

# Quantum Error Mitigation

Zhenyu Cai,<sup>1,2,\*</sup> Ryan Babbush,<sup>3</sup> Simon C. Benjamin,<sup>1,2</sup> Suguru Endo,<sup>4</sup> William J. Huggins,<sup>3</sup> Ying Li,<sup>5</sup> Jarrod R. McClean,<sup>3</sup> and Thomas E. O’Brien<sup>3</sup>

<sup>1</sup>*Department of Materials, University of Oxford, Oxford, OX1 3PH, United Kingdom*

<sup>2</sup>*Quantum Motion, 9 Sterling Way, London N7 9HJ, United Kingdom*

<sup>3</sup>*Google Quantum AI, Venice, California 90291, USA*

<sup>4</sup>*NTT Computer and Data Science Laboratories, NTT Corporation, Musashino 180-8585, Japan*

<sup>5</sup>*Graduate School of China Academy of Engineering Physics, Beijing 100193, China*

(Dated: October 4, 2022)

For quantum computers to successfully solve real-world problems, it is necessary to tackle the challenge of *noise*: the errors which occur in elementary physical components due to unwanted or imperfect interactions. The theory of quantum fault tolerance can provide an answer in the long term, but in the coming era of ‘NISQ’ machines we must seek to mitigate errors rather than completely remove them. This review surveys the diverse methods that have been proposed for quantum error mitigation, assesses their in-principle efficacy, and then describes the hardware demonstrations achieved to date. We identify the commonalities and limitations among the methods, noting how mitigation methods can be chosen according to the primary type of noise present, including algorithmic errors. Open problems in the field are identified and we discuss the prospects for realising mitigation-based devices that can deliver quantum advantage with an impact on science and business.

## CONTENTS

I. Introduction	1	VI. Open Problems	28
II. Concepts	3	A. Overarching problems	28
A. Narrative introduction to concepts and terminology	3	B. Technical questions	28
B. Error-mitigated estimators	3	VII. Conclusion	29
C. Faults in the circuit	4	Acknowledgments	29
D. Exponential scaling of the sampling overhead	5	List of Symbols	29
III. Methods	5	A. Practical Techniques in Implementations	30
A. Zero-noise extrapolation	5	1. Monte carlo sampling	30
B. Probabilistic error cancellation	7	2. Pauli twirling	30
C. Measurement error mitigation	9	3. Measurement techniques	30
D. Symmetry constraints	11	References	31
E. Purity constraints	13		
F. Subspace expansions	15		
G. $N$ -representability	16		
H. Learning-based	18		
IV. Comparisons and Combinations	20		
A. Comparison among QEM methods	20		
1. Noise calibration overhead	20		
2. Mean square errors	21		
B. Benchmarking QEM from other perspectives	22		
1. State discrimination	22		
2. State extraction	22		
C. Combinations of QEM methods	23		
D. Comparison to the other error suppression methods	24		
V. Applications	25		
A. Coherent errors	25		
B. Logical errors in fault-tolerant quantum computation	27		
C. Compilation (algorithmic) errors	27		

## I. INTRODUCTION

The central promise of quantum computing is to enable algorithms that have been shown to provide both polynomial and super-polynomial speed-ups over the best-known classical algorithms for a special set of problems. These problems range from simulating quantum mechanics (Feynman, 1982) to purely algebraic problems such as factoring integers (Shor, 1999). The strongest challenge to the viability of practical quantum computing has always been its sensitivity to errors and noise. It was realised early on that the coupling of quantum systems to their environment sets an ultimate time limit and size limit for any quantum computation (Unruh, 1995). This constraint poses a formidable challenge to the ambitions of realising a quantum computer, since it set bounds on the scalability of any algorithm. With

Submitted to *Reviews of Modern Physics*.

Please address all correspondence to Zhenyu Cai.

\* [cai.zhenyu.physics@gmail.com](mailto:cai.zhenyu.physics@gmail.com)

the advent of quantum error correction (QEC) (Calderbank and Shor, 1996; Shor, 1995; Steane, 1996), this challenge has been solved at least in theory. The celebrated threshold theorem (Aharonov and Ben-Or, 1997; Kitaev, 1997) showed that if errors in the quantum hardware could be reduced below a finite rate, known as the threshold, a fault-tolerant quantum computation could be carried out for arbitrary length even on noisy hardware. However, besides the technical challenge of building hardware that achieves the threshold, the implementation of a fault-tolerant universal gate set with current codes, such as the surface code (Fowler *et al.*, 2012), generates a qubit overhead that seems daunting at the moment. For example, recent optimised approaches show that scientific applications that are classically intractable may require hundreds of thousands of qubits (Kivlichan *et al.*, 2020), while industrial applications will require millions of qubits (Lee *et al.*, 2021). There is ongoing theoretical research to find alternative codes with a more favourable overhead, and recent progress gives reasons to be optimistic (Breuckmann and Eberhardt, 2021; Dinur *et al.*, 2022; Gottesman, 2014; Pantelev and Kalachev, 2022). Nevertheless the challenge of realising full-scale fault-tolerant quantum computing is a considerable one.

This of course motivates the question of whether other approaches, prior to the era of fully fault-tolerant systems, might achieve quantum advantage with significant practical impacts. One might hope so, given the continual and remarkable progress that has been made in quantum computational hardware. In recent years, it has become routine to see reports of experiments demonstrating high-quality control over multiple qubits [see e.g. Asavanant *et al.* (2019); Ebadi *et al.* (2021); Jurcevic *et al.* (2021); Madjarov *et al.* (2020); and Xue *et al.* (2022)], some even reaching beyond 50 qubits [see e.g. Arute *et al.* (2019) and Wu *et al.* (2021)]. Meanwhile other experiments have indeed demonstrated early-stage fault-tolerant potentials [see e.g. Abobeih *et al.* (2022); Egan *et al.* (2021); Google Quantum AI (2022); Krinner *et al.* (2022); Postler *et al.* (2022); Ryan-Anderson *et al.* (2022); and Takeda *et al.* (2022)]. Of course, the works mentioned here are far from exhaustive as it is impossible to capture the all the breakthroughs on different fronts across the diverse range of platforms, we refer the reader to Acín *et al.* (2018) and Altman *et al.* (2021) and the references therein for the key milestones in different platforms.

The primary goal of *quantum error mitigation* (QEM) is to translate this continuous progress in quantum hardware into immediate improvements for quantum information processing. While accepting that the hardware imperfections will limit the complexity of quantum algorithms, nevertheless we can expect that every advance should enable this boundary to be pushed further. As this review will demonstrate, the mitigation approach indeed proves to be both practically effective and quite

fascinating as an intellectual challenge.

When exploring the prospects for achieving quantum advantage through error mitigation, it is crucial to consider suitable forms of circuits. It is understood that in the era of noisy, intermediate-scale quantum (NISQ) devices, only certain approaches may be able to achieve meaningful and useful results. Due to the limited coherence times and the noise floor present in quantum hardware, one typically resorts to the idea of quantum computation with short-depth circuits. Motivating examples include variational quantum circuits in physics simulations (McClean *et al.*, 2016; Peruzzo *et al.*, 2014; Wecker *et al.*, 2015), approximate optimisation algorithms (Farhi *et al.*, 2014), and even heuristic algorithms for quantum machine learning (Biamonte *et al.*, 2017). Typically in applications of these kinds, the algorithm can be understood as applying a short-depth quantum circuit to a simple initial state and then estimating the expectation value of a relevant observable. Such expectation values ultimately lead to the output of the algorithm, which must be accurate enough to be useful in some context (for example, for estimating the energies of molecular states, a useful level of *chemical accuracy* corresponds to 1kcal/mol (Helgaker *et al.*, 2000)). This leads to the most essential feature of QEM: the ability to minimise the noise-induced bias in expectation values on noisy hardware. However, this can also be achieved by QEC and many other long-established tools like decoherence-free subspaces and dynamical decoupling sequences (derived from optimal quantum control) (Lidar, 2014; Suter and Álvarez, 2016). Therefore this feature alone is not sufficient to capture the QEM techniques that we wish to cover in this review.

It is challenging to find a universally acceptable definition of quantum error mitigation. For the purposes of this review, we will define the term ‘*quantum error mitigation*’ as algorithmic schemes that reduce the noise-induced bias in the expectation value by post-processing outputs from an ensemble of circuit runs, using circuits at the *same noise level* as the original unmitigated circuit *or above*. That is to say QEM will only reduce the effective damage due to noise for *the whole ensemble of circuit runs* (with the help of post-processing), but when we zoom into each individual circuit run, the circuit noise level remains unchanged or even increases. This is in contrast to the other techniques like QEC which aim to reduce the effect of noise on the output in *every single circuit run*.

Since QEM performs post-processing using data directly from noisy hardware, it will become impractical if the amount of noise in the whole circuit is so large that it completely damages the output. In practice, this usually means that for a given hardware set-up, there is a maximum circuit size (circuit depth times qubit number) beyond which QEM will become impractical, usually due to an infeasible number of circuit repetitions. In contrast

to QEC, there is no specific error threshold that one must surpass before QEM can be useful; different qubit operation error rates simply lead to different circuit sizes for which QEM will be practical. In other words, as quantum hardware continues to advance, we will be able to apply QEM to ever larger quantum circuits for more challenging applications, without requiring big jumps in the technologies.

There are certain desirable features for error mitigation; the methods reviewed here meet the following criteria to differing extents. First, the mitigation method should ideally only require a very modest qubit overhead to remain practical on current and near-term quantum hardware. Nevertheless, error mitigation techniques should provide an accuracy guarantee for the method. Such a guarantee should ideally provide a formal error bound on the mitigated expectation values that indicate how well the method works at varying levels of noise. The bounds would then indicate which concrete hardware improvements would lead to improved estimates. Needless to say, methods that are conceptually simple and easy to implement experimentally lead to practically feasible approaches. Lastly, a reliable error mitigation method should require few assumptions (or, no assumptions) about the final state that is prepared by the computation. Making strong assumptions about the final state, for example that the state is a product state, may restrict the method to scenarios where a computational advantage over classical approaches may not be given.

We will start by introducing the basic notion of QEM in Sec. II, with the details of different QEM techniques presented in Sec. III. The comparison and combinations of these individual techniques will then be discussed in Sec. IV. In Sec. V we will explore the application of QEM in different noise scenarios. Finally, we will discuss the open problems in the field in Sec. VI and offer a conclusion in Sec. VII

## II. CONCEPTS

### A. Narrative introduction to concepts and terminology

In this section we will introduce certain key concepts and terminology that are common to all the QEM methods. However, do note that the approach used in this section might not be the native way for introducing individual techniques in Sec. III. In those cases, we will keep terminology that is unique to the given technique self-contained in the respective section.

Near-term quantum devices have imperfections that degrade the desired output information. A QEM protocol will aim to minimise this degradation. We will use the term *primary circuit* to describe the process that, ideally, would produce the *perfect output state*  $\rho_0$  whose properties we are interested in. In practice, due to the

noise present in the primary circuit, the actual output state is some *noisy state*  $\rho$  instead.

Typically the ideal output information we seek is the expectation value of some *observable of interest*  $O$  of the ideal output  $\rho_0$ . Commonly we would obtain this information simply by averaging the measurement results of repeated execution, as opposed to, say, some phase estimation techniques that can obtain the result through single-shot measurements but require deeper circuits that are more relevant to the fault-tolerant computing era. Therefore, even if we had ideal hardware, we would still need to perform repeated executions to determine the average. We will use  $N_{\text{cir}}$  to denote the *number of circuit executions*, or ‘shots’, that we employ – this will include any executions of variant circuits called for in the QEM protocol. Even in the noiseless limit, the finite  $N_{\text{cir}}$  will usually imply a finite inaccuracy in our estimated average, often called *shot noise*. However, with perfect noiseless hardware, there would be zero *bias*. In other words, there would be no systematic shift to the estimated mean versus the true value (the infinite sampling limit). Given that our hardware is not perfect, there will generally be finite bias. QEM protocols aim to reduce this bias, but this often means an increase in the variance (for a fixed number of circuit executions  $N_{\text{cir}}$ ). One could increase  $N_{\text{cir}}$  to compensate but this cost should be acknowledged; the cost is the *sampling overhead* of that error-mitigation method versus the ideal noiseless case. In the following subsections we will make these terms and concepts more precise.

### B. Error-mitigated estimators

Our goal is to estimate the expectation value  $\text{Tr}[O\rho_0]$  of some observable of interest  $O$ . Using the outputs of the primary circuit and its variants, we can construct an estimator  $\hat{O}$  for our target parameter  $\text{Tr}[O\rho_0]$ . The quality of a given estimator can be assessed in different ways. One way is to use prediction intervals, which calculate the interval within which the outcome of the estimator will fall with a given probability, offering a rigorous bound on the worst-case deviation of the estimator. Here however, in order to see the different factors that contribute to the deviation of the estimator more clearly, we will instead focus on the expected (average-case) square deviation of our estimator  $\hat{O}$  from the true value  $\text{Tr}[O\rho_0]$ , which is called the *mean square error*:

$$\text{MSE}[\hat{O}] = \text{E}[(\hat{O} - \text{Tr}[O\rho_0])^2]. \quad (1)$$

The ultimate goal of error mitigation is to reduce  $\text{MSE}[\hat{O}]$  as much as possible, but this needs to be achieved using only finite resources. To quantify this, it is useful to decompose the mean square error of an estimator into two

components, the bias and the variance of the estimator:

$$\text{MSE}[\hat{O}] = \text{Bias}[\hat{O}]^2 + \text{Var}[\hat{O}]$$

with the bias and the variance defined as:

$$\text{Bias}[\hat{O}] = \text{E}[\hat{O}] - \text{Tr}[O\rho_0]$$

$$\text{Var}[\hat{O}] = \text{E}[\hat{O}^2] - \text{E}[\hat{O}]^2.$$

In this review when we say ‘bias’, sometimes we are referring to the magnitude of the bias  $|\text{Bias}[\hat{O}]|$ , the exact meaning should be obvious from the context.

The simplest way to construct the estimator  $\hat{O}$  is by directly measuring  $O$  on the *noisy* output state of the primary circuit  $\rho$ , and the measurement output is simply denoted using the random variable  $\hat{O}_\rho$ . After running the noisy primary circuit  $N_{\text{cir}}$  times, we can take the average of these noisy outputs (obtaining the noisy sample mean) to estimate the ideal expectation value  $\text{Tr}[O\rho_0]$ . This noisy sample mean estimator is denoted as  $\bar{O}_\rho$  and its mean square error is given by

$$\text{MSE}[\bar{O}_\rho] = \underbrace{(\text{Tr}[O\rho] - \text{Tr}[O\rho_0])^2}_{\text{Bias}[\bar{O}_\rho]=\text{Bias}[\hat{O}_\rho]} + \underbrace{\frac{\text{Tr}[O^2\rho] - \text{Tr}[O\rho]^2}{N_{\text{cir}}}}_{\text{Var}[\bar{O}_\rho]=\text{Var}[\hat{O}_\rho]/N_{\text{cir}}}. \quad (2)$$

We can see that the error contribution due to the variance, which is often called *shot noise*, will reduce as we increase the number of circuit runs  $N_{\text{cir}}$ . In the limit of a large number of circuit executions, the mean square error  $\text{MSE}[\bar{O}_\rho]$  will be mainly limited by the bias of the estimator  $|\text{Bias}[\bar{O}_\rho]|$ , which is a systematic error that cannot be reduced by increasing the number of circuit runs.

In order to reduce the bias, we can apply QEM using data obtained from the noisy primary circuit and its variants, as will be discussed in Sec. III. We will construct an error-mitigated estimator  $\bar{O}_{\text{em}}$  using the *same number of circuit runs*  $N_{\text{cir}}$ . We would want to construct the error-mitigated estimator  $\bar{O}_{\text{em}}$  in such a way that it can achieve a smaller bias than the naive noisy estimator  $\bar{O}_\rho$ ,

$$|\text{Bias}[\bar{O}_{\text{em}}]| \leq |\text{Bias}[\bar{O}_\rho]|.$$

This reduction in the bias is usually achieved by constructing a more complex estimator that extracts and amplifies the useful information buried within the noise. As a result, the error-mitigated estimator is also more sensitive to the variation in the sampled data, and thus its variance will usually increase,

$$\text{Var}[\bar{O}_{\text{em}}] \geq \text{Var}[\bar{O}_\rho].$$

Such a bias-variance trade-off is illustrated in Fig. 1 and can be found in almost all areas of parameter estimation. As we will see later, different ways of performing error mitigation often lead to different trade-offs between bias

and variance, giving the user a choice between a quickly-converging QEM method with large residual error, and one that is more costly but more accurate.

We can define an ‘one-shot’ error-mitigated estimator  $\hat{O}_{\text{em}}$  that will satisfy  $\text{E}[\hat{O}_{\text{em}}] = \text{E}[\bar{O}_{\text{em}}]$  and  $\text{Var}[\hat{O}_{\text{em}}] = N_{\text{cir}}\text{Var}[\bar{O}_{\text{em}}]$ . The number of circuit runs needed for given estimator  $\hat{X}$  to achieve the shot noise level  $\epsilon$  is given by  $N_{\text{shot}}^\epsilon(\hat{X}) = \text{Var}[\hat{X}]/\epsilon^2$ . To reach the same shot noise level as the original noisy estimator, the error-mitigated estimator will require more circuit runs. This factor of increase in the number of circuit runs is called the *sampling overhead*, which is given by:

$$C_{\text{em}} = \frac{N_{\text{shot}}^\epsilon(\hat{O}_{\text{em}})}{N_{\text{shot}}^\epsilon(\hat{O}_\rho)} = \frac{\text{Var}[\hat{O}_{\text{em}}]}{\text{Var}[\hat{O}_\rho]}. \quad (3)$$

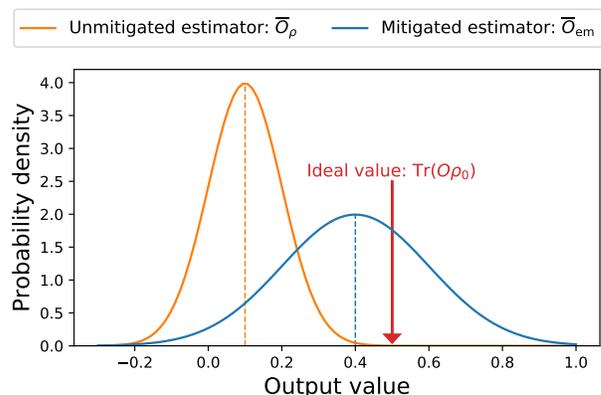


FIG. 1 The probability density distributions of the unmitigated estimator and the error-mitigated estimator. We see a decrease of bias and an increase of variance after performing error mitigation.

The sampling overhead can also be estimated using the range of the estimator through Hoeffding’s inequality. The *range* of a random variable  $\hat{X}$ , denoted as  $\text{R}[\hat{X}]$ , is the difference between the maximum and minimum possible values taken by  $\hat{X}$ . The smallest number of samples required to estimate  $\text{E}[\hat{X}]$  to  $\epsilon$ -precision with  $1 - \delta$  probability using Hoeffding’s inequality is given by

$$N_{\text{hff}}^{\epsilon,\delta}(\hat{X}) = \frac{\ln(2/\delta)}{2\epsilon^2} \text{R}[\hat{X}]^2 \quad (4)$$

which can be used to estimate the sampling overhead required for the error-mitigated estimator  $\hat{O}_{\text{em}}$  to achieve the same  $\epsilon$  and  $\delta$  as the unmitigated estimator  $\hat{O}_\rho$ :

$$C_{\text{em}} = \frac{N_{\text{shot}}^\epsilon(\hat{O}_{\text{em}})}{N_{\text{shot}}^\epsilon(\hat{O}_\rho)} \sim \frac{N_{\text{hff}}^{\epsilon,\delta}(\hat{O}_{\text{em}})}{N_{\text{hff}}^{\epsilon,\delta}(\hat{O}_\rho)} = \frac{\text{R}[\hat{O}_{\text{em}}]^2}{\text{R}[\hat{O}_\rho]^2}. \quad (5)$$

### C. Faults in the circuit

To gain intuition about the performance and costs of QEM, we need a way to quantify the damages due to

noise in the circuit. Doing this by modelling noise in complete generality on a quantum system can be very challenging (Lidar and Brun, 2013). One useful approximation, which is widely employed in the field of quantum error correction, is to model noise as discrete, probabilistic events named *faults* that can occur at the various *locations* in the circuit including gates, idling steps and measurements (Gottesman, 2009; Terhal, 2015). Assuming the error events in different locations are *independent*, if location  $f$  has the error probability of  $p_f$ , then the total probability that there is no fault in the circuit is given by

$$P_0 = \prod_f (1 - p_f), \quad (6)$$

which we will simply call the *fault-free probability* of the circuit. It decays exponentially with the increase of the number of fault locations in the circuit (and thus the number of qubits and the circuit depth). Do note that the fault-free probability is *not* the fidelity of the output state, but it is a lower bound for the fidelity. This is because the faults may not send the state to an orthogonal state, and the faults also may cancel each other.

We can also quantify the amount of noise in the circuit using the average number of faults in each circuit run, which is given by:

$$\lambda = \sum_f p_f, \quad (7)$$

and will be called the *circuit fault rate*. In the simple case that all  $M$  fault locations in the circuit have the same error rate  $p$ , we then simply have  $\lambda = Mp$ . If the circuit contains a large number (more than dozens) of fault locations, and the circuit fault rate is of the order of unity  $\lambda \sim 1$ , then the number of faults occurring in a given circuit run can be modelled using a *Poisson distribution* with mean  $\lambda$  using the Le Cam’s theorem (Cai, 2021a; Endo *et al.*, 2018; Le Cam, 1960), i.e. the probability that  $\ell$  faults occur in the circuit is given by  $P_\ell = e^{-\lambda} \lambda^\ell / \ell!$ . In this way, the fault-free probability is given by

$$P_0 = e^{-\lambda}, \quad (8)$$

which decay exponentially with the circuit fault rate.

#### D. Exponential scaling of the sampling overhead

We can perform *bias-free* QEM if we are able to post-select for the fault-free circuit runs without needing any additional circuit components. The fraction of circuit runs that are selected is simply given by the fault-free probability  $e^{-\lambda}$  in Eq. (8), which means that we will still require  $e^\lambda$  times more circuit runs to obtain the same number of “effective” circuit runs as a noise-free machine and achieve the same level of shot noise. Hence, even

allowing for the “magical” post-selection of fault-free circuit runs, the sampling overhead  $C_{\text{em}} = e^\lambda$  will still increase exponentially with the circuit fault rate (and thus the circuit size).

Now let us move beyond trying to extract the error-free state and instead focus on obtaining the right expectation value for the observable of interest. Without loss of generality, we will assume all the observables that we measure are traceless and have bounded spectrums. Cai (2021a) and Wang *et al.* (2021b) have shown that the expectation value of such observables under Pauli gate noise is bounded by an exponential decay curve against the increase of the circuit fault rate  $\lambda$ . As a result, if we are looking at an error-mitigation protocol whose error-mitigated estimator  $\hat{O}_{\text{em}}$  is the *linear combination* of the output of a set of noisy circuits with the circuit fault rate being  $\lambda$  or more, then we have:

$$\text{E}[\hat{O}_{\text{em}}] = \mathcal{O}(e^{-\beta\lambda})$$

for some positive  $\beta$  (Wang *et al.*, 2021a). In order to resolve such an exponentially small quantity, we will need an exponential number of samples, similar to what we obtain when trying to extract the error-free state.

For a given noisy circuit with the circuit fault rate  $\lambda$ , rather than performing active correction to reduce  $\lambda$  in each circuit run like in quantum error correction, QEM relies on post-processing the outputs from an ensemble of circuit runs with the same circuit fault rate  $\lambda$  or above. Hence, following the arguments above, we see that QEM cannot efficiently tackle noisy circuits with large  $\lambda$  *on its own* due to the exponential sampling overhead. However, as we will see later, due to the much lower implementation cost for QEM in terms of additional circuit components and qubits, it has become an effective means to stretch the application potential of near-term noisy devices and will be a useful tool to help alongside quantum error correction in the longer term.

### III. METHODS

After introducing the overall concept of QEM, we will now look at how the various error-mitigated estimators are actually constructed through performing different QEM methods.

#### A. Zero-noise extrapolation

In this section, we will make use of noisy states obtained at different circuit fault rates. The state obtained at the circuit fault rate  $\lambda$  will be denoted as  $\rho_\lambda$ . The noisy expectation value  $\text{Tr}[O\rho_\lambda]$  can be viewed a function of  $\lambda$ . In this way, the ideal expectation value we want is just the value of the function at  $\lambda = 0$ . Trying to obtain this zero-noise value using data points at different

circuit fault rates brings us to the concept of *zero-noise extrapolation* (also called *error extrapolation*), which was first introduced by Li and Benjamin (2017) and Temme *et al.* (2017).

Using  $\lambda_1$  to denote the smallest circuit fault rate we can achieve, we can probe  $\text{Tr}[O\rho_\lambda]$  at a range of *boosted* error rate  $\{\lambda_m\}$  with  $\lambda_m < \lambda_{m+1}$  to obtain a set of data points  $\{(\lambda_m, \text{Tr}[O\rho_{\lambda_m}])\}$ . We can model  $\text{Tr}[O\rho_\lambda]$  using a parametrised function  $f(\lambda; \vec{\theta})$  and fit it to the data points to obtain a set of optimal parameters  $\vec{\theta}^*$ . The error-mitigated estimate of the zero-noise output  $\text{Tr}[O\rho_0]$  is then given by

$$\mathbb{E}[\hat{O}_{\text{em}}] = f(0; \vec{\theta}^*).$$

A simple illustration of zero-noise extrapolation is shown in Fig. 2.

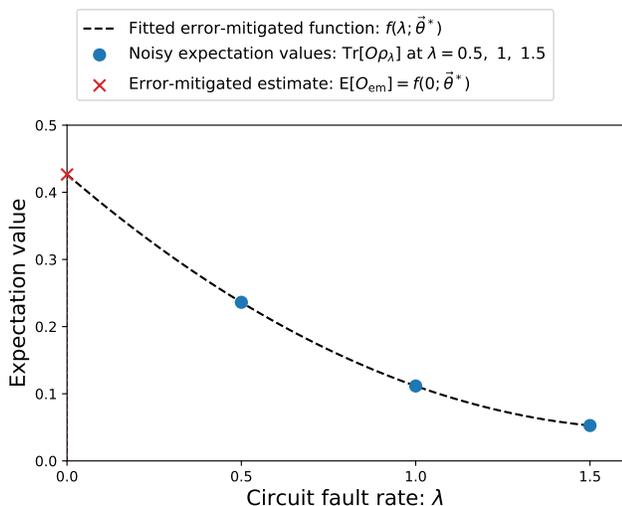


FIG. 2 Obtaining the error-mitigated estimate using zero-noise extrapolation. Here we are performing extrapolation using three noisy expectation values at the circuit fault rate of  $\lambda = 0.5, 1, 1.5$ , where 0.5 is the lowest circuit fault rate we can achieve and the other two are obtained through boosting the noise in the device.

If  $\lambda$  is small, Temme *et al.* (2017) showed that  $\text{Tr}[O\rho_\lambda]$  can be approximated using a polynomial function in the same spirit as truncated Taylor expansion:

$$\text{Tr}[O\rho_\lambda] \approx f(\lambda; \vec{\theta}) = \sum_{\ell=0}^{M-1} \theta_\ell \frac{\lambda^\ell}{\ell!}. \quad (9)$$

Here we have a polynomial of degree  $M - 1$ , which has  $M$  different free parameters and the zero-noise estimate we want is  $\theta_0^* = f(0; \vec{\theta}^*)$ . The simplest case is linear extrapolation with  $M = 2$  (Li and Benjamin, 2017). If we try to fit Eq. (9) with  $M$  data points, which is the minimal number of data points needed, we can perform Richardson extrapolation as discussed in Temme *et al.*

(2017). The corresponding error-mitigated estimate obtained using the set of data points  $\{(\lambda_m, \text{Tr}[O\rho_{\lambda_m}])\}$  is

$$\mathbb{E}[\hat{O}_{\text{em}}] = \theta_0^* = \sum_{m=1}^M \text{Tr}[O\rho_{\lambda_m}] \prod_{k \neq m} \frac{\lambda_k}{\lambda_k - \lambda_m}. \quad (10)$$

The bias in our estimate is mostly due to our omission of the higher degree terms in the polynomial approximation, and thus we should expect  $\text{Bias}[\hat{O}_{\text{em}}] = \mathcal{O}(\lambda^M)$ .

The error-mitigated expectation value in Eq. (10) is a linear combination of the set of noisy expectation values  $\{\text{Tr}[O\rho_{\lambda_m}]\}$ , thus it can be estimated using the Monte Carlo sampling method (see Appendix A.1) and the corresponding sampling overhead for Richardson extrapolation is given by:

$$C_{\text{em}} \sim \left( \sum_{m=1}^M \left| \prod_{k \neq m} \frac{\lambda_k}{\lambda_k - \lambda_m} \right| \right)^2. \quad (11)$$

We see that if any of the probed circuit fault rates  $\lambda_m$  is too big or the gap between any two data points  $|\lambda_m - \lambda_k|$  is too small, then  $C_{\text{em}}$  will blow up and the extrapolation would become infeasible. For the simple case of equal-gap Richardson extrapolation ( $\lambda_m = m\lambda_1$ ), the sampling overhead in Eq. (11) becomes  $C_{\text{em}} \sim (2^M - 1)^2$ , which grows exponentially with the number of data points  $M$ .

As mentioned, theoretically Richardson extrapolation will only be valid for small  $\lambda$  due to the approximation we made in Eq. (9). However, a recent experiment by Kim *et al.* (2021) showed that Richardson extrapolation can be effective at large  $\lambda$  in practice. There they performed a 26-qubit simulation of the 2D transverse field Ising model in the limit of strong entanglement, and showed that the error-mitigated evolution of the magnetisation is competitive in comparison to standard tensor network methods. To look for an extrapolation method that is naturally compatible with large  $\lambda$ , we can consider a large circuit with Pauli noise. Such a circuit in the limit of  $\lambda \rightarrow \infty$  will become a random circuit with *zero* expectation value for bounded traceless observables as mentioned in Sec. II.D. A polynomial extrapolation function diverges at  $\lambda \rightarrow \infty$  and thus does not fit the intuition above, which motivates the extrapolation schemes that use an exponential decay curve or even a multi-exponential decay curve instead. Exponential extrapolation has proven to be able to achieve smaller biases than Richardson and more generally polynomial extrapolation in some numerical simulations and cloud experiments (Endo *et al.*, 2018; Giurgica-Tiron *et al.*, 2020), especially for Pauli noise (Cai, 2021a). It is worth noting that going beyond Richardson extrapolation, the error-mitigated estimator obtained through least-square fitting often does not have a closed-form representation, and thus there are not yet analytical expressions for their biases and variance in most of the cases. It is possible to combine exponential and Richardson extrapolation by performing Richardson extrapolation on

the function  $e^\lambda \text{Tr}[O\rho_\lambda]$  instead of  $\text{Tr}[O\rho_\lambda]$ , which can give explicit bounds on the sampling overheads and biases (Cai, 2021b).

When we try to boost the noise in the circuit, we must know the exact factor of increase in the circuit fault rate. Ideally, we would want to increase the error strength in the various fault locations without changing their error models, which can be challenging to implement in experiments. Temme *et al.* (2017) showed that under the assumption that the noise was time-invariant, the noise could be effectively amplified by stretching and recalibrating the gate pulses, which was demonstrated on a superconducting platform (Kandala *et al.*, 2019). When targeting only single-qubit gate errors, these experiments can successfully perform Richardson extrapolation up to the fourth order, highlighting the accuracy of the noise amplification. Rescaling noise for two-qubit cross-resonance gates in these architectures can be more challenging due to reasons like more complicated drive Hamiltonian and drive-dependent coherence times (Chow *et al.*, 2011). Nevertheless, pulse-stretching experiments with slowed-down two-qubit cross-resonance gates have been shown to be effective with linear extrapolation (Kandala *et al.*, 2019), most recently demonstrated in experiments with up to 26 qubits and a depth of 120 (Kim *et al.*, 2021).

Alternatively, Dumitrescu *et al.* (2018); Giurgica-Tiron *et al.* (2020); and He *et al.* (2020) tried to boost the circuit fault rate by inserting a sequence of abundant gates that are equivalent to the identity if operated noiselessly. This is easier to implement and calibrate compared to the pulse-stretching method. However, while this will work well with depolarising gate noise, it may change the gate error model for more general gate noise such that the factor of increase in the error strength is no longer well-defined (Kim *et al.*, 2021). Note that for both pulse stretching and gate insertion, an  $M$ -time increase in the noise strength might lead to up to an  $M$ -time increase in the circuit runtime. If the error models at the various fault locations are completely known, Li and Benjamin (2017) suggested that it is possible to controllably amplify the noise by probabilistically inserting gates to simulate the faults. In fact, as noted in Cai (2021a), it would be more efficient to use these probabilistically inserted gates to perform probabilistic error cancellation (Sec. III.B) instead, which yields new data points at *reduced* error strength. Of course, learning a representative noise model can be challenging for large-scale devices, especially for correlated noise.

Due to the simplicity of zero-noise extrapolation, it is one of the most widely implemented QEM methods. In addition to the experiments mentioned above, it was also successfully demonstrated in a wide range of other experiments, especially through cloud platforms [see e.g. Garmon *et al.* (2020); Keen *et al.* (2020); Klco *et al.* (2018); Tacchino *et al.* (2020); and Yeter-Aydeniz *et al.* (2020)].

## B. Probabilistic error cancellation

An alternative quantum error mitigation method referred to as *probabilistic error cancellation* was introduced by Temme *et al.* (2017). A particular feature of this method is that it can fully remove the bias of expectation values of generic quantum circuits ( $\text{Bias}[\hat{O}_{em}] = 0$ ). This comes at the expense of a sampling overhead  $C_{em}$  that grows exponentially with the circuit fault rate  $\lambda$ . The key idea is noting that the noise-free expectation value can be written as a linear combination of expectation values from a set of noisy quantum circuits. This real-valued, linear combination can be interpreted as a *quasi-probability decomposition* which can be sampled, c.f. Appendix A.1, according to a Monte Carlo procedure. In this review we formalise the method using the superoperator representation (Gilchrist *et al.*, 2011), in which density matrices  $\rho$  are vectorised into  $|\rho\rangle\rangle$  and quantum channels are written as matrices acting on  $|\rho\rangle\rangle$ , denoted using the curly font like  $\mathcal{U}$ . Taking the trace with the observable  $O$  is then written as the inner product with the vectorised observable  $\text{Tr}[O\rho] = \langle\langle O|\rho\rangle\rangle$ .

To construct an estimator for an ideal channel  $\mathcal{U}$  from noisy operations, we need to choose a set of *noisy* basis operations  $\{\mathcal{B}_n\}$  that we can implement in the physical hardware. These operations are for example noisy gates, state preparation operations and measurements. These operations are assumed to be learned from the noisy hardware in experiments through some form of tomography. An example of a complete set of basis operations is discussed in Endo *et al.* (2018). With a sufficiently large basis, we can decompose the ideal operation into

$$\mathcal{U} = \sum_n \alpha_n \mathcal{B}_n. \quad (12)$$

with real coefficients  $\alpha_n$ . Some of the coefficients  $\alpha_n$  in this expansion can be negative, which means that this decomposition does not necessarily correspond to a probabilistic mixture of physical maps. The expansion therefore is often not a physical map that can be implemented directly. However, if we are applying  $\mathcal{U}$  on some input state  $\rho_{\text{in}}$  in order to measure some observable  $O$  and obtain the ideal expectation value  $\text{Tr}[O\rho_0] = \langle\langle O|\rho_0\rangle\rangle$ , then the ideal expectation value can be decomposed into:

$$\langle\langle O|\rho_0\rangle\rangle = \langle\langle O|\mathcal{U}|\rho_{\text{in}}\rangle\rangle = \sum_n \alpha_n \langle\langle O|\mathcal{B}_n|\rho_{\text{in}}\rangle\rangle, \quad (13)$$

i.e. the ideal expectation value  $\langle\langle O|\mathcal{U}|\rho_{\text{in}}\rangle\rangle$  can be decomposed into a linear combination of noisy expectation values  $\{\langle\langle O|\mathcal{B}_n|\rho_{\text{in}}\rangle\rangle\}$  that we can obtain individually.

When the expansion in Eq. (13) has many terms, the linear combination of noisy expectation values in Eq. (13) can be estimated using the Monte Carlo sampling method (Appendix A.1). The method samples different basis operations  $\mathcal{B}_n$  according to their weight in the expansion,

i.e. the noisy circuit corresponding to  $\mathcal{B}_n$  is chosen with the probability  $|\alpha_n|Q^{-1}$ , where  $Q = \sum_n |\alpha_n|$ , and the circuit output is multiplied with  $\text{sign}(\alpha_n)Q$  before being used to estimate the error-mitigated expectation value  $E[\hat{O}_{\text{em}}]$ . This multiplicative factor leads to an increase in the variance and correspondingly to a sampling overhead (given by Eq. (A1)):

$$\begin{aligned} \text{Var}[\hat{O}_{\text{em}}] &\sim Q^2 \text{Var}[\hat{O}_{\rho_\lambda}] \\ C_{\text{em}} &\sim Q^2 = \left( \sum_n |\alpha_n| \right)^2. \end{aligned} \quad (14)$$

In practice, we can often implement the noisy version of the target operation, denoted as  $\mathcal{U}_p$ , which can be written as  $\mathcal{U}_p = (1-p)\mathcal{U} + p\mathcal{N}$  where  $p$  is the operation error rate and  $\mathcal{N}$  is the noise element. It can be rewritten as:

$$\mathcal{U} = \frac{1}{1-p}\mathcal{U}_p - \frac{p}{1-p}\mathcal{N}. \quad (15)$$

Compared to Eq. (12), we see that  $\mathcal{U}_p$  is one of the basis operations  $\mathcal{U}_p = \mathcal{B}_1$ . In the simple case that  $\mathcal{N}$  is also one of the basis operations that we can implement:  $\mathcal{N} = \mathcal{B}_2$ , the sampling overhead for removing the noise in  $\mathcal{U}_p$  is given by:

$$C_{\text{em}} \sim Q^2 = \left( \frac{1}{1-p} + \frac{p}{1-p} \right)^2 = \left( \frac{1+p}{1-p} \right)^2. \quad (16)$$

More generally,  $\mathcal{N}$  must be decomposed into the rest of the basis  $\mathcal{N} = \sum_{n \neq 1} w_n \mathcal{B}_n$ . For the common scenario in which  $\mathcal{U}_p$  suffers from Pauli noise and we can perform high fidelity Pauli gates to be used as our basis  $\{\mathcal{B}_n\}$ , the sampling overhead is still given by Eq. (16).

Up to now we are using the basis set to cancel out the noisy component  $\mathcal{N}$  in  $\mathcal{U}_p = (1-p)\mathcal{U} + p\mathcal{N}$ . Alternatively, there is another means of decomposition to effectively “invert” the noise channel  $\mathcal{E}_p$  in  $\mathcal{U}_p = \mathcal{E}_p\mathcal{U}$  as discussed in [Temme et al. \(2017\)](#) and [Endo et al. \(2018\)](#) and its resultant sampling overhead is similar. However, this noise-inversion implementation requires inserting additional operations after every noisy operation and thus we might need up to twice the original circuit runtime.

To perform the error cancellation described above, we need to have the full description of the noisy operation  $\mathcal{U}_p$ . This can only be efficiently characterised for individual local gates. Any circuit we want to implement can be decomposed into a sequence of  $M$  ideal gates  $\{\mathcal{U}_m\}$ , and the ideal expectation value we want to obtain will be

$$\langle\langle O | \rho_0 \rangle\rangle = \langle\langle O | \prod_{m=1}^M \mathcal{U}_m | \rho_{\text{in}} \rangle\rangle.$$

Similar to Eq. (13), we decompose the individual gates into the noisy basis using Eq. (12)

$$\begin{aligned} \langle\langle O | \rho_0 \rangle\rangle &= E[\hat{O}_{\text{em}}] = \langle\langle O | \prod_{m=1}^M \left( \sum_{n_m} \alpha_{mn_m} \mathcal{B}_{n_m} \right) | \rho_{\text{in}} \rangle\rangle \\ &= \sum_{\vec{n}} \alpha_{\vec{n}} \langle\langle O | \mathcal{B}_{\vec{n}} | \rho_{\text{in}} \rangle\rangle \end{aligned} \quad (17)$$

where  $\vec{n} = \{n_1, n_2, \dots, n_M\}$  is the set of labels for a sequence of basis elements, and we have defined  $\mathcal{B}_{\vec{n}} = \prod_{m=1}^M \mathcal{B}_{n_m}$ ,  $\alpha_{\vec{n}} = \prod_{m=1}^M \alpha_{mn_m}$  and  $Q = \sum_{\vec{n}} |\alpha_{\vec{n}}|$ . Note that the set of noisy basis elements  $\{\mathcal{B}_n\}$  here includes the basis for all the gates in the circuit, it is thus over-complete and contains the noisy version of all the target gates. Again we can obtain samples of  $\hat{O}_{\text{em}}$  using Monte Carlo sampling as discussed above with the label of a single basis  $n$  replaced by the label of a sequence of basis  $\vec{n}$ .

The overall sampling overhead of mitigating the errors in the whole circuit is simply the product of the sampling overhead of each gate. As explored in [Temme et al. \(2017\)](#) and subsequently [Endo et al. \(2018\)](#), if we assume all the gates suffer from Pauli noise with the same error rate  $p$ , the circuit size  $M$  is large and the circuit fault rate  $\lambda = Mp$  is finite, then taking the product of the gate-level sampling overhead in Eq. (16) we have:

$$C_{\text{em}} = \prod_{m=1}^M \left( \frac{1+p}{1-p} \right)^2 \approx e^{4Mp} = e^{4\lambda}. \quad (18)$$

In fact, the overhead here is still valid even if the gates in the circuit have different error rates, as long as the circuit faults follow a Poisson distribution (see Sec. II.C). More generally, it is possible to only apply probabilistic error cancellation partially. Denoting the resultant circuit fault rate as  $\lambda_{\text{em}}$ , the corresponding sampling cost is simply  $C_{\text{em}} = e^{4(\lambda - \lambda_{\text{em}})}$ , which grows exponentially with the reduction in the circuit fault rate ([Cai, 2021a](#)).

As mentioned, to be able to perform the decomposition in Eq. (12), we need to have a set of basis elements that span the ideal operations, and full characterisation of this basis. [Endo et al. \(2018\)](#) have shown that such basis can be constructed using single-qubit Clifford gates and  $Z$ -measurement with reasonable fidelity, while the characterisation can be carried out efficiently using gate set tomography. In practice, we usually supplement this set of basis elements with the noisy version of the target operation, such that it is over-complete as discussed. Only noisy operations with local noise can be efficiently characterised using the protocol above.

A recent experimental implementation of the probabilistic error cancellation method used a correlated Pauli-noise model that is supported over the full gate layer on the device ([van den Berg et al., 2022](#)). The noise model is given in terms of a sparse Pauli-Lindbladian  $\mathcal{L}(\rho) = \sum_k \lambda_k (P_k \rho P_k - \rho)$  for a set of Pauli matrices  $P_k$ , which can be efficiently learned for a polynomial number of Pauli terms (usually low-weight Pauli terms). Although the noise model is correlated across the full circuit, it can be inverted efficiently and its quasi-probability distribution can be sampled exactly. A similar approach was also adopted by [Ferracin et al. \(2022\)](#) which used cycle benchmarking to characterise the low-weight correlated Pauli noise components for performing

probabilistic error cancellation. It is also possible to mitigate these correlated noise components with the help of learning-based methods (Strikis *et al.*, 2021) as will be discussed in Sec. III.H.

From Eqs. (12) and (14), we see that the sampling overhead of probabilistic error cancellation is highly dependent on the basis we choose. For the standard basis proposed in Endo *et al.* (2018), it will perform well when the gates in the circuit suffer from Pauli noise (which can be achieved via Pauli twirling). Going beyond that, Takagi (2021) has derived a lower bound for the sampling overhead under general noise models. Such a lower bound on the sampling overhead has proven to be a good measure for many properties of the noise channel that we try to mitigate (Jiang *et al.*, 2021), but the basis required to reach this lower bound may not be implementable by the given hardware. To find a better practical basis beyond the standard basis, Piveteau *et al.* (2022) proposed to use variational circuits to construct the basis and the authors numerically tested this using real hardware noise models. For actual experiments, probabilistic error cancellation has been successfully demonstrated by Song *et al.* (2019) and Zhang *et al.* (2020) on superconducting and trapped-ion platforms. Sun *et al.* (2021) has shown that probabilistic error cancellation can also be applied to continuous noise processes, for which partial mitigation is possible by expanding the noise process into a perturbation series (Hama and Nishi, 2022). It is also possible to use the non-Markovianity in noise to reduce the sampling cost (Hakoshima *et al.*, 2021).

### C. Measurement error mitigation

Depending on the error mitigation procedure it is necessary to make a distinction between the types of errors that occur during a calculation. Errors that occur during the state preparation and measurement stage are referred to as SPAM errors (Lin *et al.*, 2021; Merkel *et al.*, 2013). The error that occurs at the final measurement stage introduces an additional bias in the expectation value of interest. To put this into a more concrete form, let us continue to use the super-operator representation introduced in Sec. III.B. When we perform measurements and obtain the binary string  $x \in \{0, 1\}^N$  as the output, ideally we want to perform projective measurements in the computational basis  $\{\langle x|\}$ . However, some measurement noise  $\mathcal{A}$  might occur and transform the projective measurements into some positive operator-value measures (POVM)  $\{\langle E_x|\} = \{\langle x|\mathcal{A}\}$ , leading to a different output statistic.

The origins of the measurement errors are as diverse as the hardware that is used to implement quantum processors. For example, a dominant error in the measurement of superconducting qubits is due to thermal excitations and  $T_1$ -decay (Blais *et al.*, 2004; Wallraff *et al.*, 2004),

while for ion traps a major source of uncertainty arises from the difficulty of detecting an ion's dark state and collisions in the trap (Bergquist *et al.*, 1986; Nagourney *et al.*, 1986; Sauter *et al.*, 1986). Other architectures experience noise in the measurement stage from different sources (Haroche and Raimond, 2006). From the perspective of the measurement error protocols most frequently considered in the literature, it is sufficient to consider a simplified model that is agnostic regarding the actual origin of the noise. This model makes the assumption that the noise channel  $\mathcal{A}$  has the computational subspace  $\{\langle x| \mid x \in \{0, 1\}^N\}$  as its invariant subspace, i.e. the resultant POVM basis  $\{\langle E_x|\}$  lives within the computational subspace. This is not the most general measurement error model, but it is nonetheless the most frequently considered model as other coherent errors are usually assumed to be part of the computational stage instead of the measurement stage. Under this assumption, the POVM  $\langle E_x|$  can be decomposed into the computational basis:

$$\langle E_x| = \sum_y \langle E_x|y\rangle \langle y| = \sum_y \langle x|\mathcal{A}|y\rangle \langle y| = \sum_y A_{xy} \langle y|. \quad (19)$$

Here the *assignment matrix*  $A$  is a *transition matrix* (stochastic matrix) whose entries  $\langle x|\mathcal{A}|y\rangle$  represent the transition probability from the measurement result  $y$  to  $x$  due to the noise channel  $\mathcal{A}$  (Chow *et al.*, 2010). The entry  $A_{xy} = \langle x|\mathcal{A}|y\rangle = \langle E_x|y\rangle$  can be obtained by estimating the probability of the  $x$  outcome when we prepare the computational state  $|y\rangle$  (assumed to be almost perfect) and perform the set of noisy measurement  $\{\langle E_x|\}$ . If  $A$  is full rank, then we can invert Eq. (19) and obtain:

$$\langle y| = \sum_x A_{yx}^{-1} \langle E_x| \quad (20)$$

i.e. we can simulate the behaviour of the ideal measurement  $\{\langle y|\}$  using a linear combination of the noisy measurements  $\{\langle E_x|\}$ , just as we have done in probabilistic error cancellation in Sec. III.B. Hence, the associated sampling overhead will increase exponentially with the measurement faults rate, similar to Eq. (18) for probabilistic error cancellation.

For an incoming state  $|\rho\rangle$ , the output distribution using the ideal measurements is given by the vector  $\vec{p}_0 = \{\langle y|\rho\rangle\}$  while the output distribution using the noisy measurements is  $\vec{p}_{\text{noi}} = \{\langle E_x|\rho\rangle\}$ . Applying Eqs. (19) and (20) on  $|\rho\rangle$ , we then have:

$$\vec{p}_{\text{noi}} = A\vec{p}_0 \quad \Rightarrow \quad \vec{p}_0 = A^{-1}\vec{p}_{\text{noi}}. \quad (21)$$

For a given observable  $O$  with the spectrum  $\vec{O} = \{O_x\}$ , i.e.  $\langle O| = \sum_x O_x \langle x|$ , its ideal expectation value is:

$$\langle O|\rho\rangle = \vec{O}^T \vec{p}_0 = \vec{O}^T A^{-1} \vec{p}_{\text{noi}}. \quad (22)$$

Hence, the ideal expectation value can be obtained once we know the assignment matrix  $A$  and the noisy output distribution  $\vec{p}_{\text{noi}}$ .

In early experiments with only a few qubits, [Kandala et al. \(2017\)](#) have performed a full readout tomography of the noisy output distribution  $\vec{p}_{\text{noi}}$  in the computational basis. Then, all entries of the assignment matrix  $A$  were estimated, which can be used to estimate the ideal expectation value using Eq. (22). This approach is not efficiently scalable as the size of  $A$  scales exponentially with the number of qubits  $N$ .

The simplest way to tackle the problem is to assume that the measurement errors of different qubits are not correlated, which implies that the assignment matrix is just the tensor product of the assignment matrices of the individual qubits  $A = \bigotimes_{n=1}^N A_n$ . However, realistic noise encountered in experiments may not be captured accurately by this simplified model and it can be observed that correlations between individual bit-flips are in fact present ([Heinsoo et al., 2018](#)).

[Bravyi et al. \(2021\)](#) try to construct the assignment matrix using continuous-time Markov processes with the generators (transition rate matrices)  $\{G_i\}$  being single- and two-qubit operators:

$$A = e^G \quad \text{with} \quad G = \sum_{i=1}^{2N^2} r_i G_i. \quad (23)$$

The assignment matrix  $A$  is now determined by  $2N^2$  positive coefficients  $\{r_i\}$  that can be learned by only considering a polynomial number of input bit strings. Once the coefficients are learned, we can easily construct the inverse matrix  $A^{-1} = e^{-G}$  for error mitigation.

As mentioned in Appendix A.2, we can perform Pauli twirling on the noise channel  $\mathcal{A}$  by conjugating it with random Pauli operators, which will remove all the off-diagonal elements of  $\mathcal{A}$  in the Pauli basis and produce a Pauli channel  $\mathcal{D}$ . This can be used to simplify measurement error mitigation as discussed by [Chen et al. \(2021\)](#) and [van den Berg et al. \(2022\)](#). Since we are only interested in the action of  $\mathcal{A}$  on the computational subspace, we only need to consider Pauli basis elements that are the tensor products of  $Z$  denoted as  $\{\langle\langle Z^x | \rangle\rangle\}$  with  $x$  being the bit string that marks the qubits that are acted by  $Z$ . Suppose ideally we would like to perform the Pauli measurement  $\langle\langle Z^x | \rangle\rangle$ , however we can only perform the noisy measurement  $\langle\langle Z^x | \mathcal{A} \rangle\rangle$ . Now using Pauli twirling, we can transform the noisy measurement into

$$\langle\langle Z^x | \mathcal{D} \rangle\rangle = D_x \langle\langle Z^x | \rangle\rangle \quad (24)$$

using the fact that the twirled channel  $\mathcal{D}$  is diagonal in the Pauli basis with the entries being  $D_x = \langle\langle Z^x | \mathcal{D} | Z^x \rangle\rangle = \langle\langle Z^x | \mathcal{A} | Z^x \rangle\rangle$ . Thus the noisy measurement  $\langle\langle Z^x | \mathcal{D} \rangle\rangle$  is simply the ideal measurement  $\langle\langle Z^x | \rangle\rangle$  rescaled by

the factor  $D_x$ . For a given input state  $|\rho\rangle\rangle$ , we have:

$$\underbrace{\langle\langle Z^x | \rho \rangle\rangle}_{\text{ideal}} = D_x^{-1} \underbrace{\langle\langle Z^x | \mathcal{D} | \rho \rangle\rangle}_{\text{twirled noisy}}.$$

Hence, by transforming the observable into  $Z^x$  and performing Pauli twirling, we only need to rescale the noisy expectation value by the factor  $D_x^{-1}$  to obtain the ideal expectation value. Note that since conjugation with  $Z$  will have trivial effects within the computational subspace, we only need to conjugate the noise channel with a random operator in  $\{I, X\}^{\otimes N}$  (i.e. random bit-flips) to achieve the effect of Pauli twirling mentioned above.

In practice, the ideal output distribution  $\vec{p}_0$  is often sparse. With weak measurement noise, we would expect the corresponding noisy output distribution  $\vec{p}_{\text{noi}}$  to also be sparse and have non-zero probability at all positions that are non-zero in  $\vec{p}_0$ . Using this fact, [Nation et al. \(2021\)](#) proposed to focus on the action of measurement noise  $\mathcal{A}$  within the subspace spanned by the basis of  $\vec{p}_{\text{noi}}$  with non-zero probability, which gives an assignment matrix with a much smaller dimension. The ideal output distribution  $\vec{p}_0$  can then be obtained by inverting the assignment matrix within this subspace. The inversion can be sped up by considering a matrix-free preconditioned iteration algorithm.

Due to sampling noise, the estimation of ideal distribution  $\vec{p}_0$  we obtained using matrix inversion in Eq. (21) may contain negative values and thus is not a valid probability distribution. However, it is still able to provide an unbiased expectation value estimate using Eq. (22) ([Bravyi et al., 2021](#)). On the other hand, instead of using matrix inversion, one might try to solve a constrained optimisation problem with the cost function  $\|A\vec{p}_0 - \vec{p}_{\text{noi}}\|_2^2$  such that  $\vec{p}_0$  is a valid probability distribution. This can be solved using maximum likelihood ([Chen et al., 2019](#); [Geller, 2020](#); [Maciejewski et al., 2020](#)) or iterative Bayesian unfolding ([Nachman et al., 2020](#)).

There is a wide range of other techniques for combating measurement errors. In [Hamilton et al. \(2020\)](#), a representation based on cumulant expansion is proposed to capture correlations between observables. Measurement error protocols that are directly tailored to VQE type calculations are possible ([Barron and Wood, 2020](#)). Other approaches propose the use of final pre-measurement entangling circuits to combat noise ([Hicks et al., 2022](#)). [Tannu and Qureshi \(2019\)](#) propose to exploit a potential asymmetry in the noise strength of the assignment matrix  $A$  by flipping bit-values into a configuration less likely to be affected by noise. The experimental observations have been used to train classical neural networks to infer predictions of the correct expectation values ([Palmieri et al., 2020](#)). As measurement noise is a major obstacle in almost all experimental set-ups, and elementary measurement error mitigation schemes are easy to implement and

readily available in open-source packages (LaRose *et al.*, 2021; Treinish *et al.*, 2022), implementation of measurement error mitigation is almost *ubiquitous in all near-term experiments*.

#### D. Symmetry constraints

A simple but effective scheme for suppressing errors is to identify errors that break the symmetries of the ideal quantum state, and remove them via post-selection. This notion originates from quantum error correction (Gottesman, 1997; Terhal, 2015), in which we explicitly define a set of measurements to detect and correct all local errors at the cost of additional qubit overhead. Though explicitly correcting errors is required for scalability (Shor, 1996), *quantum error detection* of artificially-added symmetries has been widely recognised as an important milestone towards this end goal (Córcoles *et al.*, 2015; Gottesman, 2016; Kelly *et al.*, 2015; Linke *et al.*, 2017; Nigg *et al.*, 2014). In practical applications, quantum circuits often possess inherent symmetries that can be used for error mitigation without the need to execute a quantum circuit on an error-detection code. Measuring these *inherent symmetries* and discarding circuit runs that produce the wrong results produces a (post-selected) state  $\rho_{\text{sym}}$ . The broad class of schemes that directly or indirectly measure  $\text{Tr}[O\rho_{\text{sym}}]$  are known collectively as *symmetry verification* (Bonet-Monroig *et al.*, 2018; McArdle *et al.*, 2019). If the symmetry measurements are perfect,  $\rho_{\text{sym}}$  must have non-decreasing overlap on the ideal state  $\rho_0$  compared to the noisy state  $\rho$  as we have thrown away states with zero overlaps. In practice, symmetry measurements may themselves introduce errors into the state, thus choosing the right symmetries and optimising their measurements are at the core of symmetry verification.

Bonet-Monroig *et al.* (2018) and McArdle *et al.* (2019) proposed various easily-accessible symmetry operators  $S$  for symmetry verification, drawing from those that naturally emerge in physical systems. In this context, given a physical system, its symmetry operators  $S$  are operators that commute with the system Hamiltonian  $H$ :  $[H, S] = HS - SH = 0$ . When this is the case,  $H$  and  $S$  may be simultaneously diagonalized, i.e. energy eigenstates  $|\Psi_j\rangle$  can be chosen in such a way that  $S|\Psi_j\rangle = s|\Psi_j\rangle$ , where  $s$  is an eigenvalue of  $S$ . Thus, measuring  $S$  on a prepared quantum state, and post-selecting on the correct symmetry eigenvalue  $s$  for the target energy eigenstate, should project one closer to the said energy eigenstate. Furthermore, time evolution by  $e^{iHt}$  leaves the eigenspaces of  $S$  invariant, which means that dynamic properties of the physical system may be studied entirely within these eigenspaces as well. Common examples of symmetries are the parity ( $\prod_i Z_i$ ) and the  $Z$  component of the total spin ( $\sum_i Z_i$ ) of a spin sys-

tem, or the particle number  $\sum_i n_i$  of a fermionic system. Note that the Jordan-Wigner transformation maps  $\sum_i n_i$  to  $\sum_i Z_i$ , modulo a constant shift. Given an  $N$ -qubit simulation, the eigensubspaces of these symmetries have dimension  $2^{N-1}$ ,  $\binom{N}{(N-s)/2}$  and  $\binom{N}{s}$ , respectively, with  $s$  being the eigenvalue of the given eigensubspace. With the right  $s$ , these subspaces are large enough for executing classically intractable quantum algorithms. Note though that it is not necessary to enforce symmetries during an entire circuit in order to verify them at the end (Dallaire-Demers *et al.*, 2019). Even when using a circuit that does not conserve the symmetry of the target physical problem, symmetry verification can still be used for projecting back into the appropriate spin or number sector (Khamoshi *et al.*, 2020; Tsuchimochi *et al.*, 2020; Yen *et al.*, 2019), which can be viewed as mitigating algorithmic errors (Sec. V.C).

As mentioned, the process of *direct* symmetry verification is carried out by measuring both the symmetry operators  $S$  and the target observable  $O$  in every circuit run and discarding runs that produce the wrong output for  $S$  (i.e. fail the symmetry check). Since symmetries  $S$  are typically global observables, measuring them alongside the target observable  $O$  is non-trivial. The additional circuit components required for their measurements can introduce additional errors, reducing or even nullifying the effect of error mitigation. Various ways to measure multiple operators (e.g. the symmetry operator and the target observable) in the same circuit run are discussed in Appendix A.3. One way is to use a Hadamard test to measure a Pauli symmetry like the number parity (Bonet-Monroig *et al.*, 2018; McArdle *et al.*, 2019). Other practical schemes involve measuring qubit-wise commuting operators (Izmaylov *et al.*, 2019), i.e. when both  $S$  and  $O$  can be obtained through post-processing the same set of single-qubit Pauli measurements across all qubits (see Appendix A.3). This implies that if we are using only single-qubit rotations and readout to perform the measurements, symmetries such as the  $Z$ -component of the total spin  $\sum_i Z_i$  or parity  $\prod_i Z_i$  cannot be measured simultaneously with any operator that is not diagonal in the computational basis.

The issue of qubit-wise commutativity presents a specific problem in chemistry where the fermion hopping operator  $c_i^\dagger c_j + c_j^\dagger c_i$  (our target observable) commutes with the particle number  $\sum_i n_i$  (the symmetry operator), but not qubit-wise. An identical problem presents in spin physics between the operators  $X_i X_j + Y_i Y_j$  and  $\sum_i Z_i$ . This may be solved by noting that the rotation

$$\exp\left[i\frac{\pi}{8}(X_i Y_j - Y_j X_i)\right] \quad (25)$$

maps  $X_i X_j + Y_i Y_j$  to  $Z_i - Z_j$  while leaving  $Z_i + Z_j$  invariant (Huggins *et al.*, 2021b). In terms of fermionic systems, this is the operator  $e^{i\frac{\pi}{2}\text{FSWAP}_{i,j}}$ , where  $\text{FSWAP} = c_i^\dagger c_j + c_j^\dagger c_i - n_i - n_j$  (Google Quantum AI and Collabo-

rators, 2020a). In a spin system, this allows for the joint measurement of hopping terms and the  $Z$ -component of the total spin using only a constant depth circuit (assuming all-to-all coupling). However, this measurement does not parallelise efficiently; one can only measure hopping terms simultaneously between disjoint pairs of qubits, requiring  $O(N)$  distinct measurements to estimate all spin-spin hopping terms simultaneously with the  $Z$ -component of the total spin. In contrast, without the requirement to simultaneously measure the total spin- $Z$  component, all spin-spin hopping terms may be estimated by just two distinct choices of single-qubit rotation and readout. In a fermionic system, we need  $O(N)$  distinct measurements in the first place for estimating all fermionic hopping terms due to the lack of mutually commuting terms (Bonet-Monroig *et al.*, 2020), and so simultaneous measurements of the particle number do not present a significant overhead.

Instead of constructing circuits to simultaneously diagonalise and measure the symmetry operators  $S$  and the target observable  $O$ , Bonet-Monroig *et al.* (2018) showed that it is possible to perform effective post-selection via *post-processing* in symmetry verification, which turns out to be closely related to subspace expansion (McClean *et al.*, 2017) (see Sec. III.F). Let us consider the simple example in which there is only one single Pauli symmetry operator  $S$ , and the ideal state lives within the  $+1$ -eigenspace of  $S$  defined by the projector  $\Pi = \frac{1}{2}(1+S)$ . In this way, if the state prepared prior to the post-selection is  $\rho$ , the post-selected state is then

$$\rho_{\text{sym}} = \frac{\Pi\rho\Pi}{\text{Tr}[\Pi\rho\Pi]}. \quad (26)$$

The symmetry-verified expectation value for the target observable  $O$  is given by

$$\text{Tr}[O\rho_{\text{sym}}] = \frac{\text{Tr}[O\Pi\rho\Pi]}{\text{Tr}[\Pi\rho\Pi]} = \frac{\text{Tr}[O_{\text{sym}}\rho]}{\text{Tr}[\Pi\rho]}, \quad (27)$$

where  $O_{\text{sym}} = \Pi O \Pi$  is the symmetrised observable, i.e. the verified expectation value  $\text{Tr}[O\rho_{\text{sym}}]$  is simply the quotient between the noisy expectation value of  $O_{\text{sym}}$  and  $\Pi$ . The symmetrised observable  $O_{\text{sym}}$  and the symmetry projector  $\Pi$  can be further decomposed into the Pauli basis, e.g. if  $O$  is Pauli, the Pauli decomposition of  $O_{\text{sym}}$  is simply  $O_{\text{sym}} = \Pi O \Pi = \frac{1}{4}(O + SO + OS + SOS)$ . In this way, the expectation values of  $O_{\text{sym}}$  and  $\Pi$  can be obtained by measuring the expectation values of these Pauli basis operators (or via Monte Carlo sampling in Appendix A.1). In the method described above, we only need to measure one Pauli observable in a given circuit run, which can be carried out using only single-qubit Pauli measurements without needing additional circuit components. The simple examples above can be further generalised into the cases of multiple symmetries and non-Pauli symmetries with a change in the definition of

the projector  $\Pi$ , reflecting a change in the symmetry subspace. We can also decompose the observables into bases beyond the Pauli basis, as long as the components are easy to measure. Instead of combining the measurement results of the basis to reconstruct the projected observable  $O_{\text{sym}}$  and the projector  $\Pi$ , Cai (2021c) showed that it is possible to achieve smaller biases by combining these measurement results with different weights at the cost of larger sampling overhead.

Note that as we try to describe different ways of performing symmetry verification above, we clearly separate post-selection away from post-processing (for a clear distinction between different methods) even though technically post-selection is a specific type of post-processing in the most general context of QEM.

For direct in-circuit symmetry verification, the fraction of circuit runs that are “useful” is simply the “pass rate” of the symmetry checks given by  $\text{Tr}[\Pi\rho]$ . The corresponding sampling overhead is simply the *inverse* of this “pass rate”:  $C_{\text{em}} \sim \text{Tr}[\Pi\rho]^{-1}$ . The fault-free post-selection discussed in Sec. II.D can be viewed as the ideal symmetry verification that can detect all faults. It can achieve zero bias, but correspondingly its sampling overhead also sets the upper bound for all possible direct symmetry verification. As discussed above, we can greatly simplify the measurement circuit through post-processing, but this will come at a higher sampling overhead, which scales as  $C_{\text{em}} \sim \text{Tr}[\Pi\rho]^{-2}$  (Cai, 2021c; Huggins *et al.*, 2021b).

Though it is simplest to rely on the native symmetries of a system for the purposes of error mitigation, one can consider adding more symmetries artificially, or unitarily transforming a system to improve the error mitigation power of a set of symmetries. This is important as symmetry-based methods cannot mitigate against errors that commute with all symmetries of the system. It was shown in Bonet-Monroig *et al.* (2018) that one can unitarily transform chemistry Hamiltonians so that no single-qubit operator commutes with all symmetries, and that this can be preferable even to removing qubits from the system. They also described a basic scheme to add artificial symmetries to a system. However, this scheme makes local system operators highly non-local, and thus is relatively unscalable. A solution to the above was found in the Bravyi-Kitaev superfast (BKSF) transformation, where artificial symmetries are used to transform local fermionic Hamiltonians to local qubit Hamiltonians (Bravyi and Kitaev, 2002; Setia and Whitfield, 2018). These symmetries are necessary for implementing a geometrically-local fermion-to-qubit transformation in more than one dimension, and at the same time they also provide a natural boon for symmetry-based QEM. The list of fermion-to-qubit transformations has seen significant development and optimisation in recent years (Setia *et al.*, 2019; Steudtner and Wehner, 2018, 2019), with attempts to optimise the number of local errors that can be mitigated (Derby *et al.*, 2021; Jiang *et al.*, 2019). It

was pointed out in Jiang *et al.* (2019) that this need not be contained to fermionic lattice models; as time evolution on a fermionic system may be mapped to a series of operations that are local on a lattice (without any asymptotic growth in the circuit depth), one can implement this using a mapping intended for a local fermionic lattice without paying the price of some system observables being extensively large. Jiang *et al.* (2019) further found an encoding (the ‘Majorana loop stabiliser code’, or MLSC) that can mitigate or even correct all single-qubit errors, making it in effect an error correcting code of distance 3. Unfortunately, it appears that these methods cannot be extended to construct codes of arbitrarily large distances due to the Eastin-Knill theorem (Eastin and Knill, 2009).

A large number of experiments have demonstrated stabiliser measurements in error correction codes throughout the 2010’s; (Córcoles *et al.*, 2015; Kelly *et al.*, 2015; Linke *et al.*, 2017; Nigg *et al.*, 2014; Vuillot, 2018). However, to the best of our knowledge the first experimental demonstration of symmetry verification using *natural* symmetries ( $\prod_i Z_i$ ) in a quantum algorithm was by Sagastizabal *et al.* (2019) through post-processing, as low-cost techniques to measure natural symmetries were not known prior to this. Later, Google Quantum AI and Collaborators (2020a) made direct simultaneous measurement of the number operator and the fermionic 1-RDM using the FSWAP rotation (Eq. (25)) and combined this with McWeeny purification (Sec. III.E). Staniscic *et al.* (2021) has successfully demonstrated the verification of multiple symmetries (number, particle-hole symmetry, and total spin) and also combine it with learning-based methods (Sec. III.H). Due to the simplicity and effectiveness of symmetry verification, it has been employed in a wide range of other experiments [see e.g. Dborin *et al.* (2022); Google Quantum AI and Collaborators (2020b, 2022); Neill *et al.* (2021); and Staniscic *et al.* (2021)].

## E. Purity constraints

Many quantum algorithms target the preparation of an ideal state  $\rho_0$  that is pure:  $\rho_0 = |\psi_0\rangle\langle\psi_0|$ . Many common noise channels are stochastic, which will turn our ideal pure state  $\rho_0$  into some noisy mixed state  $\rho$ . At a high level, error mitigation techniques based on purity constraints attempt to reduce the bias in the expectation value by trying to approximate the pure state closest to  $\rho$ . If we look at the spectral decomposition of  $\rho$ ,

$$\rho = \sum_{i=1}^{2^N} p_i |\phi_i\rangle\langle\phi_i|, \quad (28)$$

where the eigenvalues are ordered  $p_i \geq p_j$  for  $i < j$  and we assume  $p_1 > p_2$  for simplicity, then the closest

pure state to  $\rho$  in the trace distance is simply the dominant eigenvector  $|\phi_1\rangle\langle\phi_1|$ . In principle, one could use the quantum principle component analysis to sample from the eigenbasis of  $\rho$  (Lloyd *et al.*, 2014). Combined with post-selection this would allow for the efficient preparation of the dominant eigenvector. However, the additional circuit required is too deep for mitigating errors in near-term devices and thus simpler strategies are needed.

One such strategy, referred to as *virtual distillation* (VD) (Huggins *et al.*, 2021a) or *error suppression by derangement* (ESD) (Koczor, 2021b), uses collective measurements of  $M$  copies of  $\rho$  in order to access expectation values with respect to the  $M^{\text{th}}$  degree purified state

$$\rho_{\text{pur}}^{(M)} = \frac{\rho^M}{\text{Tr}[\rho^M]} = \frac{1}{\sum_{i=1}^{2^N} p_i^M} \sum_{i=1}^{2^N} p_i^M |\phi_i\rangle\langle\phi_i|. \quad (29)$$

We see that under the assumption  $p_1$  is strictly greater than  $p_2$ , we have  $\lim_{M \rightarrow \infty} \rho_{\text{pur}}^{(M)} = |\phi_1\rangle\langle\phi_1|$ . The rate at which this is achieved is exponential in the number of copies used for purification  $M$ . The remaining bias in the VD/ESD estimator at  $M \rightarrow \infty$  comes from the deviation between  $|\phi_1\rangle\langle\phi_1|$  and the target state  $\rho_0$ , which is sometimes known as the *coherent mismatch* or *noise floor* (Huggins *et al.*, 2021a; Koczor, 2021b). As the largest sources of noise in state-of-the-art quantum devices are typically incoherent, this coherent mismatch can be expected to be significantly smaller than the error in the unmitigated state. Indeed, numerical and analytic studies have confirmed that the error suppression from VD/ESD can be of multiple orders of magnitude for large systems, even using as little as  $M = 2$  copies of the state (Huggins *et al.*, 2021a; Koczor, 2021a,b).

It was shown by Huggins *et al.* (2021a) and Koczor (2021b) that one can estimate expectation values of the purified states in Eq. (29) without ever having to prepare them on a quantum device. The expectation value of this purified state with respect to the observable of interest  $O$  is  $\text{Tr}[O\rho_{\text{pur}}^{(M)}] = \text{Tr}[O\rho^M] / \text{Tr}[\rho^M]$ . In VD/ESD, this is obtained by estimating  $\text{Tr}[O\rho^M]$  and  $\text{Tr}[\rho^M]$  in separate measurements. Let  $S_M$  be the cyclic permutation operator between  $M$  copies of  $\rho$ , the quantity  $\text{Tr}[O\rho^M]$  can be estimated using

$$\text{Tr}[O\rho^M] = \text{Tr}[S_M O_m \rho^{\otimes M}] = \text{Tr}[S_M \bar{O} \rho^{\otimes M}] \quad (30)$$

where  $O_m$  is the operator  $O$  acting on the  $m^{\text{th}}$  copy and  $\bar{O} = \frac{1}{M} \sum_m O_m$  is the observable symmetrised under copy permutation. Thus  $\text{Tr}[O\rho^M]$  can be obtained by measuring  $S_M O_m$  or  $S_M \bar{O}$  on  $M$  noisy copies of  $\rho$ . A diagrammatic proof of Eq. (30) is shown in Fig. 3. One can extend this to estimate  $\text{Tr}[\rho^M]$  by putting  $I$  in place of  $O$ .

Measuring a global operator like  $S_M$  can be challenging, but it can be decomposed into transversal operations among different copies  $S_M = \bigotimes_{n=1}^N \tilde{S}_M^{(n)}$  where  $\tilde{S}_M^{(n)}$

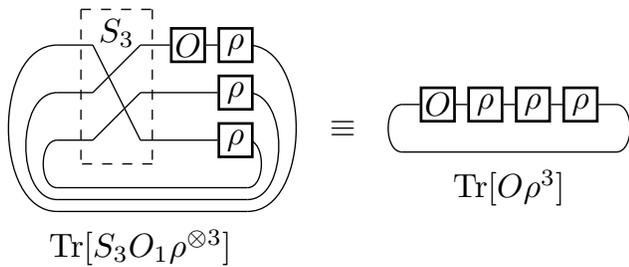


FIG. 3 A diagrammatic proof of Eq. (30) for 3 copies ( $M = 3$ ) with the observable acting on the first copy ( $m = 1$ ). The proof here uses tensor network notations (Bridgeman and Chubb, 2017) and it can be easily extended to more copies and/or with  $O$  acting on the  $m^{\text{th}}$  copy rather than the first.

cyclically permute the  $n^{\text{th}}$  qubits of different copies as shown in Fig. 4. If the observable  $O$  acts on a single qubit, then the symmetrised observable  $\bar{O}$  will commute with all  $\tilde{S}_M^{(n)}$ , and thus  $S_M \bar{O}$  can be obtained by measuring low-weight operators  $\tilde{S}_M^{(n)}$  and  $O_m$  for all  $n$  and  $m$  and then post-process. This requires only transversal operations among the identically-labelled qubits of each copy of  $\rho$ , avoiding global measurement. Explicit circuits for  $O = Z$  and  $M = 2, 3$  without ancilla qubits are given by Huggins *et al.* (2021a); more general measurement may be achieved by a Hadamard test using ancilla qubits (Huggins *et al.*, 2021a; Koczor, 2021b). If the observable  $O$  is not single-qubit, but it is a tensor product of single-qubit operators:  $O = \bigotimes_{n=1}^N G^{(n)}$  (e.g.  $O$  is Pauli), then the observable  $S_M O_m$  in Eq. (30) can be decomposed into a tensor product of low-weight operators  $\bigotimes_{n=1}^N \tilde{S}_M^{(n)} G_m^{(n)}$  which can be measured in a transversal manner (Cai *et al.*, 2022). We can use Hadamard tests to measure each  $\tilde{S}_M^{(n)} G_m^{(n)}$ , which require  $N$  ancilla qubits in total. To efficiently carry out any of the low-weight measurement schemes above, we need transversal operations among different copies, which can be challenging to implement in practice and may involve long-range interactions. A hardware architecture with native transversal operations among different copies has been proposed (Cai *et al.*, 2022), in which an example implementation of VD/ESD is shown with almost no space-time overhead.

In the more general case, one can imagine either unitarily transforming  $O$  to a single-qubit observable (if possible) or linearly decomposing it into a sum of simpler terms (which is always possible). One could additionally imagine decomposing  $S_M O_m$  or  $S_M \bar{O}$  into a linear combination of Pauli operators and measuring the whole set of operators using shadow tomography (Huang *et al.*, 2020; Seif *et al.*, 2022). This has the advantage that one may reuse a single copy of  $\rho$  rather than using multiple physical copies. However, the sampling cost in this case scale exponentially with the number of qubits and thus may be difficult to scale beyond small system sizes (Seif *et al.*, 2022).

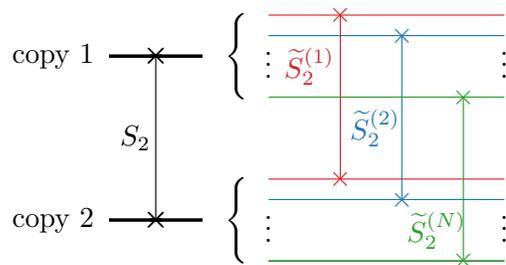


FIG. 4 Decomposition of the copy-swap operator  $S_2$  into transversal qubit-swap operators  $\tilde{S}_2$ . Similar decomposition also applies to other cyclic copy-permutation operators  $S_M$  with  $M > 2$ .

Instead of estimating  $\text{Tr}[O \rho^2]$  using two copies of  $\rho$  separated in space, it is possible to do the estimation using two copies separated in time. This technique was given multiple names since its initial incarnations (Cai, 2021d; Huo and Li, 2022; O'Brien *et al.*, 2021), but here we refer to it as *echo verification* (EV). To perform EV, ideally we would want to measure the ideal state projector  $\rho_0 = |\psi_0\rangle\langle\psi_0|$  at the end of the circuit for post-selection, which will return 1 if the output state is the ideal state and 0 otherwise, just like a symmetry check. The measurement of the projector  $\rho_0$  is carried out by applying the inverse of the primary circuit (O'Brien *et al.*, 2021) and the noise in the inverse circuit means that we are effectively measuring the noisy operator  $\bar{\rho}$  instead of  $\rho_0$ . If we measure the observable of interest  $O$  along with the noisy projector  $\bar{\rho}$  in the same circuit run and post-select according to the outcome of  $\bar{\rho}$ , we are effectively performing measurement of  $O$  on the state  $(\bar{\rho}\rho + \rho\bar{\rho})/(2 \text{Tr}[\bar{\rho}\rho])$  (Huo and Li, 2022). Compared to Eq. (29), we see that this is simply the 2<sup>nd</sup> degree purified state with  $\bar{\rho}$  in place of  $\rho$ . The measurement of  $O$  alongside  $\bar{\rho}$  is usually carried out using a Hadamard test (O'Brien *et al.*, 2021). Alternatively, it is possible to achieve a similar degree of error suppression without the ancilla by preparing a superposition of  $\rho_0$  with a known eigenstate of  $O$  and applying the gate  $O$  on the quantum state (assuming  $O$  is unitary) (Lu *et al.*, 2021; O'Brien *et al.*, 2021), though this is not formally equivalent to 2<sup>nd</sup> degree purification. The error mitigation power of EV has been demonstrated in numerical simulation (O'Brien *et al.*, 2021) and a four-qubit experiment (Huo and Li, 2022). The results differ depending on the exact circuit implementation of the methods, even in simplified noise models (O'Brien *et al.*, 2021), indicating that further optimising the circuit implementation can be an interesting direction to investigate. Gu *et al.* (2022) recently proved that applying EV to control-free phase estimation (Lu *et al.*, 2021; O'Brien *et al.*, 2021; Russo *et al.*, 2021) corrects any noise source with Hermitian Kraus operators to first-order.

Due to their similarity, EV and VD/ESD are directly

comparable in terms of performance and resource requirements. If one wishes to suppress the incoherent contributions of  $\rho$  beyond 2<sup>nd</sup> degree purification, then one can only use VD/ESD instead of EV. However, EV has a smaller qubit footprint and requires less circuit overhead for measurement (especially as it removes the need to perform operations across multiple copies of the same state). In fact, EV can be combined with VD/ESD to achieve a high degree of purification with a lower qubit and circuit overhead (Cai, 2021d). There is also a difference between these two approaches in terms of their sampling overhead, similar to the difference between the sampling overhead of direct and post-processing symmetry verification (Sec. III.D). For EV, the error-mitigation is done through direct post-selection, and thus the sampling overhead is simply the inverse of the success probability  $C_{\text{em}} \sim \text{Tr}[\bar{\rho}]^{-1} \sim \text{Tr}[\rho^2]^{-1}$ . For the two-copy version of VD/ESD, the error-mitigation is done through effective post-selection (post-processing), and thus the sampling overhead increases more steeply, scaling as  $C_{\text{em}} \sim \text{Tr}[\rho^2]^{-2}$ . More generally, the sampling cost of VD/ESD scales exponentially in the number of copies  $M$  (as  $\text{Tr}[\rho^M]$  is exponentially small in  $M$  unless  $\rho$  is pure). If the faults in the circuit follow a Poisson distribution (Sec. II.C), we then have  $\text{Tr}[\rho^2] \lesssim P_0^2 = e^{-2\lambda}$  where  $P_0$  is the fault-free probability of the circuit and  $\lambda$  is the circuit fault rate. This suggests that both EV and VD/ESD can incur a sampling cost growing exponentially with the circuit fault rate, just as we have discussed for general bias-free QEM in Sec. II.D. Furthermore, both methods (in their standard form) lack the parallelisability of unmitigated expectation value estimation. As mentioned above VD/ESD typically estimates expectation values of only  $N$  operators (rather than their products). EV is restricted further; it was shown in (Polla *et al.*, 2022) that EV cannot be efficiently parallelised for even commuting observables, reducing the information extracted by the method to a single bit per state preparation. Though this can be significantly optimised over a simple Pauli decomposition of a complex  $O$  (Polla *et al.*, 2022), it presents roughly an  $O(N)$  overhead in sampling compared to VD/ESD.

The original formulation of EV in terms of projection on the initial state gives a clear connection between purification- and (symmetry) projection-based techniques (Sec. III.D). Though they are clearly not equivalent, this raises the question of whether a similar connection can be made to VD/ESD. Some connection was originally made in Huggins *et al.* (2021a), where it was pointed out that in the eigenbasis  $\{|\phi_j\rangle\}$  of  $\rho$ , the swap operator measurement has zero expectation value on terms in  $\rho \otimes \rho$  except for those of the form  $|\phi_j\rangle\langle\phi_j| \otimes |\phi_j\rangle\langle\phi_j|$ . Cai (2021c) show that VD/ESD can naturally arise from the verification of copy-permutation symmetry by replacing the symmetry projector with

a general linear combination of permutation operators, which is also connected to subspace expansion using symmetry operators (will be discussed in Sec. III.F). Yoshioka *et al.* (2022) further exploited the connection between VD/ESD and subspace expansion and looked at the effect of performing error mitigation using a linear combination of states with different degrees of purification.

## F. Subspace expansions

In some quantum tasks, one has knowledge of not only the ideal circuit and potential noise sources, but also the structure of the task at hand. One such class of methods that can take advantage of the knowledge of the problem are *quantum subspace expansion* techniques (McClean *et al.*, 2020, 2017; Takeshita *et al.*, 2020). Many tasks in quantum computing, such as optimisation or state preparation, can be phrased as the desire to minimise the objective function  $\langle\psi|H|\psi\rangle$  with respect to the state  $|\psi\rangle$  for some known Hermitian operator  $H$ . Very often we are unable to directly prepare the optimal state, either due to a coherent error in our implementation of the ideal circuit, or a simple lack of knowledge of the ideal circuit. However, there is usually a set of  $M$  different states  $\{|\phi_i\rangle\}$  (linearly independent but not necessarily orthogonal to each other) that we can easily prepare, which can be used to construct our target state through linear combination  $|\psi_{\vec{w}}\rangle = \sum_{i=0}^{M-1} w_i |\phi_i\rangle$ . In this way, the problem of finding the optimal state becomes finding the optimal set of coefficients  $\vec{w}^*$ :

$$\vec{w}^* = \arg \min_{\vec{w}} \langle\psi_{\vec{w}}|H|\psi_{\vec{w}}\rangle \quad \text{s.t.} \quad \langle\psi_{\vec{w}}|\psi_{\vec{w}}\rangle = 1. \quad (31)$$

This has the well-known exact solution in the form of a generalised linear eigenvalue problem

$$\bar{H}W = \bar{S}WE, \quad (32)$$

where

$$\bar{H}_{ij} = \langle\phi_i|H|\phi_j\rangle, \quad \bar{S}_{ij} = \langle\phi_i|\phi_j\rangle, \quad (33)$$

i.e.  $\bar{H}$  is the  $M \times M$  matrix representations of  $H$  in the chosen basis set  $\{|\phi_i\rangle\}$  and  $\bar{S}$  is the overlap matrix for the basis set. In Eq. (32),  $W$  and  $E$  are the matrices of eigenvectors and eigenvalues, respectively, of the solved problem. The eigenvector in  $W$  with the lowest eigenvalue is precisely the optimal combination of coefficients  $\vec{w}^*$  for the state  $|\psi_{\vec{w}^*}\rangle$ .

If we want to obtain the improved expectation value of some observable  $O$  with respect to our new found state  $|\psi_{\vec{w}^*}\rangle$ , it is simply given as:

$$\langle\psi_{\vec{w}^*}|O|\psi_{\vec{w}^*}\rangle = \sum_{i,j=0}^{M-1} w_i^* w_j^* \langle\phi_i|O|\phi_j\rangle$$

i.e. we can construct it using the optimal weight  $\bar{w}^*$  and the measurement results of  $\langle \phi_i | O | \phi_j \rangle$ , without needing to explicitly prepare  $|\psi_{\bar{w}^*}\rangle$ . For the special case of  $O = H$ , the improved expectation value is simply given by the smallest eigenvalue in  $E$  when we solve Eq. (32). If desired however, one could prepare the state  $|\psi_{\bar{w}^*}\rangle$  via LCU methods (Childs and Wiebe, 2012) in order to use it as the input state for a subsequent quantum routine.

If one takes the limit of choosing  $|\phi_i\rangle$  to be a complete basis for the whole Hilbert space, then solving the optimisation problem will return the ideal state, however choosing an exponentially large space to perform classical optimisation obviously defeats the purpose of using a quantum computer to begin with. Hence, how to choose the right set of basis state  $\{|\phi_i\rangle\}$  is the key to the success of quantum subspace expansion. The first basis state  $|\phi_0\rangle$  we select is usually the best state that we can prepare before performing quantum subspace expansion. In this way, in the worst case we will simply obtain back  $|\phi_0\rangle$  through subspace expansion.

For the other basis states, the original work of McClean *et al.* (2017) suggests that we can draw inspiration from the configuration interaction (CI) expansions (Helgaker *et al.*, 2000) in quantum chemistry, which is commonly used for improving energy and properties of mean-field states as well as determining excited states for response properties. There each of the other basis states is generated by applying an *expansion basis* operator  $G_i$  on the original state such that  $G_i |\phi_0\rangle = |\phi_i\rangle$ . Knowledge of a good set of expansion basis operators  $\{G_i\}$  can come from symmetry considerations, excitation operators, or simply knowing that correcting (as opposed to replacing) the state  $|\phi_0\rangle$  requires that the additional states are connected directly or indirectly through  $H$ . In this way, the expanded state is now  $|\psi_{\bar{w}}\rangle = \Gamma_{\bar{w}} |\phi_0\rangle$  where  $\Gamma_{\bar{w}} = \sum_{i=0}^{M-1} w_i G_i$  is the *expansion operator*, which is a weighted sum of expansion basis operators. Once the expansion basis operators are determined, the matrix elements required for solving the optimisation equation in Eq. (32) can be measured on a quantum computer through

$$\begin{aligned} \bar{H}_{ij} &= \langle \phi_0 | G_i^\dagger H G_j | \phi_0 \rangle = \text{Tr} \left[ G_i^\dagger H G_j \rho \right] \\ \bar{S}_{ij} &= \langle \phi_0 | G_i^\dagger G_j | \phi_0 \rangle = \text{Tr} \left[ G_i^\dagger G_j \rho \right] \end{aligned} \quad (34)$$

without needing detailed knowledge of the original state  $\rho = |\phi_0\rangle\langle\phi_0|$ .

So far both the starting state  $|\phi_0\rangle$  and the expanded state  $|\psi_{\bar{w}^*}\rangle$  are pure states and thus any errors that we have removed are coherent errors. This is in stark contrast to our focus on incoherent errors in the last section (Sec. III.E). The right-hand side of Eq. (34) is suggestive of the fact that we can apply expansion around some mixed state  $\rho$  for removing incoherent errors. This is indeed first conjectured in the original works of McClean

*et al.* (2017) and later confirmed by several experimental implementations of the method (Colless *et al.*, 2018; Sagastizabal *et al.*, 2019; Urbanek *et al.*, 2020). These observations are put on more solid theoretical footing by Bonet-Monroig *et al.* (2018) and McClean *et al.* (2020). The effective state after performing subspace expansion on a noisy state  $\rho_\lambda$  is shown to be (McClean *et al.*, 2020):

$$\rho_{\text{sub}} = \frac{\Gamma_{\bar{w}} \rho_\lambda \Gamma_{\bar{w}}^\dagger}{\text{Tr} \left[ \Gamma_{\bar{w}} \rho_\lambda \Gamma_{\bar{w}}^\dagger \right]}. \quad (35)$$

We see that this is very similar to the symmetry-verified state in Eq. (26) with the symmetry projector  $\Pi$  replaced by the expansion operator  $\Gamma_{\bar{w}}$ . This implies that using the symmetry operators as our expansion basis operators, we can recover the symmetry subspace by performing subspace expansion. Cai (2021c) try to further generalise this by searching for expanded “states” of the form  $\Gamma_{\bar{w}} \rho / \text{Tr}[\Gamma_{\bar{w}} \rho]$  instead, which allows us to also incorporate purification-based QEM in Sec. III.E under this formalism.

A number of recent works have also looked into other possible sets of expansion basis operators  $\{G_k\}$ . For example, when the operators  $\{G_k\}$  are chosen to be powers of the Hamiltonian  $\{H^k\}$ , we see that these methods coincide with quantum Krylov subspace methods like qLanczos (Motta *et al.*, 2020) or other methods based on filtering the eigenspectrum via functions of the Hamiltonian (Suchsland *et al.*, 2021). Another recent work that was mentioned before (Yoshioka *et al.*, 2022) has included the operators that are powers of the density matrix, making close ties to the purification-based QEM methods. By doing so, optimal combinations of states can now exploit problem-specific knowledge related to purity in addition to general knowledge.

## G. $N$ -representability

Often in the context of quantum simulation (and sometimes more broadly), the goal is ultimately to measure a set of observables corresponding to a marginal of the total density matrix known as a reduced density matrix (RDM). Given a general quantum state  $\rho$  on  $N$  qubits, the set of  $p$ -qubit RDMs are obtained by integrating out  $q$ -qubits (such that  $N - q = p$ ) of the joint distribution,

$${}^p \rho_{m_1, \dots, m_p} = \text{Tr}_{n_1, n_2, \dots, n_q} [\rho] \quad (36)$$

resulting in  $\binom{n}{p}$  different RDMs each of dimension  $2^p \times 2^p$ . The coefficients  $n_1, \dots, n_q$  on the trace operator indicate which qubits are integrated out of  $\rho$  and coefficients  $m_1, \dots, m_p$  label the subsystem marginal. The result of this marginalization is a distribution over  $p$ -qubits.

The connection to error-mitigation is that these RDMs are known to have special geometric structures; not

all marginals that one can write down are consistent with having come from a valid (or in the parlance of this field, “representable”) wave function. In principle there exists a set of conditions that constrain the space of representable RDMs. Articulating and evaluating these equality and inequality constraints is known as the “ $N$ -representability problem” (Mazziotti, 2016), and the problem is formally QMA-Complete (Liu *et al.*, 2007). However, for most RDMs of interest, one can write down and evaluate at least some of the  $N$ -representability conditions. That knowledge can often be used to mitigate errors, in a spirit similar to the application of symmetry constraints, but generally using different methods.

The focus on measuring RDMs is especially common when simulating many-body systems of identical particles. For example, because real fermions interact pairwise, most properties of interest can be obtained using just the 1-particle and 2-particle reduced density matrices (and because the particles are identical, there is only a single 1-RDM and a single 2-RDM). Using the second quantized fermionic creation and annihilation operators acting on site  $p$ ,  $a_p^\dagger$  and  $a_p$ , the fermionic 1-RDM and 2-RDM can be expressed as

$${}^1D_j^i = \text{Tr} \left[ a_i^\dagger a_j {}^nD \right] = \langle \psi | a_i^\dagger a_j | \psi \rangle \quad (37)$$

$${}^2D_{rs}^{pq} = \text{Tr} \left[ a_p^\dagger a_q^\dagger a_s a_r {}^nD \right] = \langle \psi | a_p^\dagger a_q^\dagger a_s a_r | \psi \rangle \quad (38)$$

where  ${}^kD$  is the  $k$ -particle RDM and the equalities on the right correspond to the case of pure states. For a system of  $N$  sites this 2-RDM is only of dimensions  $N^2 \times N^2$  (in contrast to  $2^N \times 2^N$  for the full density matrix), and yet completely determines the energy of a fermionic system with pairwise interactions.

Some of the simpler constraints we can express on the 1- and 2-RDMs are as follows:

1. Hermiticity of the density matrices

$${}^1D_j^i = ({}^1D_j^i)^* \quad (39)$$

$${}^2D_{rs}^{pq} = ({}^2D_{pq}^{rs})^* \quad (40)$$

2. Antisymmetry of the 2-particle marginal

$${}^2D_{rs}^{pq} = -{}^2D_{sr}^{pq} = -{}^2D_{rs}^{qp} = {}^2D_{sr}^{qp} \quad (41)$$

3. The  $(p-1)$ -marginal is related to the  $p$ -marginal by contraction—e.g. the 2-marginal can be contracted to the 1-marginal

$${}^1D_j^i = \frac{1}{n-1} \sum_k {}^2D_{jk}^{ik} \quad (42)$$

4. The trace of each marginal is fixed by the number of particles in the system

$$\text{Tr} [{}^1D] = n \quad (43)$$

$$\text{Tr} [{}^2D] = n(n-1) \quad (44)$$

5. The marginals are proportional to density matrices and are thus positive semidefinite

$$\{{}^1D, {}^2D\} \succeq 0. \quad (45)$$

This is not an exhaustive list of all  $N$ -representability constraints.

The work of Rubin *et al.* (2018) first pointed out that knowledge of these constraints could be used for mitigating errors when measuring RDMs. The essential idea is that in the course of a NISQ simulation, one might measure an RDM that violates  $N$ -representability conditions as a consequence of errors corrupting the estimation of the tomography elements composing the RDM. However, one can use even a partial list of RDM conditions in order to project the noisy and unrepresentable RDM estimate back to the nearest RDM consistent with a list of RDM constraints. Because the RDM constraints all take the form of equality and inequality constraints, this can be performed using semidefinite programming. Such an approach was demonstrated numerically in Rubin *et al.* (2018) and experimentally in Smart and Mazziotti (2020).

Of special interest for error-mitigation, there is a sub-discipline of the  $N$ -representability field known as “pure state  $N$ -representability” which is concerned with describing the geometry of pure density matrices (i.e.  $N$ -representability that must hold for pure states). In principle, using pure state representability would potentially allow one to measure the 2-RDM of a partially decohered state and then project that estimate back to the nearest RDM consistent with having come from a pure state. While this idea was first discussed in Rubin *et al.* (2018), it has been difficult to realise in practice since pure  $N$ -representability conditions are extremely difficult to compute. The state-of-the-art in the field is that specialised computer algebra systems are needed to generate the pure state conditions (DePrince, 2016; Klyachko, 2006). Nevertheless, some work has succeeded in using some of these conditions in quantum simulations (Smart and Mazziotti, 2019).

In the case when the 1-RDM of a fermionic system is expected to be idempotent ( $D^2 = D$ ), a special type of purification known as McWeeny purification (McWeeny, 1960) is possible. This purification scheme is achieved by iterating on a non-idempotent 1-RDM estimate as

$${}^1D_{i+1} = 3({}^1D_i)^2 - 2({}^1D_i)^3 \quad (46)$$

until idempotency is restored. The only fermionic states with idempotent 1-RDMs are Slater determinants, which limits the applicability of this scheme in strongly-correlated systems. Despite the lack of theoretical justification, McCaskey *et al.* (2019) demonstrated moderate mitigation success from McWeeny purification in a correlated four-qubit chemistry simulation. The most notable

success however, came from the application in [Google Quantum AI and Collaborators \(2020a\)](#) to a Hartree-Fock state, for which the idempotency assumption is justified. In this work McWeeny purification demonstrated between one and two orders of magnitude of error suppression, on top of the other mitigation methods used. The hope remains that similar results can be demonstrated in other experiments by enforcing the pure-state representability constraints discussed in the previous paragraph, or by applying purification-based QEM methods discussed in [Sec. III.E](#).

## H. Learning-based

Given the primary circuit  $\mathbf{P}$ , its noisy expectation value is denoted as  $E(\mathbf{P})$  and we are trying to estimate its noiseless expectation value  $E_0(\mathbf{P})$ . To achieve this, we obtain the error-mitigated expectation value  $E_{\vec{\theta}}(\mathbf{P})$  as a function of the primary circuit and a set of parameters  $\vec{\theta}$ . In previous sections, we see that the function parameters  $\vec{\theta}$  in different QEM methods are obtained through different noise calibration processes. However, we can also obtain  $\vec{\theta}$  through *learning-based methods* ([Czarnik et al., 2021](#); [Strikis et al., 2021](#)) using training circuits. We will construct a *training circuit*  $\mathbf{T}$  that is

1. similar to  $\mathbf{P}$ , usually in terms of circuit structures, so that it contains similar circuit faults as  $\mathbf{P}$ ;
2. classically simulable, i.e. its ideal expectation value  $E_0(\mathbf{T})$  can be obtained via classical simulation.

Knowing the exact value of  $E_0(\mathbf{T})$  enables us to find a good good set of parameters  $\vec{\theta}$  for the error mitigation function  $E_{\vec{\theta}}$  by minimising the difference between  $E_0(\mathbf{T})$  and  $E_{\vec{\theta}}(\mathbf{T})$ . Then we will assume the error mitigation protocol  $E_{\vec{\theta}}$  obtained from the training circuit  $\mathbf{T}$  also works well for the primary circuit  $\mathbf{P}$  due to their similarity, which give us the error-mitigated result  $E_{\vec{\theta}}(\mathbf{P})$  as an estimate of the ideal result  $E_0(\mathbf{P})$ . More generally we can have more than one training circuit, which is denoted using the training set  $\mathbb{T}$ . The simplest loss function we can construct to obtain the optimal  $\vec{\theta}$  is

$$L_{\mathbb{T}}(\vec{\theta}) = \frac{1}{|\mathbb{T}|} \sum_{\mathbf{T} \in \mathbb{T}} (E_0(\mathbf{T}) - E_{\vec{\theta}}(\mathbf{T}))^2. \quad (47)$$

The whole process of learning-based QEM is summarised in [Fig. 5](#). If the error-mitigated estimate  $E_{\vec{\theta}}(\mathbf{T})$  is linear in  $\vec{\theta}$ , which is the case for many QEM schemes, then the optimal  $\vec{\theta}$  can be obtained using linear least squares.

The simplest error mitigation function simply rescales and shifts the noisy expectation value to approximate the ideal expectation value:

$$E_0(\mathbf{A}) \approx E_{\vec{\theta}}(\mathbf{A}) = \theta_0 + \theta_1 E(\mathbf{A}). \quad (48)$$

Here the input circuit  $\mathbf{A}$  can be the primary circuit  $\mathbf{P}$  or the training circuits  $\mathbf{T} \in \mathbb{T}$ . [Czarnik et al. \(2021\)](#) first proposed to use such a linear function to mitigate errors, whose coefficients can be obtained by training using the Clifford variants of the primary circuit. This has been shown to be effective in experiments ([Mi et al., 2021](#); [Urbanek et al., 2021](#)). When performing fermion simulations, it is also possible to train using the closest free-fermion model as proposed and demonstrated by [Google Quantum AI and Collaborators \(2020b\)](#), or using fermionic linear optics ([Montanaro and Stanisc, 2021](#)) as demonstrated experimentally by [Stanisc et al. \(2021\)](#). In some particular use cases like measuring the out-of-time-order correlator ([Mi et al., 2021](#)), by simply removing/replacing a small number of gates of the primary circuit, one can analytically derive the output of the resultant circuits, which can also be used as the training circuits.

The linear error mitigation function in [Eq. \(48\)](#) can naturally arise when we assume all noise sources in the circuit are globally depolarising. In such a case, the resultant noisy state is just a mixture of the ideal state and the completely mixed state:  $\rho = P_0 \rho_0 + (1 - P_0) I / 2^N$  ([Mi et al., 2021](#); [Vovrosh et al., 2021](#)) where  $P_0$  is the fault-free probability of the circuit ([Sec. II.C](#)). Hence, the ideal expectation value takes the simple form of

$$\underbrace{\text{Tr}[O\rho_0]}_{E_0(\mathbf{P})} = \frac{1}{P_0} \underbrace{\text{Tr}[O\rho]}_{E(\mathbf{P})} - \frac{1 - P_0}{P_0 2^N} \text{Tr}[O],$$

which is in the same form as [Eq. \(48\)](#). The assumption of global depolarising noise motivates the linear error mitigation function, but is *not* a necessary assumption for applying this error mitigation function using training circuits. There is evidence that global depolarising noise is an effective phenomenological error model emerging from gate-wise error models when the gate number is large ([Qin et al., 2021](#)). On the other hand, if indeed there is only global depolarising noise in the circuit, we can actually estimate the rescaling factor  $P_0^{-1}$  using [Eq. \(8\)](#) if we know the circuit fault rate, or obtain  $P_0^{-1}$  by measuring

$$\text{Tr}[\rho^2] = P_0^2 + P_0(1 - P_0)/2^{N-1} + (1 - P_0)^2/2^{2N}$$

and solving the quadratic equation ([Vovrosh et al., 2021](#)). Multiple ways to measure  $\text{Tr}[\rho^2]$  have been discussed in purification-based QEM ([Sec. III.E](#)). These circuits for obtaining  $P_0^{-1}$  do *not* have the same circuit structure as the primary circuit  $\mathbf{P}$ , so instead of viewing them as the training circuits for applying learning-based rescaling and shifting, it may be more appropriate to view them as the noise calibration circuits for performing probabilistic error cancellation ([Sec. III.B](#)) against global depolarising channels or for a special case of linear error extrapolation ([Sec. III.A](#)) ([Cai, 2021a](#)).

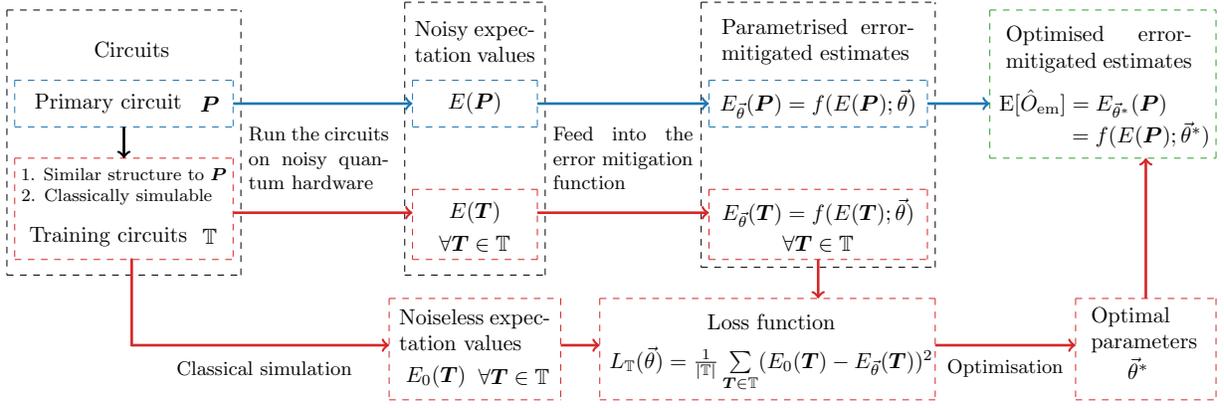


FIG. 5 A diagram showing the process of learning-based quantum error mitigation. For the cases in which we need to construct variants of the input circuit for the error mitigation function like in Eq. (49), the whole process is similar, but instead of only running the input circuits on the quantum hardware, we need to run all the different variants of the input circuits.

In Eq. (48),  $E_{\vec{\theta}}(\mathbf{A})$  is simply a function of the noisy expectation value  $E(\mathbf{A})$  of the input circuit  $\mathbf{A}$ , thus only the input circuit  $\mathbf{A}$  needs to be run on the quantum hardware. This is not the case for many of the QEM methods we discussed before. In general for a given input circuit  $\mathbf{A}$ , we need to construct a set of *response measurement circuits*  $\{\mathbf{A}_{\text{rsp},i}\}$  that are variants of the input circuit  $\mathbf{A}$  by adding/replacing gates and/or adding measurements. These circuits can be, e.g. circuits of different noise levels for zero-noise extrapolation, circuits with different added gates for probabilistic error cancellation and circuits with different added measurements for symmetry verification. The error-mitigated expectation value will be a function of the outputs of all of these response measurement circuits instead of just the input circuit:

$$E_{\vec{\theta}}(\mathbf{A}) = f(\{E(\mathbf{A}_{\text{rsp},i})\}; \vec{\theta}). \quad (49)$$

One such example would be the error mitigation estimate for probabilistic error cancellation in Eq. (17):

$$E[\hat{O}_{\text{em}}] = \sum_{\vec{n}} \alpha_{\vec{n}} \langle O | \mathcal{B}_{\vec{n}} | \rho_{\text{in}} \rangle \mapsto E_{\vec{\theta}}(\mathbf{A}) = \sum_{\vec{n}} \theta_{\vec{n}} E(\mathbf{A}_{\text{rsp},\vec{n}}).$$

Here the response measurement circuit  $\mathbf{A}_{\text{rsp},\vec{n}}$  corresponds to preparing  $\rho_{\text{in}}$ , applying the sequence of operation  $\mathcal{B}_{\vec{n}}$  and measuring  $O$ , which as mentioned differs from  $\mathbf{A}$  by additions or replacements of some subset of gates. The parameters  $\theta_{\vec{n}}$  can be obtained as  $\alpha_{\vec{n}}$  through device calibration as discussed in Sec. III.B. On the other hand, we can also obtain  $\theta_{\vec{n}}$  via learning-based methods (Strikis *et al.*, 2021). Applying learning-based methods to probabilistic error cancellation implies that we do not have enough information about the gate errors in the primary circuit (otherwise we will apply probabilistic error cancellation directly). Hence, we would need to assume Pauli gate errors in the primary circuit or apply Pauli twirling, such that the set of response measurement circuits can simply be constructed by adding Pauli gates. Other than optimising over  $\theta_{\vec{n}}$  directly, we see that the

response circuit coefficient  $\alpha_{\vec{n}}$  is actually the product of the coefficients for individual gates in Sec. III.B. In a similar way, we can write  $\theta_{\vec{n}}$  as the product of the coefficients for individual gates and optimise over the gate coefficients instead. This would greatly simplify the optimisation problem if the number of gate types in the circuit is small. By incorporating the appropriate response measurement circuit  $\mathbf{A}_{\text{rsp},\vec{n}}$ , it is also possible to mitigate spatially and temporally correlated using learning-based methods.

Let us continue to use probabilistic error cancellation with Clifford training (Strikis *et al.*, 2021) as the example to illustrate how the training circuits can be constructed. In probabilistic error cancellation, we want to remove all faults in the circuit. To find a way to mitigate these faults using the training circuits, the same faults must exist in these training circuits. Such training circuits can be constructed by replacing gates in the primary circuit with gates that have the same error channels. If we compile the primary circuit such that all the multi-qubit gates in the primary circuit are Clifford, we then only need to replace the single-qubit gates to construct Clifford training circuits that are classically simulable. If we further assume that all the single-qubit gates have the same error channel *or* they have negligible error rates compared to the multi-qubit gates, then the fault distribution of these Clifford training circuits will be the same as the primary circuit. We will use  $\mathbb{C}$  and  $\mathbb{U}$  to denote the set of all possible circuits generated by replacing the single-qubit gates in the primary circuit with random single-qubit Clifford and random single-qubit unitary, respectively (note that  $\mathbb{C} \subset \mathbb{U}$  and  $\mathbf{P} \in \mathbb{U}$ ). By constructing the training circuits in the way outlined above, the training loss function  $L_{\mathbb{T}}(\vec{\theta})$  in Eq. (47) is a homogeneous polynomial of degree 2 in matrix elements of the single-qubit gates in the circuits. Since the Clifford group is a unitary 2-design, the loss function satisfies  $L_{\mathbb{C}}(\vec{\theta}) = L_{\mathbb{U}}(\vec{\theta})$  (Wang *et al.*, 2021c). Therefore, by training over the Clifford circuits and min-

imising  $L_{\mathbb{C}}(\vec{\theta})$ , we are minimising the errors in the more general unitary circuits in  $\mathbb{U}$ . When  $L_{\mathbb{C}}(\vec{\theta})$  goes to zero, we have  $L_{\mathbb{U}}(\vec{\theta}) = 0$ , which implies  $E_{\vec{\theta}}(\mathbf{A}) = E_0(\mathbf{A})$ , i.e. all errors are perfectly mitigated, for all  $\mathbf{A} \in \mathbb{U}$  including the primary circuit  $\mathbf{P}$ .

Since the size of the Clifford set  $\mathbb{C}$  grows exponentially with the number of qubits, it is impractical to evaluate  $E_0(\mathbf{T})$  and the corresponding noisy response measurements  $\{E(\mathbf{T}_{\text{rsp},i})\}$  for all Clifford training circuits  $\mathbf{T} \in \mathbb{C}$ . Furthermore, the majority of noisy Clifford training circuits have near-zero expectation values which are costly to evaluate in terms of sampling overhead. One way to circumvent this is to truncate the Clifford training set by using only circuits with large noiseless expectation values  $|E_0(\mathbf{T})|$  due to their more significant contribution to the loss function, which was shown to be effective in numerical simulations (Czarnik *et al.*, 2022; Strikis *et al.*, 2021). It is also possible to use Monte-Carlo sampling and variational update of the parameters  $\vec{\theta}$  to overcome the large size of  $\mathbb{C}$  (Strikis *et al.*, 2021). Since it is possible to classically simulate circuits with a small number of non-Clifford gates, we can keep a few single-qubit gates in the primary circuit untouched when we define the training set. Alternative ways to truncate the Clifford set may be explored, e.g. based on their similarity to the primary circuits.

If the training set is large enough such that the training can target a wide range of application circuits (computational tasks), then the training can be carried out at the device calibration stage and the training sampling overhead may be omitted when considering a particular computational task. On the other hand, when the training set is small such that the training is targeting a small set of application circuits (computational tasks), then the training is best viewed as a part of a task itself and the training overhead needs to be included in any resource audit for that computational task. Note that a smaller training set usually also means a smaller sampling overhead for training. The relation between the size of the training set  $|\mathbb{T}|$  and the performance of applying the trained result onto the target circuit  $\mathbf{P}$  is only rigorously developed for the case of training using the full Clifford set  $\mathbb{C}$  for probabilistic error cancellation as discussed above. More general studies into the trade-off between them would be essential for different learning-based methods.

So far we have only explicitly talked about learning-based methods applied to rescaling and shifting the noisy result (Eq. (48)) and to probabilistic error cancellation. However, learning-based methods are in general compatible with almost all QEM methods that we have discussed, e.g. error extrapolation (Lowe *et al.*, 2021) and purification-based method (Bultrini *et al.*, 2021). There are also suggestions that it can be applied to symmetry-based methods (Cai, 2021c). Some of the possible roles

of learning-based methods in other QEM methods will be further discussed in the next section.

## IV. COMPARISONS AND COMBINATIONS

### A. Comparison among QEM methods

In the last section, we have provided a detailed account of the individual QEM methods. In order to see the connections and differences among them more clearly and provide a discussion of their respective costs in a more coherent framework, Cai (2021b) divides the process of QEM into two stages:

1. **Noise calibration:** measures the strength of some given noise components;
2. **Response measurement:** measures how the observable of interest responds to changes in the noise components that were calibrated in the last step.

Combining both components will inform us about how the observable of interest changes due to the presence of the calibrated noise components, and thus will enable us to construct an error-mitigated estimator protected from the calibrated noise components. We have already discussed such a structure for QEM methods when introducing the learning-based methods (Sec. III.H), there the noise calibration process is simply the training process. Of course the division between noise calibration and response measurement is not always clear-cut. For most of the QEM methods that we have discussed, the error-mitigated estimator will be a linear combination of the results obtained from the response measurement, and the way to combine them (the weightings of each term) will be determined by the noise calibration. Treating QEM as a two-stage process enables us to discuss the cost of noise calibration and response measurement separately.

#### 1. Noise calibration overhead

In the previous sections when talking about sampling overhead, we are mostly referring to the response measurement sampling overhead. We often assumed certain knowledge about the noise without discussing the cost of obtaining it. Now let us look more closely into the cost of the noise calibration for various types of QEM methods.

*a. Gate error mitigation* These QEM methods target the *gate errors* in the primary circuit. Their noise calibration can simply be carried out using standard gate noise benchmarking/characterisation techniques (Eisert *et al.*, 2020; Kliesch and Roth, 2021), e.g. gate infidelity estimation for *zero-noise extrapolation* (Sec. III.A), full gate error characterisation for *probabilistic error cancellation*

(Sec. III.B) and detector error characterisation for *measurement error mitigation* (Sec. III.C). These can all be done in the device calibration stage and if this is the case, then ideally noise calibration is effectively free at the stage of applying QEM. However, fully correlated noise models are exponentially expensive (in terms of qubit number) to characterise and various ways to circumvent this have been mentioned in Secs. III.B and III.C. Moreover, device parameters can drift in time and thus routine re-calibration might be needed. In such a case, the noise calibration cost is no longer negligible at the QEM application stage and low-cost device calibration techniques would be essential [see e.g. [Google Quantum AI and Collaborators \(2020b\)](#)].

*b. State error mitigation* These QEM methods target the *errors on the output state* of the primary circuit. Since we do not know the exact form of the ideal state, we will try to probe errors that violate known constraints on the output state like *symmetry constraints* (Sec. III.D) and *purity constraints* (Sec. III.E). Their noise calibration will measure the strength of the noise that violates these constraints, e.g. the fraction of the circuit runs that fail the symmetry verification. In these cases, both noise calibration and response measurement involve measuring additional operators on the unmitigated noisy state. Hence, in some settings, the noise calibration can be performed alongside response measurement by simply measuring additional operators without additional circuit runs ([Bonet-Monroig et al., 2018](#); [Cai, 2021b,c](#)), which means the noise calibration is essentially free.

*c. Observable error mitigation* Not all errors in the primary circuit will affect our observable of interest, and this class of QEM methods only target the *error components that are damaging to our observable*. Methods like *subspace expansion* (Sec. III.F) and *N-representability* (Sec. III.G) target error components that violate some given constraints on the noiseless observables (which can be the observable of interest or its components). For observable error mitigation, the noise calibration will be dependent on the observable of interest. Hence, the noise calibration accuracy required is highly dependent on the problems we try to solve and thus the associated cost has not been analytically derived.

As mentioned, *learning-based methods* (Sec. III.H) are a means to perform the noise calibration process using training circuits. Compared to the circuits used in the original noise calibration process in different QEM methods, the training circuits are usually more closely related to the primary circuit, so that we can target faults that are more specific to the primary circuit and hopefully can reduce the calibration cost. For example, when trying to construct the training circuit, we can make use of

the structure of the primary circuit for gate error mitigation, or we can make use of the known observable of interest for state error mitigation. The training cost for learning-based methods is thus highly problem-specific and usually hard to analytically quantify, just like the cost for noise calibration in observable error mitigation.

The categorisation of QEM methods above is mainly for providing better intuitions rather than a definitive guide. For example, the RDM measured in the  $N$ -representability method can be viewed as an observable and then the method is seen as a type of observable error mitigation; alternatively it can be viewed as a state in a reduced subspace, so that the method is a form of state error mitigation. After our discussion of the noise calibration requirements above, we can now move onto the more general comparisons between QEM methods beyond the noise calibration stage.

## 2. Mean square errors

As discussed in Sec. II.B, the most straightforward metric for comparing two different error-mitigated estimators is just to compare their mean square errors (Eq. (1)). However, general comparisons between two QEM methods cannot be made in a similar manner because for each QEM method, there is a wide range of error-mitigated estimators we can construct using different hyper-parameters (e.g. the degree of purification for purification-based QEM, the number of symmetries in symmetry-based QEM, the number of data points in zero-noise extrapolation, etc). Each of these estimators has a different trade-off between bias and variance and thus a different mean square error. Comparisons between two QEM methods are further complicated by their different sets of assumptions (e.g. whether there are known constraints to states, knowledge about the noise) and different hardware requirements (e.g. qubit number and connectivity). Hence, in order to compare two QEM methods, very often we need to know the exact primary circuit whose noise we try to mitigate and the set of experimental constraints we need to adhere to (on things like qubit numbers, runtime and hardware error rate), such that we can specify exact implementations of the two given QEM methods and deduce the corresponding mean square errors. Even in such cases, the bias and variance often can only be calculated analytically for some canonical implementations. Hence, in most of the QEM literature mentioned in this review, when trying to assess or compare the performance of QEM methods, their bias and variance are often obtained using numerical simulations or physical experiments by applying them to some specific use case. This is not a good indicator of the general performance of a given QEM method, but this can usually demonstrate key characteristics of the QEM methods like their target noise types, their scaling be-

haviour, etc.

While recognising that it is hard to make general comparisons between different QEM methods, nevertheless we endeavour to summarise the costs and performance of some *canonical implementations* of typical QEM methods in Table IV. The intention is to capture the essential distinguishing features of these QEM approaches.

## B. Benchmarking QEM from other perspectives

Besides computing the mean square error for some specific implementation of a given QEM method, we can also look into other general characteristics of the QEM method by viewing the process of QEM from another perspective.

### 1. State discrimination

QEM allows us to better estimate the expectation values of the various noisy states output from a noisy system. However, the data-processing inequality never allows the distinguishability between these noisy states to increase (Nielsen and Chuang, 2010). This fact was used by Takagi *et al.* (2022) to obtain explicit bounds on the range that the error-mitigated estimator may take, which in turn determines the number of samples needed to achieve a given level of performance.

Their work considered an error-mitigation protocol in which the first step is obtaining  $M$  noisy copies of the error-free state  $\rho_0$  through a process denoted as  $\mathcal{L}$ :

$$\mathcal{L}(\rho_0) = \bigotimes_{m=1}^M \mathcal{E}_m(\rho_0)$$

where  $\{\mathcal{E}_m\}$  are the different effective noise channels acting on different copies. The second step is constructing an error-mitigated observable  $O_{\text{em}}$  based on the QEM technique and the observable of interest  $O$ , such that measuring  $O_{\text{em}}$  on these noisy copies will output the error-mitigated estimator  $\hat{O}_{\text{em}}$ :

$$\text{E}[\hat{O}_{\text{em}}] = \text{Tr}[O_{\text{em}}\mathcal{L}(\rho_0)].$$

The *maximum bias* that can be achieved by the QEM method for all possible observables of interest  $O$  and target states  $\rho_0$  is given as:

$$B_{\text{max}} = \max_{O, \rho_0} \text{Bias}[\hat{O}_{\text{em}}] = \max_{O, \rho_0} [\text{Tr}[O_{\text{em}}\mathcal{L}(\rho_0)] - \text{Tr}[O\rho_0]].$$

Using this, Takagi *et al.* (2022) managed to prove a lower bound for the range of the error-mitigated estimator:

$$\text{R}[\hat{O}_{\text{em}}] \geq \max_{\rho_0, \sigma_0} \frac{D_{\text{tr}}(\rho_0, \sigma_0) - 2B_{\text{max}}}{D_{\text{tr}}(\mathcal{L}(\rho_0), \mathcal{L}(\sigma_0))}. \quad (50)$$

where  $D_{\text{tr}}$  denotes the trace distance. The range  $\text{R}[\hat{O}_{\text{em}}]$  can be used to obtain the number of samples required via Eq. (4).

The trace distance  $D_{\text{tr}}(\mathcal{L}(\rho_0), \mathcal{L}(\sigma_0))$  is related to the maximum probability of distinguishing between the multiple noisy copies of  $\rho_0$  and  $\sigma_0$ . It will decrease exponentially under Pauli noise as the circuit fault rate  $\lambda$  increases, causing  $\text{R}[\hat{O}_{\text{em}}]$  and thus the sampling overhead to increase exponentially. This lower bound in Eq. (50) can be further tightened if the POVM that we can perform for distinguishing the multiple noisy copies of  $\rho_0$  and  $\sigma_0$  is restricted to be local (Takagi *et al.*, 2022), which better reflects the mechanisms of most of the QEM methods. Takagi *et al.* (2022) has shown that probabilistic error cancellation can saturate the lower bound for the local version of Eq. (50) for some noise model, which means that its sampling overhead is optimal in these cases from the state discrimination perspective.

Note that the bounds obtained above are more for the purpose of demonstrating the fundamental limits on the performance of a given QEM setup, rather than to be used as a metric for comparing different QEM methods. Furthermore, the discussions here do not cover QEM techniques that non-linearly combine the results from response measurement (e.g. extrapolations that have more data points than free parameters or exponential extrapolation (Cai, 2021a; Endo *et al.*, 2018)).

### 2. State extraction

In symmetry-based or purification-based QEM, it is natural to think of QEM as a process of extracting the symmetry-verified or purified state out of the noisy state (Bonet-Monroig *et al.*, 2018; Huggins *et al.*, 2021a; Koczor, 2021b; McArdle *et al.*, 2019). More generally in most of the QEM methods, the error-mitigated expectation value  $\text{E}[\hat{O}_{\text{em}}]$  is a linear function of the observable of interest  $O$ , which can be written as

$$\text{E}[\hat{O}_{\text{em}}] = \text{Tr}[O\rho_{\text{em}}].$$

Here  $\rho_{\text{em}}$  can be viewed as the error-mitigated component that we try to extract out of the noisy state using the given QEM method (Cai, 2021b).

A higher fidelity of the error-mitigated state  $\rho_{\text{em}}$  against the ideal state  $\rho_0$ , denoted as  $F(\rho_0, \rho_{\text{em}})$ , is shown to correspond to lower biases in various numerical simulations (Cai, 2021c; Huggins *et al.*, 2021a; Koczor, 2021b). More exactly, using results in Koczor (2021a) and the Fuchs-van de Graaf inequality (Fuchs and van de Graaf, 1999), the bias after error mitigation can be bounded by:

$$\begin{aligned} \text{Bias}[\hat{O}_{\text{em}}] &\leq 2\|O\|_{\infty} D_{\text{tr}}(\rho_0, \rho_{\text{em}}) \\ &\leq 2\|O\|_{\infty} \sqrt{1 - F(\rho_0, \rho_{\text{em}})}, \end{aligned} \quad (51)$$

where  $\|O\|_{\infty}$  is the largest absolute eigenvalue of  $O$ .

As discussed in Cai (2021b), by assuming the ideal state  $\rho_0$  to be a *pure* state, we can decompose the noisy state  $\rho$  into the error-mitigated component  $\rho_{\text{em}}$  and an erroneous component  $\rho_{\text{err}}$  (not necessarily a valid density matrix) that is orthogonal to the ideal state  $\rho_0$  ( $\text{Tr}[\rho_0\rho_{\text{err}}] = 0$ ):

$$\rho = p_{\text{em}}\rho_{\text{em}} + (1 - p_{\text{em}})\rho_{\text{err}}. \quad (52)$$

Here  $p_{\text{em}}$  is the amount of error-mitigated component contained in the noisy state. Only a partial amount of this error-mitigated component can be successfully extracted using the given QEM method and this amount is denoted as  $q_{\text{em}}$  with  $q_{\text{em}} \leq p_{\text{em}}$ .

In this picture of extracting error-mitigated states, the factor of improvement in the fidelity, which we will call the *fidelity boost*, is given by:

$$B_{\text{em}} = \frac{\text{Tr}[\rho_0\rho_{\text{em}}]}{\text{Tr}[\rho_0\rho]} = p_{\text{em}}^{-1}$$

(where  $\rho_0$  is assumed to be pure) and the corresponding sampling overhead is given as:

$$C_{\text{em}} \sim q_{\text{em}}^{-2}. \quad (53)$$

Cai (2021b) has given a list of examples of how  $p_{\text{em}}$  and  $q_{\text{em}}$  can be calculated to obtain  $B_{\text{em}}$  and  $C_{\text{em}}$ , with some of the results shown in Table IV.

The fraction of the error-mitigated state that is successfully extracted using the QEM method is called the *extraction rate*:

$$r_{\text{em}} = \frac{q_{\text{em}}}{p_{\text{em}}} = \frac{B_{\text{em}}}{\sqrt{C_{\text{em}}}}.$$

This is simply the ratio between the fidelity boost and the square root of the sampling overhead, and thus is a natural indicator for the *cost-effectiveness* of a given QEM method. By definition the maximum extraction rate we can achieve is  $r_{\text{em}} = 1$ , which is achieved by symmetry verification (Cai, 2021b), indicating its cost-effectiveness. Note that if we are able to perform direct post-selection instead post-processing to extract the error-mitigated state (like in direct symmetry verification and echo verification), then the sampling overhead will be  $C_{\text{em}} \sim q_{\text{em}}^{-1}$  instead and the extraction rate will be  $r_{\text{em}} = B_{\text{em}}/C_{\text{em}}$ . Similar to Sec. IV.B.1, the arguments in this section also do not apply to QEM techniques that non-linearly combine the results from response measurement.

### C. Combinations of QEM methods

Rather than trying to pick a better QEM method, one might try to combine different QEM methods instead, in the hope of being able to target more noise components and/or achieving a better trade-off between bias

and variance. We have mentioned some possible connections and combinations of QEM methods when discussing individual techniques in Sec. III, and we will try to provide additional insights in this section.

The simplest way of combining different QEM methods is applying them in parallel such that each of them targets a different noise source in the circuit. For example using zero-noise extrapolation or probabilistic error cancellation to target noise in the main computation while using measurement error mitigation to tackle noise at the measurement stage. For a circuit consisting of Pauli-symmetry-preserving components affected by Pauli noise, we can detect and partially mitigate the circuit faults that anti-commute with the symmetry using symmetry verification, while the rest of the circuit faults that commute with the symmetry and thus are immune to symmetry verification can be mitigated using zero-noise extrapolation (through only scaling these ‘commuting’ noise) or probabilistic error cancellation (Cai, 2021a). As will be discussed in Sec. V.C, QEM can be used to mitigate errors in the circuit compilation. Hence, it is possible to use one QEM method to mitigate these compilation errors, while using another QEM method to mitigate noise in the circuit. Note that we can also apply a given QEM method multiple times, each time targeting a different noise source, which has been demonstrated for zero-noise extrapolations (Otten and Gray, 2019).

When the different QEM methods applied interfere with each other rather than working entirely independently, their order of application becomes important. The QEM method that is first applied will be called the *base QEM method*. In a sense, applying one QEM method on top of another can be viewed as “concatenating” the QEM methods, and the overall sampling overhead is simply the product of the sampling overheads of both stages of QEM. If we want to apply symmetry- or purification-based QEM on top of another QEM method, we can simply view the base QEM method as a process of extracting error-mitigated states out of the noisy state as discussed in Sec. IV.B.2, and then apply symmetry constraints or purity constraints to this error-mitigated state (Cai, 2021b). In this way, we are able to remove the remaining symmetry- or purity-violating noise that was not targeted by the base QEM method. In a similar way, for a given error-mitigated state, we can also perform subspace expansion around it or we can measure its RDM, which means that we can directly perform subspace expansion or  $N$ -representability on top of other QEM methods. Google Quantum AI and Collaborators (2020a) has experimentally demonstrated the application of  $N$ -representability on top of symmetry verification.

One may apply zero-noise extrapolation on top of another QEM method by performing extrapolation using error-mitigated expectation values (McArdle *et al.*, 2019). However, the error models of the circuit can be altered by the base QEM method, such that the noise

scaling factor might not be well-defined anymore. Furthermore, the shape of the extrapolation curves may be altered by the base QEM method. It is possible to circumvent this by using probabilistic error cancellation as the base QEM method since it can be applied partially without changing the error models, yielding data points of reduced noise strength for extrapolation as mentioned before in Sec. III.A (Cai, 2021a). For other base QEM methods, as long as we know the effective noise scaling factor and the circuit fault rate is small after the base QEM, Richardson extrapolation will still be applicable, e.g. it has been applied on top of purification-based methods in numerical simulations (Koczor, 2021b) and on top of instances of probabilistic error cancellation that changes the error models (Sun *et al.*, 2021). In some specific use cases, the new shape of the extrapolation curve can be analytically deduced, one such example is given in Cai (2021a) with symmetry verification being the base QEM method. More generally, we can try to use learning-based methods to predict the shape of the extrapolation curves after performing the base QEM, especially for the special case of rescaling and shifting the noisy expectation value (see Sec. III.H). Applying rescaling on top of symmetry verification has been demonstrated in numerical simulations (Stanisic *et al.*, 2021) and experiments (Google Quantum AI and Collaborators, 2020b). As mentioned in Sec. III.H, rescaling and shifting can be viewed as performing probabilistic error cancellation for mitigating global depolarising noise. However, probabilistic error cancellation in a more general context is difficult to perform on top of other QEM methods because the effective error model in the circuit will change due to the base QEM method.

Learning-based methods can be viewed as a way to perform the noise calibration process rather than a stand-alone QEM method as mentioned in Sec. IV.A.1. Hence, it can be easily combined with the other methods by replacing their noise calibration process (Cai, 2021b; Lowe *et al.*, 2021), which we have mentioned above in several examples. We can also use learning-based methods to obtain the optimal hyper-parameters for various QEM methods (Bultrini *et al.*, 2021).

There were also attempts to construct a unified framework containing hyper-parameters that can lead to different existing QEM methods when taking different values. Such frameworks give us access to a wide range of other possible QEM implementations by taking hyper-parameter values sitting in between existing QEM methods. We can think of this as “interpolating” between different QEM methods instead of “concatenating” different QEM methods. These new “interpolated” QEM implementations can achieve different bias-variance trade-offs beyond the existing QEM methods and often also target a different set of noise components. This allows us to choose the best implementation that fits our target problem and experimental constraints rather than be-

ing restricted to the existing QEM methods. We have mentioned some of these examples in the previous section (Sec. III), e.g. the combination of purification-based methods with subspace expansion (Yoshioka *et al.*, 2022), the generalisation of symmetry verification and subspace expansion (Cai, 2021c) or the combination of zero-noise extrapolation and probabilistic error cancellation (Mari *et al.*, 2021). Some QEM combinations mentioned earlier in this section can also be viewed as attempts to construct a unified framework (Bultrini *et al.*, 2021; Lowe *et al.*, 2021).

#### D. Comparison to the other error suppression methods

Quantum error mitigation is one family of strategies for suppressing errors. It is natural to consider how it relates to other well-known error-suppression methods, such as decoherence-free subspace/subsystem (DFS), dynamical decoupling (DD) and quantum error correction (QEC). Symmetry is at the heart of all the methods mentioned here, thus we will first try to compare QEC against symmetry verification to gain some intuition.

In QEC, we perform regular symmetry (stabiliser) measurements to detect and correct errors accumulated in the quantum system. If we discard the quantum states that violate these symmetry checks instead of correcting them, we will have quantum error detection instead. Symmetry verification (Sec. III.D) can usually be seen as an error-*detection* scheme whose symmetries are given by the physical problem of interest and the detection can be carried out via post-processing. Such native symmetries and the possible use of post-processing detection typically mean that much fewer qubits and a much lower gate fidelity are required for symmetry verification compared to QEC to achieve the “break-even” point of performing better than the uncorrected/unmitigated circuit. Furthermore, performing logical operations in QEC code usually come with higher space-time overhead than the physical operations performed in QEM, leading to longer circuit run time and even more qubit overhead for QEC.

Even though we are talking about symmetry verification in the above comparison, many of the intuitions provided are applicable to general QEM methods. If we are trying to remove (almost) all bias as in fault-tolerant QEC, the sampling overhead of general QEM will grow *exponentially* with the circuit size as discussed in Sec. II.D. Using all the arguments above, we can compile a rough guide for the differences between QEC and QEM as outlined in Table I. Note that there are intersections between QEC and symmetry-based QEM like post-processing quantum error detection and adding more symmetries to QEM (at the cost of adding more qubits) as discussed in Sec. III.D.

Without knowing the specific application task, it will be difficult to compare the amount of resources required

	QEC	QEM
Qubit overhead	High	Low
Circuit runtime	Often scale with code distance	Similar to the unmitigated circuit
Sampling overhead	Const.	Exponential in the circuit size <sup>1</sup>
Gate error rate	Must be below code threshold	Important to keep it low; no threshold <sup>2</sup>
Mid-circuit measurement	Essential and frequent	Infrequent or not required

<sup>1</sup> Assuming we try to construct an (nearly) unbiased estimator like in QEC and with fixed gate error rates.

<sup>2</sup> Lower error rate means smaller sampling overheads. No threshold for the gate noise in the unmitigated circuit, but there might be requirements on the additional gates needed for QEM.

TABLE I A rough guide on the differences between QEC and QEM. The exact difference will be dependent on the QEC codes, the QEM method and the exact application scenario.

for QEC and QEM. There have been attempts to estimate the resource required to obtain *classically intractable* Fermi-Hubbard ground state energy using variational eigensolver (VQE) with the help of symmetry verification and error extrapolation (Cai, 2020). To overcome the sampling overhead, which is partly due to QEM, it was estimated that more than 100 independent quantum processors are needed for parallelisation, each containing 50 qubits and with a gate error rate of  $10^{-4}$ . The gate fidelity and the total number of qubits required here are actually similar to those required for solving the problem using a fault-tolerant algorithm (Kivlichan *et al.*, 2020). However, given the same total number of qubits, building hundreds of identical quantum processors (cores) using commercial fabrication techniques is usually easier than building the single big integrated processor required for the fault-tolerant algorithm. Note that what we have discussed here is not a sheer comparison between QEC and QEM, but more of a comparison between a NISQ algorithm with the help of QEM and a fault-tolerant algorithm for a specific application.

In practice, QEC and QEM are more likely to be complementary rather than competing methods since in many practical applications we can apply QEM on top of QEC as will be further discussed in Sec. V.B. It is also possible to use QEM to mitigate compilation errors, which is not possible using QEC as will be discussed in Sec. V.C.

Unlike QEC and QEM, both dynamical decoupling (DD) and decoherence-free subspace/subsystem (DFS) methods are *open-loop* quantum control techniques whose action does not depend on the status of the quantum state and thus no measurements are needed for their

implementation (Lidar, 2014; Suter and Álvarez, 2016). In fact, DFS is a passive technique that requires no action at all, we only need to choose the right symmetry subspace that is immune to the target set of noise. QEC and symmetry-based QEM all have some element of DFS in them in the sense that they are all immune to noise generated by the symmetry operators. DD on the other hand is achieved by applying a sequence of decoupling pulses to “average” out the harmful interaction between the quantum system and the environment. In the limit of many rounds of instantaneous decoupling pulse, group-based DD is shown to be equivalent to making continuous symmetry measurements to achieve the quantum Zeno effect (Burgarth *et al.*, 2019; Facchi *et al.*, 2004). Such a DD sequence can be applied using the native symmetry of the physical problem just like in symmetry-based QEM, which is shown to be effective in quantum simulation (Tran *et al.*, 2021).

## V. APPLICATIONS

So far we have discussed QEM mostly in the general context of expectation value estimation without overspecifying the application scenarios. Through numerical experiments, QEM has been proven to be effective in a wide range of applications like linear equation solvers (Vazquez *et al.*, 2022), quantum metrology (Yamamoto *et al.*, 2022), Monte-carlo simulations (Yang *et al.*, 2021) and in numerous other examples that we mentioned before in Sec. III. Moreover, many physical experiments have successfully employed QEM for noise suppression, with some pioneering and state-of-the-art examples summarised in Table II. Different application scenarios and hardware platforms will give rise to different types of noise. In this section, we will look into how to adapt the implementation of QEM to different noise types and the expected results.

### A. Coherent errors

The degree of coherence in an error channel can be quantified by how well it preserves the purity of an average incoming state (Wallman *et al.*, 2015). In this sense, the “most coherent” errors are just unitary errors, while incoherent errors usually refer to stochastic Pauli channels. More generally, the term *coherent errors* refers to a broad spectrum of noise that sits closer to the unitary errors than the Pauli errors. There are cases in which coherent errors can be mitigated using coherent control solutions like dynamical decoupling (Lidar, 2014; Suter and Álvarez, 2016). The remnant coherent errors can be mitigated using most of the QEM techniques mentioned before (with the exception of purity-based methods). In fact, subspace expansion (Sec. III.F) was originally con-

QEM Methods	Platform	Applications	Reference
ZNE	SC	Variational eigensolver	Dumitrescu <i>et al.</i> (2018) and Kandala <i>et al.</i> (2019)
ZNE	SC	Real-time dynamic simulation	Kim <i>et al.</i> (2021)
ZNE	Ion	Entanglement entropy measurement	Foss-Feig <i>et al.</i> (2022)
PEC	SC	Deterministic quantum computation with pure states	Song <i>et al.</i> (2019)
PEC	SC	Real-time dynamic simulation	van den Berg <i>et al.</i> (2022)
PEC	Ion	Gate fidelity estimation	Zhang <i>et al.</i> (2020)
SYM	SC	Variational eigensolver	Sagastizabal <i>et al.</i> (2019)
SYM	Ion	Variational state preparation	Zhu <i>et al.</i> (2020)
SUB	SC	Variational eigensolver	Colless <i>et al.</i> (2018)
LEA	SC	Variational eigensolver	Dborin <i>et al.</i> (2022)
LEA	SC	Quantum information scrambling	Mi <i>et al.</i> (2021)
SYM & NRP	SC	Variational state preparation	Google Quantum AI and Collaborators (2020a)
SYM & LEA	SC	Variational eigensolver	Stanisic <i>et al.</i> (2021)
SYM & LEA	SC	Real-time dynamic simulation	Google Quantum AI and Collaborators (2020b)

SC, Ion: superconducting qubits and trapped-ion qubits.

ZNE, PEC, SYM, PUR, SUB, NRP, LEA: zero-noise extrapolation, probabilistic error cancellation, symmetry constraints, purity constraints, subspace expansion,  $N$ -representability and learning-based methods. Note that measurement error mitigation is present in almost all experimental works and thus is not explicitly included here.

TABLE II A list of examples of pioneering and state-of-the-art experimental applications of QEM.

structured to target coherent errors (McClellan *et al.*, 2017). However for many other QEM methods, Pauli errors can be far easier to mitigate compared to coherent errors as discussed in the previous sections. The advantages of mitigating Pauli noise are summarised below:

- Zero-noise extrapolation: arguments can be made that the expectation value of Pauli observable should decay following a multi-exponential curve under Pauli noise. Knowing the form of the extrapolation curve can lead to lower sampling costs and smaller biases.
- Probabilistic error cancellation: Pauli noise requires less overhead to be characterised and removed using the standard basis, similarly for measurement error mitigation.
- Symmetry constraints: when dealing with symmetry-preserving circuit components and Pauli symmetry, analytical arguments can be made about the proportion of noise removed when considering Pauli noise. This also enables us to select a good set of symmetries to use.
- Purity constraints: errors that are less coherent can improve the ultimate accuracy that can be reached using purity constraints.

- Learning-based: standardising error channels by transforming them into Pauli noises will result in a more similar error model between the training circuits and target circuits. The effects of Pauli noise on the observable are usually easier to be characterised and thus require less training.

Even for subspace expansion, though Pauli errors might not be easier to mitigate, they are certainly easier to analyse and thus enable us to select a better set of expansion operators. Furthermore, coherent errors usually accumulate at a faster rate than their Pauli counterpart (Gutiérrez and Brown, 2015; Kueng *et al.*, 2016; Sanders *et al.*, 2015). Hence, *one of the most effective ways to deal with coherent errors in the context of QEM is simply transforming them into Pauli errors.*

Any arbitrary quantum channel can be transformed into a Pauli channel using *Pauli twirling*, which originated from entanglement purification (Bennett *et al.*, 1996a,b; Knill, 2004) and is widely used in areas like quantum benchmarking (Kliesch and Roth, 2021) for transforming quantum states or noise channels into a standardised form. When applying Pauli twirling to environmental noise or noise from Clifford gates, we simply insert random Pauli gates during the circuit compilation (as discussed in Appendix A.2). Most of these random

Pauli gates can in fact be absorbed into the existing Pauli gates (Wallman and Emerson, 2016) and thus it can often be implemented with a relatively small gate cost. It is also possible to twirl over the Clifford group, which will further homogenise the probability of different Pauli errors, resulting in a depolarising channel, which is even easier to analyse and mitigate. However, Clifford gates, especially multi-qubit ones, can be much harder to implement than Pauli gates.

## B. Logical errors in fault-tolerant quantum computation

The ultimate goal of quantum computation is to have a fully scalable fault-tolerant quantum system in which we can suppress the logical errors to an arbitrary small level by increasing the size of the system. However, before we achieve that, it is likely that there will be an extended period of time in which quantum error correction is successfully implemented, but the minimum logical error rate achievable is still substantial due to reasons like hardware limitations in the system size, challenges in large-scale decoding, and/or insufficient resources for magic state distillation. In this ‘early fault-tolerant’ era, many tasks of interest requiring an effectively zero error rate will be impossible to perform. *As long as the results of the fault-tolerant algorithm are obtained through expectation value estimation, the residual logical errors can be mitigated using all of the QEM methods mentioned before.* We can follow the exact same arguments outlined in the previous sections and simply replace the qubit registers, the operations and the error channels with their logical equivalents.

In particular, the role of probabilistic error cancellation in the Clifford+ $T$  paradigm of universal fault-tolerant quantum computation was studied extensively in several works (Lostaglio and Ciani, 2021; Piveteau *et al.*, 2021; Suzuki *et al.*, 2022). The logical noise associated with Clifford gates can be transformed into logical Pauli channels using Pauli twirling by inserting random Pauli gates as mentioned in Sec. V.A. To mitigate the resultant logical Pauli noise using probabilistic error cancellation, we only need to insert Pauli gates into the unmitigated circuit. These additional logical Pauli gates required for error mitigation can be applied virtually and noiselessly through updating the Pauli frame (Suzuki *et al.*, 2022), boosting the effectiveness of probabilistic error cancellation in the fault-tolerant setting. The errors in the noisy logical  $T$  gate can be mitigated in a similar way, but to twirl them we need to apply additional logical Clifford gates instead of just Pauli gates (see Appendix A.2). Suzuki *et al.* (2022) show that to achieve a logical circuit fault rate of  $10^{-3}$  with a sampling overhead of 100, probabilistic error cancellation can reduce the physical qubit overhead by 80% for some classically intractable problems and more than 45% in many practical fault-tolerant

applications.

The magic state distillation process needed for implementing fault-tolerant  $T$  gates accounts for significant costs in fault-tolerant computation even with recent advances (Litinski, 2019; O’Gorman and Campbell, 2017). Therefore, it makes sense to study the details of error mitigation for imperfect logical  $T$  gates. Piveteau *et al.* (2021) consider logical  $T$  gates implemented through either gate teleportation or code switching, and found that it is sufficient to consider the effective logical error as a dephasing channel (with the help of logical twirling), which is much easier to characterise and mitigate using probabilistic error cancellation. Assuming perfect Clifford gates, they showed that one can use probabilistic error cancellation to remove errors in a circuit that has 2000  $T$  gates and a physical error rate of  $10^{-3}$  at a sampling overhead of 1000. This is well beyond the classically tractable regime of  $\sim 50$   $T$  gates (Bravyi and Gosset, 2016). Lostaglio and Ciani (2021) also discuss applying QEM to noisy logical  $T$  gates but with more emphasis on the resource theoretic aspect. There a resource measure called quantum-assisted robustness of magic is introduced, which indicates the speed up of quantum circuit simulation with noisy non-Clifford gates using probabilistic error cancellation.

Besides gate errors, there will also be compilation errors when we try to approximate an arbitrary unitary gate using the Clifford+ $T$  gate set. The form of the resultant compilation errors can be analytically computed for local gates, which can then be mitigated using probabilistic error cancellation (Suzuki *et al.*, 2022). The resultant compilation error will decrease exponentially with the increase of the number of  $T$  gates used (Dawson and Nielsen, 2006; Kitaev, 1997; Kliuchnikov *et al.*, 2013; Ross and Selinger, 2016). Hence, there is a trade-off between the number of  $T$  gates used and the cost of the QEM we need to apply. We will discuss more instances of compilation errors beyond the context of gate synthesis in the next section.

## C. Compilation (algorithmic) errors

To execute a unitary operation, we need to first compile it into a sequence of gates that the quantum computer can efficiently execute. The compiled circuit does not always exactly represent the target unitary, and any resultant mismatches will be called *compilation errors* or sometimes *algorithmic errors*. We have discussed one such instance in the context of gate synthesis in Sec. V.B. Compilation errors will be present even if all gates can be executed perfectly without any noise, and thus they *cannot* be removed through QEC. On the other hand, compilation errors can be removed through some of the QEM techniques that we have mentioned and this is one of the areas where QEM is non-trivially different from

QEC.

Endo *et al.* (2019) tried to remove such compilation errors via error extrapolation. There they look at the specific problem of using Trotterisation (Suzuki, 1990, 1991) for Hamiltonian simulation. Higher-order Trotterisation will have fewer compilation errors but deeper circuits. Balancing these two aspects means there is an optimal order of Trotterisation in a given practical setting. By applying Richardson extrapolation to the results of different orders of Trotterisation up to the optimal order, we can estimate the result of infinite-order Trotterisation (Endo *et al.*, 2019), which corresponds to the result of zero compilation errors.

Mitigating compilation errors is even more natural for QEM methods that are based on the known constraints of the ideal state or the ideal observables since compilation errors can also violate these constraints just like gate errors. One such example was discussed by Huggins *et al.* (2021a). There they look at the randomised Trotterisation (Campbell, 2019; Childs *et al.*, 2019; Ouyang *et al.*, 2020) in which the compiled circuit is inherently probabilistic while the ideal state is known to be pure. Hence, it is natural to apply purification-based QEM here to remove the stochastic errors due to the randomised compilation process. Another such example is subspace expansion applied in the context of variational eigensolver (McClean *et al.*, 2017). There one of the main goals of subspace expansion was overcoming the limited representation power of the ansatz circuit, which can be viewed as a generalisation of algorithmic errors.

## VI. OPEN PROBLEMS

### A. Overarching problems

Having thus surveyed the diverse concepts, implementations and applications of QEM in the previous sections, it is appropriate to now reflect on the key unanswered questions in the field. We identify several such questions below.

#### 1. What is the full landscape of the zoo of QEM?

In Sec. III, we have discussed a set of QEM methods and their variants, which are grouped into three categories in Sec. IV.A.1. However, this is by no means the full landscape of QEM. Are there better classifications of the different QEM methods? Moreover, are there new methods waiting to be discovered?

#### 2. What are some good performance metrics for QEM?

In Sec. II.B, we identify the bias and variance of the error-mitigated estimator as its key performance metrics and in Sec. IV, we further define some performance metrics by viewing QEM from other perspectives. However, these metrics are greatly dependent on the exact problem we try to solve and many

of them can only be obtained analytically for some canonical cases. Are there other performance metrics for QEM that can better reflect their practical performance and/or can be analytically calculated for a wider range of methods? Can these metrics take into account the various additional information/resources required by different QEM methods? Instead of performance metrics, should we identify a set of well-defined use cases which can then be used to benchmark QEM methods through numerics or experiments, much like the MNIST or ImageNet database in computer vision research?

#### 3. What is the optimal QEM strategy for some practical use case?

This is a natural question that follows from the first two. After we have a full view of all possible QEM methods out there and good performance metrics for their practical performance, we can then identify the optimal QEM strategy for some practical use cases. This is most likely to be a hybrid of different QEM methods.

#### 4. What is the connection between QEM and QEC?

We have attempted to connect QEM to QEC in Sec. IV.D. There, most of the connection is made specifically for symmetry-based QEM. It would be interesting to see a more systemic approach to connect general QEM methods to QEC, or even to merge them under a larger common framework of error suppression. For this to happen, of course we would need a good description of the framework for QEM first, which is the first question we presented above.

#### 5. Can we extend QEM beyond expectation value estimation?

While expectation value estimation is at the heart of many useful algorithms to date, there are still many algorithms fall outside this paradigm, e.g. a repeat-until-success algorithm like Shor's algorithm, or algorithms that output a single-shot measurement, such as quantum phase estimation using the quantum Fourier transform. Can we identify an equivalent of QEM for these algorithms and usefully apply concepts such as those described in the present review? For example, it should be possible to improve single-shot algorithms via post-selection, in the same spirit of quantum error detection.

### B. Technical questions

There are also many other more technical questions about the implementations and applications of different QEM techniques, many of which we have mentioned in this review when talking about the individual techniques. We will further list some example questions here.

1. Are there general protocols to scale the noise without changing the error model? If yes, can we deduce the shape of the extrapolation curve from the error model? (Sec. III.A)
2. How can we efficiently handle drifts in the error model when applying probabilistic error cancellation in practice? (Sec. III.B)
3. What is the training cost for learning-based QEM? (Sec. III.H)
4. Various quantum computing hardware architectures have been realised, at least at the prototype level, with differing connectivities and native gate sets. What is the interplay between the hardware feature set and the suitable choices of QEM strategy?

## VII. CONCLUSION

In this review, we have provided a comprehensive survey of QEM which has ranged from the basic concepts and motivations, through to the implementation details of specific techniques. Due to the low hardware requirements of QEM compared to QEC, as well as the broad range of mitigation techniques available to target diverse application scenarios, QEM has already become an integral part of many recent experimental demonstrations of quantum hardware. Even though we cannot rely on QEM alone to suppress *all* errors in all cases, especially when the circuit fault rate is large, we can always aim for a sweet spot between the amount of noise removed and the resource required. Consequently we can expect that QEM techniques will continue to establish themselves as indispensable enablers, vital to maximising the reach of each generation of hardware. We anticipate that this will remain true even into the fault-tolerant era, since recent works have shown that it is possible to reduce the hardware requirements of fault-tolerance by applying QEM alongside QEC. Indeed there are applications of QEM concepts entirely beyond handling physical device imperfections and instead mitigating compilation (algorithmic) imperfections, i.e. honing the performance of algorithms which of necessity produce only approximate answers.

Despite already playing a significant role in practical applications, the current landscape of QEM is still dynamic and complex, with many unexplored territories. While we have made every effort to present the rather tangled threads of this topic in a clear way in this review, it is evident that a more systematic and unified structure for QEM is desirable as the field matures further. There are still many open problems left unanswered as summarised in Sec. VI. Solving these problems may be the key to a clearer and more structured view of QEM and its role in the grand scheme of error suppression. We

hope that the community of theoretical and experimental researchers who will drive the field forwards, may find that the survey of ideas and methods presented in this review provides useful guidance along the way.

## ACKNOWLEDGMENTS

We thank Kristan Temme, Abhinav Kandala, and Jay Gambetta for their valuable insights and useful discussions. ZC is grateful to Bálint Koczor for helpful discussions. ZC is supported by the Junior Research Fellowship from St John’s College, Oxford. RB thanks Nicholas Rubin for helpful discussions about pure state  $N$ -representability. SCB acknowledges financial support from EPSRC Hub grants under agreement No. EP/T001062/1, and from the IARPA funded LogiQ project. SE is supported by Moonshot R&D, JST, Grant No. JPMJMS2061; MEXT Q-LEAP Grant No. JPMXS0120319794, and PRESTO, JST, Grant No. JPMJPR2114, and is grateful for useful discussions with Ryuji Takagi, Nobuyuki Yoshioka, Yasunari Suzuki, and Kento Tsubouchi. YL acknowledges the support of the National Natural Science Foundation of China (Grants No. 11875050 and No. 12088101) and NSAF (Grant No. U1930403).

## LIST OF SYMBOLS

$C_{\text{em}}$	error mitigation sampling overhead
$N$	number of qubits
$N_{\text{cir}}$	number of circuit runs
$O$	observable of interest
$\hat{O}_{\text{em}}$	error-mitigated estimator
$\hat{O}_{\rho}$	random variable from measuring $O$ on state $\rho$
$p$	physical gate error rate
$P_0$	circuit fault-free probability
$\lambda$	circuit fault rate, the average number of faults per circuit run
$\rho$	unmitigated noisy state
$\rho_0$	ideal noiseless state

## Appendix A: Practical Techniques in Implementations

### 1. Monte carlo sampling

In many QEM techniques, very often the error-mitigated expectation value  $E[\hat{O}_{\text{em}}]$  is a linear sum of the expectation values of a set of random variables  $\{\hat{O}_n\}$  (that are the outputs of the set of response measurement circuits for the QEM technique):

$$E[\hat{O}_{\text{em}}] = \sum_n \alpha_n E[\hat{O}_n],$$

where  $\{\alpha_n\}$  are real coefficients. A naive way for estimating  $E[\hat{O}_{\text{em}}]$  would be performing estimation for the individual terms  $E[\hat{O}_n]$  up to a certain precision, then combine the results. Such a method is not scalable if the number of terms  $\{E[\hat{O}_n]\}$  is large, which is the case for many QEM methods. Instead, we can construct the estimator  $\hat{O}_{\text{em}}$  using Monte Carlo methods:

$$\hat{O}_{\text{em}} = A \sum_n \frac{|\alpha_n|}{A} \text{sgn}(\alpha_n) \hat{O}_n = A \hat{O}_{\text{mix}},$$

where  $A = \sum_n |\alpha_n|$ . Here we see that each sample of  $\hat{O}_{\text{em}}$  is just a sample of  $\hat{O}_{\text{mix}}$  scaled by a factor  $A$ , and  $\hat{O}_{\text{mix}}$  is the probabilistic mixture of the set of random variables  $\{\text{sgn}(\alpha_n) \hat{O}_n\}$  which can be sampled by picking  $\text{sgn}(\alpha_n) \hat{O}_n$  with the probability  $\frac{|\alpha_n|}{A}$ .

The component random variables  $\hat{O}_n$ , being generated from circuits that are variants of the primary circuit, can be expected to have a similar variance as the unmitigated estimator  $\hat{O}_\rho$  generated from the noisy primary circuit. In such a case, we then have  $\text{Var}[\hat{O}_n] \sim \text{Var}[\hat{O}_\rho] \sim \text{Var}[\hat{O}_{\text{mix}}]$  and thus

$$\text{Var}[\hat{O}_{\text{em}}] = A^2 \text{Var}[\hat{O}_{\text{mix}}] \sim A^2 \text{Var}[\hat{O}_\rho].$$

Hence, the sampling overhead (Eq. (3)) of performing error-mitigated estimation using  $\hat{O}_{\text{em}}$  instead of  $\hat{O}_\rho$  is then

$$C_{\text{em}} = A^2 = \left( \sum_n |\alpha_n| \right)^2. \quad (\text{A1})$$

Instead of making assumptions about the variance of the component random variables  $\hat{O}_n$ , we can also obtain a similar sampling overhead using Hoeffding's inequality as shown in Eq. (5). This uses the fact that the component random variables  $\hat{O}_n$  usually have the same range as  $\hat{O}_\rho$  since they are usually obtained from the measurement of the same observable, or even if the observables are different, they are often all Pauli observables which have the same range.

Here we have talked about using Monte Carlo sampling for estimating the error-mitigated expectation value. Similar arguments can also be applied to the estimation of loss functions in learning-based methods and other similar situations.

### 2. Pauli twirling

Given a noise process  $\mathcal{N}$ , twirling it over a symmetry group  $\mathbb{G}$  means conjugating  $\mathcal{N}$  with random elements in  $\mathbb{G}$ :

$$\mathbf{T}_{\mathbb{G}}(\mathcal{N}) = \frac{1}{|\mathbb{G}|} \sum_{\mathcal{G} \in \mathbb{G}} \mathcal{G} \mathcal{N} \mathcal{G}.$$

Note that here we are using the super-operator formalism. Twirling over the Pauli group, which is called Pauli twirling, will remove all the off-diagonal elements of  $\mathcal{N}$  in the Pauli basis, transforming the error channel into Pauli noise.

Now suppose we have a noisy Clifford gate  $\mathcal{C}_\epsilon$  that is simply the ideal Clifford Channel  $\mathcal{C}$  followed by the noise channel  $\mathcal{N}$ :  $\mathcal{C}_\epsilon = \mathcal{N} \mathcal{C}$ . If we want to twirl the noise channel  $\mathcal{N}$  of the noisy Clifford gate  $\mathcal{C}_\epsilon$ , we can apply  $\mathcal{C}_\epsilon$  with a random Pauli  $\mathcal{G}$  and its corresponding Pauli  $\mathcal{C}^\dagger \mathcal{G} \mathcal{C}$  in each circuit run:

$$\begin{aligned} \mathbf{T}_{\mathbb{G}}(\mathcal{C}_\epsilon) &= \frac{1}{|\mathbb{G}|} \sum_{\mathcal{G} \in \mathbb{G}} (\mathcal{C} \mathcal{G} \mathcal{C}^\dagger) \mathcal{C}_\epsilon \mathcal{G} = \frac{1}{|\mathbb{G}|} \sum_{\mathcal{G}' \in \mathbb{G}} \mathcal{G}' \mathcal{C}_\epsilon (\mathcal{C}^\dagger \mathcal{G}' \mathcal{C}) \\ &= \left( \frac{1}{|\mathbb{G}|} \sum_{\mathcal{G} \in \mathbb{G}} \mathcal{G} \mathcal{N} \mathcal{G} \right) \mathcal{C}. \end{aligned}$$

To twirl the error of a sequence of Clifford gates  $\prod_{m=M}^1 \mathcal{C}_m$ , the random Pauli for twirling consecutive Clifford gates can be merged.

To twirl a noisy  $T$  gate:  $\mathcal{T}_\epsilon = \mathcal{N} \mathcal{T}$ , we then need to apply the following circuit

$$\mathbf{T}_{\mathbb{G}}(\mathcal{T}_\epsilon) = \frac{1}{|\mathbb{G}|} \sum_{\mathcal{G}' \in \mathbb{G}} \mathcal{G}' \mathcal{T}_\epsilon (\mathcal{T}^\dagger \mathcal{G}' \mathcal{T}) = \left( \frac{1}{|\mathbb{G}|} \sum_{\mathcal{G} \in \mathbb{G}} \mathcal{G} \mathcal{N} \mathcal{G} \right) \mathcal{T},$$

where  $\mathcal{T}^\dagger \mathcal{G}' \mathcal{T}$  is some Clifford gates.

### 3. Measurement techniques

To measure an arbitrary operator, we can always measure its Pauli basis and then combine the results. Hence, without loss of generality, we will be mostly focusing on Pauli measurements in this section.

In practical experiments, it is often the case that we can only perform single-qubit  $Z$  measurements to high accuracy. Hence, one way to measure a Pauli operator is by transforming it into single-qubit  $Z$  using Clifford circuits. Given linear qubit connectivity, the additional Clifford circuits needed will require long-range gates of depth  $\mathcal{O}(\log(N))$  or local gates of depth  $\mathcal{O}(N)$ , which is discussed in the context of symmetry verification by [Bonet-Monroig et al. \(2018\)](#). Alternatively, for a given Pauli operator  $O$ , if we can implement controlled- $O$  gates, we can also *indirectly* measure  $O$  through the Hadamard test as shown in Fig. 6. The controlled- $O$  gate can be

implemented using a Clifford circuit that transforms  $O$  to single-qubit  $Z$  along with a controlled- $Z$  gate. The cost of this is similar to the direct measurement discussed above.

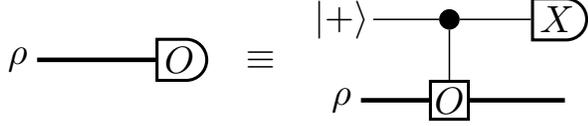


FIG. 6 Hadamard test circuit for performing Pauli  $O$  measurement. If  $O$  is some general unitary, then the Hadamard test circuit above in which we measure  $X \otimes I$  will output the expectation value  $\text{Re}(\text{Tr}[O\rho])$ , while measuring  $Y \otimes I$  will output the expectation value  $\text{Im}(\text{Tr}[O\rho])$ .

Any given Pauli operator  $O$  can be written as the tensor product of single-qubit Pauli operators  $O = \bigotimes_{n=1}^N G_n$  where  $G_n$  is the action of  $O$  on the  $n^{\text{th}}$  qubit. Hence, the controlled- $O$  in the Hadamard test can also be decomposed into low-weight controlled- $G_n$  gates instead. In fact, we can measure  $O$  directly by performing single-qubit Pauli measurements of its components  $\{G_n\}$  and multiplying the results. In this way, the additional circuitry needed for measuring  $O$  is only one layer of single-qubit Clifford for changing the measurement basis.

Now suppose we want to measure *two commuting* Pauli operators in the same circuit run, which can be useful in for example symmetry verification, we can use the single-qubit measurements plus post-processing scheme mentioned above, but we must make sure the two Pauli operators *qubit-wise* commute (Izmaylov *et al.*, 2019), i.e. for every qubit, the single-qubit Pauli components acting on it from different observables commute. For example,  $XXI$  and  $IXZ$  are qubit-wise commuting, but  $XX$  and  $YY$  are not. Note that for a set of operators to qubit-wise commute, their actions on a given qubit must be the same Pauli operator or the identity. More generally, the set of operators that are linear combinations of a set of qubit-wise commuting Pauli operators are also qubit-wise commuting and can be measured simultaneously using single-qubit Pauli measurements in the same circuit run. Whenever two Pauli operators commute, we can make them qubit-wise commuting using a suitable choice of Clifford circuit. Examples of this have been discussed in direct symmetry verification in Sec. III.D.

In some other cases, we are more interested in the expectation values of a set of commuting operators instead of their exact measurement results in any given circuit run. In these scenarios, instead of trying to measure multiple observables in every circuit run, it is possible to obtain these expectation values through *shadow tomography* (Huang *et al.*, 2020) using random single-qubit Pauli measurement and post-processing.

Now let us move beyond measuring commuting Pauli operators and consider the case in which we want to measure the *product* of a Pauli operator  $O$  and some general

Hermitian operator  $S$  that does not necessarily commute with  $O$ . For the circuits shown in Fig. 7, if we perform a projective measurement of the Pauli operator  $O$  first, and then measure the operator  $S$ , then the latter measurement is equivalent to measuring the components of  $S$  that commute with  $O$ , which is simply  $S_+ = \frac{S+OSO}{2}$  (Cai, 2021d; Huo and Li, 2022; Mitarai and Fujii, 2019). Taking the product of the two measurements, we can obtain the expectation value of the symmetrised product  $OS_+ = \frac{OS+S O}{2}$ , which is useful in symmetry verification in Sec. III.D (in which  $S$  is the symmetry) and echo verification in Sec. III.E (in which  $S$  is the dual state).

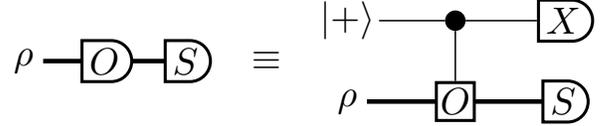


FIG. 7 Circuits for measuring  $OS_+ = \frac{OS+S O}{2}$  in which  $O$  is Pauli and  $S$  is Hermitian. On the left circuit, we first perform a non-destructive measurement of  $O$  and then measure  $S$  and take the product of the measurement results. On the right circuit, the non-destructive measurement of  $O$  is carried out using Hadamard tests and the expectation value of  $OS_+$  is simply obtained by measuring  $X \otimes S$  at the end.

The component of  $S$  that anti-commute with  $O$ , denoted as  $S_- = \frac{S-OSO}{2}$ , can be obtained by first applying the Pauli rotation  $\exp(-i\frac{\pi}{4}O)$  and then measuring  $S$  (Cai, 2021d; Huo and Li, 2022; Mitarai and Fujii, 2019). Combined with the measurement of  $S_+$  above, we can obtain the expectation value of the product observable  $OS$ . Alternatively, if we can implement controlled- $O$  and controlled- $S$ , then they can be composed to give controlled- $OS$ , which can be used to measure  $OS$  using the Hadamard test.

## REFERENCES

- Abobeih, M H, Y. Wang, J. Randall, S. J. H. Loenen, C. E. Bradley, M. Markham, D. J. Twitchen, B. M. Terhal, and T. H. Taminiau (2022), ‘‘Fault-tolerant operation of a logical qubit in a diamond quantum processor,’’ *Nature* **606** (7916), 884–889.
- Acín, Antonio, Immanuel Bloch, Harry Buhrman, Tommaso Calarco, Christopher Eichler, Jens Eisert, Daniel Esteve, Nicolas Gisin, Steffen J. Glaser, Fedor Jelezko, Stefan Kuhr, Maciej Lewenstein, Max F. Riedel, Piet O. Schmidt, Rob Thew, Andreas Wallraff, Ian Walmsley, and Frank K. Wilhelm (2018), ‘‘The quantum technologies roadmap: A European community view,’’ *New Journal of Physics* **20** (8), 080201.
- Aharonov, D, and M. Ben-Or (1997), ‘‘Fault-tolerant Quantum Computation with Constant Error,’’ in *Proceedings of the Twenty-ninth Annual ACM Symposium on Theory of Computing*, STOC ’97 (ACM, New York, NY, USA) pp. 176–188.
- Altman, Ehud, Kenneth R. Brown, Giuseppe Carleo, Lincoln D. Carr, Eugene Demler, Cheng Chin, Brian De-

- Marco, Sophia E. Economou, Mark A. Eriksson, Kai-Mei C. Fu, Markus Greiner, Kaden R.A. Hazzard, Randall G. Hulet, Alicia J. Kollár, Benjamin L. Lev, Mikhail D. Lukin, Ruichao Ma, Xiao Mi, Shashank Misra, Christopher Monroe, Kater Murch, Zaira Nazario, Kang-Kuen Ni, Andrew C. Potter, Pedram Roushan, Mark Saffman, Monika Schleier-Smith, Irfan Siddiqi, Raymond Simmonds, Meenakshi Singh, I.B. Spielman, Kristan Temme, David S. Weiss, Jelena Vučković, Vladan Vuletić, Jun Ye, and Martin Zwerlein (2021), “Quantum Simulators: Architectures and Opportunities,” *PRX Quantum* **2** (1), 017003.
- Arute, Frank, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C. Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando G. S. L. Brandao, David A. Buell, Brian Burkett, Yu Chen, Zijun Chen, Ben Chiaro, Roberto Collins, William Courtney, Andrew Dunsworth, Edward Farhi, Brooks Foxen, Austin Fowler, Craig Gidney, Marissa Giustina, Rob Graff, Keith Guerín, Steve Habegger, Matthew P. Harrigan, Michael J. Hartmann, Alan Ho, Markus Hoffmann, Trent Huang, Travis S. Humble, Sergei V. Isakov, Evan Jeffrey, Zhang Jiang, Dvir Kafri, Kostyantyn Kechedzhi, Julian Kelly, Paul V. Klimov, Sergey Knysh, Alexander Korotkov, Fedor Kostritsa, David Landhuis, Mike Lindmark, Erik Lucero, Dmitry Lyakh, Salvatore Mandrà, Jarrod R. McClean, Matthew McEwen, Anthony Megrant, Xiao Mi, Kristel Michielsen, Masoud Mohseni, Josh Mutus, Ofer Naaman, Matthew Neeley, Charles Neill, Murphy Yuezhen Niu, Eric Ostby, Andre Petukhov, John C. Platt, Chris Quintana, Eleanor G. Rieffel, Pedram Roushan, Nicholas C. Rubin, Daniel Sank, Kevin J. Satzinger, Vadim Smelyanskiy, Kevin J. Sung, Matthew D. Trevithick, Amit Vainsencher, Benjamin Vallonga, Theodore White, Z. Jamie Yao, Ping Yeh, Adam Zalcman, Hartmut Neven, and John M. Martinis (2019), “Quantum supremacy using a programmable superconducting processor,” *Nature* **574** (7779), 505–510.
- Asavanant, Warit, Yu Shiozawa, Shota Yokoyama, Baramjee Charoensombutamon, Hiroki Emura, Rafael N. Alexander, Shuntaro Takeda, Jun-ichi Yoshikawa, Nicolas C. Menicucci, Hidehiro Yonezawa, and Akira Furusawa (2019), “Generation of time-domain-multiplexed two-dimensional cluster state,” *Science* **366** (6463), 373–376.
- Barron, George S, and Christopher J. Wood (2020), “Measurement Error Mitigation for Variational Quantum Algorithms,” arXiv:2010.08520 [quant-ph].
- Bennett, Charles H, Gilles Brassard, Sandu Popescu, Benjamin Schumacher, John A. Smolin, and William K. Wootters (1996a), “Purification of Noisy Entanglement and Faithful Teleportation via Noisy Channels,” *Physical Review Letters* **76** (5), 722–725.
- Bennett, Charles H, David P. DiVincenzo, John A. Smolin, and William K. Wootters (1996b), “Mixed-state entanglement and quantum error correction,” *Physical Review A* **54** (5), 3824–3851.
- van den Berg, Ewout, Zlatko K. Mineev, Abhinav Kandala, and Kristan Temme (2022), “Probabilistic error cancellation with sparse Pauli-Lindblad models on noisy quantum processors,” arXiv:2201.09866 [quant-ph].
- Bergquist, J C, Randall G. Hulet, Wayne M. Itano, and D. J. Wineland (1986), “Observation of Quantum Jumps in a Single Atom,” *Physical Review Letters* **57** (14), 1699–1702.
- Biamonte, Jacob, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd (2017), “Quantum machine learning,” *Nature* **549** (7671), 195–202.
- Blais, Alexandre, Ren-Shou Huang, Andreas Wallraff, S. M. Girvin, and R. J. Schoelkopf (2004), “Cavity quantum electrodynamics for superconducting electrical circuits: An architecture for quantum computation,” *Physical Review A* **69** (6), 062320.
- Bonet-Monroig, X, R. Sagastizabal, M. Singh, and T. E. O’Brien (2018), “Low-cost error mitigation by symmetry verification,” *Physical Review A* **98** (6), 062339.
- Bonet-Monroig, Xavier, Ryan Babbush, and Thomas E. O’Brien (2020), “Nearly Optimal Measurement Scheduling for Partial Tomography of Quantum States,” *Physical Review X* **10** (3), 031064.
- Bravyi, Sergey, and David Gosset (2016), “Improved Classical Simulation of Quantum Circuits Dominated by Clifford Gates,” *Physical Review Letters* **116** (25), 250501.
- Bravyi, Sergey, and Alexei Kitaev (2002), “Fermionic quantum computation,” *Annals of Physics* **298** (1), 210–226.
- Bravyi, Sergey, Sarah Sheldon, Abhinav Kandala, David C. McKay, and Jay M. Gambetta (2021), “Mitigating measurement errors in multiqubit experiments,” *Physical Review A* **103** (4), 042605.
- Breuckmann, Nikolas P, and Jens Niklas Eberhardt (2021), “Quantum Low-Density Parity-Check Codes,” *PRX Quantum* **2** (4), 040101.
- Bridgeman, Jacob C, and Christopher T Chubb (2017), “Hand-waving and interpretive dance: An introductory course on tensor networks,” *Journal of Physics A: Mathematical and Theoretical* **50** (22), 223001.
- Bultrini, Daniel, Max Hunter Gordon, Piotr Czarnik, Andrew Arrasmith, Patrick J. Coles, and Lukasz Cincio (2021), “Unifying and benchmarking state-of-the-art quantum error mitigation techniques,” arXiv:2107.13470 [quant-ph].
- Burgarth, Daniel, Paolo Facchi, Giovanni Gramegna, and Saverio Pascazio (2019), “Generalized product formulas and quantum control,” *Journal of Physics A: Mathematical and Theoretical* **52** (43), 435301.
- Cai, Zhenyu (2020), “Resource Estimation for Quantum Variational Simulations of the Hubbard Model,” *Physical Review Applied* **14** (1), 014059.
- Cai, Zhenyu (2021a), “Multi-exponential error extrapolation and combining error mitigation techniques for NISQ applications,” *npj Quantum Information* **7** (1), 1–12.
- Cai, Zhenyu (2021b), “A Practical Framework for Quantum Error Mitigation,” arXiv:2110.05389 [quant-ph].
- Cai, Zhenyu (2021c), “Quantum Error Mitigation using Symmetry Expansion,” *Quantum* **5**, 548.
- Cai, Zhenyu (2021d), “Resource-efficient Purification-based Quantum Error Mitigation,” arXiv:2107.07279 [quant-ph].
- Cai, Zhenyu, Adam Siegel, and Simon Benjamin (2022), “Looped Pipelines Enabling Effective 3D Qubit Lattices in a Strictly 2D Device,” arXiv:2203.13123 [quant-ph].
- Calderbank, A R, and Peter W. Shor (1996), “Good quantum error-correcting codes exist,” *Physical Review A* **54** (2), 1098–1105.
- Campbell, Earl (2019), “Random Compiler for Fast Hamiltonian Simulation,” *Physical Review Letters* **123** (7), 070503.
- Chen, Senrui, Wenjun Yu, Pei Zeng, and Steven T. Flammia (2021), “Robust Shadow Estimation,” *PRX Quantum* **2** (3), 030348.
- Chen, Yanzhu, Maziar Farahzad, Shinjae Yoo, and Tzu-Chieh Wei (2019), “Detector tomography on IBM quantum computers and mitigation of an imperfect measurement,” *Physical Review A* **100** (5), 052315.
- Childs, Andrew M, Aaron Ostrander, and Yuan Su (2019),

- “Faster quantum simulation by randomization,” *Quantum* **3**, 182.
- Childs, Andrew M, and Nathan Wiebe (2012), “Hamiltonian simulation using linear combinations of unitary operations,” *Quantum Information and Computation* **12** (11&12), 901–924.
- Chow, J M, L. DiCarlo, J. M. Gambetta, A. Nunnenkamp, Lev S. Bishop, L. Frunzio, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf (2010), “Detecting highly entangled states with a joint qubit readout,” *Physical Review A* **81** (6), 062325.
- Chow, Jerry M, A. D. Córcoles, Jay M. Gambetta, Chad Rigetti, B. R. Johnson, John A. Smolin, J. R. Rozen, George A. Keefe, Mary B. Rothwell, Mark B. Ketchen, and M. Steffen (2011), “Simple All-Microwave Entangling Gate for Fixed-Frequency Superconducting Qubits,” *Physical Review Letters* **107** (8), 080502.
- Colless, J I, V. V. Ramasesh, D. Dahlen, M. S. Blok, M. E. Kimchi-Schwartz, J. R. McClean, J. Carter, W. A. de Jong, and I. Siddiqi (2018), “Computation of Molecular Spectra on a Quantum Processor with an Error-Resilient Algorithm,” *Physical Review X* **8** (1), 011021.
- Córcoles, A D, Easwar Magesan, Srikanth J. Srinivasan, Andrew W. Cross, M. Steffen, Jay M. Gambetta, and Jerry M. Chow (2015), “Demonstration of a quantum error detection code using a square lattice of four superconducting qubits,” *Nature Communications* **6** (1), 6979.
- Czarnik, Piotr, Andrew Arrasmith, Patrick J. Coles, and Lukasz Cincio (2021), “Error mitigation with Clifford quantum-circuit data,” *Quantum* **5**, 592.
- Czarnik, Piotr, Michael McKerns, Andrew T. Sornborger, and Lukasz Cincio (2022), “Improving the efficiency of learning-based error mitigation,” arXiv:2204.07109 [quant-ph].
- Dallaire-Demers, Pierre-Luc, Jonathan Romero, Libor Veis, Sukin Sim, and Alán Aspuru-Guzik (2019), “Low-depth circuit ansatz for preparing correlated fermionic states on a quantum computer,” *Quantum Science and Technology* **4** (4), 045005.
- Dawson, CM, and M.A. Nielsen (2006), “The Solovay-Kitaev algorithm,” *Quantum Information and Computation* **6** (1), 81–95.
- Dborin, James, Vinul Wimalaweera, Fergus Barratt, Eric Ostby, Thomas E. O’Brien, and Andrew G. Green (2022), “Simulating groundstate and dynamical quantum phase transitions on a superconducting quantum computer,” arXiv:2205.12996 [quant-ph].
- DePrince, A Eugene (2016), “Variational optimization of the two-electron reduced-density matrix under pure-state N-representability conditions,” *The Journal of Chemical Physics* **145** (16), 164109.
- Derby, Charles, Joel Klassen, Johannes Bausch, and Toby Cubitt (2021), “Compact fermion to qubit mappings,” *Physical Review B* **104** (3), 035118.
- Dinur, Irit, Shai Evra, Ron Livne, Alexander Lubotzky, and Shahar Mozes (2022), “Locally testable codes with constant rate, distance, and locality,” in *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2022 (Association for Computing Machinery, New York, NY, USA) pp. 357–374.
- Dumitrescu, E F, A. J. McCaskey, G. Hagen, G. R. Jansen, T. D. Morris, T. Papenbrock, R. C. Pooser, D. J. Dean, and P. Lougovski (2018), “Cloud Quantum Computing of an Atomic Nucleus,” *Physical Review Letters* **120** (21), 210501.
- Eastin, Bryan, and Emanuel Knill (2009), “Restrictions on Transversal Encoded Quantum Gate Sets,” *Physical Review Letters* **102** (11), 10.1103/PhysRevLett.102.110502.
- Ebadi, Sepehr, Tout T. Wang, Harry Levine, Alexander Keesling, Giulia Semeghini, Ahmed Omran, Dolev Bluvstein, Rhine Samajdar, Hannes Pichler, Wen Wei Ho, Soonwon Choi, Subir Sachdev, Markus Greiner, Vladan Vuletić, and Mikhail D. Lukin (2021), “Quantum phases of matter on a 256-atom programmable quantum simulator,” *Nature* **595** (7866), 227–232.
- Egan, Laird, Dripto M. Debroy, Crystal Noel, Andrew Risinger, Daiwei Zhu, Debopriyo Biswas, Michael Newman, Muyuan Li, Kenneth R. Brown, Marko Cetina, and Christopher Monroe (2021), “Fault-tolerant control of an error-corrected qubit,” *Nature* **598** (7880), 281–286.
- Eisert, Jens, Dominik Hangleiter, Nathan Walk, Ingo Roth, Damian Markham, Rhea Parekh, Ulysse Chabaud, and Elham Kashefi (2020), “Quantum certification and benchmarking,” *Nature Reviews Physics* **2** (7), 382–390.
- Endo, Suguru, Simon C. Benjamin, and Ying Li (2018), “Practical Quantum Error Mitigation for Near-Future Applications,” *Physical Review X* **8** (3), 031027.
- Endo, Suguru, Qi Zhao, Ying Li, Simon Benjamin, and Xiao Yuan (2019), “Mitigating algorithmic errors in a Hamiltonian simulation,” *Physical Review A* **99** (1), 012334.
- Facchi, P, D. A. Lidar, and S. Pascazio (2004), “Unification of dynamical decoupling and the quantum Zeno effect,” *Physical Review A* **69** (3), 032314.
- Farhi, Edward, Jeffrey Goldstone, and Sam Gutmann (2014), “A Quantum Approximate Optimization Algorithm,” arXiv:1411.4028 [quant-ph].
- Ferracin, Samuele, Akel Hashim, Jean-Loup Ville, Ravi Naik, Arnaud Carignan-Dugas, Hammam Qassim, Alexis Morvan, David I. Santiago, Irfan Siddiqi, and Joel J. Wallman (2022), “Efficiently improving the performance of noisy quantum computers,” arXiv:2201.10672 [quant-ph].
- Feynman, Richard P (1982), “Simulating physics with computers,” *International Journal of Theoretical Physics* **21** (6-7), 467–488.
- Foss-Feig, Michael, Stephen Ragole, Andrew Potter, Joan Dreiling, Caroline Figgatt, John Gaebler, Alex Hall, Steven Moses, Juan Pino, Ben Spaun, Brian Neyenhuis, and David Hayes (2022), “Entanglement from Tensor Networks on a Trapped-Ion Quantum Computer,” *Physical Review Letters* **128** (15), 150504.
- Fowler, Austin G, Matteo Mariantoni, John M. Martinis, and Andrew N. Cleland (2012), “Surface codes: Towards practical large-scale quantum computation,” *Physical Review A* **86** (3), 032324.
- Fuchs, CA, and J. van de Graaf (1999), “Cryptographic distinguishability measures for quantum-mechanical states,” *IEEE Transactions on Information Theory* **45** (4), 1216–1227.
- Garmon, J W O, R. C. Pooser, and E. F. Dumitrescu (2020), “Benchmarking noise extrapolation with the OpenPulse control framework,” *Physical Review A* **101** (4), 042308.
- Geller, Michael R (2020), “Rigorous measurement error correction,” *Quantum Science and Technology* **5** (3), 03LT01.
- Gilchrist, Alexei, Daniel R. Terno, and Christopher J. Wood (2011), “Vectorization of quantum operations and its use,” arXiv:0911.2539 [quant-ph].
- Giurgica-Tiron, Tudor, Yousef Hindy, Ryan LaRose, Andrea Mari, and William J. Zeng (2020), “Digital zero noise extrapolation for quantum error mitigation,” in *2020 IEEE*

- International Conference on Quantum Computing and Engineering (QCE)*, pp. 306–316.
- Google Quantum AI, (2022), “Suppressing quantum errors by scaling a surface code logical qubit,” arXiv:2207.06431 [quant-ph].
- Google Quantum AI and Collaborators, (2020a), “Hartree-Fock on a superconducting qubit quantum computer,” *Science* **369** (6507), 1084–1089.
- Google Quantum AI and Collaborators, (2020b), “Observation of separated dynamics of charge and spin in the Fermi-Hubbard model,” arXiv:2010.07965 [quant-ph].
- Google Quantum AI and Collaborators, (2022), “Formation of robust bound states of interacting photons,” arXiv:2206.05254 [cond-mat, physics:quant-ph].
- Gottesman, Daniel (2009), “An Introduction to Quantum Error Correction and Fault-Tolerant Quantum Computation,” arXiv:0904.2557 [quant-ph].
- Gottesman, Daniel (2014), “Fault-Tolerant quantum computation with constant overhead,” *Quantum Information and Computation* **14** (15&16), 1339–1371.
- Gottesman, Daniel (2016), “Quantum fault tolerance in small experiments,” arXiv:1610.03507 [quant-ph].
- Gottesman, Daniel Eric (1997), *Stabilizer Codes and Quantum Error Correction*, Ph.D. thesis (California Institute of Technology).
- Gu, Yanwu, Yunheng Ma, Nicolo Forcellini, and Dong E. Liu (2022), “Noise-resilient phase estimation with randomized compiling,” arXiv:2208.04100 [quant-ph].
- Gutiérrez, Mauricio, and Kenneth R. Brown (2015), “Comparison of a quantum error-correction threshold for exact and approximate errors,” *Physical Review A* **91** (2), 022335.
- Hakoshima, Hideaki, Yuichiro Matsuzaki, and Suguru Endo (2021), “Relationship between costs for quantum error mitigation and non-Markovian measures,” *Physical Review A* **103** (1), 012611.
- Hama, Yusuke, and Hirofumi Nishi (2022), “Quantum Error Mitigation via Quantum-Noise-Effect Circuit Groups,” arXiv:2205.13907 [quant-ph].
- Hamilton, Kathleen E, Tyler Kharazi, Titus Morris, Alexander J. McCaskey, Ryan S. Bennink, and Raphael C. Pooser (2020), “Scalable quantum processor noise characterization,” in *2020 IEEE International Conference on Quantum Computing and Engineering (QCE)*, pp. 430–440.
- Haroche, Serge, and Jean-Michel Raimond (2006), *Exploring the Quantum: Atoms, Cavities, and Photons*, Oxford Graduate Texts (Oxford University Press, Oxford).
- He, Andre, Benjamin Nachman, Wibe A. de Jong, and Christian W. Bauer (2020), “Zero-noise extrapolation for quantum-gate error mitigation with identity insertions,” *Physical Review A* **102** (1), 012426.
- Heinsoo, Johannes, Christian Kraglund Andersen, Ants Remm, Sebastian Krinner, Theodore Walter, Yves Salathé, Simone Gasparinetti, Jean-Claude Besse, Anton Potočnik, Andreas Wallraff, and Christopher Eichler (2018), “Rapid High-fidelity Multiplexed Readout of Superconducting Qubits,” *Physical Review Applied* **10** (3), 034040.
- Helgaker, Trygve, Poul Jorgensen, and Jeppe Olsen (2000), *Molecular Electronic-Structure Theory* (John Wiley & Sons, Ltd).
- Hicks, Rebecca, Bryce Kobrin, Christian W. Bauer, and Benjamin Nachman (2022), “Active readout-error mitigation,” *Physical Review A* **105** (1), 012419.
- Huang, Hsin-Yuan, Richard Kueng, and John Preskill (2020), “Predicting many properties of a quantum system from very few measurements,” *Nature Physics* **16** (10), 1050–1057.
- Huggins, William J, Sam McArdle, Thomas E. O’Brien, Joonho Lee, Nicholas C. Rubin, Sergio Boixo, K. Birgitta Whaley, Ryan Babbush, and Jarrod R. McClean (2021a), “Virtual Distillation for Quantum Error Mitigation,” *Physical Review X* **11** (4), 041036.
- Huggins, William J, Jarrod R. McClean, Nicholas C. Rubin, Zhang Jiang, Nathan Wiebe, K. Birgitta Whaley, and Ryan Babbush (2021b), “Efficient and noise resilient measurements for quantum chemistry on near-term quantum computers,” *npj Quantum Information* **7** (1), 1–9.
- Huo, Mingxia, and Ying Li (2022), “Dual-state purification for practical quantum error mitigation,” *Physical Review A* **105** (2), 022427.
- Izmaylov, Artur F, Tzu-Ching Yen, and Ilya G. Ryabinkin (2019), “Revising the measurement process in the variational quantum eigensolver: Is it possible to reduce the number of separately measured operators?” *Chemical Science* **10** (13), 3746–3755.
- Jiang, Jiaqing, Kun Wang, and Xin Wang (2021), “Physical Implementability of Linear Maps and Its Application in Error Mitigation,” *Quantum* **5**, 600.
- Jiang, Zhang, Jarrod McClean, Ryan Babbush, and Hartmut Neven (2019), “Majorana Loop Stabilizer Codes for Error Mitigation in Fermionic Quantum Simulations,” *Physical Review Applied* **12** (6), 064041.
- Jurcevic, Petar, Ali Javadi-Abhari, Lev S. Bishop, Isaac Lauer, Daniela F. Bogorin, Markus Brink, Lauren Capelluto, Oktay Günlük, Toshinari Itoko, Naoki Kanazawa, Abhinav Kandala, George A. Keefe, Kevin Krsulich, William Landers, Eric P. Lewandowski, Douglas T. McClure, Giacomo Nannicini, Adinath Narasgond, Hasan M. Nayfeh, Emily Pritchett, Mary Beth Rothwell, Srikanth Srinivasan, Neereja Sundaresan, Cindy Wang, Ken X. Wei, Christopher J. Wood, Jeng-Bang Yau, Eric J. Zhang, Oliver E. Dial, Jerry M. Chow, and Jay M. Gambetta (2021), “Demonstration of quantum volume 64 on a superconducting quantum computing system,” *Quantum Science and Technology* **6** (2), 025020.
- Kandala, Abhinav, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M. Chow, and Jay M. Gambetta (2017), “Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets,” *Nature* **549** (7671), 242–246.
- Kandala, Abhinav, Kristan Temme, Antonio D. Córcoles, Antonio Mezzacapo, Jerry M. Chow, and Jay M. Gambetta (2019), “Error mitigation extends the computational reach of a noisy quantum processor,” *Nature* **567** (7749), 491–495.
- Keen, Trevor, Thomas Maier, Steven Johnston, and Pavel Lougovski (2020), “Quantum-classical simulation of two-site dynamical mean-field theory on noisy quantum hardware,” *Quantum Science and Technology* **5** (3), 035001.
- Kelly, J. R. Barends, A. G. Fowler, A. Megrant, E. Jeffrey, T. C. White, D. Sank, J. Y. Mutus, B. Campbell, Yu Chen, Z. Chen, B. Chiaro, A. Dunsworth, I.-C. Hoi, C. Neill, P. J. J. O’Malley, C. Quintana, P. Roushan, A. Vainsencher, J. Wenner, A. N. Cleland, and John M. Martinis (2015), “State preservation by repetitive error detection in a superconducting quantum circuit,” *Nature* **519** (7541), 66–69.
- Khamoshi, Armin, Francesco A. Evangelista, and Gustavo E.

- Scuseria (2020), “Correlating AGP on a quantum computer,” *Quantum Science and Technology* **6** (1), 014004.
- Kim, Youngseok, Christopher J. Wood, Theodore J. Yoder, Seth T. Merkel, Jay M. Gambetta, Kristan Temme, and Abhinav Kandala (2021), “Scalable error mitigation for noisy quantum circuits produces competitive expectation values,” arXiv:2108.09197 [cond-mat, physics:quant-ph].
- Kitaev, A Yu (1997), “Quantum computations: Algorithms and error correction,” *Russian Mathematical Surveys* **52** (6), 1191.
- Kivlichan, Ian D, Craig Gidney, Dominic W. Berry, Nathan Wiebe, Jarrod McClean, Wei Sun, Zhang Jiang, Nicholas Rubin, Austin Fowler, Alán Aspuru-Guzik, Hartmut Neven, and Ryan Babbush (2020), “Improved Fault-Tolerant Quantum Simulation of Condensed-Phase Correlated Electrons via Trotterization,” *Quantum* **4**, 296.
- Klco, N, E. F. Dumitrescu, A. J. McCaskey, T. D. Morris, R. C. Pooser, M. Sanz, E. Solano, P. Lougovski, and M. J. Savage (2018), “Quantum-classical computation of Schwinger model dynamics using quantum computers,” *Physical Review A* **98** (3), 032331.
- Kliesch, Martin, and Ingo Roth (2021), “Theory of Quantum System Certification,” *PRX Quantum* **2** (1), 010201.
- Kliuchnikov, Vadym, Dmitri Maslov, and Michele Mosca (2013), “Asymptotically Optimal Approximation of Single Qubit Unitaries by Clifford and  $T$  Circuits Using a Constant Number of Ancillary Qubits,” *Physical Review Letters* **110** (19), 190502.
- Klyachko, Alexander A (2006), “Quantum marginal problem and  $N$ -representability,” *Journal of Physics: Conference Series* **36**, 72–86.
- Knill, E (2004), “Fault-Tolerant Postselected Quantum Computation: Threshold Analysis,” arXiv:quant-ph/0404104.
- Koczor, Bálint (2021a), “The dominant eigenvector of a noisy quantum state,” *New Journal of Physics* **23** (12), 123047.
- Koczor, Bálint (2021b), “Exponential Error Suppression for Near-Term Quantum Devices,” *Physical Review X* **11** (3), 031057.
- Krinner, Sebastian, Nathan Lacroix, Ants Remm, Agustin Di Paolo, Elie Genois, Catherine Leroux, Christoph Hellings, Stefania Lazar, Francois Swiadek, Johannes Herrmann, Graham J. Norris, Christian Kraglund Andersen, Markus Müller, Alexandre Blais, Christopher Eichler, and Andreas Wallraff (2022), “Realizing repeated quantum error correction in a distance-three surface code,” *Nature* **605** (7911), 669–674.
- Kueng, Richard, David M. Long, Andrew C. Doherty, and Steven T. Flammia (2016), “Comparing Experiments to the Fault-Tolerance Threshold,” *Physical Review Letters* **117** (17), 170502.
- LaRose, Ryan, Andrea Mari, Sarah Kaiser, Peter Karalekas, Andre Alves, Piotr Czarnik, Mohamed El Mandouh, Max Gordon, Yousef Hindy, Aaron Robertson, Purva Thakre, Nathan Shammah, and William Zeng (2021), “Mitiq: A software package for error mitigation on noisy quantum computers,” .
- Le Cam, Lucien (1960), “An approximation theorem for the Poisson binomial distribution,” *Pacific Journal of Mathematics* **10** (4), 1181–1197.
- Lee, Joonho, Dominic W. Berry, Craig Gidney, William J. Huggins, Jarrod R. McClean, Nathan Wiebe, and Ryan Babbush (2021), “Even More Efficient Quantum Computations of Chemistry Through Tensor Hypercontraction,” *PRX Quantum* **2** (3), 030305.
- Li, Ying, and Simon C. Benjamin (2017), “Efficient Variational Quantum Simulator Incorporating Active Error Minimization,” *Physical Review X* **7** (2), 021050.
- Lidar, Daniel A (2014), “Review of Decoherence-Free Subspaces, Noiseless Subsystems, and Dynamical Decoupling,” in *Quantum Information and Computation for Chemistry* (John Wiley & Sons, Ltd) pp. 295–354.
- Lidar, Daniel A, and Todd A. Brun (2013), “Introduction to decoherence and noise in open quantum systems,” in *Quantum Error Correction*, edited by Daniel A. Lidar and Todd A. Brun (Cambridge University Press, Cambridge) pp. 3–45.
- Lin, Junan, Joel J. Wallman, Ian Hincks, and Raymond Laflamme (2021), “Independent state and measurement characterization for quantum computers,” *Physical Review Research* **3** (3), 033285.
- Linke, Norbert M, Mauricio Gutierrez, Kevin A. Landsman, Caroline Figgatt, Shantanu Debnath, Kenneth R. Brown, and Christopher Monroe (2017), “Fault-tolerant quantum error detection,” *Science Advances* **3** (10), e1701074.
- Litinski, Daniel (2019), “Magic State Distillation: Not as Costly as You Think,” *Quantum* **3**, 205.
- Liu, Yi-Kai, Matthias Christandl, and F. Verstraete (2007), “Quantum Computational Complexity of the  $\mathbb{N}$ -Representability Problem: QMA Complete,” *Physical Review Letters* **98** (11), 110503.
- Lloyd, Seth, Masoud Mohseni, and Patrick Rebentrost (2014), “Quantum principal component analysis,” *Nature Physics* **10** (9), 631–633.
- Lostaglio, M, and A. Ciani (2021), “Error Mitigation and Quantum-Assisted Simulation in the Error Corrected Regime,” *Physical Review Letters* **127** (20), 200506.
- Lowe, Angus, Max Hunter Gordon, Piotr Czarnik, Andrew Arrasmith, Patrick J. Coles, and Lukasz Cincio (2021), “Unified approach to data-driven quantum error mitigation,” *Physical Review Research* **3** (3), 033098.
- Lu, Sirui, Mari Carmen Bañuls, and J. Ignacio Cirac (2021), “Algorithms for Quantum Simulation at Finite Energies,” *PRX Quantum* **2** (2), 020321.
- Maciejewski, Filip B, Zoltán Zimborás, and Michał Oszmaniec (2020), “Mitigation of readout noise in near-term quantum devices by classical post-processing based on detector tomography,” *Quantum* **4**, 257.
- Madjarov, Ivaylo S, Jacob P. Covey, Adam L. Shaw, Joonhee Choi, Anant Kale, Alexandre Cooper, Hannes Pichler, Vladimir Schkolnik, Jason R. Williams, and Manuel Endres (2020), “High-fidelity entanglement and detection of alkaline-earth Rydberg atoms,” *Nature Physics* **16** (8), 857–861.
- Mari, Andrea, Nathan Shammah, and William J. Zeng (2021), “Extending quantum probabilistic error cancellation by noise scaling,” *Physical Review A* **104** (5), 052607.
- Mazziotti, David A (2016), “Pure- $\mathbb{N}$ -representability conditions of two-fermion reduced density matrices,” *Physical Review A* **94** (3), 032516.
- McArdle, Sam, Xiao Yuan, and Simon Benjamin (2019), “Error-Mitigated Digital Quantum Simulation,” *Physical Review Letters* **122** (18), 180501.
- McCaskey, Alexander J, Zachary P. Parks, Jacek Jakowski, Shirley V. Moore, Titus D. Morris, Travis S. Humble, and Raphael C. Pooser (2019), “Quantum chemistry as a benchmark for near-term quantum computers,” *npj Quantum Information* **5** (1), 1–8.
- McClean, Jarrod R, Zhang Jiang, Nicholas C. Rubin, Ryan

- Babbush, and Hartmut Neven (2020), “Decoding quantum errors with subspace expansions,” *Nature Communications* **11** (1), 636.
- McClean, Jarrod R, Mollie E. Kimchi-Schwartz, Jonathan Carter, and Wibe A. de Jong (2017), “Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states,” *Physical Review A* **95** (4), 042308.
- McClean, Jarrod R, Jonathan Romero, Ryan Babbush, and Alán Aspuru-Guzik (2016), “The theory of variational hybrid quantum-classical algorithms,” *New Journal of Physics* **18** (2), 023023.
- McWeeny, R (1960), “Some Recent Advances in Density Matrix Theory,” *Reviews of Modern Physics* **32** (2), 335–369.
- Merkel, Seth T, Jay M. Gambetta, John A. Smolin, Stefano Poletto, Antonio D. Córcoles, Blake R. Johnson, Colm A. Ryan, and Matthias Steffen (2013), “Self-consistent quantum process tomography,” *Physical Review A* **87** (6), 062119.
- Mi, Xiao, Pedram Roushan, Chris Quintana, Salvatore Mandrà, Jeffrey Marshall, Charles Neill, Frank Arute, Kunal Arya, Juan Atalaya, Ryan Babbush, Joseph C. Bardin, Rami Barends, Joao Basso, Andreas Bengtsson, Sergio Boixo, Alexandre Bourassa, Michael Broughton, Bob B. Buckley, David A. Buell, Brian Burkett, Nicholas Bushnell, Zijun Chen, Benjamin Chiaro, Roberto Collins, William Courtney, Sean Demura, Alan R. Derk, Andrew Dunsworth, Daniel Eppens, Catherine Erickson, Edward Farhi, Austin G. Fowler, Brooks Foxen, Craig Gidney, Marissa Giustina, Jonathan A. Gross, Matthew P. Harrigan, Sean D. Harrington, Jeremy Hilton, Alan Ho, Sabrina Hong, Trent Huang, William J. Huggins, L. B. Ioffe, Sergei V. Isakov, Evan Jeffrey, Zhang Jiang, Cody Jones, Dvir Kafri, Julian Kelly, Seon Kim, Alexei Kitaev, Paul V. Klimov, Alexander N. Korotkov, Fedor Kostritsa, David Landhuis, Pavel Laptev, Erik Lucero, Orion Martin, Jarrod R. McClean, Trevor McCourt, Matt McEwen, Anthony Megrant, Kevin C. Miao, Masoud Mohseni, Shirin Montazeri, Wojciech Mruczkiewicz, Josh Mutus, Ofer Naaman, Matthew Neeley, Michael Newman, Murphy Yuezhen Niu, Thomas E. O’Brien, Alex Opremcak, Eric Ostby, Balint Pato, Andre Petukhov, Nicholas Redd, Nicholas C. Rubin, Daniel Sank, Kevin J. Satzinger, Vladimir Shvarts, Doug Strain, Marco Szalay, Matthew D. Trevithick, Benjamin Villalonga, Theodore White, Z. Jamie Yao, Ping Yeh, Adam Zalcman, Hartmut Neven, Igor Aleiner, Kostyantyn Kechedzhi, Vadim Smelyanskiy, and Yu Chen (2021), “Information scrambling in quantum circuits,” *Science* **374** (6574), 1479–1483.
- Mitarai, Kosuke, and Keisuke Fujii (2019), “Methodology for replacing indirect measurements with direct measurements,” *Physical Review Research* **1** (1), 013006.
- Montanaro, Ashley, and Stasja Stanisic (2021), “Error mitigation by training with fermionic linear optics,” arXiv:2102.02120 [quant-ph].
- Motta, Mario, Chong Sun, Adrian T. K. Tan, Matthew J. O’Rourke, Erika Ye, Austin J. Minnich, Fernando G. S. L. Brandão, and Garnet Kin-Lic Chan (2020), “Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution,” *Nature Physics* **16** (2), 205–210.
- Nachman, Benjamin, Miroslav Urbanek, Wibe A. de Jong, and Christian W. Bauer (2020), “Unfolding quantum computer readout noise,” *npj Quantum Information* **6** (1), 1–7.
- Nagourney, Warren, Jon Sandberg, and Hans Dehmelt (1986), “Shelved optical electron amplifier: Observation of quantum jumps,” *Physical Review Letters* **56** (26), 2797–2799.
- Nation, Paul D, Hwajung Kang, Neereja Sundaresan, and Jay M. Gambetta (2021), “Scalable Mitigation of Measurement Errors on Quantum Computers,” *PRX Quantum* **2** (4), 040326.
- Neill, C, T. McCourt, X. Mi, Z. Jiang, M. Y. Niu, W. Mruczkiewicz, I. Aleiner, F. Arute, K. Arya, J. Atalaya, R. Babbush, J. C. Bardin, R. Barends, A. Bengtsson, A. Bourassa, M. Broughton, B. B. Buckley, D. A. Buell, B. Burkett, N. Bushnell, J. Campero, Z. Chen, B. Chiaro, R. Collins, W. Courtney, S. Demura, A. R. Derk, A. Dunsworth, D. Eppens, C. Erickson, E. Farhi, A. G. Fowler, B. Foxen, C. Gidney, M. Giustina, J. A. Gross, M. P. Harrigan, S. D. Harrington, J. Hilton, A. Ho, S. Hong, T. Huang, W. J. Huggins, S. V. Isakov, M. Jacob-Mitos, E. Jeffrey, C. Jones, D. Kafri, K. Kechedzhi, J. Kelly, S. Kim, P. V. Klimov, A. N. Korotkov, F. Kostritsa, D. Landhuis, P. Laptev, E. Lucero, O. Martin, J. R. McClean, M. McEwen, A. Megrant, K. C. Miao, M. Mohseni, J. Mutus, O. Naaman, M. Neeley, M. Newman, T. E. O’Brien, A. Opremcak, E. Ostby, B. Pató, A. Petukhov, C. Quintana, N. Redd, N. C. Rubin, D. Sank, K. J. Satzinger, V. Shvarts, D. Strain, M. Szalay, M. D. Trevithick, B. Villalonga, T. C. White, Z. Yao, P. Yeh, A. Zalcman, H. Neven, S. Boixo, L. B. Ioffe, P. Roushan, Y. Chen, and V. Smelyanskiy (2021), “Accurately computing the electronic properties of a quantum ring,” *Nature* **594** (7864), 508–512.
- Nielsen, Michael A, and Isaac L. Chuang (2010), *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, Cambridge).
- Nigg, D, M. Müller, E. A. Martinez, P. Schindler, M. Hennrich, T. Monz, M. A. Martin-Delgado, and R. Blatt (2014), “Quantum computations on a topologically encoded qubit,” *Science* **345** (6194), 302–305.
- O’Brien, Thomas E, Stefano Polla, Nicholas C. Rubin, William J. Huggins, Sam McArdle, Sergio Boixo, Jarrod R. McClean, and Ryan Babbush (2021), “Error Mitigation via Verified Phase Estimation,” *PRX Quantum* **2** (2), 020317.
- O’Gorman, Joe, and Earl T. Campbell (2017), “Quantum computation with realistic magic-state factories,” *Physical Review A* **95** (3), 032338.
- Otten, Matthew, and Stephen K. Gray (2019), “Accounting for errors in quantum algorithms via individual error reduction,” *npj Quantum Information* **5** (1), 11.
- Ouyang, Yingkai, David R. White, and Earl T. Campbell (2020), “Compilation by stochastic Hamiltonian sparsification,” *Quantum* **4**, 235.
- Palmieri, Adriano Macarone, Egor Kovlakov, Federico Bianchi, Dmitry Yudin, Stanislav Straupe, Jacob D. Biamonte, and Sergei Kulik (2020), “Experimental neural network enhanced quantum tomography,” *npj Quantum Information* **6** (1), 1–5.
- Panteleev, Pavel, and Gleb Kalachev (2022), “Asymptotically good Quantum and locally testable classical LDPC codes,” in *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2022 (Association for Computing Machinery, New York, NY, USA) pp. 375–388.
- Peruzzo, Alberto, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J. Love, Alán Aspuru-Guzik, and Jeremy L. O’Brien (2014), “A variational eigenvalue solver on a photonic quantum processor,” *Nature*

- Communications* **5**, 4213.
- Piveteau, Christophe, David Sutter, Sergey Bravyi, Jay M. Gambetta, and Kristan Temme (2021), “Error Mitigation for Universal Gates on Encoded Qubits,” *Physical Review Letters* **127** (20), 200505.
- Piveteau, Christophe, David Sutter, and Stefan Woerner (2022), “Quasiprobability decompositions with reduced sampling overhead,” *npj Quantum Information* **8** (1), 1–9.
- Polla, Stefano, Gian-Luca R. Anselmetti, and Thomas E. O’Brien (2022), “Optimizing the information extracted by a single qubit measurement,” arXiv:2207.09479 [quant-ph].
- Postler, Lukas, Sascha Heußen, Ivan Pogorelov, Manuel Risper, Thomas Feldker, Michael Meth, Christian D. Marciniak, Roman Stricker, Martin Ringbauer, Rainer Blatt, Philipp Schindler, Markus Müller, and Thomas Monz (2022), “Demonstration of fault-tolerant universal quantum gate operations,” *Nature* **605** (7911), 675–680.
- Qin, Dayue, Yanzhu Chen, and Ying Li (2021), “Error statistics and scalability of quantum error mitigation formulas,” arXiv:2112.06255 [quant-ph].
- Ross, Neil J, and Peter Selinger (2016), “Optimal ancilla-free Clifford+T approximation of z-rotations,” *Quantum Information & Computation* **16** (11-12), 901–953.
- Rubin, Nicholas C, Ryan Babbush, and Jarrod McClean (2018), “Application of fermionic marginal constraints to hybrid quantum algorithms,” *New Journal of Physics* **20** (5), 053020.
- Russo, A E, K. M. Rudinger, B. C. A. Morrison, and A. D. Baczewski (2021), “Evaluating Energy Differences on a Quantum Computer with Robust Phase Estimation,” *Physical Review Letters* **126** (21), 210501.
- Ryan-Anderson, C, N. C. Brown, M. S. Allman, B. Arkin, G. Asa-Attuah, C. Baldwin, J. Berg, J. G. Bohnet, S. Braxton, N. Burdick, J. P. Campora, A. Chernoguzov, J. Esposito, B. Evans, D. Francois, J. P. Gaebler, T. M. Gatterman, J. Gerber, K. Gilmore, D. Gresh, A. Hall, A. Hankin, J. Hostetter, D. Lucchetti, K. Mayer, J. Myers, B. Neyenhuis, J. Santiago, J. Sedlacek, T. Skripka, A. Slattery, R. P. Stutz, J. Tait, R. Tobey, G. Vittorini, J. Walker, and D. Hayes (2022), “Implementing Fault-tolerant Entangling Gates on the Five-qubit Code and the Color Code,” arXiv:2208.01863 [quant-ph].
- Sagastizabal, R, X. Bonet-Monroig, M. Singh, M. A. Rol, C. C. Bultink, X. Fu, C. H. Price, V. P. Ostroukh, N. Muthusubramanian, A. Bruno, M. Beekman, N. Haider, T. E. O’Brien, and L. DiCarlo (2019), “Experimental error mitigation via symmetry verification in a variational quantum eigensolver,” *Physical Review A* **100** (1), 010302.
- Sanders, Yuval R, Joel J. Wallman, and Barry C. Sanders (2015), “Bounding quantum gate error rate based on reported average fidelity,” *New Journal of Physics* **18** (1), 012002.
- Sauter, Th, W. Neuhauser, R. Blatt, and P. E. Toschek (1986), “Observation of Quantum Jumps,” *Physical Review Letters* **57** (14), 1696–1698.
- Seif, Alireza, Ze-Pei Cian, Sisi Zhou, Senrui Chen, and Liang Jiang (2022), “Shadow Distillation: Quantum Error Mitigation with Classical Shadows for Near-Term Quantum Processors,” arXiv:2203.07309 [quant-ph].
- Setia, Kanav, Sergey Bravyi, Antonio Mezzacapo, and James D. Whitfield (2019), “Superfast encodings for fermionic quantum simulation,” *Physical Review Research* **1** (3), 033033.
- Setia, Kanav, and James D. Whitfield (2018), “Bravyi-Kitaev Superfast simulation of fermions on a quantum computer,” *The Journal of Chemical Physics* **148** (16), 164104.
- Shor, P (1999), “Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer,” *SIAM Review* **41** (2), 303–332.
- Shor, Peter W (1995), “Scheme for reducing decoherence in quantum computer memory,” *Physical Review A* **52** (4), R2493–R2496.
- Shor, PW (1996), “Fault-tolerant quantum computation,” in *Proceedings of 37th Conference on Foundations of Computer Science*, pp. 56–65.
- Sidi, Avram (2003), *Practical Extrapolation Methods: Theory and Applications*, Cambridge Monographs on Applied and Computational Mathematics (Cambridge University Press, Cambridge).
- Smart, Scott E, and David A. Mazziotti (2019), “Quantum-classical hybrid algorithm using an error-mitigating  $\mathbb{N}\mathbb{S}$ -representability condition to compute the Mott metal-insulator transition,” *Physical Review A* **100** (2), 022517.
- Smart, Scott E, and David A. Mazziotti (2020), “Efficient two-electron ansatz for benchmarking quantum chemistry on a quantum computer,” *Physical Review Research* **2** (2), 023048.
- Song, Chao, Jing Cui, H. Wang, J. Hao, H. Feng, and Ying Li (2019), “Quantum computation with universal error mitigation on a superconducting quantum processor,” *Science Advances* **5** (9), eaaw5686.
- Stanisic, Stasja, Jan Lukas Bosse, Filippo Maria Gambetta, Raul A. Santos, Wojciech Mruzekiewicz, Thomas E. O’Brien, Eric Ostby, and Ashley Montanaro (2021), “Observing ground-state properties of the Fermi-Hubbard model using a scalable algorithm on a quantum computer,” arXiv:2112.02025 [cond-mat, physics:quant-ph].
- Steane, A M (1996), “Error Correcting Codes in Quantum Theory,” *Physical Review Letters* **77** (5), 793–797.
- Steuertner, Mark, and Stephanie Wehner (2018), “Fermion-to-qubit mappings with varying resource requirements for quantum simulation,” *New Journal of Physics* **20** (6), 063010.
- Steuertner, Mark, and Stephanie Wehner (2019), “Quantum codes for quantum simulation of fermions on a square lattice of qubits,” *Physical Review A* **99** (2), 022308.
- Strikis, Armands, Dayue Qin, Yanzhu Chen, Simon C. Benjamin, and Ying Li (2021), “Learning-Based Quantum Error Mitigation,” *PRX Quantum* **2** (4), 040330.
- Suchsland, Philippe, Francesco Tacchino, Mark H. Fischer, Titus Neupert, Panagiotis Kl Barkoutsos, and Ivano Tavernelli (2021), “Algorithmic Error Mitigation Scheme for Current Quantum Processors,” *Quantum* **5**, 492.
- Sun, Jinzhao, Xiao Yuan, Takahiro Tsunoda, Vlatko Vedral, Simon C. Benjamin, and Suguru Endo (2021), “Mitigating Realistic Noise in Practical Noisy Intermediate-Scale Quantum Devices,” *Physical Review Applied* **15** (3), 034026.
- Suter, Dieter, and Gonzalo A. Álvarez (2016), “Colloquium: Protecting quantum information against environmental noise,” *Reviews of Modern Physics* **88** (4), 041001.
- Suzuki, Masuo (1990), “Fractal decomposition of exponential operators with applications to many-body theories and Monte Carlo simulations,” *Physics Letters A* **146** (6), 319–323.
- Suzuki, Masuo (1991), “General theory of fractal path integrals with applications to many-body theories and sta-

- tistical physics,” *Journal of Mathematical Physics* **32** (2), 400–407.
- Suzuki, Yasunari, Suguru Endo, Keisuke Fujii, and Yuuki Tokunaga (2022), “Quantum Error Mitigation as a Universal Error Reduction Technique: Applications from the NISQ to the Fault-Tolerant Quantum Computing Eras,” *PRX Quantum* **3** (1), 010345.
- Tacchino, Francesco, Panagiotis Barkoutsos, Chiara Macchiavello, Ivano Tavernelli, Dario Gerace, and Daniele Bajoni (2020), “Quantum implementation of an artificial feed-forward neural network,” *Quantum Science and Technology* **5** (4), 044010.
- Takagi, Ryuji (2021), “Optimal resource cost for error mitigation,” *Physical Review Research* **3** (3), 033178.
- Takagi, Ryuji, Suguru Endo, Shintaro Minagawa, and Mile Gu (2022), “Fundamental limits of quantum error mitigation,” *npj Quantum Information* **8** (1), 1–11.
- Takeda, Kenta, Akito Noiri, Takashi Nakajima, Takashi Kobayashi, and Seigo Tarucha (2022), “Quantum error correction with silicon spin qubits,” *Nature* **608** (7924), 682–686.
- Takeshita, Tyler, Nicholas C. Rubin, Zhang Jiang, Eunseok Lee, Ryan Babbush, and Jarrod R. McClean (2020), “Increasing the Representation Accuracy of Quantum Simulations of Chemistry without Extra Quantum Resources,” *Physical Review X* **10** (1), 011004.
- Tannu, Swamit S, and Moinuddin K. Qureshi (2019), “Mitigating Measurement Errors in Quantum Computers by Exploiting State-Dependent Bias,” in *Proceedings of the 52nd Annual IEEE/ACM International Symposium on Microarchitecture*, MICRO ’52 (Association for Computing Machinery, New York, NY, USA) pp. 279–290.
- Temme, Kristan, Sergey Bravyi, and Jay M. Gambetta (2017), “Error Mitigation for Short-Depth Quantum Circuits,” *Physical Review Letters* **119** (18), 180509.
- Terhal, Barbara M (2015), “Quantum error correction for quantum memories,” *Reviews of Modern Physics* **87** (2), 307–346.
- Tran, Minh C, Yuan Su, Daniel Carney, and Jacob M. Taylor (2021), “Faster Digital Quantum Simulation by Symmetry Protection,” *PRX Quantum* **2** (1), 010323.
- Treinish, Matthew, Jay Gambetta, Paul Nation, Paul Kassebaum, Diego M. Rodríguez, Salvador de la Puente González, Shaohan Hu, Kevin Krsulich, Jake Lishman, Jim Garrison, Laura Zdanski, Jessie Yu, Manoel Marques, Julien Gacon, David McKay, Juan Gomez, Lauren Capelluto, Travis-S-IBM, Ashish Panigrahi, Ilerongil, Rafey Iqbal Rahman, Steve Wood, Luciano Bello, Toshinari Itoko, Christopher J. Wood, Divyanshu Singh, Drew, Eli Arbel, and Glen (2022), “Qiskit,” Zenodo.
- Tsuchimochi, Takashi, Yuto Mori, and Seiichiro L. Ten-no (2020), “Spin-projection for quantum computation: A low-depth approach to strong correlation,” *Physical Review Research* **2** (4), 043142.
- Unruh, W G (1995), “Maintaining coherence in quantum computers,” *Physical Review A* **51** (2), 992–997.
- Urbanek, Miroslav, Daan Camps, Roel Van Beeumen, and Wibe A. de Jong (2020), “Chemistry on Quantum Computers with Virtual Quantum Subspace Expansion,” *Journal of Chemical Theory and Computation* **16** (9), 5425–5431.
- Urbanek, Miroslav, Benjamin Nachman, Vincent R. Pascuzzi, Andre He, Christian W. Bauer, and Wibe A. de Jong (2021), “Mitigating Depolarizing Noise on Quantum Computers with Noise-Estimation Circuits,” *Physical Review Letters* **127** (27), 270502.
- van den Berg, Ewout, Zlatko K. Mineev, and Kristan Temme (2022), “Model-free readout-error mitigation for quantum expectation values,” *Physical Review A* **105** (3), 032620.
- Vazquez, Almudena Carrera, Ralf Hiptmair, and Stefan Woerner (2022), “Enhancing the Quantum Linear Systems Algorithm Using Richardson Extrapolation,” *ACM Transactions on Quantum Computing* **3** (1), 2:1–2:37.
- Vovrosh, Joseph, Kiran E. Khosla, Sean Greenaway, Christopher Self, M. S. Kim, and Johannes Knolle (2021), “Simple mitigation of global depolarizing errors in quantum simulations,” *Physical Review E* **104** (3), 035309.
- Vuillot, Christophe (2018), “Is error detection helpful on IBM 5Q chips?” *Quantum Information and Computation* **18** (11&12), 949–964.
- Wallman, Joel, Chris Granade, Robin Harper, and Steven T. Flammia (2015), “Estimating the coherence of noise,” *New Journal of Physics* **17** (11), 113020.
- Wallman, Joel J, and Joseph Emerson (2016), “Noise tailoring for scalable quantum computation via randomized compiling,” *Physical Review A* **94** (5), 052325.
- Wallraff, A, D. I. Schuster, A. Blais, L. Frunzio, R.-S. Huang, J. Majer, S. Kumar, S. M. Girvin, and R. J. Schoelkopf (2004), “Strong coupling of a single photon to a superconducting qubit using circuit quantum electrodynamics,” *Nature* **431** (7005), 162–167.
- Wang, Samson, Piotr Czarnik, Andrew Arrasmith, M. Cerezo, Lukasz Cincio, and Patrick J. Coles (2021a), “Can Error Mitigation Improve Trainability of Noisy Variational Quantum Algorithms?” arXiv:2109.01051 [quant-ph].
- Wang, Samson, Enrico Fontana, M. Cerezo, Kunal Sharma, Akira Sone, Lukasz Cincio, and Patrick J. Coles (2021b), “Noise-induced barren plateaus in variational quantum algorithms,” *Nature Communications* **12** (1), 6961.
- Wang, Zhen, Yanzhu Chen, Zixuan Song, Dayue Qin, Hekang Li, Qiujiang Guo, H. Wang, Chao Song, and Ying Li (2021c), “Scalable Evaluation of Quantum-Circuit Error Loss Using Clifford Sampling,” *Physical Review Letters* **126** (8), 080501.
- Wecker, Dave, Matthew B. Hastings, and Matthias Troyer (2015), “Progress towards practical quantum variational algorithms,” *Physical Review A* **92** (4), 042303.
- Wu, Yulin, Wan-Su Bao, Sirui Cao, Fusheng Chen, Ming-Cheng Chen, Xiawei Chen, Tung-Hsun Chung, Hui Deng, Yajie Du, Daojin Fan, Ming Gong, Cheng Guo, Chu Guo, Shaojun Guo, Lianchen Han, Linyin Hong, He-Liang Huang, Yong-Heng Huo, Liping Li, Na Li, Shaowei Li, Yuan Li, Futian Liang, Chun Lin, Jin Lin, Haoran Qian, Dan Qiao, Hao Rong, Hong Su, Lihua Sun, Liangyuan Wang, Shiyu Wang, Dachao Wu, Yu Xu, Kai Yan, Weifeng Yang, Yang Yang, Yangsen Ye, Jianghan Yin, Chong Ying, Jiale Yu, Chen Zha, Cha Zhang, Haibin Zhang, Kaili Zhang, Yiming Zhang, Han Zhao, Youwei Zhao, Liang Zhou, Qingling Zhu, Chao-Yang Lu, Cheng-Zhi Peng, Xiaobo Zhu, and Jian-Wei Pan (2021), “Strong Quantum Computational Advantage Using a Superconducting Quantum Processor,” *Physical Review Letters* **127** (18), 180501.
- Xue, Xiao, Maximilian Russ, Nodar Samkharadze, Brennan Undseth, Amir Sammak, Giordano Scappucci, and Lieven M. K. Vandersypen (2022), “Quantum logic with spin qubits crossing the surface code threshold,” *Nature* **601** (7893), 343–347.
- Yamamoto, Kaoru, Suguru Endo, Hideaki Hakoshima, Yuichiro Matsuzaki, and Yuuki Tokunaga (2022), “Error-

- mitigated quantum metrology,” arXiv:2112.01850 [quant-ph].
- Yang, Yongdan, Bing-Nan Lu, and Ying Li (2021), “Accelerated Quantum Monte Carlo with Mitigated Error on Noisy Quantum Computer,” *PRX Quantum* **2** (4), 040361.
- Yen, Tzu-Ching, Robert A. Lang, and Artur F. Izmaylov (2019), “Exact and approximate symmetry projectors for the electronic structure problem on a quantum computer,” *The Journal of Chemical Physics* **151** (16), 164111.
- Yeter-Aydeniz, Kübra, Raphael C. Pooser, and George Siopsis (2020), “Practical quantum computation of chemical and nuclear energy levels using quantum imaginary time evolution and Lanczos algorithms,” *npj Quantum Information* **6** (1), 1–8.
- Yoshioka, Nobuyuki, Hideaki Hakoshima, Yuichiro Matsuzaki, Yuuki Tokunaga, Yasunari Suzuki, and Suguru Endo (2022), “Generalized Quantum Subspace Expansion,” *Physical Review Letters* **129** (2), 020502.
- Zhang, Shuaining, Yao Lu, Kuan Zhang, Wentao Chen, Ying Li, Jing-Ning Zhang, and Kihwan Kim (2020), “Error-mitigated quantum gates exceeding physical fidelities in a trapped-ion system,” *Nature Communications* **11** (1), 587.
- Zhu, D., S. Johri, N. M. Linke, K. A. Landsman, C. Huerta Alderete, N. H. Nguyen, A. Y. Matsuura, T. H. Hsieh, and C. Monroe (2020), “Generation of thermofield double states and critical ground states with a quantum computer,” *Proceedings of the National Academy of Sciences* **117** (41), 25402–25406.

Methods	Probabilistic Error Cancellation	Richardson Extrapolation (Equal-gap <sup>#</sup> )	Symmetry Verification	Virtual Distillation	Echo Verification
Main assumptions	Full knowledge of the noise.	Ability to scale the noise. Small $\lambda$ . <sup>¶</sup>	The ideal state contains inherent symmetry.	The ideal state $\rho_0$ is pure. The noise is stochastic such that $\rho_0$ is the dominant eigenvector of the noisy state $\rho$ . <sup>§</sup>	
Hyper-parameters	Circuit fault rate <i>after</i> mitigation: $\lambda_{\text{em}}$	Number of data points: $M$	The projector of the symmetry subspace: $\Pi$	Degree of purification: $M$	Nil
Qubit overhead	1	1	1 <sup>‡</sup>	$M$	1
Circuit runtime overhead	up to $\sim 2$	up to $\sim M$	1	$\sim 1$ <sup>  </sup>	2
Sampling overhead ( $C_{\text{em}}$ )	$e^{4(\lambda-\lambda_{\text{em}})}$	$(2^M - 1)^2$	Post-selection: $\text{Tr}[\Pi\rho]^{-1}$ Post-processing: $\text{Tr}[\Pi\rho]^{-2}$	$\text{Tr}[\rho^M]^{-2} \gtrsim \frac{e^{2M\lambda}}{[1+(e^\lambda-1)^M]^2}$	$\text{Tr}[\rho^2]^{-1} \gtrsim \frac{e^{2\lambda}}{1+(e^\lambda-1)^2}$
Fidelity boost*	$e^{\lambda-\lambda_{\text{em}}}$	$1 + \mathcal{O}(\lambda)$	$\text{Tr}[\Pi\rho]^{-1}$	$\frac{\text{Tr}[\rho_0\rho]^{M-1}}{\text{Tr}[\rho^M]} \gtrsim \frac{e^\lambda}{1+(e^\lambda-1)^M}$	Same as VD with $M = 2$
Bias	Can reach 0 when $\lambda_{\text{em}} = 0$ .	$\mathcal{O}(\lambda^M)$	Can be upper-bounded using the error-mitigated fidelity using Eq. (51), which in turns is related to the fidelity boost achieved.		

\* Fidelity boost is the factor of increase in the fidelity against the ideal state  $\rho_0$  after applying error mitigation (Sec. IV.B.2).

¶ There are successful experiments operating beyond the small- $\lambda$  assumption. The small- $\lambda$  assumption is not needed for some other extrapolation methods. (Sec. III.A)

# There are other variants of Richardson extrapolation that may provide better scalings for the overheads (Sidi, 2003).

§ If  $\rho_0$  is not the dominant eigenvector of the noisy state, the ultimate fidelity achieved will be limited by coherent mismatch (Sec. III.E).

‡ Assuming only the inherent symmetries of the physical problems are used. Additional qubit overhead might be needed if we want to introduce additional symmetries.

|| Additional circuit components are needed to swap in between the noisy copies and the corresponding additional runtime required will depend on the connectivity of the hardware.

TABLE IV A table summarising the assumptions, costs and performances associated with some of the QEM methods mentioned in this review. Some of the expressions of the sampling overhead and the fidelity boost are derived under the assumption that the occurrence of faults in the circuit follows a Poisson distribution with the circuit fault rate being  $\lambda$  as discussed in Sec. II.C. We use  $\rho$  and  $\rho_0$  to denote the unmitigated noisy state and the ideal noiseless state, respectively. Only a specific instance of error extrapolation is included here. Measurement error mitigation is not included here since it can be viewed as the special case of probabilistic error cancellation focusing on the measurement noise. Quantum subspace expansion and  $N$ -representability are also not included in here since their implementation costs and performance are highly problem-specific.