_n^\dagger} \right)^i \boxed{\tilde{U}_n} = \left(\boxed{U_n} \boxed{\mathcal{E}} \boxed{U_n^\dagger} \boxed{\mathcal{E}} \right)^i \boxed{U_n} \boxed{\mathcal{E}} \\ &= \boxed{U_n} \left(\boxed{\mathcal{E}} \right)^{2i+1}, \end{aligned} \quad (\text{B.221})$$

is the effective noisy gate obtained after i identity insertions for the n th noisy gate in the circuit. Here we have assumed that the noisy gate and its noisy inverse have the same noise channel, which commutes with them both:

$$\boxed{\tilde{U}_n} = \boxed{U_n} \boxed{\mathcal{E}} = \boxed{\mathcal{E}} \boxed{U_n}, \quad \boxed{\tilde{U}_n^\dagger} = \boxed{U_n^\dagger} \boxed{\mathcal{E}} = \boxed{\mathcal{E}} \boxed{U_n^\dagger}. \quad (\text{B.222})$$

So the effective noise channel of the noise-amplified gate $\tilde{U}_{n,i}$, after i identity insertions is $\mathcal{E}_i$, where (see Eqs (B.199) and (B.221)):

$$\boxed{\mathcal{E}}_i = \left(\boxed{\mathcal{E}} \right)^{2i+1} = \begin{cases} \text{---} & \text{with probability } 1 - x_i, \\ \boxed{Z} & \text{with probability } x_i, \end{cases} \quad (\text{B.223})$$

where (see Eqs (B.200) and (B.206)):

$$x_i = \frac{1}{2} \left(1 - (1 - 2x)^{2i+1} \right), \quad (\text{B.224})$$

is the amplified noise amplitude, where $x_0 = x$ is the original noise amplitude. So, to leading order, the noise amplitude of the i th circuit is $2i + 1$ times the original noise amplitude:

$$x_i = (2i + 1)x + \mathcal{O}_0(x^2). \quad (\text{B.225})$$

B.5.2 Zero-Noise Extrapolation

Zero-noise extrapolation or Richardson extrapolation [16] is a mitigation method that involves finding a weighted sum of noisy observables that is equivalent to the noise-free observable up to some power in the noise amplitude:

$$\langle \tilde{O} \rangle_M = \sum_{i=0}^{N_M} c_{\text{GI};i} \langle \tilde{O} \rangle_i = \langle \hat{O} \rangle + \mathcal{O}_0(x^{N_M+1}), \quad (\text{B.226})$$

Appendix

where (see Eq. (B.215)):

$$\langle \tilde{O} \rangle_M = \text{Tr} \left[\hat{O} \sum_{i=0}^{N_M} c_{\text{Gl};i} \tilde{\mathcal{C}}_i(\rho_0) \right], \quad \langle \tilde{O} \rangle_i = \text{Tr} \left[\hat{O} \tilde{\mathcal{C}}_i(\rho_0) \right], \quad \langle \hat{O} \rangle = \text{Tr} \left[\hat{O} \mathcal{C}(\rho_0) \right], \quad (\text{B.227})$$

are the mitigated observable, i th noisy observable, and noise-free observable, respectively; and N_M is the order of the mitigation and the number of extra noisy circuit variants required for the mitigation. Each of the noisy circuit variants should be equivalent to the original circuit except for having an amplified noise amplitude.

We Taylor expand our noisy observables in terms of the noise amplitude:

$$\langle \tilde{O} \rangle_i = \langle \tilde{O} \rangle(x_i) = \sum_{m=0}^{\infty} d_m x_i^m = \langle \hat{O} \rangle + \mathcal{O}_0(x_i), \quad (\text{B.228})$$

where d_m are coefficients independent of x_i . For dephasing noise we have (Eq. (B.208)):

$$\langle \tilde{O} \rangle_i = \text{Tr} \left[\hat{O} \tilde{\mathcal{C}}_i(\rho_0) \right] = \sum_{k=0}^{N_G} (1 - x_i)^{N_G - k} x^k \text{Tr} \left[\hat{O} \check{\chi}_{\text{SN},k}(\rho_0) \right], \quad (\text{B.229})$$

and therefore:

$$d_m = \begin{cases} \sum_{k=0}^m \frac{(N_G - k)!}{(N_G - m)! (m - k)!} (-1)^{m-k} \text{Tr} \left[\hat{O} \check{\chi}_{\text{SN},k}(\rho_0) \right] & \text{for } m \leq N_G, \\ 0 & \text{for } m > N_G. \end{cases} \quad (\text{B.230})$$

This means Eq. (B.226) is equivalent to:

$$\langle \tilde{O} \rangle_M = \sum_{i=0}^{N_M} c_{\text{Gl};i} \langle \tilde{O} \rangle_i = \sum_{m=0}^{\infty} d_m \sum_{i=0}^{N_M} c_{\text{Gl};i} x_i^m = \langle \hat{O} \rangle + \mathcal{O}_0(x^{N_M+1}), \quad (\text{B.231})$$

or equivalently:

$$\sum_{i=0}^{N_M} c_{\text{Gl};i} = 1, \quad \sum_{i=0}^{N_M} c_{\text{Gl};i} x_i^k = x^k \sum_{i=0}^{N_M} c_{\text{Gl};i} \left(\frac{x_i}{x} \right)^k = 0 \text{ for } 1 \leq k \leq N_M. \quad (\text{B.232})$$

So the coefficients depend only on the relative quantity x_i/x and not on the absolute value of the noise amplitude. If we solve Eq. (B.232) we find (see Ref. [34]):

$$c_{\text{Gl};i} = \prod_{m=0: m \neq i}^{N_M} \frac{x_m}{x_m - x_i}, \quad \sum_{i=0}^{N_M} c_{\text{Gl};i} x_i^{N_M+1} = (-1)^{N_M} \prod_{n=0}^{N_M} x_n. \quad (\text{B.233})$$

Thus (Eq. (B.231)):

$$\langle \tilde{O} \rangle_M = \langle \hat{O} \rangle + (-1)^{N_M} \left(\prod_{n=0}^{N_M} x_n \right) d_{N_M+1} + \mathcal{O}_0(x^{N_M+2}). \quad (\text{B.234})$$

Appendix

B.5.3 Bias, Proxy Bias, and Noise Level after Mitigation

The bias is given by (see Eqs (B.213), (B.214), (B.217), and (B.230)):

$$\begin{aligned}
\varepsilon_o &= \left| \langle \tilde{\mathcal{O}} \rangle_M - \langle \hat{\mathcal{O}} \rangle \right| = \left(\prod_{n=0}^{N_M} x_n \right) |d_{N_M+1}| + \mathcal{O}_0(x^{N_M+2}) \\
&= \left(\prod_{n=0}^{N_M} x_n \right) \sum_{k=0}^{N_M+1} \frac{(N_G - k)!}{(N_G - N_M - 1)! (N_M + 1 - k)!} \left| \text{Tr} \left[\hat{\mathcal{O}}_{\check{\chi}_{SN,k}}(\rho_0) \right] \right| + \mathcal{O}_0(x^{N_M+2}) \\
&\leq \left(\prod_{n=0}^{N_M} x_n \right) \sum_{k=0}^{N_M+1} \frac{N_G!}{(N_G - N_M - 1)! (N_M + 1 - k)! k!} + \mathcal{O}_0(x^{N_M+2}) \\
&= 2^{N_M+1} \left(\prod_{n=0}^{N_M} x_n \right) \frac{N_G!}{(N_G - N_M - 1)! (N_M + 1)!} + \mathcal{O}_0(x^{N_M+2}), \tag{B.235}
\end{aligned}$$

where we have assumed that the observable is a Pauli word operator ($\hat{\mathcal{O}} = P_i$), as in App. B.4.4, and therefore the modulus of its expectation for any physical operator is bounded by unity. We have also assumed that $N_M < N_G$, if $N_M \geq N_G$ then the mitigation is unbiased. So the mitigated proxy bias (an upper bound for the bias) can be given by:

$$\varepsilon_{\heartsuit;M} = 2^{N_M+1} \left(\prod_{n=0}^{N_M} (2i + 1) \right) \frac{N_G!}{(N_G - N_M - 1)! (N_M + 1)!} x^{N_M+1} + \mathcal{O}_0(x^{N_M+2}), \tag{B.236}$$

where we have used the expression for the noise amplitude of the i th circuit variant given in Eq. (B.225). Since we have:

$$\frac{N_G!}{(N_M + 1)! (N_G - N_M - 1)!} = \frac{N_G^{N_M+1}}{(N_M + 1)!} + \mathcal{O}_L(N_G^{N_M}), \tag{B.237}$$

in the large circuit ($N_G \ll 1$) and small noise ($N_G x \ll 1$) regime we have:

$$\varepsilon_{\heartsuit;M} = 2^{N_M+1} \left(\prod_{n=0}^{N_M} (2i + 1) \right) \frac{(N_G x)^{N_M+1}}{(N_M + 1)!} + \mathcal{O}_{LS}(N_G^{N_M} x^{N_M+1}, N_G^{N_M+2} x^{N_M+2}), \tag{B.238}$$

where we have used Eqs (B.230) and (B.231) in order to determine the correct scaling with N_G and x . This means the noise level (after mitigation) is given by (Eq. (B.236)):

$$e_{G;M} = \left(\prod_{n=0}^{N_M} (2i + 1) \right) \frac{N_G!}{N_G^{N_M+1} (N_G - N_M - 1)! (N_M + 1)!} e_G^{N_M+1}, \tag{B.239}$$

in the small noise regime ($N_G x \ll 1$) and²⁸:

$$e_{G;M} = \frac{(2N_M + 1)!!}{(N_M + 1)!} e_G^{N_M+1} < \frac{(2N_M + 2)!!}{2(N_M + 1)!} e_G^{N_M+1} = \frac{(2e_G)^{N_M+1}}{2}, \tag{B.240}$$

in the small noise and large circuit regime ($xN_G, N_G^{-1} \ll 1$). Where $!!$ is the double factorial, defined by:

$$n!! = n(n-2)!!, \quad 1!! = 0!! = 1, \tag{B.241}$$

and $e_G = 2xN_G$ is defined in Eq. (B.220).

²⁸The ratio of the value used to bound the mitigated noise level in Eq. (B.240) to the true mitigated noise level is roughly $\sqrt{\pi(N_M + 1)}/2$ (this value has been obtained numerically) for large N_M .

Appendix

B.5.4 Sampling Cost, Length Scale Factor, and Runtime Scaling

The most efficient way to implement our mitigation procedure (Eq. (B.231)) is to use Monte-Carlo sampling [14]. To implement this we define a probability ($q_{\text{Gl};i}$) and sign ($s_{\text{Gl};i}$) associated with each of circuit variants, so the i th circuit has:

$$q_{\text{Gl};i} = \frac{|c_{\text{Gl};i}|}{C_{\text{Gl}}}, \quad s_{\text{Gl};i} = \frac{c_{\text{Gl};i}}{|c_{\text{Gl};i}|}, \quad C_{\text{Gl}} = \sum_{i=0}^{N_{\text{M}}} |c_{\text{Gl};i}|, \quad (\text{B.242})$$

where C_{Gl} is known as the sampling cost [36], for reasons that will become apparent. We simply make N_{R} runs of our circuit and each time we randomly choose one of our $N_{\text{M}} + 1$ circuits to implement and multiply the resulting output by the sign $s_{\text{Gl};i}$ corresponding to the circuit i that we choose for that run. We sample circuit i with probability $q_{\text{Gl};i}$. The expectation of this procedure is $\langle \tilde{O} \rangle_{\text{M}} / C_{\text{Gl}}$ so we simply multiply our expectation by the sampling cost in order to obtain our mitigated value. The variance of the Monte Carlo sampling procedure is given by the shot noise:

$$\sigma_{\text{MC}}^2 \propto \frac{1}{N_{\text{R}}}. \quad (\text{B.243})$$

However, because we multiply our final result by C_{Gl} we will increase the variance by a factor of C_{Gl}^2 . This means the ratio of the number of runs needed for the mitigated circuit to the number of runs needed for the unmitigated circuit, to yield the same variance is approximately [14]:

$$\frac{N_{\text{R,M}}}{N_{\text{R}}} \approx C_{\text{Gl}}^2. \quad (\text{B.244})$$

So C_{Gl} is the (square root of the) cost for sampling quasi-stochastic ensembles, if $c_{\text{Gl};i} > 0$ for all i then $C_{\text{Gl}} = 1$ (see Eq. (B.232)). For the coefficients given in Eq. (B.233) (see Eq. (B.225)):

$$c_{\text{Gl};i} = \prod_{m=0:m \neq i}^{N_{\text{M}}} \frac{2m+1 + \mathcal{O}_0(x)}{2(m-i) + \mathcal{O}_0(x)} = \frac{(2N_{\text{M}}+1)!!}{2^{N_{\text{M}}} (2i+1) (i!) ((N_{\text{M}}-i)!)} + \mathcal{O}_0(x). \quad (\text{B.245})$$

So the sampling cost for using IISM:NA to mitigate dephasing noise is given by²⁹:

$$C_{\text{Gl}} = \sum_{i=0}^{N_{\text{M}}} |c_{\text{Gl};i}| = \frac{(2N_{\text{M}}+1)!! {}_2F_1\left(\frac{1}{2}, -N_{\text{M}}; \frac{3}{2}; -1\right)}{2^{N_{\text{M}}} (N_{\text{M}}!)} + \mathcal{O}_0(x), \quad (\text{B.246})$$

where $!!$ is the double factorial,

$${}_2F_1(a, b; c; z) = \sum_{k=0}^{\infty} \frac{\Gamma(a+k)}{\Gamma(a)} \frac{\Gamma(b+k)}{\Gamma(b)} \frac{\Gamma(c)}{\Gamma(c+k)} \frac{z^k}{k!}, \quad (\text{B.247})$$

is a Hypergeometric function and:

$$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt, \quad (\text{B.248})$$

is the Euler gamma function (see [43]).

²⁹The sampling cost sum (see Eqs (B.245) and (B.246)) was calculated using Mathematica [43].

Appendix

As well as increasing the number of samples required, each run will take longer, on average, to implement because we are increasing the number of gates we use. The i th circuit has $2i + 1$ times the gates of the original noisy circuit and so will take $2i + 1$ times longer (neglecting initialisation and measurement time). Thus the average runtime of the mitigated circuit ($\tau_{E,M}$) relative to the original noisy circuit (τ_E) is:

$$F_{G1,L} = \frac{\tau_{E,M}}{\tau_E} = \sum_{i=0}^{N_M} q_{G1,i} (2i + 1) = \frac{\sum_{i=0}^{N_M} |c_{G1,i}| (2i + 1)}{C_{G1}}, \quad (B.249)$$

which we denote the length scale factor (again neglecting initialisation and measurement time). So the length scale factor for using IISM:NA to mitigate dephasing noise is given by:

$$\begin{aligned} F_{G1,L} &= \frac{(2N_M + 1)!!}{2^{N_M} C_{G1}} \sum_{i=0}^{N_M} \frac{1}{(i!) ((N_M - i)!)} + \mathcal{O}_0(x) = \frac{(2N_M + 1)!!}{N_M! C_{G1}} + \mathcal{O}_0(x) \\ &= \frac{2^{N_M}}{{}_2F_1\left(\frac{1}{2}, -N_M; \frac{3}{2}; -1\right)} + \mathcal{O}_0(x). \end{aligned} \quad (B.250)$$

Thus the ratio of the total mitigation procedure runtime (t_M) compared to that for the original noisy circuit (t_0) to achieve the same variance is given by:

$$S_{G1} = \frac{t_M}{t_0} \approx C_{G1}^2 F_{G1,L} = \frac{((2N_M + 1)!!)^2 {}_2F_1\left(\frac{1}{2}, -N_M; \frac{3}{2}; -1\right)}{2^{N_M} (N_M!)^2} + \mathcal{O}_0(x), \quad (B.251)$$

which we denote the runtime scaling (again we assume negligible initialisation and measurement times). The width scale factor doesn't feature here because we use the same number of qubits as the original noisy circuit.

B.5.5 Noise Boundary

The noise boundary is the maximum noise that can be mitigated for a fixed maximum runtime scaling ($S_{G1} = S_*$) and target bias ($\varepsilon_O = \varepsilon_*$). To ensure we do not exceed the bias we choose an order of mitigation (N_M) that will reduce the proxy bias to below the target bias (ε_*). To make our analytical analysis more straightforward we will use the mitigated noise level bound derived in Eq. (B.240) rather than the full proxy bias. So we set the target bias (ε_*) equal to the mitigated noise level bound and then invert the expression to find the required mitigation order (N_M):

$$\varepsilon_* = \frac{(2e_G)^{N_M+1}}{2} \implies N_M = \frac{\ln(2\varepsilon_*)}{\ln(2e_G)} - 1. \quad (B.252)$$

Now, we want to find the largest possible noise level (e_G) that we can mitigate so we will assume that $N_M \gg 1$, i.e. our target bias is much smaller than our original noise level. In this case we can use the fact that:

$${}_2F_1\left(\frac{1}{2}, -N_M; \frac{3}{2}; -1\right) = \frac{2^{N_M}}{N_M} \left(1 + \mathcal{O}\left(\frac{1}{N_M^2}\right)\right), \quad (B.253)$$

$$\frac{(2N_M + 1)!!}{2^{N_M} (N_M!)} = 2\sqrt{\frac{N_M}{\pi}} \left(1 + \frac{3}{8N_M} + \mathcal{O}\left(\frac{1}{N_M^2}\right)\right). \quad (B.254)$$

Appendix

(found by numerical and analytical analysis) to simplify the form of our sampling cost and length scale factor (Eqs (B.246) and (B.250)):

$$C_{\text{Gl}} = \frac{(2N_{\text{M}} + 1)!! {}_2F_1\left(\frac{1}{2}, -N_{\text{M}}; \frac{3}{2}; -1\right)}{2^{N_{\text{M}}} (N_{\text{M}}!)} + \mathcal{O}_0(x) = \frac{2^{N_{\text{M}}+1}}{\sqrt{N_{\text{M}}}\pi} \left(1 + \mathcal{O}\left(\frac{1}{N_{\text{M}}}\right)\right) + \mathcal{O}_0(x), \quad (\text{B.255})$$

$$F_{\text{Gl,L}} = \frac{2^{N_{\text{M}}}}{{}_2F_1\left(\frac{1}{2}, -N_{\text{M}}; \frac{3}{2}; -1\right)} + \mathcal{O}_0(x) = N_{\text{M}} \left(1 - \mathcal{O}\left(\frac{1}{N_{\text{M}}^2}\right)\right) + \mathcal{O}_0(x), \quad (\text{B.256})$$

which means the runtime scaling is given by (Eq. (B.251)):

$$S_{\text{Gl}} \approx C_{\text{Gl}}^2 F_{\text{Gl,L}} = \frac{4^{N_{\text{M}}+1}}{\pi} \left(1 + \mathcal{O}\left(\frac{1}{N_{\text{M}}}\right)\right) + \mathcal{O}_0(x). \quad (\text{B.257})$$

to leading order. Now, we set our runtime scaling equal to our maximum runtime scaling (S_{Φ}) and substitute our mitigation order from Eq. (B.252), then rearrange to determine the maximum noise level we can mitigate:

$$S_{\Phi} = \frac{4^{\frac{\ln(2\varepsilon_{\Phi})}{\ln(2e_{\text{G}})}}}{\pi} \implies e_{\text{G}} = 2^{\frac{2^{\frac{\ln(2\varepsilon_{\Phi})}{\ln(\pi S_{\Phi})}} - 1}{1}}. \quad (\text{B.258})$$

So the noise boundary is given by:

$$e_{\text{B}} = 2^{\frac{2^{\frac{\ln(2\varepsilon_{\Phi})}{\ln(\pi S_{\Phi})}} - 1}{1}} = \frac{(2\varepsilon_{\Phi})^{\frac{\ln(4)}{\ln(\pi S_{\Phi})}}}{2}. \quad (\text{B.259})$$

So we require a runtime scaling greater than $4/\pi$ to achieve any mitigation at all and the larger our tolerated runtime scaling the larger noise boundary we can tolerate. However large our tolerated runtime scaling is, the noise boundary will always be smaller than 0.5.

B.5.6 Robustness

A method is robust if the runtime scaling and the proxy bias are unchanged (to leading order) if there are some imperfections in the noise amplification procedure. For identity insertions (App. B.5.1), the source of imperfections would be differences between the noise model of the inverse gate and the original noisy gate, i.e. we have (c.f. Eq. (B.222)):

$$\boxed{\widetilde{U_n}} = \boxed{U_n} \boxed{\varepsilon} = \boxed{\varepsilon} \boxed{U_n}, \quad \boxed{\widetilde{U_n^\dagger}} = \boxed{U_n^\dagger} \boxed{\varepsilon_\dagger} = \boxed{\varepsilon_\dagger} \boxed{U_n^\dagger}. \quad (\text{B.260})$$

where:

$$\boxed{\varepsilon} = \begin{cases} \text{---} & \text{with prob. } 1 - x \\ \boxed{Z} & \text{with prob. } x \end{cases}, \quad \boxed{\varepsilon_\dagger} = \begin{cases} \text{---} & \text{with prob. } 1 - y_\dagger x \\ \boxed{Z} & \text{with prob. } y_\dagger x. \end{cases} \quad (\text{B.261})$$

In this case the amplified gate noise amplitudes are (see Eqs (B.200) and (B.206)):

$$x_i = \frac{1 - (1 - 2(1 + y_\dagger)x + 4y_\dagger x^2)^i (1 - 2x)}{2} = (1 + i(1 + y_\dagger))x + \mathcal{O}_0(x^2). \quad (\text{B.262})$$

Appendix

This means the coefficients are given by (Eq. (B.233)):

$$c_{\text{GI};i} = \prod_{m=0:m \neq i}^{N_M} \frac{x_m}{x_m - x_i} = (-1)^i \frac{\prod_{m=0}^{N_M} (1 + m(1 + y_{\dagger}))}{(1 + y_{\dagger})^{N_M} (1 + i(1 + y_{\dagger})) i! (N_M - i)!} + \mathcal{O}_0(x). \quad (\text{B.263})$$

This means the sampling cost is given by (Eq. (B.246)):

$$C_{\text{GI}} = \sum_{j=0}^{N_M} |c_{\text{GI};j}| = \frac{\prod_{m=0}^{N_C} (1 + m(1 + y_{\dagger})) {}_2F_1\left(\frac{1}{1+y_{\dagger}}, -N_M; \frac{y_{\dagger}+2}{y_{\dagger}+1}; -1\right)}{(1 + y_{\dagger})^{N_M} (N_M!)} + \mathcal{O}_0(x), \quad (\text{B.264})$$

and the length scale factor by (Eq. (B.249)):

$$F_{\text{GI}} = \frac{\sum_{i=0}^{N_M} |c_{\text{GI};i}| (2i + 1)}{C_{\text{GI}}} = 1 + 2N_M \frac{{}_2F_1\left(\frac{2+y_{\dagger}}{1+y_{\dagger}}, 1 - N_M; \frac{2y_{\dagger}+3}{y_{\dagger}+1}; -1\right)}{(y_{\dagger} + 2) {}_2F_1\left(\frac{1}{1+y_{\dagger}}, -N_M; \frac{y_{\dagger}+2}{y_{\dagger}+1}; -1\right)} + \mathcal{O}_0(x). \quad (\text{B.265})$$

The proxy bias is given by (see Eqs (B.235) and (B.262)):

$$\begin{aligned} \varepsilon_{\heartsuit;M} &= 2^{N_M+1} \left(\prod_{n=0}^{N_M} x_n \right) \frac{N_G!}{(N_G - N_M - 1)! (N_M + 1)!} + \mathcal{O}_0(x^{N_M+2}) \\ &= \frac{\prod_{n=0}^{N_M} (2N_G x_n)}{(N_M + 1)!} + \mathcal{O}_{\text{LS}}\left(N_G^{N_M} x^{N_M+1}, N_G^{N_M+2} x^{N_M+2}\right) \\ &= (2N_G x)^{N_M+1} \frac{\prod_{n=0}^{N_M} ((1 + n(1 + y_{\dagger})))}{(N_M + 1)!} + \mathcal{O}_{\text{LS}}\left(N_G^{N_M} x^{N_M+1}, N_G^{N_M+2} x^{N_M+2}\right). \end{aligned} \quad (\text{B.266})$$

So the mitigated noise level is given by:

$$e_{\text{G};M} = e_G^{N_M+1} \frac{\prod_{n=0}^{N_M} ((1 + n(1 + y_{\dagger})))}{(N_M + 1)!}. \quad (\text{B.267})$$

Thus we see that the imperfection in the inverse noisy gate's noise channel will change the sampling cost and length scale factor, and therefore the runtime scaling; and also the proxy bias and therefore the mitigated noise level. This means that IISM:NA is not robust but it is quasi-robust. Since, though the metrics depend on the value of $y_{\dagger}$ to leading order the order of the leading order term is independent of $y_{\dagger}$, i.e. it is still zeroth order in the noise amplitude (x) for the sampling cost and the length scale factor and $N_M + 1$ th order in the noise level (e_G) for the proxy bias. Here we have assumed that, despite the imperfect control, we are able to achieve perfect characterisation. So, we cannot generate noisy gate variants precisely but we know the impact of the errors that we make during the generation procedure.

B.5.7 Qualitative Certification

Now we have all the information we need (see Eqs (B.239), (B.257), and (B.259)) to provide a qualitative evaluation of IISM:NA. We will assess each of the certification criteria for mitigating dephasing noise with IISM:NA. We include the proxy bias ($\varepsilon_{\heartsuit;M}$, Eqs (B.238)

Appendix

and (B.240)), runtime scaling (S_{GI} , Eqs (B.257), (B.257), (B.249), (B.246), (B.233), and (B.224)), and noise boundary (e_{B} , Eq. (B.259)) again for illustrative purposes.

$$\varepsilon_{\text{Q};\text{M}} < \frac{(2e_{\text{G}})^{N_{\text{M}}+1}}{2} + \mathcal{O}_{\text{LS}} \left(\frac{e_{\text{G}}^{N_{\text{M}}+1}}{N_{\text{G}}}, e_{\text{G}}^{N_{\text{M}}+2} \right), \quad (\text{B.268})$$

$$S_{\text{GI}} \approx C_{\text{GI}}^2 F_{\text{GI},\text{L}} = \frac{4^{N_{\text{M}}+1}}{\pi} \left(1 + \mathcal{O} \left(\frac{1}{N_{\text{M}}} \right) \right) + \mathcal{O}_0(x), \quad F_{\text{GI},\text{L}} = \frac{\sum_{i=0}^{N_{\text{M}}} |c_{\text{GI};i}| (2i+1)}{C_{\text{GI}}},$$

$$C_{\text{GI}} = \sum_{i=0}^{N_{\text{M}}} |c_{\text{GI};i}|, \quad c_{\text{GI};i} = \prod_{m=0:m \neq i}^{N_{\text{M}}} \frac{\frac{1}{2} \left(1 - (1-2x)^{2m+1} \right)}{\frac{1}{2} \left(1 - (1-2x)^{2m+1} \right) - \frac{1}{2} \left(1 - (1-2x)^{2i+1} \right)}, \quad (\text{B.269})$$

$$e_{\text{B}} = 2^{\frac{2 \ln(2\varepsilon_{\text{+}})}{\ln(\pi S_{\text{+}})} - 1} = \frac{(2\varepsilon_{\text{+}})^{\frac{\ln(4)}{\ln(\pi S_{\text{+}})}}}{2}. \quad (\text{B.270})$$

- **Scalable:** Yes (see Eq. (B.269)); the coefficients of the mitigation are independent of the number of gates in the circuit (see Eq. (B.233)), which means so is the sampling cost (Eq. (B.246)). The length scale factor is also independent of the number of gates in the circuit as it depends only on the coefficients, the sampling cost, and the number of identity insertions (Eq. (B.249)). Since the runtime scaling depends only on the length scale factor and the sampling cost, it is independent of the number of gates in the circuit and therefore the mitigation method is scalable.
- **Unbounded:** No (see Eq. (B.270)); the maximum noise level we can mitigate is 0.5, no matter how large we allow the runtime scaling to be so we cannot mitigate arbitrarily large noise levels.
- **Precise:** Yes (see Eq. (B.268)); we can make the proxy bias (and therefore the bias) as small as we like by increasing N_{M} (the proxy bias is exponentially suppressed), provided that $e_{\text{G}} < 0.5$. So we can achieve arbitrary accuracy with this method if we are willing to pay the price (of a long runtime).
- **Efficient:** No (see Eq. (B.269)); even if the noise level (and therefore the noise amplitude, x) tends to zero the runtime scaling will not be unity. Even a first order ($N_{\text{M}} = 1$) mitigation has $S_{\text{GI}} = 6$ in the zero-noise limit ($x \rightarrow 0$) and all higher order mitigations would have a larger runtime scaling. So this method does not make good use of reductions in the noise level before mitigation.
- **Robust:** Yes (well quasi-robust, see App. B.5.6); the runtime scaling and the proxy bias are modified to leading order if the inverse noisy gate has a different noise amplitude to the original noisy gate but the order of the leading order term is unchanged. This would not be an issue for Hermitian gates as the inverse noisy gate is the same as the original noisy gate.

So IISM:NA is not SUPER but it is SPqR.

C Mitigation Methods

Here we will provide the parameters and usage directions for all the mitigation methods studied in this work. We will not provide the derivations of these results here but the full derivations can be found in [1] and a worked example of the derivation is given in App. B for IISM:NA mitigating dephasing noise.

Appendix

C.1 Noise Models

We consider three types of noise:

- Stochastic noise:

$$\boxed{\mathcal{E}_n} = \begin{cases} \text{---} & \text{with probability } 1 - p \\ \boxed{\mathcal{N}_n} & \text{with probability } p \end{cases} \quad (\text{C.271})$$

where $\mathcal{E}_n$ is the error channel of the n th gate, $\mathcal{N}_n$ is a CPTP channel, and p is the noise amplitude. The noise level is given by:

$$e_G = e_{\text{SN}} = 2pN_G, \quad (\text{C.272})$$

where N_G is the number of gates in the circuit. $\mathcal{E}_n$ is assumed to be closed so:

$$\mathcal{N}_n^2 = a\mathcal{I} + (1 - a)\mathcal{N}_n, \quad 0 \leq a \leq 1, \quad (\text{C.273})$$

where a is determined by the precise structure of $\mathcal{N}_n$ (our numerical results use $a = 1$).

- Rotational errors:

$$\boxed{\mathcal{E}_n} = \boxed{e^{i\frac{\phi}{2}P_n}}, \quad (\text{C.274})$$

where $\mathcal{E}_n$ is the error channel of the n th gate, P_n is a Hermitian operator, and ϕ is the noise amplitude. The noise level is given by:

$$e_G = e_{\text{RE}} = |\phi| N_G. \quad (\text{C.275})$$

- (Stochastic) over-rotation errors:

$$\boxed{\mathcal{E}_n} = \boxed{e^{i\frac{\phi}{2}P_n}} \begin{cases} \text{---} & \text{with probability } 1 - p \\ \boxed{P_n} & \text{with probability } p \end{cases}, \quad (\text{C.276})$$

where $\mathcal{E}_n$ is the error channel of the n th gate, P_n is the Hermitian operator that generates the n th gate, and ϕ and p are the rotational and stochastic noise amplitudes respectively. The full noise level is given by (see Eqs (C.272) and (C.275)):

$$e_G = e_{\text{RE}} + e_{\text{SN}}. \quad (\text{C.277})$$

C.2 Linear Mitigation

Linear mitigation obtains a mitigated observable estimate from a linear combination of noisy observable estimates:

$$\langle \tilde{O} \rangle_{\text{M}} = \sum_{i=0}^{N_C} c_{\text{GI};i} \langle \tilde{O} \rangle_i = \sum_{i=0}^{N_C} c_{\text{GI};i} \text{Tr} \left[\hat{O} \tilde{\mathcal{C}}_{;i}(\rho_0) \right] = \text{Tr} \left[\hat{O} \tilde{\mathcal{C}}_{;\text{M}}(\rho_0) \right], \quad (\text{C.278})$$

where $\langle \tilde{O} \rangle_{\text{M}}$ is the mitigated observable, N_C is the number of extra circuit variants we use in the mitigation (generally $\tilde{\mathcal{C}}_0$ is the original noisy circuit), $c_{\text{GI};i}$ is the coefficient used for the i th noisy observable estimate (or equivalently the i th noisy circuit variant), $\hat{O}$ is the observable we wish to measure, $\tilde{\mathcal{C}}_{;i}$ is the (map corresponding to) the i th noisy circuit variant, ρ_0 is the initial state our circuit acts on, and $\tilde{\mathcal{C}}_{;\text{M}}$ is the effective mitigated circuit.

Appendix

C.2.1 Scope

Mitigation methods can have different scopes:

- Synchronous methods use the same variant generation strategy for each gate in a given variant circuit:

$$\boxed{\mathcal{C}}_{\text{SM}} = \left\{ c_{\text{Gl};j} : \boxed{\tilde{\mathcal{C}}}_j \right\}_{j=0}^{N_C} = \left\{ c_{\text{Gl};j} : \boxed{\tilde{U}_1}_{j_1(j)} \text{---} \boxed{\tilde{U}_2}_{j_2(j)} \text{---} \dots \text{---} \boxed{\tilde{U}_{N_G}}_{j_{N_G}(j)} \right\}_{j=0}^{N_C}, \quad (\text{C.279})$$

where $\mathcal{C}_{\text{SM}}$ is the (channel equivalent to the) synchronously mitigated circuit, $c_{\text{Gl};j}$ is the coefficient of the j th noisy circuit variant, $\tilde{\mathcal{C}}_j$ is the j th noisy circuit variant, N_C is the number of circuit variants used, $\tilde{U}_{n;j}$ is the j th variant of the n th gate.

- Asynchronous methods can use different variant generation strategies for each gate in a given variant circuit:

$$\boxed{\mathcal{C}}_{\text{AM}} = \left\{ c_{\text{Gl};j} : \boxed{\tilde{\mathcal{C}}}_j \right\}_{j=0}^{N_C} = \left\{ c_{\text{Gl};j} : \boxed{\tilde{U}_1}_{J_1(j)} \text{---} \boxed{\tilde{U}_2}_{J_2(j)} \text{---} \dots \text{---} \boxed{\tilde{U}_{N_G}}_{J_{N_G}(j)} \right\}_{j=0}^{N_C}, \quad (\text{C.280})$$

where $\mathcal{C}_{\text{AM}}$ is the (channel equivalent to the) asynchronously mitigated circuit, $c_{\text{Gl};j}$ is the coefficient of the j th noisy circuit variant, $\tilde{\mathcal{C}}_j$ is the j th noisy circuit variant, N_C is the number of circuit variants used, $\tilde{U}_{n;J_n(j)}$ is the $J_n(j)$ th variant of the n th gate. In this case J_n is some function of j that depends on n and it need not be the case that $J_n(j) = J_m(j)$ if $n \neq m$.

- Local methods effectively mitigate each gate independently and then these mitigated gates are concatenated in order to form circuit variants and their coefficients:

$$\begin{aligned} \boxed{\mathcal{C}}_{\text{LM}} &= \left\{ c_{\text{Gl};j} : \boxed{\tilde{\mathcal{C}}}_j \right\}_{j=0}^{N_C} = \left\{ c_{\text{Gl};j} : \boxed{\tilde{U}_1}_{J_1(j)} \text{---} \boxed{\tilde{U}_2}_{J_2(j)} \text{---} \dots \text{---} \boxed{\tilde{U}_{N_G}}_{J_{N_G}(j)} \right\}_{j=0}^{N_C}, \\ &= \boxed{\tilde{U}_1}_{\text{M}} \text{---} \boxed{\tilde{U}_2}_{\text{M}} \text{---} \dots \text{---} \boxed{\tilde{U}_{N_G}}_{\text{M}}, \quad \boxed{\tilde{U}_n}_{\text{M}} = - \left\{ c_{\text{L};n,i} : \boxed{\tilde{U}_n}_i \right\}_{i=0}^{N_{\text{L},\text{C},n}}, \end{aligned} \quad (\text{C.281})$$

where $\mathcal{C}_{\text{LM}}$ is the (channel equivalent to the) locally mitigated circuit; $c_{\text{Gl};j}$ is the coefficient of the j th noisy circuit variant; $\tilde{\mathcal{C}}_j$ is the j th noisy circuit variant; N_C is the number of circuit variants used; $\tilde{U}_{n;J_n(j)}$ is the $J_n(j)$ th variant of the n th gate. For local mitigation:

$$J_n(j) = \left\lfloor \frac{j}{N_{\text{C},n-1} + 1} \right\rfloor \bmod (N_{\text{L},\text{C},n} + 1), \quad N_{\text{C},n} = \left(\prod_{m=1}^n (N_{\text{L},\text{C},m} + 1) \right) - 1, \quad (\text{C.282})$$

where $N_{\text{C},n}$ is the number of circuits required to mitigate n gates ($N_C = N_{\text{C},N_G}$, where N_G is the number of gates in the circuit), $N_{\text{L},\text{C},n}$ is the number of noisy variants we use to mitigate the n th noisy gate, $\lfloor x/y \rfloor$ is the integer part of the division of x by y , and $x \bmod y$ is the remainder of x divided by y . $\tilde{U}_{n;\text{M}}$ is the mitigated version of the n th gate; $c_{\text{L};n,i}$ is the coefficient of the i th variant of the n th gate; and $\tilde{U}_{n;i}$ is the i th noisy variant of the n th gate. The global circuit coefficients are defined by:

$$c_{\text{Gl};j} = \prod_{n=1}^{N_G} c_{\text{L};n,J_n(j)}. \quad (\text{C.283})$$

Appendix

C.2.2 Mitigation Method Implementation

We implement our mitigation methods using Monte Carlo sampling [35].

For a global mitigation method we have a mitigated circuit given by:

$$\tilde{\mathcal{C}}_{\text{M}} = \left\{ c_{\text{Gl};j} : \tilde{\mathcal{C}}_j \right\}_{j=0}^{N_{\text{C}}}, \quad C_{\text{Gl}} = \sum_{i=0}^{N_{\text{C}}} |c_{\text{Gl};i}|, \quad q_{\text{Gl};j} = \frac{|c_{\text{Gl};j}|}{C_{\text{Gl}}}, \quad s_{\text{Gl};j} = \frac{c_{\text{Gl};j}}{|c_{\text{Gl};j}|}. \quad (\text{C.284})$$

where $\tilde{\mathcal{C}}_{\text{M}}$ is the effective mitigated circuit, $c_{\text{Gl};j}$ is the coefficient of the j th circuit variant ($\tilde{\mathcal{C}}_j$), C_{Gl} is the sampling cost, $q_{\text{Gl};j}$ is the probability we select the j th circuit for a given run of the mitigated circuit, and $s_{\text{Gl};j}$ is the sign associated with the j th circuit. To implement this mitigated circuit we make N_{R} runs of the circuit, for each run we select at random from the mitigation method's component circuit variants, selecting circuit j with probability $q_{\text{Gl};j}$. We multiply the output of each run by $s_{\text{Gl};j}$ (where j matches the circuit we implement for that run). We average all the N_{R} (signed) outputs we obtain and multiply the result by C_{Gl} to obtain our mitigated observable estimate.

For a local mitigation method we have a mitigated circuit given by:

$$\tilde{\mathcal{C}}_{\text{M}} = \left(\tilde{\mathcal{U}}_{\text{M}} \right)_{n=1}^{N_{\text{G}}}, \quad \tilde{\mathcal{U}}_{\text{M}} = \left\{ c_{\text{L};n,i} : \tilde{\mathcal{U}}_{n,i} \right\}_{i=0}^{N_{\text{L},\text{C},n}}, \quad C_{\text{L},n} = \sum_{i=0}^{N_{\text{L},\text{C},n}} |c_{\text{L};n,i}|, \\ q_{\text{L};n,j} = \frac{|c_{\text{L};n,j}|}{C_{\text{L},n}}, \quad s_{\text{L};n,j} = \frac{c_{\text{L};n,j}}{|c_{\text{L};n,j}|}, \quad C_{\text{Gl}} = \prod_{n=1}^{N_{\text{G}}} C_{\text{L},n}. \quad (\text{C.285})$$

where $\tilde{\mathcal{C}}_{\text{M}}$ is the effective mitigated circuit, $\tilde{\mathcal{U}}_{\text{M}}$ is the n th effective mitigated gate in the circuit, N_{G} is the number of gates in the circuit, $c_{\text{L};n,i}$ is the coefficient of the i th noisy gate variant of the n th gate ($\tilde{\mathcal{U}}_{n,i}$), $N_{\text{L},\text{C},n}$ is the number of gate variants we use for the n th gate, $C_{\text{L},n}$ is the local sampling cost of the n th gate, $q_{\text{L};n,j}$ is the probability we select the j th variant for the n th gate, $s_{\text{L};n,j}$ is the sign associated with the j th variant of the n th gate, and C_{Gl} is the global sampling cost. To implement this mitigated circuit we make N_{R} runs of the circuit. For each gate in each run we select at random from the mitigated gate's component gate variants, selecting variant j with probability $q_{\text{L};n,j}$ (for the n th gate). We multiply the output of each run by the product of the signs associated for each gate using in that run, so for the m th run we have:

$$\prod_{n=1}^{N_{\text{G}}} s_{\text{L};n,j_{n,m}}, \quad (\text{C.286})$$

where $j_{n,m}$ is the index of the gate variant selected for the n th gate for the m th run of the circuit. We take the average of the N_{R} signed outputs and multiply the result by C_{Gl} to obtain our mitigated observable estimate.

C.3 CHILM

Custom-channel-and-hidden-inverse-assisted local mitigation (CHILM) is a local method (see App. C.2.1) that uses a mixture of custom channels and hidden inverses to amplify the noise in the noisy gate variants:

$$\tilde{\mathcal{C}}_{\text{M}} = \left(\tilde{\mathcal{U}}_{\text{M}} \right)_{n=1}^{N_{\text{G}}}, \quad \tilde{\mathcal{U}}_{\text{M}} = \left\{ c_{\text{L};n,j} : \tilde{\mathcal{U}}_{n,j} \right\}_{j=0}^{N_{\text{L},\text{C},n}}, \quad (\text{C.287})$$

$$\tilde{\mathcal{U}}_{n,j} = \tilde{\mathcal{U}}_n \tilde{\mathcal{E}}_{\text{C},j}, \quad \tilde{\mathcal{U}}_{n,N_{\text{L},\text{C},n}} = \tilde{\mathcal{U}}_{\text{HI}}, \quad (\text{C.288})$$

Appendix

where $\tilde{\mathcal{C}}_{\text{M}}$ is the mitigated circuit, $\tilde{U}_{n;\text{M}}$ is n th mitigated gate, N_{G} is the number of gates in the circuit, $c_{\text{L};n,j}$ is the mitigation coefficient for the j th gate variant of the n th gate, $\tilde{U}_{n;j}$ is the j th gate variant of the n th gate, $N_{\text{L},\text{C},n}$ is the number of gate variants for the n th gate, $\tilde{U}_n$ is the n th noisy gate before mitigation, $\mathcal{E}_{n;\text{C},j}$ is the custom noise channel used to generate the j th variant of the n th noisy gate, and $\tilde{U}_{n;\text{HI}}$ is the hidden inverse of the n th noisy gate (used for the final of the $N_{\text{L},\text{C},n}$ noisy gate variants, not counting the original noisy gate).

CHILM is a special instantiation of probabilistic error cancellation (see Ref. [16]). The use of custom channels for variant generation is discussed in Ref. [23] and hidden inverses are introduced in Refs [21, 22] (but they are not used therein for probabilistic error cancellation).

The execution time for the i th variant of the n th gate is assumed to be:

$$\tau_{\text{L},n;i} = \begin{cases} \tau_{\text{L},n}, & \text{for } i = 0, \\ 2\tau_{\text{L},n}, & \text{for } 0 < i < N_{\text{L},\text{C},n}, \\ \tau_{\text{L},n}, & \text{for } i = N_{\text{L},\text{C},n}, \end{cases} \quad (\text{C.289})$$

where $\tau_{\text{L},n}$ is the execution time for the original n th noisy gate ($\tilde{U}_{n;0} = \tilde{U}_n$). That is, we assume that the execution time of every custom error channel is the same as the original gate and we assume the execution time of the hidden inverse is the same as the original noisy gate.

We do not implement CHILM for stochastic noise.

For rotational errors we use $N_{\text{L},\text{C},n} = 2$ where:

$$\begin{aligned} \boxed{\tilde{U}_n}_0 &= \boxed{\tilde{U}_n} = \boxed{U_n} \boxed{e^{i\frac{\phi}{2}P_n}}, & \boxed{\tilde{U}_n}_1 &= \boxed{\tilde{U}_n} \boxed{e^{i\frac{\pi}{2}P_n}}, \\ \boxed{\tilde{U}_n}_2 &= \boxed{\tilde{U}_n}_{\text{HI}} = \boxed{U_n} \boxed{e^{-i\frac{\phi}{2}P_n}}, \end{aligned} \quad (\text{C.290})$$

with the coefficients:

$$c_{\text{L};0} = \frac{1}{2}, \quad c_{\text{L};1} = -\frac{1 - \cos \phi}{2 \cos \phi} = -\frac{\phi^2}{4} + \mathcal{O}_0(\phi^4), \quad c_{\text{L};2} = \frac{1}{2 \cos \phi} = \frac{1}{2} + \frac{\phi^2}{4} + \mathcal{O}_0(\phi^4), \quad (\text{C.291})$$

where ϕ is the noise amplitude of the original noisy gate (see App. C.1). This method is unbiased so the proxy bias is zero.

For (stochastic) over-rotation errors we use the same variant generation strategy as for rotational errors ($N_{\text{L},\text{C},n} = 2$):

$$\begin{aligned} \boxed{\tilde{U}_n}_0 &= \boxed{\tilde{U}_n} = \boxed{U_n} \boxed{e^{i\frac{\phi}{2}P_n}} \begin{cases} \text{---} & \text{with prob. } 1 - p \\ \boxed{P_n} & \text{with prob. } p \end{cases}, \\ \boxed{\tilde{U}_n}_1 &= \boxed{\tilde{U}_n} \boxed{e^{i\frac{\pi}{2}P_n}}, \\ \boxed{\tilde{U}_n}_2 &= \boxed{\tilde{U}_n}_{\text{HI}} = \boxed{U_n} \boxed{e^{-i\frac{\phi}{2}P_n}} \begin{cases} \text{---} & \text{with prob. } 1 - p \\ \boxed{P_n} & \text{with prob. } p \end{cases}, \end{aligned} \quad (\text{C.292})$$

Appendix

and the coefficients:

$$c_{L;0} = \frac{1}{2}, \quad c_{L;1} = -\frac{1 - (1 - 2p) \cos \phi}{2(1 - 2p) \cos \phi} = -p - \frac{\phi^2}{4} + \mathcal{O}_0(p^2, \phi^2 p, \phi^4), \quad (\text{C.293})$$

$$c_{L;2} = \frac{1}{2(1 - 2p) \cos \phi} = \frac{1}{2} + p + \frac{\phi^2}{4} + \mathcal{O}_0(p^2, \phi^2 p, \phi^4), \quad (\text{C.294})$$

where p and ϕ are the noise amplitudes of the stochastic and rotational components of the original noisy gate's error channel, respectively (see App. C.1). This method is unbiased so the proxy bias is zero. It is worth noting that the hidden inverse used in the mitigation of stochastic over-rotation errors has a noise channel that is not the inverse of the noise channel of the original noisy gate. Only the coherent (rotational) part of the noise channel is inverted. The stochastic component is assumed to be the same as the original gate.

C.4 CHISM

Custom-channel-and-hidden-inverse-assisted synchronous mitigation (CHISM) is a synchronous method (see App. C.2.1) that uses a mixture of custom channels and hidden inverses to amplify the noise in the noisy gate variants:

$$\tilde{\mathcal{C}}_{\text{M}} = \left\{ c_{\text{GL};j} : \tilde{\mathcal{C}}_j \right\}_{j=0}^{N_C} = \left\{ c_{\text{GL};j} : \left(-\tilde{U}_n \right)_{n=1}^{N_G} \right\}_{j=0}^{N_C}, \quad (\text{C.295})$$

$$-\tilde{U}_n \Big|_{j < N_C} = -\tilde{U}_n - \mathcal{E}_{n;\text{C};j}, \quad -\tilde{U}_n \Big|_{N_C} = -\tilde{U}_n \Big|_{\text{HI}}, \quad (\text{C.296})$$

where $\tilde{\mathcal{C}}_{\text{M}}$ is the mitigated circuit, $c_{\text{GL};j}$ is the mitigation coefficient for the j th circuit variant, $\tilde{\mathcal{C}}_j$ is the j th circuit variant, N_C is the number of circuit variants, $\tilde{U}_{n;j}$ is the j th variant of the n th noisy gate, N_G is the number of gates in the circuit, $\tilde{U}_n$ is the n th noisy gate before mitigation, $\mathcal{E}_{n;\text{C};j}$ is the custom noise channel used to generate the j th variant of the n th noisy gate, and $\tilde{U}_n \Big|_{\text{HI}}$ is the hidden inverse of the n th noisy gate (used for the final of the N_C noisy gate variants, not counting the original noisy gate).

CHISM requires all the mitigated noisy gates to have noise of the same type with the same noise amplitude. CHISM is a generalisation of the zero-noise extrapolation methods of Refs [16, 23] to target rotational errors. The use of custom channels for variant generation is discussed in Ref. [23] and hidden inverses are introduced in Refs [21, 22].

The execution time for the j th circuit is given by:

$$\tau_{\text{E},j} = \begin{cases} \tau_{\text{E}}, & \text{for } i = 0, \\ 2\tau_{\text{E}}, & \text{for } 0 < i < N_C, \\ \tau_{\text{E}}, & \text{for } i = N_C, \end{cases} \quad (\text{C.297})$$

where τ_{E} is the execution time for the original noisy circuit ($\tilde{\mathcal{C}}_0 = \tilde{\mathcal{C}}$). We assume that the execution time of every custom error channel is the same as the original gate and that all the gates take the same time to implement. We also assume the execution time of the hidden inverse is the same as the original noisy gate.

We do not implement CHISM for stochastic noise.

For rotational errors we use the gate variants:

$$\begin{aligned} -\tilde{U}_n \Big|_0 &= -\tilde{U}_n = -U_n - e^{i\frac{\phi}{2}P_n}, & -\tilde{U}_n \Big|_{0 < i < N_C} &= -\tilde{U}_n - e^{i\frac{i\pi}{N_C}P_n}, \\ -\tilde{U}_n \Big|_{N_C} &= -\tilde{U}_n \Big|_{\text{HI}} = -U_n - e^{-i\frac{\phi}{2}P_n}, \end{aligned} \quad (\text{C.298})$$

Appendix

which means we have the amplified noise amplitudes:

$$\phi_0 = \phi, \quad \phi_{0 < i < N_C} = \phi + \frac{2i}{N_C} \pi, \quad \phi_{N_C} = -\phi. \quad (\text{C.299})$$

We use the coefficients:

$$c_{\text{GI};i} = \prod_{j=0, j \neq i}^{N_C} \frac{\sin \frac{\phi_j}{2}}{\sin \frac{\phi_j - \phi_i}{2}}, \quad (\text{C.300})$$

where ϕ is the noise amplitude of the original noisy gate (see App. C.1) and N_C is the number of gates used for the mitigation. We do not have an analytic form for the proxy bias in general but if we set $N_C = 2N_G$ (as we do in the main text) the method is unbiased.

C.5 CIILM

Custom-channel-and-identity-insertion-assisted local mitigation (CIILM) is a local method (see App. C.2.1) that uses custom channels and identity insertions to amplify the noise in the noisy gate variants:

$$\tilde{\mathcal{C}}_{\text{M}} = \left(\tilde{U}_{\text{M}} \right)_{n=1}^{N_G}, \quad \tilde{U}_{\text{M}} = \left\{ c_{\text{L};n,j} : \tilde{U}_{\text{M}} \right\}_{j=0}^{N_{\text{L},\text{C},n}}, \quad (\text{C.301})$$

$$\tilde{U}_{\text{M}} = \left(\tilde{U}_{\text{T}} \tilde{U}_{\text{T}}^\dagger \right)^j \tilde{U}_{\text{T}}, \quad \tilde{U}_{\text{T}} = \tilde{U}_{\text{M}} \mathcal{E}_{\text{C}}, \quad \tilde{U}_{\text{T}}^\dagger = \tilde{U}_{\text{M}}^\dagger \mathcal{E}_{\text{C}}, \quad (\text{C.302})$$

where $\tilde{\mathcal{C}}_{\text{M}}$ is the mitigated circuit, $\tilde{U}_{\text{M}}$ is n th mitigated gate, N_G is the number of gates in the circuit, $c_{\text{L};n,j}$ is the mitigation coefficient for the j th gate variant of the n th gate, $\tilde{U}_{\text{M};j}$ is the j th gate variant of the n th gate, $N_{\text{L},\text{C},n}$ is the number of gate variants for the n th gate, $\tilde{U}_{\text{T}}$ is the tailored version of the n th noisy gate, $\tilde{U}_{\text{T}}^\dagger$ is the tailored version of the noisy inverse of the n th gate (assumed to have the same error channel as $\tilde{U}_{\text{T}}$), $\tilde{U}_{\text{M}}$ is the n th noisy gate before mitigation, $\mathcal{E}_{\text{C}}$ is the custom noise channel used to tailor the n th noisy gate, and $\tilde{U}_{\text{M}}^\dagger$ is the noisy inverse of the n th noisy gate. To save time in characterisation one need not ever characterise the gates $\tilde{U}_{\text{M}}$ and $\tilde{U}_{\text{M}}^\dagger$ since the control channel can be chosen independently of the noise amplitude. It suffices to characterise $\tilde{U}_{\text{T}}$ and $\tilde{U}_{\text{T}}^\dagger$.

The version of CIILM presented here requires the error channels associated with the noisy gate and its noisy inverse to be the same and that the noisy gate, the noisy inverse, their error channels, and the custom error channel all commute with each other. CIILM could be considered to be an example of NEPEC (see Ref. [18]) a variant of probabilistic error cancellation (see Ref. [16]). The use of custom channels for noise amplification is discussed in Ref. [23].

The execution time for the i th variant of the n th gate is assumed to be:

$$\tau_{\text{L},n;i} = 2(2i + 1) \tau_{\text{L},n}, \quad (\text{C.303})$$

where $\tau_{\text{L},n}$ is the execution time for the original n th noisy gate ($\tilde{U}_{\text{M}}$). That is, we assume that the execution time of every custom error channel is the same as the original gate and that the forward and inverse gates have the same execution time.

We do not implement CIILM for stochastic noise.

Appendix

For rotational errors we use $N_{L,C,n} = 2$ and the custom channel is given by:

$$\boxed{\mathcal{E}_n}_C = \boxed{e^{i\frac{\pi}{3}P_n}}, \quad (\text{C.304})$$

where P_n is the generator of the rotational error of the n th noisy gate. The coefficients are given by:

$$c_{L;0} = \frac{\sin \frac{\phi_1}{2} \sin \frac{\phi_2}{2}}{\sin \frac{\phi_1 - \phi_0}{2} \sin \frac{\phi_2 - \phi_0}{2}} = -\sqrt{3}\phi + \mathcal{O}_0(\phi^2), \quad (\text{C.305})$$

$$c_{L;1} = \frac{\sin \frac{\phi_0}{2} \sin \frac{\phi_2}{2}}{\sin \frac{\phi_0 - \phi_1}{2} \sin \frac{\phi_2 - \phi_1}{2}} = 1 + \mathcal{O}_0(\phi^2), \quad (\text{C.306})$$

$$c_{L;2} = \frac{\sin \frac{\phi_0}{2} \sin \frac{\phi_1}{2}}{\sin \frac{\phi_0 - \phi_2}{2} \sin \frac{\phi_1 - \phi_2}{2}} = \sqrt{3}\phi + \mathcal{O}_0(\phi^2), \quad (\text{C.307})$$

where the effective noise amplitudes are given by:

$$\phi_0 = \frac{2\pi}{3} + \phi, \quad \phi_1 = 2\pi + 3\phi, \quad \phi_2 = \frac{10\pi}{3} + 5\phi, \quad (\text{C.308})$$

and ϕ is the noise amplitude of the original noisy gate (see App. C.1). This method is unbiased so the proxy bias is zero.

C.6 CLM

Custom-channel-assisted local mitigation (CLM) is a local method (see App. C.2.1) that uses custom channels to amplify the noise in the noisy gate variants:

$$\boxed{\tilde{\mathcal{C}}}_M = \left(\boxed{\tilde{U}_n}_M \right)_{n=1}^{N_G}, \quad \boxed{\tilde{U}_n}_M = \left\{ c_{L;n,j} : \boxed{\tilde{U}_n}_j \right\}_{j=0}^{N_{L,C,n}}, \quad (\text{C.309})$$

$$\boxed{\tilde{U}_n}_j = \boxed{\tilde{U}_n} \boxed{\mathcal{E}_n}_{C,j} = \boxed{U_n} \boxed{\mathcal{E}_n} \boxed{\mathcal{E}_n}_{C,j} = \boxed{U_n} \boxed{\mathcal{E}_n}_j, \quad (\text{C.310})$$

where $\tilde{\mathcal{C}}_M$ is the mitigated circuit, $\tilde{U}_{n;M}$ is n th mitigated gate, N_G is the number of gates in the circuit, $c_{L;n,j}$ is the mitigation coefficient for the j th gate variant of the n th gate, $\tilde{U}_{n;j}$ is the j th gate variant of the n th gate, $N_{L,C,n}$ is the number of gate variants for the n th gate and the order of the mitigation used for that gate, $\tilde{U}_n$ is the n th noisy gate before mitigation, $\mathcal{E}_{n;C,j}$ is the custom noise channel used to generate the j th variant of the n th noisy gate, $\mathcal{E}_n$ is the original error channel of the n th noisy gate, U_n is the noise-free component of the n th gate, and $\mathcal{E}_{n;j}$ is the error channel of the j th variant of the n th gate.

CLM could be considered to be an example of NEPEC (see Ref. [18]). CLM:SN (used to mitigate stochastic noise) is functionally equivalent to the probabilistic error cancellation procedure that is declared optimal for dephasing noise in Ref. [6]. Though, in fact, it is possible to obtain a slightly lower sampling cost—as noted by Ref. [3], but the difference is negligible.

The execution time for the i th variant of the n th gate is assumed to be:

$$\tau_{L,n;i} = \begin{cases} \tau_{L,n}, & \text{for } i = 0, \\ 2\tau_{L,n}, & \text{for } i > 0, \end{cases} \quad (\text{C.311})$$

where $\tau_{L,n}$ is the execution time for the original n th noisy gate ($\tilde{U}_{n;0} = \tilde{U}_n$). That is, we assume that the execution time of every custom error channel is the same as the

Appendix

original gate. An exception is the case of (stochastic) over-rotation errors where we break our custom channels into Clifford gates and assume that each Clifford gate has the same execution time (see below).

For stochastic noise we use $N_{L,C,n} = 1$, with the noisy gate variants:

$$\begin{aligned} \boxed{\widetilde{U}_n}_0 &= \boxed{\widetilde{U}_n}, \quad \boxed{\widetilde{U}_n}_1 = \boxed{\widetilde{U}_n} \begin{cases} \text{with probability } 1 - \sin^2 \frac{\pi}{2}, \\ \boxed{\mathcal{N}_n} \text{ with probability } \sin^2 \frac{\pi}{2}, \end{cases} \end{aligned} \quad (\text{C.312})$$

and the corresponding coefficients:

$$c_{L;0} = \frac{p_1}{p_1 - p_0} = 1 + p + \mathcal{O}_0(p^2), \quad c_{L;1} = -\frac{p_0}{p_1 - p_0} = -p + \mathcal{O}_0(p^2). \quad (\text{C.313})$$

where the amplified noise amplitudes are given by:

$$p_0 = p, \quad p_1 = p + \sin^2 \frac{\pi}{2} (1 - (1 + a)p), \quad (\text{C.314})$$

where p is the noise amplitude of the original noisy gate and $0 \leq a \leq 1$ is determined by the nature of the noise channel (see Eq. (C.273)). This method is unbiased so the proxy bias is zero.

For rotational errors we use $N_{L,C,n} = 2$ and the noisy gate variants:

$$\boxed{\widetilde{U}_n}_0 = \boxed{\widetilde{U}_n}, \quad \boxed{\widetilde{U}_n}_1 = \boxed{\widetilde{U}_n} \boxed{e^{i\frac{\pi}{4}P_n}}, \quad \boxed{\widetilde{U}_n}_2 = \boxed{\widetilde{U}_n} \boxed{e^{i\frac{3\pi}{4}P_n}}, \quad (\text{C.315})$$

with the coefficients:

$$c_{L;0} = \cos \phi = 1 - \frac{\phi^2}{2} + \mathcal{O}_0(\phi^4), \quad c_{L;1} = -\frac{\cos \phi - 1 + \sin \phi}{2} = -\frac{\phi}{2} + \frac{\phi^2}{4} + \mathcal{O}_0(\phi^4), \quad (\text{C.316})$$

$$c_{L;2} = \frac{1 - \cos \phi + \sin \phi}{2} = \frac{\phi}{2} + \frac{\phi^2}{4} + \mathcal{O}_0(\phi^4), \quad (\text{C.317})$$

where ϕ is the noise amplitude of the original noisy gate (see App. C.1). This method is unbiased so the proxy bias is zero.

For (stochastic) over-rotation errors we use $N_{L,C,n} = 2$ and the same custom channels as for rotational errors (Eq. (C.315)). When we make use of this method in the main text we decompose our (non-trivial) custom channels into Clifford gates:

$$\boxed{\mathcal{E}_n}_{C,1} = \boxed{e^{i\frac{\pi}{4}P_n}} = \boxed{S_{P_n}}, \quad \boxed{\mathcal{E}_n}_{C,2} = \boxed{S_{P_n}} \boxed{P_n}, \quad S_{P_n}^2 = P_n, \quad (\text{C.318})$$

($\mathcal{E}_{n;C,0}$ is the identity channel) and we assume every gate takes the same amount of time to implement. So:

$$\tau_{L,n;i} = (i + 1) \tau_{L,n}. \quad (\text{C.319})$$

We use the coefficients:

$$c_{L;0} = \frac{\cos \phi}{1 - 2p} = 1 + 2p - \frac{\phi^2}{2} + \mathcal{O}_0(p^2, p\phi^2, \phi^4), \quad (\text{C.320})$$

$$c_{L;1} = -\frac{\cos \phi - 1 + 2p + \sin \phi}{2(1 - 2p)} = -p - \frac{\phi}{2} + \frac{\phi^2}{4} + \mathcal{O}_0(p^2, \phi p, \phi^4), \quad (\text{C.321})$$

$$c_{L;2} = \frac{1 - 2p - \cos \phi + \sin \phi}{2(1 - 2p)} = -p + \frac{\phi}{2} + \frac{\phi^2}{4} + \mathcal{O}_0(p^2, \phi p, \phi^4), \quad (\text{C.322})$$

Appendix

where p and ϕ are the noise amplitudes of the stochastic and rotational components of the original noisy gate's error channel, respectively (see App. C.1). This method is unbiased so the proxy bias is zero.

C.7 CSM

Custom-channel-assisted synchronous mitigation (CSM) is a synchronous method (see App. C.2.1) that uses custom channels to amplify the noise in the noisy gate variants:

$$\tilde{\mathcal{C}}_{\text{M}} = \left\{ c_{\text{GL};j} : \tilde{\mathcal{C}}_j \right\}_{j=0}^{N_{\text{C}}} = \left\{ c_{\text{GL};j} : \left(-\tilde{U}_n \right)_{n=1}^{N_{\text{G}}} \right\}_{j=0}^{N_{\text{C}}}, \quad (\text{C.323})$$

$$-\tilde{U}_n = -\tilde{U}_n - \mathcal{E}_{n;\text{C},j} = -U_n - \mathcal{E}_n - \mathcal{E}_{n;\text{C},j} = -U_n - \mathcal{E}_n, \quad (\text{C.324})$$

where $\tilde{\mathcal{C}}_{\text{M}}$ is the mitigated circuit, $c_{\text{GL};j}$ is the mitigation coefficient for the j th circuit variant, $\tilde{\mathcal{C}}_j$ is the j th circuit variant, $N_{\text{C}} = N_{\text{M}}$ is the number of circuit variants and the order of the mitigation, $\tilde{U}_{n;j}$ is the j th variant of the n th noisy gate, N_{G} is the number of gates in the circuit, $\tilde{U}_n$ is the n th noisy gate before mitigation, $\mathcal{E}_{n;\text{C},j}$ is the custom noise channel used to generate the j th variant of the n th noisy gate, $\mathcal{E}_n$ is the original error channel of the n th noisy gate, U_n is the noise-free component of the n th gate, and $\mathcal{E}_{n;j}$ is the error channel of the j th variant of the n th gate.

CSM requires all error channels to have the same type (e.g. they must all be stochastic or rotational) and have the same noise amplitude. All the custom channels have the same noise amplitude but may have different generators (their generator should match that of the noise channel they are tailoring). CSM is a zero-noise extrapolation method akin to Refs [16, 23]. Akin to Ref. [23] it uses custom noise channels in order to amplify the noise of gates and produce circuit variants with differing noise levels to use for the mitigation. However, it generalises their approach by adding the ability to perform unbiased mitigation of rotational errors.

The execution time for the j th circuit is given by:

$$\tau_{\text{E},j} = \begin{cases} \tau_{\text{E}}, & \text{for } i = 0, \\ 2\tau_{\text{E}}, & \text{for } i > 0, \end{cases} \quad (\text{C.325})$$

where τ_{E} is the execution time for the original noisy circuit. We assume that the execution time of every custom error channel is the same as the original gate and that all the gates take the same time to implement.

For stochastic noise we use the gate variants:

$$-\tilde{U}_n = -U_n, \quad -\tilde{U}_n = -U_n \begin{cases} \text{with probability } 1 - \sin^2 \frac{i\pi}{2N_{\text{M}}}, \\ -\mathcal{N}_n \text{ with probability } \sin^2 \frac{i\pi}{2N_{\text{M}}}, \end{cases} \quad (\text{C.326})$$

with the coefficients:

$$c_{\text{GL};i} = \prod_{m=0:m \neq i}^{N_{\text{M}}} \frac{p_m}{p_m - p_i}, \quad p_m = p + \sin^2 \frac{m\pi}{2N_{\text{M}}} (1 - (1+a)p), \quad (\text{C.327})$$

where p is the noise amplitude of the original noisy gate and $0 \leq a \leq 1$ is determined by the nature of the noise channel (see Eq. (C.273)). We do not have an analytic form for

Appendix

the proxy bias in general but if we set $N_M = N_G$ (as we do in the main text) the method is unbiased.

For rotational errors we use the gate variants:

$$\widetilde{U_n}_0 = \widetilde{U_n}, \quad \widetilde{U_n}_{0 < i} = \widetilde{U_n} \boxed{e^{i \frac{\theta_i}{2} P_n}}, \quad \theta_{0 < i} = \frac{2i-1}{N_M} \pi. \quad (\text{C.328})$$

with the coefficients:

$$c_{\text{GL};i} = \prod_{j=0, j \neq i}^{N_M} \frac{\sin \frac{\phi_j}{2}}{\sin \frac{\phi_j - \phi_i}{2}}, \quad \phi_0 = \phi, \quad \phi_{0 < i} = \phi + \theta_i, \quad (\text{C.329})$$

where ϕ is the noise amplitude of the original noisy gate (see App. C.1) and N_M is the order of the mitigation. We do not have an analytic form for the proxy bias in general but if we set $N_M = 2N_G$ (as we do in the main text) the method is unbiased.

C.8 IIAM

Identity-insertion-assisted asynchronous mitigation (IIAM) is an asynchronous method (see App. C.2.1) that uses identity insertions to amplify the noise in the noisy gate variants:

$$\widetilde{\mathcal{C}}_M = \left\{ c_{\text{GL};i} : \widetilde{\mathcal{C}}_i \right\}_{i=0}^{N_C} = \left\{ c_{\text{GL};i} : \left(\widetilde{U_n}_{J_n(i)} \right)_{n=1}^{N_G} \right\}_{i=0}^{N_C}, \quad (\text{C.330})$$

$$\widetilde{U_n}_{J_n(i)} = \left(\widetilde{U_n} \widetilde{U_n^\dagger} \right)^{J_n(i)} \widetilde{U_n} = \widetilde{U_n} \boxed{\mathcal{E}_n}_{J_n(i)}, \quad (\text{C.331})$$

where $\widetilde{\mathcal{C}}_M$ is the mitigated circuit, $c_{\text{GL};i}$ is the mitigation coefficient for the i th circuit variant, $\widetilde{\mathcal{C}}_i$ is the i th circuit variant, N_C is the number of circuit variants, $\widetilde{U_n}_{J_n(i)}$ is the $J_n(i)$ th variant of the n th noisy gate, N_G is the number of gates in the circuit, $\widetilde{U_n}$ is the n th noisy gate before mitigation, $\widetilde{U_n^\dagger}$ is the noisy inverse gate, U_n is the noise-free component of the n th gate, and $\mathcal{E}_n$ is the error channel of the $J_n(i)$ th variant of the n th gate.

IIAM requires all error channels to be identical and to commute with the gates they affect. IIAM is a generalisation of RIIM to more diverse noise models, RIIM is introduced in Ref. [34] (see also Ref. [5]) as a method for dealing with local depolarising noise. We only have a knowledge-free version of this method.

The definition of $J_n(i)$ is rather complex so it is easier to represent the mitigation in another, equivalent way:

$$\widetilde{\mathcal{C}}_M = \left\{ 1 : \widetilde{\mathcal{C}}_A \right\}_{m=0}^{N_M}, \quad \widetilde{\mathcal{C}}_A = \left\{ c_{A,m;j} : \widetilde{\mathcal{C}}_{A,m,j} \right\}_{j=0}^{N_{A,m}-1}, \quad (\text{C.332})$$

where N_M is the order of the mitigation, $\widetilde{\mathcal{C}}_{A,m}$ is a linear combination of all the circuit variants that can be produced by applying exactly m identity insertions anywhere in the circuit. This could be m identity insertions applied to any one gate, 1 identity insertion applied to any m gates, or anything in between. $\widetilde{\mathcal{C}}_{A,m,j}$ is a linear combination (with unit coefficients) of the j th class of ways to include m identity insertions and $N_{A,m}$ is the number of distinct configurations of m identity insertions that can be applied, i.e. the number of integer partitions of m . Each class has a list of instructions associated with it:

Appendix

$L_{A;m,j}$, where each element of this list denotes a number of identity insertions followed by a number of gates to apply that number of identity insertions to. For example, we have:

$$L_{A;0,0} = \{\{0,0\}\}, \quad L_{A;1,0} = \{\{1,1\}\}, \quad L_{A;4,3} = \{\{2,1\}, \{1,2\}\}, \quad L_{A;5,4} = \{\{2,2\}, \{1,1\}\}, \quad (\text{C.333})$$

the first ($L_{A;0,0}$) indicates we use no identity insertions so $\tilde{\tilde{C}}_{A;0,0}$ is just our original noisy circuit ($\tilde{C}$); the second ($L_{A;1,0}$) indicates we use one identity insertion so $\tilde{\tilde{C}}_{A;1,0}$ consists of all the noisy circuits that can be made using a single identity insertion so it is made up of N_G distinct circuits, each one applies the identity insertion to a different gate; the third ($L_{A;4,3}$) indicates each circuit that makes up $\tilde{\tilde{C}}_{A;4,3}$ has four identity insertions in total: one gate has two identity insertions and two gates have one identity insertion, we include all possible circuits that satisfy this requirement; and the fourth ($L_{A;5,4}$) indicates each circuit that makes up $\tilde{\tilde{C}}_{A;5,4}$ has five identity insertions in total: two gates have two identity insertions and one gate has one identity insertion, we include all possible circuits that satisfy this requirement. In Table 6 we provide the configuration lists for $m = 0$ to $m = 5$; the multiplicity of $\tilde{\tilde{C}}_{A;m,j}$ ($\# [\tilde{\tilde{C}}_{A;m,j}]$, i.e. the number of distinct circuits $\tilde{\tilde{C}}_{A;m,j}$ it is made up of); all the coefficients required for the mitigations; and the equivalent i index of each coefficient (see Eq. (C.330)):

$$i = j + \sum_{n=0}^{m-1} N_{A,n}, \quad c_{\text{GI};i} = \# [\tilde{\tilde{C}}_{A;m,j}] c_{A;m,j}, \quad \boxed{\tilde{C}}_i = \left\{ \frac{1}{\# [\tilde{\tilde{C}}_{A;m,j}]} : \boxed{\tilde{C}}_{A;m,j} \right\}. \quad (\text{C.334})$$

The implementation of IIAM is similar to the standard procedure (see App. C.2.2), the only difference is that $\tilde{C}_i$ is now a stochastic mixture of circuits. So each time we should sample $\tilde{C}_i$ we choose one of the constituent circuits of $\tilde{\tilde{C}}_{A;m,j}$ uniformly at random and then divide the output by the multiplicity ($\# [\tilde{\tilde{C}}_{A;m,j}]$).

The execution time for every circuit contained within $\tilde{\tilde{C}}_{A;m}$ is:

$$\tau_{E,j} = \left(1 + \frac{2m}{N_G} \right) \tau_E \quad (\text{C.335})$$

where τ_E is the execution time for the original noisy circuit and we have assumed that all gates take the same amount of time to implement.

Numerical and analytical calculations have indicated that the proxy bias is given by:

$$\varepsilon_{\heartsuit;M} = \begin{cases} \frac{(2N_G p)^{N_M+1}}{(N_M+1)!} & \text{for stochastic noise} \\ \frac{(N_G |\phi|)^{N_M+1}}{(N_M+1)!} & \text{for rotational errors} \end{cases} \quad (\text{C.336})$$

to leading order.

It is worth noting that Table 6 is slightly different from the corresponding table in Ref. [34] (see Table I of Ref. [34] or Table II of the Arxiv version: Ref. [44]). The reason for this difference is that there is a degree of freedom in the choice of coefficients when $N_M = 4$ and one wishes to mitigate local depolarising noise. We choose coefficients that will also mitigate rotational errors and the authors of Ref. [34] make a different choice. In addition it appears there may be a typographic error for the coefficient: $a_{\{3,3\}}$ in the

Appendix

fourth order mitigation (corresponding to $c_{A,2;1}$ in this work). We obtain consistent results using:

$$a_{\{3,3\}} = \frac{N_c^2 + 14N_c + \mathbf{58}}{32}, \tag{C.337}$$

whereas in Table I of Ref. [\[34\]](#) 59 is written in place of 58 in the numerator.

i (m, j)	0 (0, 0)	1 (1, 0)	2 (2, 0)	3 (2, 1)
$L_{A;m,j}$	$\{\{0, 0\}\}$	$\{\{1, 1\}\}$	$\{\{2, 1\}\}$	$\{\{1, 2\}\}$
$\# \left[\tilde{\mathcal{C}}_{A;m,j} \right]$	1	N_G	N_G	$\frac{N_G(N_G-1)}{2}$
$N_M : 0; c_{A,m,j}$	1			
$N_M : 1; c_{A,m,j}$	$\frac{2+N_G}{2}$	$-\frac{1}{2}$		
$N_M : 2; c_{A,m,j}$	$\frac{(2+N_G)(4+N_G)}{8}$	$-\frac{4+N_G}{4}$	$\frac{3}{8}$	$\frac{1}{4}$
$N_M : 3; c_{A,m,j}$	$\frac{(2+N_G)(4+N_G)(6+N_G)}{48}$	$-\frac{(4+N_G)(6+N_G)}{16}$	$\frac{3(6+N_G)}{16}$	$\frac{6+N_G}{8}$
$N_M : 4; c_{A,m,j}$	$\frac{(2+N_G)(4+N_G)(6+N_G)(8+N_G)}{384}$	$-\frac{(4+N_G)(6+N_G)(8+N_G)}{96}$	$\frac{3(6+N_G)(8+N_G)}{64}$	$\frac{(6+N_G)(8+N_G)}{32}$
$N_M : 5; c_{A,m,j}$	$\frac{(2+N_G)(4+N_G)(6+N_G)(8+N_G)(10+N_G)}{3840}$	$-\frac{(4+N_G)(6+N_G)(8+N_G)(10+N_G)}{768}$	$\frac{(6+N_G)(8+N_G)(10+N_G)}{128}$	$\frac{(6+N_G)(8+N_G)(10+N_G)}{192}$

(a) $m = 0$ to $m = 2$.

i (m, j)	4 (3, 0)	5 (3, 1)	6 (3, 2)	7 (4, 0)	8 (4, 1)	9 (4, 2)	10 (4, 3)	11 (4, 4)
$L_{A;m,j}$	$\{\{3, 1\}\}$	$\{\{2, 1\}, \{1, 1\}\}$	$\{\{1, 3\}\}$	$\{\{4, 1\}\}$	$\{\{3, 1\}, \{1, 1\}\}$	$\{\{2, 2\}\}$	$\{\{2, 1\}, \{1, 2\}\}$	$\{\{1, 4\}\}$
$\# \left[\tilde{\mathcal{C}}_{A;m,j} \right]$	N_G	$N_G (N_G - 1)$	$\frac{N_G(N_G-1)(N_G-2)}{3!}$	N_G	$N_G (N_G - 1)$	$\frac{N_G(N_G-1)}{2}$	$\frac{N_G(N_G-1)(N_G-2)}{2}$	$\frac{N_G!}{4!(N_G-4)!}$
$N_M : 3; c_{A,m,j}$	$-\frac{5}{16}$	$-\frac{3}{16}$	$-\frac{1}{8}$					
$N_M : 4; c_{A,m,j}$	$-\frac{5(8+N_G)}{32}$	$-\frac{3(8+N_G)}{32}$	$-\frac{8+N_G}{16}$	$\frac{35}{128}$	$\frac{5}{32}$	$\frac{9}{64}$	$\frac{3}{32}$	$\frac{1}{16}$
$N_M : 5; c_{A,m,j}$	$-\frac{5(8+N_G)(10+N_G)}{128}$	$-\frac{3(8+N_G)(10+N_G)}{128}$	$-\frac{(8+N_G)(10+N_G)}{64}$	$\frac{35(10+N_G)}{256}$	$\frac{5(10+N_G)}{64}$	$\frac{9(10+N_G)}{128}$	$\frac{3(10+N_G)}{64}$	$\frac{10+N_G}{32}$

(b) $m = 3$ to $m = 4$.

i (m, j)	12 (5, 0)	13 (5, 1)	14 (5, 2)	15 (5, 3)	16 (5, 4)	17 (5, 5)	18 (5, 6)
$L_{A;m,j}$	$\{\{5, 1\}\}$	$\{\{4, 1\}, \{1, 1\}\}$	$\{\{3, 1\}, \{2, 1\}\}$	$\{\{3, 1\}, \{1, 2\}\}$	$\{\{2, 2\}, \{1, 1\}\}$	$\{\{2, 1\}, \{1, 3\}\}$	$\{\{1, 5\}\}$
$\# \left[\tilde{\mathcal{C}}_{A;m,j} \right]$	N_G	$N_G (N_G - 1)$	$N_G (N_G - 1)$	$\frac{N_G(N_G-1)(N_G-2)}{2}$	$\frac{N_G(N_G-1)(N_G-2)}{2}$	$\frac{N_G(N_G-1)(N_G-2)(N_G-3)}{3!}$	$\frac{N_G(N_G-1)(N_G-2)(N_G-3)(N_G-4)}{5!}$
$N_M : 5; c_{A,m,j}$	$-\frac{63}{256}$	$-\frac{35}{256}$	$-\frac{15}{128}$	$-\frac{5}{64}$	$-\frac{9}{128}$	$-\frac{3}{64}$	$-\frac{1}{32}$

(c) $m = 5$.

Table 6: Table of IIAM coefficients and parameters (see App. C.8).

Appendix

C.9 IILM

Identity-insertion-assisted local mitigation (IILM) is a local method (see App. C.2.1) that uses identity insertions to amplify the noise in the noisy gate variants:

$$\tilde{\mathcal{C}}_{\text{M}} = \left(\tilde{U}_n \right)_{n=1}^{N_{\text{G}}}, \quad \tilde{U}_n = \left\{ c_{\text{L};n,j} : \tilde{U}_n \right\}_{j=0}^{N_{\text{L},\text{C},n}}, \quad (\text{C.338})$$

$$\tilde{U}_n = \left(\tilde{U}_n \tilde{U}_n^\dagger \right)^j \tilde{U}_n = \tilde{U}_n \tilde{\mathcal{E}}_n, \quad (\text{C.339})$$

where $\tilde{\mathcal{C}}_{\text{M}}$ is the mitigated circuit, $\tilde{U}_{n;\text{M}}$ is n th mitigated gate, N_{G} is the number of gates in the circuit, $c_{\text{L};n,j}$ is the mitigation coefficient for the j th gate variant of the n th gate, $\tilde{U}_{n;j}$ is the j th gate variant of the n th gate, $N_{\text{L},\text{C},n} = N_{\text{M}}$ is the number of gate variants for the n th gate and the order of the mitigation, $\tilde{U}_n$ is the n th noisy gate before mitigation, $\tilde{U}_n^\dagger$ is the noisy inverse of the n th gate, U_n is the noise-free component of the n th gate, and $\mathcal{E}_{n,j}$ is the error channel of the j th variant of the n th gate.

The version of IILM presented here requires the error channels associated with the noisy gate and its noisy inverse to be the same and that the noisy gate, the noisy inverse, and their error channels all commute with each other. IILM could be considered to be a special case of NEPEC (see Ref. [18]).

The execution time for the i th variant of the n th gate is given by:

$$\tau_{\text{L},n;i} = (2i + 1) \tau_{\text{L},n}, \quad (\text{C.340})$$

where $\tau_{\text{L},n}$ is the execution time for the original n th noisy gate.

The knowledge-free version (IILM:KF) uses:

$$c_{\text{L};n,i} = \prod_{m=0:m \neq i}^{N_{\text{M}}} \frac{2m + 1}{2(m - i)}. \quad (\text{C.341})$$

The proxy bias for stochastic noise (for IILM:KF) is given by:

$$\varepsilon_{\heartsuit;\text{M}} = 2N_{\text{G}} p^{N_{\text{M}}+1} (1 + a)^{N_{\text{M}}} \frac{(2N_{\text{M}} + 1)!!}{(N_{\text{M}} + 1)!} + N_{\text{G}} \mathcal{O}_0(p^{N_{\text{M}}+2}), \quad (\text{C.342})$$

where p is the noise amplitude of the original noisy gates ($\tilde{U}_{n;0} = \tilde{U}_n$), and $0 \leq a \leq 1$ is defined in Eq. (C.273). The proxy bias for rotational errors (for IILM:KF) is given by:

$$\varepsilon_{\heartsuit;\text{M}} = N_{\text{G}} |\phi|^{N_{\text{M}}+1} \frac{(2N_{\text{M}} + 1)!!}{(N_{\text{M}} + 1)!} + N_{\text{G}} \mathcal{O}_0(\phi^{N_{\text{M}}+2}). \quad (\text{C.343})$$

The noise-aware version (IILM:NA) for stochastic noise uses $N_{\text{M}} = 1$ and:

$$c_{\text{L};0} = \frac{3 - 3(1 + a)p + (1 + a)^2 p^2}{2 - 3(1 + a)p + (1 + a)^2 p^2} = \frac{3}{2} + \frac{3}{4}(1 + a)p + \mathcal{O}_0(p^2), \quad (\text{C.344})$$

$$c_{\text{L};1} = -\frac{1}{2 - 3(1 + a)p + (1 + a)^2 p^2} = -\frac{1}{2} - \frac{3}{4}(1 + a)p + \mathcal{O}_0(p^2). \quad (\text{C.345})$$

This method is unbiased so the proxy bias is zero.

Appendix

The noise-aware version (IILM:NA) for rotational errors uses ($N_M = 2$):

$$c_{L;0} = \frac{\sin \frac{3\phi}{2} \sin \frac{5\phi}{2}}{\sin \phi \sin (2\phi)} = \frac{15}{8} - \frac{35}{32} \phi^2 + \mathcal{O}_0(\phi^4), \quad (\text{C.346})$$

$$c_{L;1} = -\frac{\sin \frac{\phi}{2} \sin \frac{5\phi}{2}}{\sin^2 \phi} = -\frac{5}{4} + \frac{15}{16} \phi^2 + \mathcal{O}_0(\phi^4), \quad (\text{C.347})$$

$$c_{L;2} = \frac{\sin \frac{\phi}{2} \sin \frac{3\phi}{2}}{\sin \phi \sin (2\phi)} = \frac{3}{8} + \frac{5}{32} \phi^2 + \mathcal{O}_0(\phi^4). \quad (\text{C.348})$$

This method is unbiased so the proxy bias is zero.

C.10 IISM

Identity-insertion-assisted synchronous mitigation (IISM) is a synchronous method (see App. C.2.1) that uses identity insertions to amplify the noise in the noisy gate variants:

$$\tilde{\mathcal{C}}_M = \left\{ c_{\text{GI};j} : \tilde{\mathcal{C}}_j \right\}_{j=0}^{N_C} = \left\{ c_{\text{GI};j} : \left(-\tilde{U}_n \right)_{n=1}^{N_G} \right\}_{j=0}^{N_C}, \quad (\text{C.349})$$

$$-\tilde{U}_n = \left(-\tilde{U}_n - \tilde{U}_n^\dagger \right)^j - \tilde{U}_n = -U_n - \mathcal{E}_{n;i}, \quad (\text{C.350})$$

where $\tilde{\mathcal{C}}_M$ is the mitigated circuit, $c_{\text{GI};j}$ is the mitigation coefficient for the j th circuit variant, $\tilde{\mathcal{C}}_j$ is the j th circuit variant, $N_C = N_M$ is the number of circuit variants and the order of the mitigation, $\tilde{U}_{n;j}$ is the j th variant of the n th noisy gate, N_G is the number of gates in the circuit, $\tilde{U}_n$ is the n th noisy gate before mitigation, $\tilde{U}_n^\dagger$ is the noisy inverse gate, U_n is the noise-free component of the n th gate, and $\mathcal{E}_{n;i}$ is the error channel of the i th variant of the n th gate.

IISM requires all the mitigated noisy gates and their noisy inverses to have noise of the same type with the same noise amplitude and the noise channel of each gate must commute with the noisy gates and its noisy inverses. IISM is a special case of zero-noise extrapolation (see Refs [16, 23]).

The execution time for the j th circuit is given by:

$$\tau_{E,j} = (2j + 1) \tau_E \quad (\text{C.351})$$

where τ_E is the execution time for the original noisy circuit.

The knowledge-free version (IISM:KF) uses:

$$c_{\text{GI};i} = \prod_{m=0:m \neq i}^{N_M} \frac{2m + 1}{2(m - i)}. \quad (\text{C.352})$$

The proxy bias for stochastic noise (for IISM:KF) is given by:

$$\varepsilon_{\heartsuit;M} = (2N_G p)^{N_M+1} \frac{(2N_M + 1)!!}{(N_M + 1)!} + \mathcal{O}_{\text{LS}} \left(N_G^{N_M} p^{N_M+1}, N_G^{N_M+2} p^{N_M+2} \right), \quad (\text{C.353})$$

where p is the noise amplitude of the original noisy gates ($\tilde{U}_{n;0} = \tilde{U}_n$). The proxy bias for rotational errors (for IISM:KF) is given by:

$$\varepsilon_{\heartsuit;M} = (N_G |\phi|)^{N_M+1} \frac{(2N_M + 1)!!}{(N_M + 1)!} + \mathcal{O}_{\text{LS}} \left(N_G^{N_M} \phi^{N_M+1}, N_G^{N_M+2} \phi^{N_M+2} \right). \quad (\text{C.354})$$

Appendix

The noise-aware version (IISM:NA) for stochastic noise uses:

$$c_{\text{GL};i} = \prod_{m=0:m \neq i}^{N_M} \frac{p_m}{p_m - p_i} = \prod_{m=0:m \neq i}^{N_M} \frac{2m+1}{2(m-i)} + \mathcal{O}_0(p), \quad p_i = \frac{1 - (1 - p(1+a))^{2i+1}}{1+a}, \quad (\text{C.355})$$

where p_m is the noise amplitude of the m th noisy variant, N_M is the order of the mitigation, and $0 \leq a \leq 1$ is a parameter of the noise model (see Eq. (C.273)). The proxy bias is the same to leading order as IISM:KF (Eq. (C.353)) for $N_M \ll N_G$.

The noise-aware version (IISM:NA) for rotational errors uses:

$$c_{\text{GL};i} = \prod_{j=0,j \neq i}^{N_M} \frac{\sin \frac{\phi_j}{2}}{\sin \frac{\phi_j - \phi_i}{2}} = \prod_{j=0,j \neq i}^{N_M} \frac{1+2j}{2(j-i)} + \mathcal{O}_0(\phi), \quad \phi_i = (1+2i)\phi, \quad (\text{C.356})$$

where ϕ_j is the noise amplitude of the j th noisy variant and N_M is the order of the mitigation. The proxy bias is the same to leading order as IISM:KF (Eq. (C.354)) for $N_M \ll 2N_G$.

C.11 LC

Local cancellation (LC) is a local tailoring method (see App. C.2.1). It is similar to CLM but instead of producing unbiased gates it removes the coherent components of error channels so that the effective (locally cancelled) noise channel is a Pauli channel (or some other stochastic channel if another basis is used):

$$\tilde{\square}_U^{\text{LC}} = \left\{ c_{\text{LC};j} : \tilde{\square}_j \right\}_{j=0}^{N_{\text{LC}}} = \square_U \left\{ c_{\text{LC};j} : \square_{\mathcal{E}_j} \right\}_{j=0}^{N_{\text{LC}}} = \square_U \square_{\mathcal{E}_{\text{SN}}}, \quad (\text{C.357})$$

where $\tilde{\square}_{\text{LC}}$ is the locally cancelled version of the noisy gate: $\tilde{U}$, $c_{\text{LC};j}$ is the cancelling coefficient for the j th gate variant of the noisy gate, $\tilde{U}_{n;j}$ is the j th gate variant of the noisy gate, N_{LC} is the number of extra gate variants used to cancel the noisy gate, U is the noise-free component of the noisy gate, $\mathcal{E}_j$ is the error channel of the j th variant of the noisy gate, and $\mathcal{E}_{\text{SN}}$ is the stochastic noise channel of the locally cancelled noisy gate, which will generally be a Pauli error channel (i.e. $\mathcal{E}_{\text{SN}} = \mathcal{E}_{\text{P}}$):

$$\square_{\mathcal{E}_{\text{P}}} = \left\{ p_k : \square_{\text{P}_k} \right\}_{k=0}^{4^{N_Q}-1}, \quad \sum_{k=0}^{4^{N_Q}-1} p_k = 1, \quad (\text{C.358})$$

where $p_k \geq 0$ is the probability an error channel equivalent to applying the k th Pauli operator (P_k) occurs if a gate with this error channel is implemented, and N_Q is the number of qubits that the error channel affects.

LC is an alternative noise tailoring method to Pauli twirling (Ref. [28–31]), which can be generalised to randomised compiling (RC, Ref. [32]). Both techniques convert coherent error channels to Pauli noise channels, when averaged over many runs. RC stochastically sandwiches noisy operators between Pauli and Clifford operators, which are redundant in the absence of noise. Whereas LC uses a stochastic mixture of the original gate and noisy variants of this gate to achieve the same ends. RC has the advantage that it can be implemented knowledge free, whereas LC can only be implemented knowledge free for rotational errors if the hidden inverse can be implemented perfectly. Otherwise it requires

Appendix

quantitative knowledge of the coherent error. The choice between the techniques may depend on whether noisy variants or Clifford operators are easier to implement.

The execution time for the i th variant of the noisy gate is assumed to be:

$$\tau_{L;i} = \tau_L, \quad (\text{C.359})$$

where τ_L is the execution time for the original noisy gate. That is, we assume that the execution time of every noisy gate variant is the same.

The coefficients are calculated using the transfer matrix representation of the noisy gate variant³⁰ error channels (see Eq. (B.194) of App. B.1.2):

$$\left(\hat{T}_{\mathcal{E};j}\right)_{k,l} = \frac{1}{2^{N_Q}} \text{Tr} [P_k \mathcal{E}_{;j} (P_l)], \quad (\text{C.360})$$

where $\hat{T}_{\mathcal{E};j}$ is the transfer matrix corresponding to the error channel of the j th noisy gate variant, P_k is the k th Pauli word operator (see App. B.1.1), and N_Q is the number of qubits the error channel acts on. The Pauli error channel has a diagonal Pauli transfer matrix, so the coefficients are chosen to cancel the off-diagonal elements. The coefficients are stochastic so we require:

$$\sum_{j=0}^{N_{LC}} c_{LC;j} = \sum_{j=0}^{N_{LC}} |c_{LC;j}| = 1 \quad \implies c_{LC;j} > 0, \quad (\text{C.361})$$

for all j . So the resulting channel will be CPTP since it is a stochastic combination of CPTP channels. It is possible to relax the requirement that $c_{LC;j} > 0$ but this will incur a sampling cost, whereas conventional LC has a sampling cost of unity, so it is cost free. To cancel the off-diagonal elements of the error channel we require:

$$\sum_{j=0}^{N_{LC}} c_{LC;j} \left(\hat{T}_{\mathcal{E};j}\right)_{k,l} = 0 \text{ for } k \neq l. \quad (\text{C.362})$$

We place no restriction on the diagonal elements of the $\hat{T}_{\mathcal{E};j}$ except those that result from Eq. (C.361).

We do not implement LC for stochastic noise.

For rotational errors generated by the Hermitian operator P we use $N_{LC} = 1$. We can produce a stochastic channel using local cancellation, which will be a Pauli channel if the Hermitian operator has a diagonal Pauli transfer matrix. If we have access to two noisy gate variants with rotational errors:

$$\tilde{U}_0 = U \left[e^{i \frac{\phi_0}{2} P} \right], \quad \tilde{U}_1 = U \left[e^{i \frac{\phi_1}{2} P} \right], \quad (\text{C.363})$$

where ϕ_0 and ϕ_1 are the noise amplitudes of the two noisy gate variants, then we use the coefficients:

$$c_{L;0} = -\frac{\sin \phi_1}{\sin \phi_0 - \sin \phi_1}, \quad c_{L;1} = \frac{\sin \phi_0}{\sin \phi_0 - \sin \phi_1}, \quad (\text{C.364})$$

³⁰For the LC derivation we present we use the Pauli transfer matrix but we could use a transfer matrix in another basis and rid ourselves of coherences in that basis if we prefer, the derivation is analogous.

Appendix

to obtain the locally cancelled gate. These coefficients will be stochastic provided that $\phi_0\phi_1 < 0$, i.e. the rotational errors act in opposing directions. The resulting locally cancelled gate is given by:

$$\boxed{\widetilde{U}}_{\text{LC}} = \left\{ c_{\text{LC};j} : \boxed{\widetilde{U}}_j \right\}_{j=0}^1 = \boxed{U} \begin{cases} \text{---} & \text{with probability } 1 - p_{\text{LC}} \\ \boxed{P} & \text{with probability } p_{\text{LC}} \end{cases} \quad (\text{C.365})$$

where:

$$p_{\text{LC}} = \frac{1}{2} \left(1 - \frac{\sin(\phi_0 - \phi_1)}{\sin\phi_0 - \sin\phi_1} \right). \quad (\text{C.366})$$

So (if we had a circuit of N_{G} gates with the same noise channels that we locally cancelled in this way) the proxy bias would be given by (see App. C.1):

$$\varepsilon_{\heartsuit;\text{M}} = N_{\text{G}} \left(1 - \frac{\sin(\phi_0 - \phi_1)}{\sin\phi_0 - \sin\phi_1} \right). \quad (\text{C.367})$$

If $\phi = \phi_0 = -\phi_1$ everything simplifies and we find:

$$c_{\text{LC};0} = \frac{1}{2} = c_{\text{LC};1}, \quad \varepsilon_{\heartsuit;\text{M}} = N_{\text{G}} (1 - \cos\phi) = N_{\text{G}} \left(\frac{\phi^2}{2} + \mathcal{O}_0(\phi^4) \right) = \frac{e_{\text{RE}}^2}{2N_{\text{G}}} + \mathcal{O} \left(\frac{e_{\text{RE}}^4}{N_{\text{G}}^3} \right), \quad (\text{C.368})$$

where (see Eq. (C.275)):

$$e_{\text{RE}} = |\phi| N_{\text{G}}, \quad (\text{C.369})$$

is the original noise level. So, although this is designed as a tailoring rather than a mitigation method it can also mitigate the noise in the small noise ($e_{\text{RE}} \ll 1$) or large circuit ($N_{\text{G}} \gg 1$) regime.

C.12 TIILM

Tuned identity-insertion-assisted local mitigation (TIILM) is a local (see App. C.2.1) method that uses identity insertions to amplify the noise in the noisy gate variants. The number of identity insertions used for each gate variant is tuned to minimise the runtime scaling. The procedure is given by:

$$\boxed{\widetilde{\mathcal{C}}}_{\text{M}} = \left(\boxed{\widetilde{U}}_{\text{M}} \right)_{n=1}^{N_{\text{G}}}, \quad \boxed{\widetilde{U}}_{\text{M}} = \left\{ c_{\text{L};n,j} : \boxed{\widetilde{U}}_j \right\}_{j=0}^{N_{\text{L},\text{C},n}}, \quad (\text{C.370})$$

$$\boxed{\widetilde{U}}_j = \left(\boxed{\widetilde{U}}_n \boxed{\widetilde{U}}_n^\dagger \right)^{m_{n,j}} \boxed{\widetilde{U}}_n = \boxed{U}_n \boxed{\mathcal{E}}_j, \quad (\text{C.371})$$

where $\widetilde{\mathcal{C}}_{\text{M}}$ is the mitigated circuit, $\widetilde{U}_{n;\text{M}}$ is n th mitigated gate, N_{G} is the number of gates in the circuit, $c_{\text{L};n,j}$ is the mitigation coefficient for the j th gate variant of the n th gate, $\widetilde{U}_{n;j}$ is the j th gate variant of the n th gate, $N_{\text{L},\text{C},n}$ is the number of gate variants for the n th gate, $\widetilde{U}_n$ is the n th noisy gate before mitigation, $\widetilde{U}_n^\dagger$ is the noisy inverse of the n th gate, $m_{n,j}$ is the number of identity insertions used for the j th variant of the n th gate, U_n is the noise-free component of the n th gate, and $\mathcal{E}_{n,j}$ is the error channel of the j th variant of the n th gate.

Appendix

Our presentation of TIILM requires all error channels to commute with the gates they affect and that the error channels are the same for each gate and its inverse. TIILM could be considered to be a special case of NEPEC (see Ref. [18]). For our numerical results we numerically optimise the values of $m_{n,j}$ but for our analytic results we give two example choices, the first for the small noise regime and the second for the large circuit regime.

The execution time for the i th variant of the n th gate is given by:

$$\tau_{L,n;i} = (2m_{n,i} + 1) \tau_{L,n}, \quad (\text{C.372})$$

where $\tau_{L,n}$ is the execution time for the original n th noisy gate and $m_{n,i}$ is the number of identity insertions for the i th variant.

For stochastic noise we use $N_{L,C,n} = 1$ and (for the analytic results):

$$\kappa = \sqrt{8 - \frac{2}{N_G}}, \quad m_{n,0} = 0, \quad m_{n,1} \approx \begin{cases} \frac{\kappa}{\sqrt{1+a}} \frac{N_G}{\sqrt{e_{SN}}} & \text{for } e_{SN} \ll 1 \\ \frac{\ln 2}{1+a} \frac{N_G}{e_{SN}} & \text{for } N_G \gg 1. \end{cases} \quad (\text{C.373})$$

where e_{SN} is the stochastic noise level (see Eq. (C.272)), N_G is the number of gates in the circuit, and $0 \leq a \leq 1$ is determined by the noise channel (see Eq. (C.273)). The coefficients³¹ are given by:

$$c_{L;0} = \frac{1 - \lambda_1^{2m_{n,1}+1}}{\lambda_1^{2m_{n,0}+1} - \lambda_1^{2m_{n,1}+1}} = \begin{cases} 1 + \frac{\sqrt{1+a}}{2\kappa N_G} \sqrt{e_{SN}} + \mathcal{O}_S(e_{SN}), & \text{for } e_{SN} \ll 1 \\ 1 + (1+a) \frac{e_{SN}}{N_G} + \mathcal{O}_L\left(\frac{1}{N_G^2}\right) & \text{for } N_G \gg 1, \end{cases} \quad (\text{C.374})$$

$$c_{L;1} = -\frac{1 - \lambda_1^{2m_{n,0}+1}}{\lambda_1^{2m_{n,0}+1} - \lambda_1^{2m_{n,1}+1}} = \begin{cases} -\frac{\sqrt{1+a}}{2\kappa N_G} \sqrt{e_{SN}} (1 + \mathcal{O}_S(\sqrt{e_{SN}})) & \text{for } e_{SN} \ll 1 \\ (1+a) \frac{e_{SN}}{N_G} + \mathcal{O}_L\left(\frac{1}{N_G^2}\right) & \text{for } N_G \gg 1, \end{cases} \quad (\text{C.375})$$

where:

$$\lambda_0 = 1, \quad \lambda_1 = 1 - (1+a)p, \quad (\text{C.376})$$

are the eigenvalues of the stochastic noise channel. This method is unbiased so the proxy bias is zero.

For rotational errors we use $N_M = 2$ and (for the analytic results):

$$m_{n,0} = 0, \quad m_{n,1} = \begin{cases} \kappa_1 \frac{N_G}{\sqrt{e_{RE}}} & \text{for } e_G \ll 1 \\ \frac{\pi}{4} \frac{N_G}{e_{RE}} & \text{for } N_G \gg 1. \end{cases}, \quad m_{n,2} = \begin{cases} \frac{\pi}{2} \frac{N_G}{e_{RE}} & \text{for } e_G \ll 1 \\ \frac{3\pi}{4} \frac{N_G}{e_{RE}} & \text{for } N_G \gg 1. \end{cases} \quad (\text{C.377})$$

where we define:

$$\kappa_1 = \sqrt{\frac{2}{\pi}} \sqrt{4 - \frac{1}{N_G}}. \quad (\text{C.378})$$

Here e_{RE} is the rotational noise level (see Eq. (C.275)) and N_G is the number of gates in

³¹The first equalities for the coefficients given in Eq. (C.374) for TIILM are used for both numerical and analytic results. However the Taylor expansions are valid only for the analytic results where we use Eq. (C.373) to define the number of identity insertions we use. For numerical results we optimise $m_{n,0}$ and $m_{n,1}$ to give the lowest possible runtime scaling.

Appendix

the circuit. The coefficients³² are given by:

$$c_{L;0} = \frac{\sin \frac{(2m_{n,1}+1)\phi}{2} \sin \frac{(2m_{n,2}+1)\phi}{2}}{\sin((m_{n,1} - m_{n,0})\phi) \sin((m_{n,2} - m_{n,0})\phi)}$$

$$= \begin{cases} 1 + \frac{1}{2\kappa_1} \frac{\sqrt{e_{\text{RE}}}}{N_G} + \mathcal{O}_S(e_{\text{RE}}) & \text{for } e_{\text{RE}} \ll 1, \\ 1 + \mathcal{O}_L\left(\frac{1}{N_G^2}\right) & \text{for } N_G \gg 1, \end{cases} \quad (\text{C.379})$$

$$c_{L;1} = - \frac{\sin \frac{(2m_{n,0}+1)\phi}{2} \sin \frac{(2m_{n,2}+1)\phi}{2}}{\sin((m_{n,1} - m_{n,0})\phi) \sin((m_{n,2} - m_{n,1})\phi)}$$

$$= \begin{cases} -\frac{1}{2\kappa_1} \frac{\sqrt{e_{\text{RE}}}}{N_G} + \mathcal{O}_S\left(e_{\text{RE}}^{\frac{3}{2}}\right) & \text{for } e_{\text{RE}} \ll 1, \\ -\frac{e_{\text{RE}}}{2N_G} + \mathcal{O}_L\left(\frac{1}{N_G^2}\right) & \text{for } N_G \gg 1, \end{cases} \quad (\text{C.380})$$

$$c_{L;2} = \frac{\sin \frac{(2m_{n,0}+1)\phi}{2} \sin \frac{(2m_{n,1}+1)\phi}{2}}{\sin((m_{n,2} - m_{n,0})\phi) \sin((m_{n,2} - m_{n,1})\phi)}$$

$$= \begin{cases} \frac{\kappa_1}{2} \frac{e_{\text{RE}}^{\frac{3}{2}}}{N_G} + \mathcal{O}_S(e_{\text{RE}}^2) & \text{for } e_{\text{RE}} \ll 1 \\ \frac{e_{\text{RE}}}{2N_G} + \mathcal{O}_L\left(\frac{1}{N_G^2}\right) & \text{for } N_G \gg 1. \end{cases} \quad (\text{C.381})$$

where ϕ is the noise amplitude of the rotational error channels (Eq. (C.274)). This method is unbiased so the proxy bias is zero.

D Numerical

In this appendix we will discuss the details of the simulation procedure (App. D.1) used to generate the numerical plots in Sec. 5.3, along with details of the precision (App. D.2) and parameters (App. D.3) used.

D.1 Simulation

All our numerical plots are the result of simulations rather than implementation on a quantum computer. We simulate the circuits using the Pauli transfer matrix formalism (see App. B.1.2). First we produce the Bloch vector corresponding to the initial state of the qubits (Eq. (D.382)). Then we multiply this Bloch vector by matrices corresponding to each of the gates and noise channels in our circuit (see App. D.1.2), in the order they appear in our circuit (Eq. (136) of Sec. 5.3.1). Finally, we extract the observable expectation by combining the relevant elements of the final Bloch vector obtained after the matrix multiplications (see App. D.1.1). For local mitigation procedures we simulate a circuit where every noisy gate is replaced by the equivalent mitigated gate, where the mitigated gate transfer matrices are linear combinations of noisy gates as specified by the mitigation procedure (see App. D.1.4). For non-local mitigation procedures we simulate every circuit that appears in the linear combination required for the mitigated circuit and then linearly combine the observable estimates to find the mitigated observable estimates (see App. D.1.3).

³²The first equalities for the coefficients given in Eq. (C.379) for TIILM are used for both numerical and analytic results. However, the Taylor expansions are valid only for the analytic results where we use Eq. (C.377) to define the number of identity insertions we use. For numerical results we optimise $m_{n,0}$, $m_{n,1}$, and $m_{n,2}$ to give the lowest possible runtime scaling.

Appendix

D.1.1 Initial States and Measurements

The Bloch vector corresponding to our initial state is (see App. B.1.1):

$$\vec{b}_0 = \bigotimes_{n=1}^{N_Q} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \quad (\text{D.382})$$

where N_Q is the number of qubits in the circuit. This Bloch vector corresponds to the state: $\bigotimes_{n=1}^{N_Q} |0\rangle$ (i.e. the tensor product of the +1 eigenstates of the Pauli Z operators for each qubit). After we have multiplied this vector by all the transfer matrices corresponding to the gates in the circuit we obtain the output Bloch vector $\vec{b}$. For our observable estimates we measure:

$$\langle I_{[1]} \rangle = (\vec{b})_0, \quad \langle X_{[1]} \rangle = (\vec{b})_1, \quad \langle Y_{[1]} \rangle = (\vec{b})_2, \quad \langle Z_{[1]} \rangle = (\vec{b})_3. \quad (\text{D.383})$$

Where $(\vec{b})_i$ is the $i + 1$ th element of the (0-indexed) Bloch vector. $\langle I_{[1]} \rangle$ can be used to verify the state is properly normalised. $\langle X_{[1]} \rangle$, $\langle Y_{[1]} \rangle$, and $\langle Z_{[1]} \rangle$ are the expectation of the X, Y, and Z Pauli operators for the first qubit.

D.1.2 Transfer Matrices

We use a basis set of single qubit Pauli transfer matrices:

$$\hat{T}_{R_1}(\theta, p) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & (1-2p)\cos\theta & (1-2p)\sin\theta \\ 0 & 0 & -(1-2p)\sin\theta & (1-2p)\cos\theta \end{pmatrix}, \quad (\text{D.384})$$

$$\hat{T}_{R_2}(\theta, p) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & (1-2p)\cos\theta & 0 & -(1-2p)\sin\theta \\ 0 & 0 & 1 & 0 \\ 0 & (1-2p)\sin\theta & 0 & (1-2p)\cos\theta \end{pmatrix}, \quad (\text{D.385})$$

$$\hat{T}_{R_3}(\theta, p) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & (1-2p)\cos\theta & (1-2p)\sin\theta & 0 \\ 0 & -(1-2p)\sin\theta & (1-2p)\cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (\text{D.386})$$

These three are sufficient to define all the single qubit gates and operators that we use within this work. For example we have:

$$\hat{T}_X = \hat{T}_{R_1}(\pi, 0), \quad \hat{T}_Y = \hat{T}_{R_2}(\pi, 0), \quad \hat{T}_Z = \hat{T}_{R_3}(\pi, 0), \quad (\text{D.387})$$

$$\hat{T}_{\hat{S}_1} = \hat{T}_{R_1}\left(\frac{\pi}{2}, 0\right), \quad \hat{T}_{\hat{S}_2} = \hat{T}_{R_2}\left(\frac{\pi}{2}, 0\right), \quad \hat{T}_{\hat{S}_3} = \hat{T}_{R_3}\left(\frac{\pi}{2}, 0\right), \quad (\text{D.388})$$

$$\hat{T}_{\hat{T}_1} = \hat{T}_{R_1}\left(\frac{\pi}{4}, 0\right), \quad \hat{T}_{\hat{T}_2} = \hat{T}_{R_2}\left(\frac{\pi}{4}, 0\right), \quad \hat{T}_{\hat{T}_3} = \hat{T}_{R_3}\left(\frac{\pi}{4}, 0\right), \quad (\text{D.389})$$

where the top three are the single qubit Pauli operators and the second and third rows are the square and fourth roots of the Pauli operators, respectively. Inverse gates can be found by inverting the angle:

$$\hat{T}_{\hat{S}_1^\dagger} = \hat{T}_{R_1}\left(-\frac{\pi}{2}, 0\right), \quad \hat{T}_{\hat{S}_2^\dagger} = \hat{T}_{R_2}\left(-\frac{\pi}{2}, 0\right), \quad \hat{T}_{\hat{S}_3^\dagger} = \hat{T}_{R_3}\left(-\frac{\pi}{2}, 0\right), \quad (\text{D.390})$$

$$\hat{T}_{\hat{T}_1^\dagger} = \hat{T}_{R_1}\left(-\frac{\pi}{4}, 0\right), \quad \hat{T}_{\hat{T}_2^\dagger} = \hat{T}_{R_2}\left(-\frac{\pi}{4}, 0\right), \quad \hat{T}_{\hat{T}_3^\dagger} = \hat{T}_{R_3}\left(-\frac{\pi}{4}, 0\right), \quad (\text{D.391})$$

Appendix

The only two-qubit transfer matrix that we make use of directly is $\hat{T}_{R_{15}}(\theta, p)$, which corresponds to rotations of θ followed by a stochastic error channel with noise amplitude p , both generated by $P_{15} = Z \otimes Z$. This is too large to display in all its glory. So, instead, we shall note the elements that differ from those of the two qubit identity matrix (which is all zeros except for the diagonal elements, which are equal to unity):

$$\begin{aligned} \left(\hat{T}_{R_{15}}(\theta, p)\right)_{1,1} &= \left(\hat{T}_{R_{15}}(\theta, p)\right)_{2,2} = \left(\hat{T}_{R_{15}}(\theta, p)\right)_{4,4} = \left(\hat{T}_{R_{15}}(\theta, p)\right)_{7,7} = \left(\hat{T}_{R_{15}}(\theta, p)\right)_{8,8} \\ &= \left(\hat{T}_{R_{15}}(\theta, p)\right)_{11,11} = \left(\hat{T}_{R_{15}}(\theta, p)\right)_{13,13} = \left(\hat{T}_{R_{15}}(\theta, p)\right)_{14,14} = (1 - 2p) \cos \theta, \end{aligned} \quad (\text{D.392})$$

$$\begin{aligned} \left(\hat{T}_{R_{15}}(\theta, p)\right)_{1,14} &= \left(\hat{T}_{R_{15}}(\theta, p)\right)_{4,11} = \left(\hat{T}_{R_{15}}(\theta, p)\right)_{7,8} = \left(\hat{T}_{R_{15}}(\theta, p)\right)_{13,2} = (1 - 2p) \sin \theta \\ &= -\left(\hat{T}_{R_{15}}(\theta, p)\right)_{14,1} = -\left(\hat{T}_{R_{15}}(\theta, p)\right)_{11,4} = -\left(\hat{T}_{R_{15}}(\theta, p)\right)_{8,7} = -\left(\hat{T}_{R_{15}}(\theta, p)\right)_{2,13}, \end{aligned} \quad (\text{D.393})$$

where we use the zero index notation so the indices run from 0 to 15. We have:

$$\hat{T}_{\hat{S}_{15}} = \hat{T}_{R_{15}}\left(\frac{\pi}{2}, 0\right), \quad \hat{T}_{\hat{S}_{15}^\dagger} = \hat{T}_{R_{15}}\left(-\frac{\pi}{2}, 0\right), \quad (\text{D.394})$$

which are the Pauli transfer matrices corresponding to $\hat{S}_{15}$ (where $\hat{S}_{15}^2 = P_{15} = Z \otimes Z$) and $\hat{S}_{15}^\dagger$.

For error channels we use:

$$\hat{T}_{\mathcal{E}_{\text{SN},i}} = \hat{T}_{R_i}(0, p), \quad \hat{T}_{\mathcal{E}_{\text{RE},i}} = \hat{T}_{R_i}(\phi, 0), \quad \hat{T}_{\mathcal{E}_{\text{ORE},i}} = \hat{T}_{R_i}(\phi, p), \quad (\text{D.395})$$

which correspond to stochastic noise, rotational error, and (stochastic) over-rotation error channels generated by P_i , respectively.

Before applying a gate to our state (by multiplying the state's Bloch vector (see App. D.1.1) by our gate's Pauli transfer matrix) we must promote the gate to the correct dimensionality. In our case the state has 4^{N_Q} dimensions. We only use nearest neighbour gates in our circuit, which simplifies this procedure. We use the tensor product operator, to promote single and two qubit operators to the full Hilbert space, such that the first qubit is the furthest right and we move left for gates corresponding to higher numbered qubits. Thus, for example, a noisy T gate (i.e. $\hat{T}_3$) suffering from pure over-rotation errors ($\mathcal{E}_{\text{RE}}$) that acts on the second qubit has a transfer matrix given by:

$$I \otimes I \otimes \hat{T}_{R_3}(\phi, 0) \hat{T}_{\hat{T}_3} \otimes I, \quad (\text{D.396})$$

where I is the two dimensional identity operator. Whereas the noisy two qubit gate: $\hat{S}_{15}$ that acts on the third and fourth qubits and suffers from stochastic over-rotation errors has a transfer matrix given by:

$$\hat{T}_{R_{15}}(\phi, p) \hat{T}_{\hat{S}_{15}} \otimes I \otimes I. \quad (\text{D.397})$$

For our simulations the noise channels follow the gates, but since the two commute the order is irrelevant.

D.1.3 Non-Local Mitigation

For the synchronous and (explicitly) asynchronous strategies, which mitigate the circuit as a whole rather than (effectively) mitigating each gate independently, we obtain an

Appendix

observable estimate (for each observable) for each circuit variant and then calculate the linear combination of the observable estimates for each circuit weighted by the mitigation coefficients, to obtain the mitigated observable estimates. The mitigated bias is given by the difference between this mitigated observable estimate and the observable expectation of the noise-free circuit:

$$\varepsilon_{P_j} = |\text{Tr}[P_j(\rho - \tilde{\rho}_M)]| = \left| \left(\vec{b} \right)_j - \sum_{i=0}^{N_C} c_{\text{Gl};i} \left(\vec{b}_{;i} \right)_j \right|, \quad (\text{D.398})$$

where ε_{P_j} is the mitigated bias for the j th Pauli observable (P_j), ρ is the noise-free state, $\tilde{\rho}_M$ is the effective mitigated state, $\left(\vec{b} \right)_j$ is the j th element of the noise-free Bloch vector (corresponding to the expectation of ρ with respect to P_j), $\left(\vec{b}_{;i} \right)_j$ is the j th element of the Bloch vector output by the i th noisy circuit variant, N_C is the number of circuits used in the mitigation, and $c_{\text{Gl};i}$ is the mitigation coefficient for the i th noisy circuit variant.

For example, we can consider first order ($N_M = N_C = 1$) IISM (see App. C.10) for a circuit suffering from stochastic noise with noise amplitude p . Here we have two circuit variants: the original noisy circuit and the circuit obtained by adding one identity insertion before each gate. The noise-free state is given by (see Eqs (D.384)-(D.394)):

$$\begin{aligned} \vec{b} = & \dots \left(\text{I} \otimes \hat{T}_{\hat{T}_1} \otimes \text{I} \otimes \text{I} \right) \left(\text{I} \otimes \hat{T}_{\hat{S}_{15}} \otimes \text{I} \right) \left(\text{I} \otimes \text{I} \otimes \hat{T}_{\hat{T}_1} \otimes \text{I} \right) \left(\text{I} \otimes \text{I} \otimes \hat{T}_{\hat{S}_{15}} \right) \\ & \times \left(\text{I} \otimes \text{I} \otimes \text{I} \otimes \hat{T}_{\hat{T}_1} \hat{T}_Y \right) \vec{b}_0, \end{aligned} \quad (\text{D.399})$$

where we have explicitly noted the expression up to the 6th gate in the circuit (see Eq. (136) of Sec. 5.3.1). To obtain the state (Bloch vector) output by the original noisy circuit we perform the calculation (see Eqs (D.384)-(D.395)):

$$\begin{aligned} \vec{b}_{;0} = & \dots \left(\text{I} \otimes \hat{T}_{\mathcal{E}_{\text{SN},1}} \hat{T}_{\hat{T}_1} \otimes \text{I} \otimes \text{I} \right) \left(\text{I} \otimes \hat{T}_{\mathcal{E}_{\text{SN},15}} \hat{T}_{\hat{S}_{15}} \otimes \text{I} \right) \left(\text{I} \otimes \text{I} \otimes \hat{T}_{\mathcal{E}_{\text{SN},1}} \hat{T}_{\hat{T}_1} \otimes \text{I} \right) \\ & \times \left(\text{I} \otimes \text{I} \otimes \hat{T}_{\mathcal{E}_{\text{SN},15}} \hat{T}_{\hat{S}_{15}} \right) \left(\text{I} \otimes \text{I} \otimes \text{I} \otimes \hat{T}_{\mathcal{E}_{\text{SN},1}} \hat{T}_{\hat{T}_1} \hat{T}_{\mathcal{E}_{\text{SN},2}} \hat{T}_Y \right) \vec{b}_0. \end{aligned} \quad (\text{D.400})$$

To obtain the state output by the circuit obtained after a single identity insertion for each gate we perform the calculation (see Eqs (D.384)-(D.395)):

$$\begin{aligned} \vec{b}_{;1} = & \dots \left(\text{I} \otimes \hat{T}_{\mathcal{E}_{\text{SN},15}} \hat{T}_{\hat{S}_{15}} \hat{T}_{\mathcal{E}_{\text{SN},15}} \hat{T}_{\hat{S}_{15}}^\dagger \hat{T}_{\mathcal{E}_{\text{SN},15}} \hat{T}_{\hat{S}_{15}} \otimes \text{I} \right) \\ & \times \left(\text{I} \otimes \text{I} \otimes \hat{T}_{\mathcal{E}_{\text{SN},1}} \hat{T}_{\hat{T}_1} \hat{T}_{\mathcal{E}_{\text{SN},1}} \hat{T}_{\hat{T}_1}^\dagger \hat{T}_{\mathcal{E}_{\text{SN},1}} \hat{T}_{\hat{T}_1} \otimes \text{I} \right) \\ & \times \left(\text{I} \otimes \text{I} \otimes \hat{T}_{\mathcal{E}_{\text{SN},15}} \hat{T}_{\hat{S}_{15}} \hat{T}_{\mathcal{E}_{\text{SN},15}} \hat{T}_{\hat{S}_{15}}^\dagger \hat{T}_{\mathcal{E}_{\text{SN},15}} \hat{T}_{\hat{S}_{15}} \right) \\ & \times \left(\text{I} \otimes \text{I} \otimes \text{I} \otimes \hat{T}_{\mathcal{E}_{\text{SN},1}} \hat{T}_{\hat{T}_1} \hat{T}_{\mathcal{E}_{\text{SN},1}} \hat{T}_{\hat{T}_1}^\dagger \hat{T}_{\mathcal{E}_{\text{SN},1}} \hat{T}_{\hat{T}_1} \left(\hat{T}_{\mathcal{E}_{\text{SN},2}} \hat{T}_Y \right)^3 \right) \vec{b}_0. \end{aligned} \quad (\text{D.401})$$

This means our biases after mitigation are given by (see Eqs (C.352) and (D.398)):

$$\begin{aligned} \varepsilon_X = & \left| \left(\vec{b} \right)_1 - \left(\frac{3}{2} \left(\vec{b}_{;0} \right)_1 - \frac{1}{2} \left(\vec{b}_{;1} \right)_1 \right) \right|, & \varepsilon_Y = & \left| \left(\vec{b} \right)_2 - \left(\frac{3}{2} \left(\vec{b}_{;0} \right)_2 - \frac{1}{2} \left(\vec{b}_{;1} \right)_2 \right) \right|, \\ \varepsilon_Z = & \left| \left(\vec{b} \right)_3 - \left(\frac{3}{2} \left(\vec{b}_{;0} \right)_3 - \frac{1}{2} \left(\vec{b}_{;1} \right)_3 \right) \right|. \end{aligned} \quad (\text{D.402})$$

Appendix

D.1.4 Local Mitigation

Since the exponential number of circuits required to implement a local mitigation is impractical to simulate, we instead simulate the mitigated circuit directly. We replace each noisy gate in the circuit with its mitigated equivalent in the original noisy circuit and then simulate this mitigated circuit to obtain the mitigated Bloch vector ($\vec{b}_M$). The mitigated transfer matrix for the n th gate (used to simulate the n th mitigated gate) is given by:

$$\hat{T}_{n;M} = \sum_{i=0}^{N_{L,C,n}} c_{L;n,i} \hat{T}_{n,i} \quad (D.403)$$

where $\hat{T}_{n;M}$ is the transfer matrix for the n th mitigated gate, $N_{L,C,n}$ is the number of noisy gate variants used to mitigate the n th gate, $c_{L;n,i}$ is the mitigation coefficient for the i th noisy gate variant of the n th gate, and $\hat{T}_{n,i}$ is the transfer matrix of the i th noisy gate variant of the n th gate. The mitigated bias is given by the difference between the mitigated observable estimate (derived from $\vec{b}_M$) and the observable expectation of the noise-free circuit:

$$\varepsilon_{P_j} = |\text{Tr}[P_j(\rho - \tilde{\rho}_M)]| = \left| \left(\vec{b} \right)_j - \left(\vec{b}_M \right)_j \right|, \quad (D.404)$$

where ε_{P_j} is the mitigated bias for the j th Pauli observable (P_j), ρ is the noise-free state, $\tilde{\rho}_M$ is the effective mitigated state, $\left(\vec{b} \right)_j$ is the j th element of the noise-free Bloch vector (corresponding to the expectation of ρ with respect to P_j), and $\left(\vec{b}_M \right)_j$ is the j th element of the mitigated Bloch vector (corresponding to the expectation of $\tilde{\rho}_M$ with respect to P_j).

For example, we can consider CLM (App. C.6) for a circuit suffering from rotational errors with noise amplitude ϕ . The noise-free Bloch vector ($\vec{b}$) is calculated using the expression in Eq. (D.399). The mitigated Bloch vector ($\vec{b}_M$) is calculated by replacing all the transfer matrices in Eq. (D.399) by their mitigated equivalents: (see Eqs (D.384)-(D.394)):

$$\begin{aligned} \vec{b}_M = & \dots \left(\text{I} \otimes \hat{T}_{\hat{T}_{1;M}} \otimes \text{I} \otimes \text{I} \right) \left(\text{I} \otimes \hat{T}_{\hat{S}_{15;M}} \otimes \text{I} \right) \left(\text{I} \otimes \text{I} \otimes \hat{T}_{\hat{T}_{1;M}} \otimes \text{I} \right) \left(\text{I} \otimes \text{I} \otimes \hat{T}_{\hat{S}_{15;M}} \right) \\ & \times \left(\text{I} \otimes \text{I} \otimes \text{I} \otimes \hat{T}_{\hat{T}_{1;M}} \hat{T}_{\hat{Y};M} \right) \vec{b}_0, \end{aligned} \quad (D.405)$$

where we have explicitly noted the expression up to the 6th gate in the circuit (see Eq. (136) of Sec. 5.3.1). The mitigated transfer matrix for the first gate is given by (see Eq. (C.316)):

$$\begin{aligned} \hat{T}_{Y;M} = & \cos \phi \hat{T}_{\mathcal{E}_{RE,2}} \hat{T}_Y - \frac{\cos \phi - 1 + \sin \phi}{2} \hat{T}_{R_2} \left(\frac{\pi}{2}, 0 \right) \hat{T}_{\mathcal{E}_{RE,2}} \hat{T}_Y \\ & + \frac{1 - \cos \phi + \sin \phi}{2} \hat{T}_{R_2} \left(\frac{3\pi}{2}, 0 \right) \hat{T}_{\mathcal{E}_{RE,2}} \hat{T}_Y, \end{aligned} \quad (D.406)$$

where $\hat{T}_{Y;M}$ is the mitigated transfer matrix; ϕ is the original noise amplitude of the rotational error channel; $\hat{T}_{\mathcal{E}_{RE,2}}$ is the error channel of the original noisy gate; $\hat{T}_Y$ is the noise-free component of the original noisy gate (so the transfer matrix corresponding to the original noisy gate is $\hat{T}_{\mathcal{E}_{RE,2}} \hat{T}_Y$); and $\hat{T}_{R_2} \left(\frac{\pi}{2}, 0 \right)$ and $\hat{T}_{R_2} \left(\frac{3\pi}{2}, 0 \right)$ are the transfer matrices equivalent to the custom noise channels used to generate the first and second noisy gate variants, respectively. The other mitigated gate transfer matrices are defined analogously,

Appendix

i.e. the indices of the error and custom channels should be changed to match that of the gate, the coefficients are unchanged (if the noise amplitude is unchanged). Our biases after mitigation are given by (see Eq. (D.404)):

$$\varepsilon_X = \left| \left(\vec{b} \right)_1 - \left(\vec{b}_M \right)_1 \right|, \quad \varepsilon_Y = \left| \left(\vec{b} \right)_2 - \left(\vec{b}_M \right)_2 \right|, \quad \varepsilon_Z = \left| \left(\vec{b} \right)_3 - \left(\vec{b}_M \right)_3 \right|. \quad (\text{D.407})$$

D.1.5 Local Cancellation

Local cancellation is implemented in a similar fashion to the local mitigation strategies (see App. D.1.4). We replace each of the noisy gate transfer matrices in our simulation with the locally cancelled equivalents:

$$\hat{T}_{n;\text{LC}} = \sum_{i=0}^{N_{\text{LC},n}} c_{\text{LC};n,i} \hat{T}_{n,i}, \quad (\text{D.408})$$

where $\hat{T}_{n;\text{LC}}$ is the transfer matrix corresponding to the locally cancelled gate, $N_{\text{LC},n}$ is the number of noisy gate variants used for the cancellation, $c_{\text{LC};n,i}$ is the coefficient for the i th variant of the n th gate, and $\hat{T}_{n,i}$ is the transfer matrix corresponding to the i th variant of the n th gate. We then treat the locally cancelled gates as if they were the original noisy gates for any further mitigation.

For example, if we are locally cancelling a gate suffering from a pure over-rotational error of noise amplitude ϕ then we can use the average of the original noisy gate and its hidden inverse. To be more explicit we can consider the $\hat{S}_{15}$ gate. The transfer matrices for the noisy gate and its hidden inverse are given by:

$$\hat{T}_{\mathcal{E}_{\text{RE},15}} \hat{T}_{\hat{S}_{15}} = \hat{T}_{R_{15}}(\phi, 0) \hat{T}_{\hat{S}_{15}}, \quad \hat{T}_{\mathcal{E}_{\text{RE},15}}^\dagger \hat{T}_{\hat{S}_{15}} = \hat{T}_{R_{15}}(-\phi, 0) \hat{T}_{\hat{S}_{15}}, \quad (\text{D.409})$$

respectively. Here $\hat{T}_{\hat{S}_{15}}$ is the noise-free component of the noisy gate, $\hat{T}_{\mathcal{E}_{\text{RE},15}} = \hat{T}_{R_{15}}(\phi, 0)$ is the original error channel of the noisy gate, and $\hat{T}_{\mathcal{E}_{\text{RE},15}}^\dagger = \hat{T}_{R_{15}}(-\phi, 0)$ is the error channel of the hidden inverse of the original noisy gate. So the locally cancelled gate has a transfer matrix given by (Eq. (D.384)):

$$\frac{1}{2} \hat{T}_{\mathcal{E}_{\text{RE},15}} \hat{T}_{\hat{S}_{15}} + \frac{1}{2} \hat{T}_{\mathcal{E}_{\text{RE},15}}^\dagger \hat{T}_{\hat{S}_{15}} = \hat{T}_{R_{15}} \left(0, \frac{1 - \cos \phi}{2} \right) \hat{T}_{\hat{S}_{15}}, \quad (\text{D.410})$$

i.e. the locally cancelled gate has an effective error channel that is a stochastic noise channel with noise amplitude $\frac{1 - \cos \phi}{2}$. So we have successfully cancelled the coherent components of the channel.

D.2 Precision and Code

For all our noise-aware mitigation strategies we consider mitigation orders (N_M) up to those sufficient to achieve unbiased mitigation. This serves to verify our analytic analyses and gives us confidence in our numerical results. However, to achieve unbiased mitigation for large circuits using synchronous mitigation methods we require coefficients and circuit simulations calculated to very high precision, to avoid rounding errors. We use up to 1500 decimal places for the calculations (see Table 7). We use Python for our simulations and make use of the high precision library mpmath and the symbolic library sympy (since the native operations such as tensor products and matrix-vector multiplications seem to run faster using sympy than mpmath matrices).

Appendix

D.3 Parameters

All the plots we use are characterised by two parameters: the noise level (e_G), and the number of gates (N_G). For simplicity we split the noise level into two components, only one of which is non-zero for any one plot:

$$e_G = e_{\text{SN}} + e_{\text{RE}}, \quad (\text{D.411})$$

where (Eqs (C.272) and (C.275) of App. C.1, respectively):

$$e_{\text{SN}} = 2N_G p, \quad e_{\text{RE}} = N_G |\phi|, \quad (\text{D.412})$$

where p is the gate-wise error probability, and ϕ is the angle of the rotation error. For simplicity, all the noise models we simulate are uniform. So, every noisy gate has the same noise amplitude (p and ϕ). The parameters for the plots in Sec. 5.3 are given in the Table 7. For all these plots we find the TIILM (App. C.12) parameters: $m_{n,0} = m_0$, $m_{n,1} = m_1$, and $m_{n,2} = m_2$ (i.e. the numbers of identity insertions to use for each noisy variant gate) using an optimisation procedure. We use the values that lead to the smallest runtime scaling. These optimised values are also presented in Table 7. We use 4 qubit ($N_Q = 4$) circuits. Since we are studying over-rotation errors, we have a Hermitian generator and therefore $a = 1$ for our stochastic noise channels (see Eq. (59) of Sec. 3.3).

Table 7: Plot Parameters: Table of plot parameters used for the figures in Sec. 5.3. We also include the optimised values used for $m_{n,0} = m_0$, $m_{n,1} = m_1$, and/or $m_{n,2} = m_2$ in TIILM (see App. C.12). The final two columns give the precision used to calculate the mitigation coefficients and perform the simulation, respectively, in terms of the number of decimal places.

Plot Ref.	N_G	N_Q	e_{SN}	e_{RE}	e_{GDN}	m_0	m_1	m_2	P. Coef	P. Sim.
Fig. 2a	18	4	0.02	0	0	0	246	N/A	500	1500
Fig. 2b	180	4	0.02	0	0	0	2478	N/A	1500	1500
Fig. 3a	18	4	2	0	0	0	16	N/A	500	1500
Fig. 3b	180	4	2	0	0	0	169	N/A	1500	1500
Fig. 4a	18	4	0	0.2	0	0	58	200	500	1500
Fig. 4b	180	4	0	0.2	0	0	589	2002	1500	1500
Fig. 5a	18	4	0	1.5	0	0	15	34	500	1500
Fig. 5b	180	4	0	1.5	0	0	152	341	1500	1500
Fig. 6a	18	4	0	0.2	0	0	1066	N/A	500	1500
Fig. 6b	180	4	0	0.2	0	0	34068	N/A	1500	1500
Fig. 7a	18	4	0	1.5	0	0	135	N/A	500	1500
Fig. 7b	180	4	0	1.5	0	0	4487	N/A	1500	1500