

Extending quantum probabilistic error cancellation by noise scaling

Andrea Mari,¹ Nathan Shammah,¹ and William J. Zeng^{1,2}

¹*Unitary Fund*

²*Goldman, Sachs & Co, New York, NY*

We propose a general framework for quantum error mitigation that combines and generalizes two techniques: probabilistic error cancellation (PEC) and zero-noise extrapolation (ZNE). Similarly to PEC, the proposed method represents ideal operations as linear combinations of noisy operations that are implementable on hardware. However, instead of assuming a fixed level of hardware noise, we extend the set of implementable operations by noise scaling. By construction, this method encompasses both PEC and ZNE as particular cases and allows us to investigate a larger set of hybrid techniques. For example, gate extrapolation can be used to implement PEC without requiring knowledge of the device’s noise model, e.g., avoiding gate set tomography. Alternatively, probabilistic error *reduction* can be used to estimate expectation values at intermediate *virtual* noise strengths (below the hardware level), obtaining partially mitigated results at a lower sampling cost. Moreover, multiple results obtained with different noise reduction factors can be further post-processed with ZNE to better approximate the zero-noise limit.

CONTENTS

I. Introduction	1
II. Overview of probabilistic error cancellation	2
A. Error cancellation	4
B. Monte Carlo estimation	4
III. Overview of zero-noise extrapolation	4
IV. NEPEC: Noise-extended probabilistic error cancellation	5
V. Extrapolating gates for noise-agnostic PEC	6
A. Reducing the sampling cost of gate extrapolation	8
VI. Probabilistic error reduction and virtual ZNE	8
A. Probabilistic error reduction	8
B. Virtual ZNE	9
C. Example: depolarizing-channel error reduction	9
D. Canonical noise scaling	10
VII. Minimum sampling cost of exact NEPEC representations	11
A. A no-go test	11
B. Example: NEPEC representations with amplitude damping noise	12
VIII. Conclusions	12
Acknowledgements	13
References	13

I. INTRODUCTION

The design of efficient techniques for reducing errors in quantum processors is an important and pressing re-

search problem for near-term quantum computing. Standard quantum error correction codes [1–5] are theoretically known, but can require significant resources (e.g. qubits, gates) that are unavailable on near-term quantum devices [6]. On the one hand, the most direct way to reduce physical errors is to improve the existing hardware, e.g., by realizing more stable qubits and less noisy operations. On the other hand, given the existing noisy hardware, large improvements can be achieved at the “software level” by using several techniques which have been recently called error mitigation methods [7–9] and which are the focus of this work.

One of such methods is zero-noise extrapolation (ZNE) [7, 10, 11], where a quantum observable is measured at different noise levels (by artificially increasing the hardware noise) and extrapolated to the zero-noise limit. Another promising method is probabilistic error cancellation (PEC) [7, 8, 12], where ideal circuits are approximated with a Monte Carlo average over different noisy circuits. Even if apparently different, both the ZNE and PEC methods involve the estimation of an ideal expectation value of interest from a suitable combination of noisy expectation values which are experimentally measurable. As schematically shown in Fig. 1, the main difference lies in the choice of the measured quantities: in ZNE they correspond to a fixed circuit evaluated at different noise levels while in PEC they correspond to a set of different circuits evaluated at the same level of noise (that of the hardware).

The aim of this work is to study a more general framework in which ideal expectation values are estimated by post-processing the results of *different* circuits evaluated at *different* noise levels. We refer to this method as NEPEC (noise-extended probabilistic error cancellation). By construction, NEPEC includes both PEC and ZNE as limit cases. As we will show, intermediate techniques in which both degrees of freedom—the gates of the circuit and the noise level—are simultaneously exploited can be advantageous with respect to different figures of merit such as the simplicity of experimental implementation,

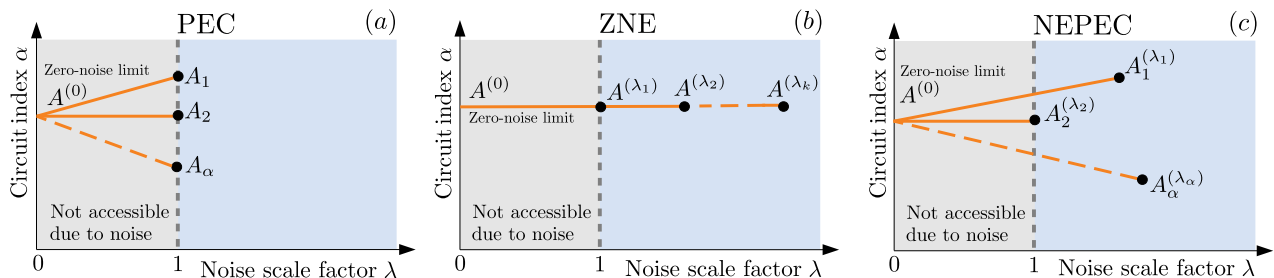


FIG. 1. Pictorial representation of different error mitigation techniques. (a) In probabilistic error cancellation (PEC), an ideal expectation value is estimated from different circuits executed at the same hardware noise $\lambda = 1$. (b) In zero-noise extrapolation (ZNE), an ideal expectation value is estimated from the same circuit executed at different noise levels. (c) In NEPEC (noise-extended probabilistic error cancellation)—the error mitigation framework introduced in this work—an ideal expectation value is estimated from a general combination of different circuits evaluated at different noise levels. In fact, the two-dimensional diagram used in this figure is a simplified representation of the more general linear combination involved in the NEPEC method. Indeed, according to Eq. (25), the two axes of the diagram can be multi-index arrays ($\vec{\lambda}$ and $\vec{\alpha}$), since the noise of each gate of a circuit could be scaled differently.

the robustness to noise characterization errors, and the sampling cost.

Our work is complementary to the recent results of Refs. [13, 14], where it is shown how noise-scaling can enhance the performances of the Clifford data regression (CDR) technique introduced in Ref. [15] and of the virtual distillation method introduced in Refs. [16, 17]. Similarly to NEPEC, Refs. [13, 14] combine different circuits and different noise levels. However, Refs. [13, 14] use a machine learning approach in which an inference model is first *trained* with a dataset of classically-simulable circuits and, as a second step, is applied to the circuit of interest. NEPEC instead requires the construction of (noise-aware or noise-agnostic) quasi-probability representations of ideal gates but, once such representations are determined, it can be directly applied to any circuit of interest without training. In Ref. [18], a modified version of the PEC technique is proposed in which the statistical variance of the estimator can be reduced at the cost of introducing a bias error, similarly to what we study in Sec. VI. In fact, the aim of both techniques is the reduction of the negative volume of quasi-probability representations, imposing lower levels of approximation. The main difference is that in Ref. [18], for each approximation level, a new representation must be numerically determined. In Sec. VI instead, different levels of noise reductions are analytically obtained from a suitable re-scaling of a single quasi-probability representation, exploiting the notion of canonical noise scaling introduced in Sec. VID. While the analytical noise scaling approach is simpler and intrinsically robust to numerical instabilities, the brute-force numerical approach of Ref. [18] could achieve a lower sampling variance. Finally, it is important to remark that the idea of combining probabilistic error reduction with ZNE was already proposed also in Ref. [19]. In Ref. [19], the derivation is focused on Pauli channels and up to first order in the error probabilities. In Sec. VI we generalize this approach beyond

the weak-noise limit and for arbitrary noise models. Differences and similarities between the NEPEC framework introduced in this work and other existing techniques are summarized in Tables I and II.

Before introducing NEPEC, we present an overview of PEC and ZNE in the next two sections, respectively. This sets the notation out and clarifies the relationship with the existing literature. The reader who is already familiar with PEC and ZNE, could directly jump to Sec. IV where NEPEC is defined. In Sec. V the concept of noise-agnostic PEC is introduced, while in Sec. VI we study two techniques that involve virtual noise scaling: probabilistic error reduction (PER) and virtual ZNE. In Sec. VII we consider the minimal sampling cost of exact NEPEC representations, finding a no-go test for a subset of cases, and providing an example in which noise scaling is beneficial. A summary of these techniques and their key features can be found in Table I. Finally, in Sec. VIII we provide concluding remarks.

II. OVERVIEW OF PROBABILISTIC ERROR CANCELLATION

We first review PEC [7, 8, 20]. We consider a typical gate-based quantum computing paradigm in which the expectation value of an observable $A = A^\dagger$ is evaluated after applying a unitary circuit \mathcal{U} to n qubits initially prepared in the product state $\rho_0 = |0\rangle\langle 0|^{\otimes n}$:

$$\langle A \rangle_{\text{ideal}} = \text{tr}[A\mathcal{U}(\rho_0)], \quad (1)$$

where the calligraphic symbol \mathcal{U} stands for the super-operator which acts on a density matrix as $\mathcal{U}(\rho) = U\rho U^\dagger$, where U is a unitary matrix.

The circuit is assumed to be composed of a polynomial number t of local unitary gates:

$$\mathcal{U} = \mathcal{G}_t \circ \dots \circ \mathcal{G}_2 \circ \mathcal{G}_1, \quad (2)$$

NEPEC technique	Bias	Statistical variance	Noise scaling method	Gate tomography	Reference
Probabilistic error cancellation (PEC)	Zero	Typically large	Not required	Required	Sec. II Temme <i>et al.</i> [7] Endo <i>et al.</i> [8] Zhang <i>et al.</i> [12]
Zero-noise extrapolation (ZNE)	Potentially large	Typically low	Real	Not required	Sec. III Li <i>et al.</i> [10] Temme <i>et al.</i> [7] Kandala <i>et al.</i> [11]
Noise-agnostic PEC	Intermediate	Typically large	Real	Not required	Sec. V
Probabilistic error reduction (PER)	Intermediate	< PEC variance	Virtual	Required	Sec. VIA (see also Cai [19])
Virtual ZNE	Intermediate	Intermediate	Virtual	Required	Sec. VIB (see also Cai [19])
Exact NEPEC representations	Zero	\leq PEC variance	Real	Required	Sec. VII

TABLE I. The table provides a qualitative summary with the key features of the different quantum error mitigation techniques reviewed (ZNE and PEC) and introduced in this work. All techniques can be seen as particular cases of the general error mitigation framework introduced in Sec. IV, *i.e.*, noise-extended probabilistic error cancellation (NEPEC). The entries of the table aim to guide the reader about typical aspects of each technique but are intentionally not rigorous and not quantitative. For example, high-order ZNE can have a large variance, while PEC applied to a short-depth circuit can have a small variance.

Technique	Training	Auxiliary qubits	Quasi-prob. representations	Noise scaling	Gate tomography	Reference
NEPEC	Not required	Not required	Required	Real or virtual	Optional	This work
vnCDR	Required	Not required	Not required	Real	Not required	Lowe <i>et al.</i> [13]
UNITED	Required	Required	Not required	Real	Not required	Bultrini <i>et al.</i> [14]
Bias-variance tradeoff	Not required	Not required	Required	Not required	Required	Piveteau <i>et al.</i> [18]

TABLE II. Qualitative table comparing the main requirements of the NEPEC framework proposed in this work to the requirements of other similar techniques existing in the literature (excluding PEC and ZNE). vnCDR stands for variable-noise Clifford data regression [13]. UNITED stands for UNified Technique for Error mitigation with Data [14]. A more detailed comparison between the different implementations of NEPEC, including the particular limit cases of PEC and ZNE, is given in Table I.

where each gate \mathcal{G}_i typically acts non-trivially only on 1 or 2 qubits. Now assume that, because of the hardware noise, we are actually able to apply only a set \mathcal{I} of noisy implementable operations. An implementable operation $\mathcal{O} \in \mathcal{I}$ corresponds to a local quantum channel which is usually non-unitary and so the ideal gates of the circuit cannot be exactly implemented in a direct way. For simplicity, in this work we ignore state-preparation and measurement (SPAM) errors but, in principle, they can be taken into account in a similar fashion [7, 8].

The main idea of PEC is to represent each ideal gate \mathcal{G}_i of the circuit as a linear combination of noisy implementable operations $\{\mathcal{O}_\alpha\}$:

$$\mathcal{G}_i = \sum_{\alpha} \eta_{i,\alpha} \mathcal{O}_{i,\alpha}, \quad \mathcal{O}_{i,\alpha} \in \mathcal{I}, \quad \eta_{i,\alpha} \in \mathbb{R}. \quad (3)$$

In principle, the index i of $\mathcal{O}_{i,\alpha}$ in Eq. (3) could be dropped. However, in this work, we keep it to explicitly define a gate-dependent basis of implementable op-

erations.

Note that this representation exists provided that: (i) the set of implementable operations is large enough to approximately or exactly represent \mathcal{G}_i (a sufficient condition is that \mathcal{I} forms a basis for the full space of quantum channels) and (ii) the coefficients $\eta_{i,\alpha}$ are allowed to take negative values. Since each operation is trace-preserving we always have:

$$\sum_{\alpha} \eta_{i,\alpha} = 1, \quad \gamma_i = \sum_{\alpha} |\eta_{i,\alpha}| \geq 1. \quad (4)$$

The normalization condition implies that one can consider $\{\eta_{i,\alpha}\}$ as a *quasi-probability* distribution [7, 21] with respect to the index α , while its one-norm γ_i is related to the amount of *negativity*, *i.e.*, the total volume of the negative coefficients. The minimum value of γ_i is 1 and is achieved only if all the coefficients are positive, corresponding to the special case in which $\{\eta_{i,\alpha}\}$ is a valid probability distribution.

As we are going to see in the next subsections, the parameter γ_i is also related to the PEC sampling cost associated to the gate \mathcal{G}_i [7, 20, 21]. Therefore, among all the possible representations given in Eq. (3), one is typically interested in those minimizing γ_i :

$$\gamma_i^{\text{opt}} = \min_{\substack{\{\eta_{i,\alpha}\} \\ \{\mathcal{O}_{i,\alpha}\}}} \left[\sum_{\alpha} |\eta_{i,\alpha}| \right] \text{ s.t. Eq. (3) holds.} \quad (5)$$

A detailed analysis of the optimal PEC sampling cost defined in Eq. (5) can be found in Ref. [20].

A. Error cancellation

Now, if we first replace Eq. (3) into Eq. (2) and then substitute the result into Eq. (1), we obtain an expression for the desired ideal expectation value as a linear combination of many noisy expectation values:

$$\langle A \rangle_{\text{ideal}} = \sum_{\bar{\alpha}} \eta_{\bar{\alpha}} \langle A_{\bar{\alpha}} \rangle_{\text{noisy}}, \quad (6)$$

where:

$$\eta_{\bar{\alpha}} := \eta_{t,\alpha_t} \dots \eta_{2,\alpha_2} \eta_{1,\alpha_1}, \quad (7)$$

$$\langle A_{\bar{\alpha}} \rangle_{\text{noisy}} := \text{tr}[A \Phi_{\bar{\alpha}}(\rho_0)], \quad (8)$$

$$\Phi_{\bar{\alpha}} := \mathcal{O}_{t,\alpha_t} \circ \dots \circ \mathcal{O}_{2,\alpha_2} \circ \mathcal{O}_{1,\alpha_1}. \quad (9)$$

By linearity of the sum, the coefficients $\eta_{\bar{\alpha}}$ form again a quasi-probability distribution for the global circuit whose one-norm γ is the product of those of the gates:

$$\sum_{\bar{\alpha}} \eta_{\bar{\alpha}} = 1, \quad \gamma = \sum_{\bar{\alpha}} |\eta_{\bar{\alpha}}| = \prod_i \gamma_i. \quad (10)$$

In principle, by running all the noisy circuits $\Phi_{\bar{\alpha}}$ and evaluating all the corresponding expectation values $\langle A_{\bar{\alpha}} \rangle$, one can exactly compute the ideal result $\langle A \rangle_{\text{ideal}}$. However, the number of terms in Eq. (6) grows exponentially with respect to the number of gates t and, in most practical situations, this approach is unfeasible. A possible solution to avoid this issue is to replace the sum in Eq. (6) with a Monte Carlo approximation [7, 8, 20, 21].

B. Monte Carlo estimation

Let us define the probability distribution $p_i(\alpha) = |\eta_{i,\alpha}|/\gamma_i$ associated to the quasi-probability representation of the gate \mathcal{G}_i . It is easy to check that we can always re-write Eq. (3) as:

$$\mathcal{G}_i = \sum_{\alpha} p_i(\alpha) \gamma_i \text{sign}(\eta_{i,\alpha}) \mathcal{O}_{i,\alpha}. \quad (11)$$

If we sample an index α at random from the distribution $p_i(\alpha)$, we obtain a random variable $\hat{\alpha}$ such that

$$\hat{\mathcal{G}}_i = \gamma_i \text{sign}(\eta_{i,\hat{\alpha}}) \mathcal{O}_{i,\hat{\alpha}}, \quad (12)$$

is an unbiased estimator for the ideal gate \mathcal{G}_i , i.e., $\mathcal{G}_i = \mathbb{E}\{\hat{\mathcal{G}}_i\}$ where the average $\mathbb{E}(\cdot)$ is with respect to the sampling distribution $p_i(\alpha)$.

If, for each gate \mathcal{G}_i of the circuit described in Eq. (2), we independently sample an index $\hat{\alpha}$ corresponding to the noisy operation $\mathcal{O}_{i,\hat{\alpha}}$, we can define an unbiased estimator for the full circuit as:

$$\hat{\mathcal{U}} = \hat{\mathcal{G}}_t \circ \dots \circ \hat{\mathcal{G}}_2 \circ \hat{\mathcal{G}}_1 = \gamma \sigma_{\bar{\alpha}} \Phi_{\bar{\alpha}}, \quad (13)$$

where $\Phi_{\bar{\alpha}}$ is the sampled noisy circuit, γ is the constant introduced in Eq. (10) and $\sigma_{\bar{\alpha}} = \prod_i \text{sign}(\eta_{i,\hat{\alpha}}) = \text{sign}(\eta_{\bar{\alpha}})$. In other words, by sampling individual gates one-by-one, we are effectively sampling from the global quasi-probability representation of the full circuit such that $\hat{\mathcal{U}} = \mathbb{E}\{\hat{\mathcal{U}}\}$. This directly implies that by measuring the observable A on the sampled circuits, one can estimate the ideal expectation value with a Monte Carlo sampling average over noisy expectation values:

$$\langle A \rangle_{\text{ideal}} = \mathbb{E} \left\{ \text{tr}[A \hat{\mathcal{U}}(\rho_0)] \right\} = \mathbb{E} \left\{ \gamma \sigma_{\bar{\alpha}} \langle A_{\bar{\alpha}} \rangle_{\text{noisy}} \right\}. \quad (14)$$

Differently from Eq. (6) which is exact but contains an exponential number of terms, the right-hand-side of Eq. (14) can be approximated by averaging over a finite number of samples. How many samples are necessary to approximate $\langle A \rangle_{\text{ideal}}$ up to a precision δ ? From the standard (classical) theory of Monte Carlo sampling it can be shown that [20, 21]:

$$\# \text{ of samples} \propto \frac{\gamma^2}{\delta^2}. \quad (15)$$

This gives a clear operational meaning to the parameter γ : the larger the negativity of the quasi-distribution the higher the PEC sampling cost. Since $\gamma = \prod_i \gamma_i$, its value scales exponentially with respect to the number of gates. Nonetheless, PEC can still be very advantageous with the medium-size circuits which can run on near-term quantum computers [12].

III. OVERVIEW OF ZERO-NOISE EXTRAPOLATION

Zero-noise extrapolation [7, 10, 11] is an error mitigation method that relies on the assumption that the noise of the physical hardware can be artificially increased beyond the base level such that, after measuring an expectation value for different strengths of the noise, it is possible to extrapolate the zero-noise limit.

More precisely, it is assumed that the strength of the noise can be quantified by some parameter Γ (e.g. the decoherence rate of some noise channel) and that it is possible to scale the noise by a dimensionless parameter $\lambda \geq 1$ such that the effective noise channel acting on the system has a larger strength $\Gamma' = \lambda\Gamma$. Importantly, differently from PEC, in this case it is not necessary to know

the details of the noise model. In fact, there are many situations in which the noise level Γ is unknown and/or the associated noise channel is hard to characterize, nonetheless, it is still possible to correctly scale Γ by some given factor λ . For example, the pulse-stretching technique introduced in [7] and several gate-level methods proposed in [22–24] allow to effectively scale the hardware noise, without actually “knowing” it (within the limits and assumptions of these methods).

We now try to express ZNE with a formalism that is as similar as possible to the one previously used for describing PEC. We consider an ideal expectation value $\langle A \rangle_{\text{ideal}}$ that we would like to infer from a set of noisy expectation values $\langle A \rangle_{\text{noisy}}^{(\lambda)}$ evaluated at different values of the noise scale factor $\lambda \geq 1$:

$$\langle A \rangle_{\text{noisy}}^{(\lambda)} := \text{tr}[A\mathcal{U}^{(\lambda)}(\rho_0)], \quad (16)$$

$$\mathcal{U}^{(\lambda)} := \mathcal{G}_t^{(\lambda)} \circ \dots \circ \mathcal{G}_2^{(\lambda)} \circ \mathcal{G}_1^{(\lambda)}, \quad (17)$$

where $\mathcal{G}_i^{(\lambda)}$ represents the noise-scaled implementation of the ideal unitary gate $\mathcal{G}_i = \mathcal{G}_i^{(\lambda=0)}$. If a polynomial fit [22] is used to evaluate the zero-noise limit (Richardson extrapolation [7] is a special case), the ideal result can always be expressed as linear combination of the measured noisy expectation values:

$$\langle A \rangle_{\text{ideal}} = \sum_{\lambda \in S} \eta_\lambda \langle A \rangle_{\text{noisy}}^{(\lambda)} + \Delta, \quad (18)$$

where $S = \{\lambda_1, \lambda_2, \dots, \lambda_m\}$ is the chosen set of m noise scale factors and where η_λ are real coefficients which are completely determined by S and by the degree of the polynomial but are independent of the measured results (because a polynomial fit is a linear regression problem) [25]. Differently from the exact linear combination introduced in Eq. (6), in ZNE the result is subject to a some bias error Δ , which depends on the amount of noise Γ and on the quality of the extrapolation model.

It is easy to check that, also in this case, $\{\eta_\lambda \mid \lambda \in S\}$ can be interpreted as a quasi-probability distribution:

$$\sum_{\lambda \in S} \eta_\lambda = 1, \quad \gamma = \sum_{\lambda \in S} |\eta_\lambda| \geq 1. \quad (19)$$

Similarly to PEC, the parameter γ quantifies how much the statistical uncertainty on the expectation values $\langle A \rangle_{\text{noisy}}^{(\lambda)}$ gets amplified by the linear combination in Eq. (18) and so how large is the mitigation overhead. For example, in the case of Richardson extrapolation, γ scales exponentially with respect to the size of S [22, 24] which in practice implies that only a few noise scale factors must be used to avoid numerical instabilities. In general, the number of terms in the linear combination is small and one can directly measure all the noisy expectation values. This means that, in the case of ZNE, it is not necessary to use any probabilistic Monte Carlo sampling.

To summarize, in both the PEC and ZNE methods, an ideal expectation value is represented as an average over noisy expectation values with respect to a quasi probability distribution. The only difference lies in the domain of the distribution as shown in Fig. 1. Can we consider the two domains of PEC and ZNE as particular cases of a more general configuration space? This is the main idea proposed in this work and we formalize it in the next section.

IV. NEPEC: NOISE-EXTENDED PROBABILISTIC ERROR CANCELLATION

In this section we introduce NEPEC (noise-extended probabilistic error cancellation), a generalization of PEC in which the set of implementable operations is extended by noise scaling (see Fig. 1).

We begin with the set of implementable operations \mathcal{I} which appears in Eq. (3) and which is at the basis of PEC. Let us also assume that for each operation $\mathcal{O} \in \mathcal{I}$, one can scale the noise by a factor $\lambda \geq 1$, obtaining a noise-scaled operation $\mathcal{O}^{(\lambda)}$. In practice this implies that, thanks to noise scaling, we have at our disposal an extended set of implementable operations:

$$\tilde{\mathcal{I}} = \{\mathcal{O}^{(\lambda)} \mid \mathcal{O} \in \mathcal{I}, \lambda \geq 1\}. \quad (20)$$

By construction, $\tilde{\mathcal{I}}$ includes \mathcal{I} as a subset. A noise-scaled implementable operation $\mathcal{O}^{(\lambda)}$ can be expressed as:

$$\mathcal{O}^{(\lambda)} := \mathcal{G}^{(\lambda)} = \mathcal{E}^{(\lambda)} \circ \mathcal{G}, \quad (21)$$

where \mathcal{G} is some ideal unitary operation and $\mathcal{E}^{(\lambda)}$ is some noise channel such that $\mathcal{E}^{(1)}$ corresponds to the hardware base noise and $\mathcal{E}^{(0)} = \text{Id}$. For example, if a gate \mathcal{G} acting on k qubits is affected by a depolarizing channel \mathcal{D}_p with error probability p [20], the associated noise-scaled operation can be defined as in Eq. (21) with:

$$\mathcal{E}^{(\lambda)} = \mathcal{D}_{\lambda p} = (1 - \lambda p)\text{Id} + \lambda p \sum_{\mathcal{P} \in P_k} \frac{\mathcal{P}}{4^k - 1}, \quad \lambda \in [1, \lambda_{\text{max}}], \quad (22)$$

where P_k is the set of all Pauli strings of length k with the exclusion of the identity string $\text{Id}_1 \otimes \text{Id}_2 \otimes \dots \otimes \text{Id}_k$ and λ_{max} is the maximum scale factor. If we impose λp to be a valid probability we obtain $\lambda_{\text{max}} = p^{-1}$, while if we more realistically require that the maximum scaling is achieved when the output state is completely mixed we get $\lambda_{\text{max}} = (1 - 4^{-k})p^{-1} < p^{-1}$.

In practice, noise scaling can be experimentally achieved by acting on the physical control pulses [7] or by acting on the circuit at a gate-level [22–24]. For example, we anticipate that in the simulation shown in Fig. 2 (presented in the next section), we will apply *unitary folding* [22] to approximately scale the noise with odd integers values of λ . According to this method, instead of using

Eq. (21), a noise-scaled operation associated to a gate \mathcal{G} is defined as:

$$\begin{aligned} \mathcal{O}^{(\lambda)} &:= \mathcal{G}^{(\lambda)} = \mathcal{G}^{(1)} \circ \left[\mathcal{G}^{\dagger(1)} \circ \mathcal{G}^{(1)} \right]^{(\lambda-1)/2}, \\ \lambda &= 1, 3, \dots, \end{aligned} \quad (23)$$

where $\mathcal{G}^{(1)}$, corresponds to the gate \mathcal{G} executed at the hardware base noise (i.e., at $\lambda = 1$). More details on the practical advantages and limitations of unitary folding can be found in Refs. [22, 26]. Even if the final results may depend on the choice of the noise scaling method, the abstract formulation of the NEPEC technique is actually independent from this choice.

If we expand an ideal gate \mathcal{G}_i of a given circuit in the new basis $\tilde{\mathcal{I}}$, we obtain:

$$\mathcal{G}_i = \sum_{\alpha, \lambda} \eta_{i, \alpha, \lambda} \mathcal{O}_{i, \alpha}^{(\lambda)}, \quad \mathcal{O}_{i, \alpha}^{(\lambda)} \in \tilde{\mathcal{I}}, \quad \eta_{i, \alpha, \lambda} \in \mathbb{R}, \quad (24)$$

which is the natural generalization of Eq. (3) to a noise-extended basis. The associated mitigation cost is given by the one-norm of the quasi-probability distribution $\{\eta_{i, \alpha, \lambda}\}$, i.e., $\gamma_i = \sum_{\alpha, \lambda} |\eta_{i, \alpha, \lambda}| \geq 1$.

Similarly to PEC and ZNE, our goal is to approximate an ideal expectation value as a linear combination of noisy expectation values associated to implementable circuits:

$$\langle A \rangle_{\text{ideal}} = \sum_{\vec{\alpha}, \vec{\lambda}} \eta_{\vec{\alpha}, \vec{\lambda}} \langle A_{\vec{\alpha}} \rangle_{\text{noisy}}^{(\vec{\lambda})} \quad (25)$$

$$\langle A_{\vec{\alpha}} \rangle_{\text{noisy}}^{(\vec{\lambda})} := \text{tr}[A \Phi_{\vec{\alpha}}^{(\vec{\lambda})}(\rho_0)], \quad (26)$$

$$\Phi_{\vec{\alpha}}^{(\vec{\lambda})} := \mathcal{O}_{t, \alpha_t}^{(\lambda_t)} \circ \dots \circ \mathcal{O}_{2, \alpha_2}^{(\lambda_2)} \circ \mathcal{O}_{1, \alpha_1}^{(\lambda_1)}, \quad (27)$$

where $\mathcal{O}_{i, \alpha}^{(\lambda)} \in \tilde{\mathcal{I}}$ is i -th noisy gate of the circuit, α enumerates different implementable operations and λ is the noise scale factor.

Since in NEPEC we have two degrees of freedom—the choice of gates and their noise scale factors—the set of coefficients $\{\eta_{\vec{\alpha}, \vec{\lambda}}\}$ should be considered as a quasi-distribution with respect to the pair of indices $(\vec{\alpha}, \vec{\lambda})$:

$$\sum_{\vec{\alpha}, \vec{\lambda}} \eta_{\vec{\alpha}, \vec{\lambda}} = 1, \quad \gamma = \sum_{\vec{\alpha}, \vec{\lambda}} |\eta_{\vec{\alpha}, \vec{\lambda}}|. \quad (28)$$

In most practical situations, it is not possible to measure all the terms in Eq. (25) because their number is too large. In these cases, one can sample a random pair of indices $(\vec{\alpha}, \vec{\lambda})$ from the probability distribution $p_{\vec{\alpha}, \vec{\lambda}} = |\eta_{\vec{\alpha}, \vec{\lambda}}|/\gamma$, such that the ideal expectation value can be estimated probabilistically as an average over many samples:

$$\langle A \rangle_{\text{ideal}} = \mathbb{E} \left\{ \gamma \sigma_{\vec{\alpha}, \vec{\lambda}} \langle A_{\vec{\alpha}} \rangle_{\text{noisy}}^{(\vec{\lambda})} \right\}, \quad (29)$$

where $\sigma_{\vec{\alpha}, \vec{\lambda}} = \text{sign}(\eta_{\vec{\alpha}, \vec{\lambda}})$. This is the NEPEC version of Eq. (14) and, also in this case, the sampling cost scales as γ^2 .

In Eq. (25), the sum is over the two variables $\vec{\alpha}$ and $\vec{\lambda}$. If we fix $\lambda_j = 1$, we re-obtain the PEC linear combination given in Eq. (6). If we instead fix the choice of the noisy gates to match that of the ideal circuit and impose a uniform scale factor $\lambda_j = \lambda$, the sum over $\vec{\alpha}$ disappears and we recover the ZNE linear combination introduced in Eq. (18). A graphical representation showing how PEC and ZNE are particular cases of NEPEC is given in Fig. 1.

Is there any advantage in using the more general NEPEC framework compared to the particular limit cases represented by PEC and ZNE? The answer to this question depends on the considered figure of merit and so it is better to address more specific questions:

1. Can NEPEC be used without the full tomographic knowledge of the gates which is instead necessary in PEC? We address this question in Sec. V.
2. Can NEPEC be used to effectively implement virtual noise scaling methods for ZNE? We address this question in Sec. VI.
3. Can NEPEC allow for a smaller sampling cost compared to PEC? We address this question in Sec. VII.

In the next sections we show, with theoretical arguments and with explicit numerical examples, that all the previous questions can have a positive answer, even though this depends on the specific circuit, gate set, and device under consideration. The techniques introduced hereafter are qualitatively compared to ZNE and PEC in Table I.

V. EXTRAPOLATING GATES FOR NOISE-AGNOSTIC PEC

In the previous section we presented the general formulation of NEPEC, introducing the quasi-probability representations for expectation values [Eqs. (25) and (29)], and for individual gates (Eq. (24)). However, among all the possible representations which could be used, what are the good ones for practical applications? Are there specific NEPEC representations that have some good physical motivation or that are practically easier to implement?

In this section we consider a particular technique where the quasi-probability representation for individual gates is inspired by ZNE while the full circuit is sampled according to the standard PEC algorithm. The main advantage of this approach is the possibility of obtaining an approximate probabilistic error cancellation technique which is *noise-agnostic*, i.e., which does not require the full characterization of the noise model or the full tomography of the noisy gate set.

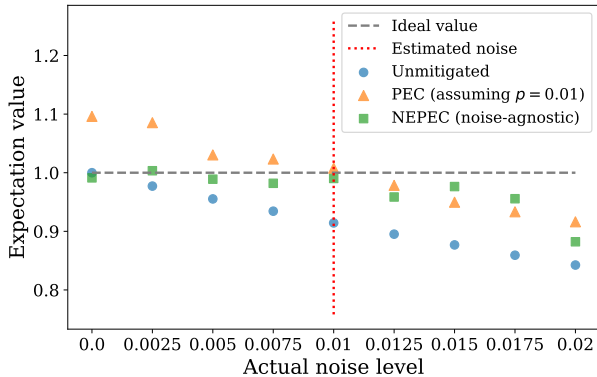


FIG. 2. The expectation value of the observable $A = |0\rangle\langle 0|$ is estimated with different techniques for a single-qubit randomized benchmarking circuit of depth 14 such that $\langle A \rangle_{\text{ideal}} = 1$. The PEC results are based on quasi-probability representations built assuming a fixed estimated noise level of $p = 0.01$. When this assumption matches the actual noise level of the hardware, PEC achieves an optimal cancellation of the noise. On the other hand, when the actual noise of the system is different from the assumed value (noise characterization error), PEC results are more biased than NEPEC results. The NEPEC results are obtained using the noise-agnostic representations defined according to Eq. (30) with scale factors $S = \{\lambda_1, \lambda_2\} = \{1, 51\}$ and coefficients given by Eq. (31). Unitary folding is used for digitally scaling the noise as described in Eq. (23) without assuming a specific noise model or noise level. All the noisy expectation values are directly evaluated from simulated density matrices and therefore shot noise is absent in this figure (corresponding to the limit of infinite shots). The random fluctuations of the PEC and NEPEC points are due to the finite number (5000) of Monte Carlo samples. For the numerical simulation of this example, we used the error mitigation software package Mitiq [26].

The idea is to represent each ideal gate \mathcal{G}_i of a circuit as a linear combination of the same noisy gate executed at different noise scale factors $\lambda \geq 1$. In practice we drop the α index which appears in the NEPEC representation of a generic gate given in Eq. (24) and keep only the sum over a finite set of noise scale factors $S_i = \{\lambda_1, \lambda_2, \dots\}$:

$$\mathcal{G}_i = \sum_{\lambda \in S_i} \eta_{i,\lambda} \mathcal{G}_i^{(\lambda)} + \Delta_i, \quad \eta_{i,\lambda} \in \mathbb{R}, \quad (30)$$

where we also add a small bias error—the super-operator Δ_i in (30)—since the restricted noisy basis may not allow for an exact representation.

Note that Eq. (30) could be considered as a kind of “extrapolation” of the noisy gate $\mathcal{G}_i^{(\lambda)}$ to the zero-noise limit $\mathcal{G}_i^{(\lambda=0)}$. Differently from standard ZNE, where the extrapolation is applied to some scalar expectation value, here we are instead extrapolating the super-operator of a single gate with the aim of obtaining a good quasi-probability representation.

Crucially, the scalar coefficients $\eta_{i,\lambda}$ can be determined in exactly the same way as in standard ZNE. More precisely, for any polynomial model (including Richardson extrapolation), the coefficients $\eta_{i,\lambda}$ depend only on the set of noise scale factors S_i . This means that, given the extrapolation order and the set S_i , we can compute the coefficients $\eta_{i,\lambda}$ with the same standard methods which are often used for ZNE. For example, in the case of Richardson extrapolation, the explicit formula for the coefficients is [22]:

$$\eta_{i,\lambda} = \prod_{\substack{\lambda' \in S_i \\ \lambda' \neq \lambda}} \frac{\lambda'}{\lambda' - \lambda}, \quad \lambda \in S_i. \quad (31)$$

Alternatively, only in the case in which one has the full tomographic knowledge of the noise scaled gates, one could directly optimize the coefficients $\eta_{i,\lambda}$ in Eq. (30) to better represent the ideal gate.

Once we have the quasi-probability representation described in Eq. (30) for each gate of an ideal circuit, we can estimate expectation values via the usual probabilistic error cancellation approach. In practice one has to stochastically modulate the effective noise along the circuit by sampling a noise scale factor for each gate according to the probability distribution $p_i(\lambda) = |\eta_{i,\lambda}|/\gamma_i$. Eventually, to estimate the ideal result one should combine all the measured samples according to Eq. (29).

Compared to the standard decomposition used in PEC and introduced in Eq. (3), in order to use the specific NEPEC representation given in Eq. (30), one needs some direct or indirect way of scaling the noise of individual gates. On the other hand, the main advantage of this method is that it does not require the full knowledge of the noise model which would instead be necessary to evaluate Eq. (3).

A simple example demonstrating noise-agnostic error mitigation with NEPEC is shown in Fig. 2 for a single-qubit circuit subject to different levels of depolarizing noise. In Fig. 2, one can also observe that standard PEC (with fixed gate representations) is quite sensible to noise characterization errors (*i.e.* the mismatch between the assumed and the actual noise-models). On the contrary, exploiting the noise-independent representations defined in Eq. (30), NEPEC is by construction more robust with respect to the noise level and to the noise type.

Note: Instead of considering the unitary operator \mathcal{G}_i in Eq. (30) as the i -th gate of the circuit, one may also associate it to the i -th layer or to the i -th sub-circuit. The proposed procedure would be exactly the same, with the only constraint of using a uniform noise scale factor for each i -th part of the circuit. Depending on the noise scaling method, acting at the level of layers may be more practical with respect to scaling the noise of individual gates. For example, it is probably simpler to apply the pulse-stretching [7, 11] technique layer-wise

instead of gate-wise.

A. Reducing the sampling cost of gate extrapolation

The sampling cost parameters γ_i of the quasi-probability representation defined in Eq. (30) can be quite large compared to the PEC representation of Eq. (3) and this fact can be a practical limitation for large circuits. We propose two solutions to reduce the sampling cost in the case of large and highly noisy circuits:

- 1 Considering \mathcal{G}_i as the unitary matrix which represents a large block of the circuit instead of a single gate or of a thin layer of gates.
- 2 Applying large noise scaling to increase the distance between the different elements of S_i (the set of noise scale factors).

An example of the first option is to divide a long circuit into a few (say 2 or 3) parts of approximately equal size. Each part could be represented as in Eq. (30). In this setting, the full circuit is expressible as a linear combination of a limited number of terms which can be easily measured.

The second option instead is motivated by the fact that, the more the values of λ are different, the smaller the one-norm of the extrapolation coefficients. For example, for a linear extrapolation with scale factors $S_i = \{1, \lambda_2\}$, the two extrapolation coefficients would be $\eta_{i,1} = \lambda_2/(\lambda_2 - 1)$ and $\eta_{i,\lambda_2} = 1/(1 - \lambda_2)$, corresponding to a mitigation cost of $\gamma_i = (1 + \lambda_2)/(\lambda_2 - 1)$ which tends to one for a large λ_2 . In practice, however, there is usually a trade-off for the optimal amount of noise scaling: large noise scaling values are convenient for reducing the sampling variance, but too large noise scaling may give a bad extrapolation bias (large Δ in Eq. (30)). Moreover, depending on how noise scaling is defined, there can be a maximum value of lambda λ above which it is impossible to scale the noise. For example, for the noise-scaled depolarizing channel $\mathcal{D}_{\lambda p}$ acting on k qubits defined in Eq. (22), a physically motivated maximum value is $\lambda_{\max} = (1 - 4^{-k})p^{-1}$, where p is the base error rate of the hardware. In this case, linear extrapolation of a single-qubit gate would give a mitigation cost of:

$$\gamma_i = \frac{1 + \lambda_{\max}}{\lambda_{\max} - 1} = \frac{1 + \epsilon}{1 - \epsilon}, \quad \epsilon = (4/3)p. \quad (32)$$

This is larger than the PEC optimal cost $\gamma_{\text{opt}} = (1 + \epsilon/2)(1 - \epsilon)$ [20], consistent with the no-go result of Sec. VII. Interestingly, if we instead set $\lambda_{\max} = p^{-1}$, we obtain $\gamma_i = (1 + p)/(1 - p)$ which is slightly smaller than the optimal PEC cost of Ref. [20]. However one should take into account that, for $\lambda > (1 - 4^{-k})p^{-1}$, the corresponding noise-scaled depolarizing channel may be

impossible to implement on hardware even if mathematically well-defined. For this reason, Eq. (32) is a more realistic and prudent estimate of the actual error mitigation cost.

VI. PROBABILISTIC ERROR REDUCTION AND VIRTUAL ZNE

In this section we show how one can use a probabilistic sampling approach for indirectly implementing a “virtual” noise scaling process instead of aiming for a complete cancellation of errors. Differently from common noise scaling methods like pulse-stretching or unitary folding, virtual noise scaling can be used to effectively reduce the noise below the hardware level instead of scaling it up. Similarly to the general NEPEC framework discussed in Sec. IV, also in this section we extend the space of implementable operations to arbitrary (virtual) noise levels and we show how this can be useful for reducing the statistical variance of the mitigated results.

We comment that a similar notion of probabilistic noise reduction was recently proposed also in Ref. [19].

A. Probabilistic error reduction

For noise scale factors λ larger than 1 (noise amplification), probabilistic noise scaling is not new. Indeed the probabilistic application of Pauli gates for amplifying the noise, was already used in one of the first experimental applications of ZNE [10]. In this work instead we explore the possibility of virtually scaling down the noise in the interval $\lambda \in [0, 1]$ and we refer to this technique as probabilistic error reduction (PER), which can be considered both as noise scaling method and as an error mitigation method.

Given a generic definition of noise scaling (see Eq. (21)), PER can be obtained generalizing the PEC quasi-probability representation given in Eq. (3). Specifically, we replace the unitary gate on the l.h.s. of Eq. (3) with a non-unitary channel corresponding to the i -th noise-scaled gate $\mathcal{G}_i^{(\lambda)}$ of the circuit of interest:

$$\begin{aligned} \mathcal{G}_i^{(\lambda)} &= \sum_{\alpha} \eta_{i,\alpha}(\lambda) \mathcal{O}_{i,\alpha}, \quad \mathcal{O}_{i,\alpha} \in \mathcal{I}, \\ \sum_{\alpha} \eta_{i,\alpha}(\lambda) &= 1 \quad \forall \lambda, \quad \gamma_i(\lambda) := \sum_{\alpha} |\eta_{i,\alpha}(\lambda)|. \end{aligned} \quad (33)$$

Explicit quasi-probability representations for the particular case of depolarizing noise are derived in subsection VIC and a more general scenario is considered subsection VID. An example of PER is shown in Fig. 3(a), where an expectation value is estimated at different virtual noise scale factors, interpolating between the unmitigated result ($\lambda = 1$) and PEC ($\lambda = 0$).

Here we stress an important aspect which is evident in Fig. 3(a): the statistical uncertainty associated to the

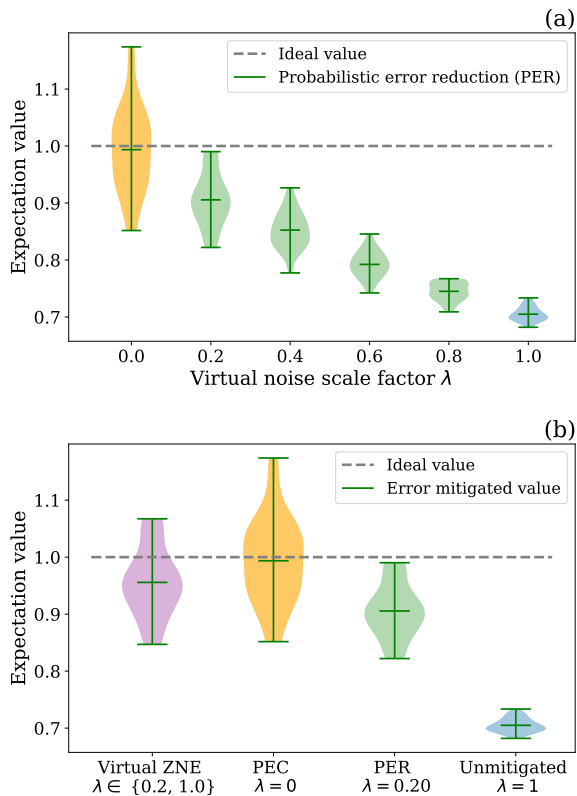


FIG. 3. (a) Probabilistic error reduction by virtual noise scaling. A single-qubit randomized-benchmarking circuit composed of 46 gates is simulated assuming a depolarizing noise model with $p = 0.015$. The expectation value of $A = |0\rangle\langle 0|$ is estimated at different values of the noise scale factor λ via PER according to Eq. (33) and, more specifically, (35). For increasing values of λ , the results continuously interpolate between PEC ($\lambda = 0$) and no mitigation ($\lambda = 1$), while the associated statistical variance gets reduced. (b) Error mitigation by linear zero-noise extrapolation with virtual scale factors $S = \{\lambda_1, \lambda_2\} = \{0.2, 1\}$, compared to standard PEC and to PER (with $\lambda = 0.2$). We observe that all error mitigation methods improve the unmitigated result. As expected, virtual ZNE and PER are on average more biased than PEC, however, their statistical uncertainty can be significantly smaller. For each noise scale factor, 5×10^4 measurements (shots) have been used. For each shot, a different circuit was sampled according to PER. To estimate the statistical distribution, the samples have been divided into 25 batches of 2×10^3 elements each. The “violin plots” show the statistical distributions with respect to independent batches. The central horizontal segment in each “violin” corresponds to the mean of all the samples, which is equal to the mean of the batches. For the numerical simulation of this example, we used the error mitigation software package Mitiq [26].

PER estimates decreases for larger values of λ . For this reason PER can be considered as a low-cost error mitigation method in the sense that, compared to PEC, it requires a smaller number of samples at the price of a partial mitigation of the noise. Equivalently, for a fixed number of samples, PER results are characterized by a

smaller statistical uncertainty compared to PEC. For this reason PER could be useful in real-world scenarios characterized by a large level of noise, in which the sampling cost of PEC would be too large. We mention that a similar bias-variance trade off in probabilistic error cancellation has been recently studied also in Eq. [18], although with a different approach in which noise scaling is not involved.

B. Virtual ZNE

A possible way to get a better estimate of the ideal expectation value is to consider PER as a noise scaling method and to apply ZNE as a second step (see also Ref. [19]). A demonstration of this virtual zero-noise extrapolation technique is reported in Fig. 3(b).

The potential advantage with respect to standard ZNE is quite evident: we can now explore the region $0 < \lambda < 1$ and this makes the extrapolation to $\lambda = 0$ less biased. But what is the advantage with respect to standard PEC? In PEC, the ideal expectation value ($\lambda = 0$) is directly computed with an unbiased estimator such that no extrapolation is necessary. As discussed in the case of PER, also virtual ZNE can have a smaller statistical error compared to PEC, as shown the example of Fig. 3(b). A notable fact is that the zero-noise limit could also be extrapolated with a non-linear function of the noisy expectation values (e.g. when using an exponential fit), while PEC is always linear by construction.

In conclusion, we may summarize both PER and virtual ZNE as hybrid techniques interpolating between two standard inference methods: ZNE (characterized by low variance but biased) and PEC (unbiased but affected by a large variance).

C. Example: depolarizing-channel error reduction

To clarify the probabilistic error reduction (PER) method, we explicitly present a simple single-qubit example. We consider the noisy gate set corresponding to all single-qubit unitaries followed by depolarizing noise: $\mathcal{I} = \{\mathcal{D}_p \circ \mathcal{G} \mid \mathcal{G} \text{ is unitary}\}$. Instead of completely cancelling the noise as in PEC, in PER we are interested in the effective implementation of noise scaled gates:

$$\mathcal{G}^{(\lambda)} = D_{\lambda p} \circ \mathcal{G}, \quad 0 \leq \lambda \leq \lambda_{\max} = 3/4p^{-1}. \quad (34)$$

Extending the same calculations of Refs. [7, 20] to the case $\lambda \neq 0$ and using the notation $\epsilon = 4/3p$, it is easy to derive the following quasi-probability representation which is valid for any noise-scaled gate:

$$\mathcal{G}^{(\lambda)} = \eta_1 \mathcal{O}_1 + \eta_2 \mathcal{O}_2 + \eta_3 \mathcal{O}_3 + \eta_4 \mathcal{O}_4, \quad (35)$$

where:

$$\begin{aligned}\eta_1 &= \left(1 + \frac{3}{4} \frac{\epsilon(1-\lambda)}{1-\epsilon}\right), & \mathcal{O}_1 &= \mathcal{D}_p \circ \mathcal{G}, \\ \eta_2 &= -\frac{1}{4} \frac{\epsilon(1-\lambda)}{1-\epsilon}, & \mathcal{O}_2 &= \mathcal{D}_p \circ (\mathcal{X} \circ \mathcal{G}), \\ \eta_3 &= -\frac{1}{4} \frac{\epsilon(1-\lambda)}{1-\epsilon}, & \mathcal{O}_3 &= \mathcal{D}_p \circ (\mathcal{Y} \circ \mathcal{G}), \\ \eta_4 &= -\frac{1}{4} \frac{\epsilon(1-\lambda)}{1-\epsilon}, & \mathcal{O}_4 &= \mathcal{D}_p \circ (\mathcal{Z} \circ \mathcal{G}).\end{aligned}\quad (36)$$

In the definition of the noisy operations \mathcal{O}_j given in the equation above, the composition of the unitary channels applied before \mathcal{D}_p represent a single elementary gate. For $\lambda = 0$, Eq. (36) reduces to the PEC quasi-probability representation obtained in Refs. [7, 20]. For $1 \leq \lambda \leq \lambda_{\max} = \epsilon^{-1}$, all coefficients are positive corresponding to the direct application of depolarizing errors to scale up the noise (see *e.g.* [10]). For $0 \leq \lambda \leq 1$ instead the situation is less trivial and corresponds to the probabilistic error reduction regime considered in this work. In this case, the last 3 coefficients of Eq. (36) are negative and the negative volume of the quasi-distribution decreases linearly with λ , interpolating between the two extremes of full error cancellation (at $\lambda = 0$) and no-mitigation (at $\lambda = 1$). More explicitly the one-norm of the quasi-probability given in Eq. (36), as a function of λ , is given by:

$$\gamma^{(\lambda)} = \begin{cases} \gamma - \lambda(\gamma - 1), & \lambda \in [0, 1], \\ 1, & \lambda \in [1, \epsilon^{-1}], \end{cases}\quad (37)$$

where $\gamma = 1 + \frac{3}{2} \frac{\epsilon}{1-\epsilon}$ is the PEC one-norm (at $\lambda = 0$). For the sake of completeness, we mention that Eqs. (36) are also valid in the region $\lambda = [\epsilon^{-1}, p^{-1}]$ where the noise scaled depolarizing channel is completely positive (physical). However the one-norm $\gamma^{(\lambda)}$ remains equal to 1 only for $\lambda = [\epsilon^{-1}, \frac{\gamma+1}{\gamma-1}]$, while it increases again for $\lambda \in [\frac{\gamma+1}{\gamma-1}, p^{-1}]$. This is consistent with the more general analysis that we will present in the next subsection (Sec. VID, Eq. (44)).

For a fixed number of samples, Eq. (37) implies that PER is affected by a decreasing statistical uncertainty for increasing values of λ , at the price of a larger bias error. For the same reason, the statistical variance of the virtual ZNE technique is reduced too. Both phenomena are clearly visible in the numerical example shown in Fig. 3 and these qualitative features are reported in Table I.

D. Canonical noise scaling

In the previous example we considered a depolarizing noise model which depends on a single parameter p and therefore admits a natural notion of noise reduction $p \rightarrow \lambda p$. The same approach could be applied when the noise model is an amplitude damping channel [7, 20]. However, how can one meaningfully scale an arbitrary

noise model? In this section we show that a generic quasi-probability representation of an ideal gate induces an abstract canonical noise model that we can associate to that gate. This abstract noise model is a well-defined trace-preserving and completely positive channel which depends linearly on a single noise parameter p . In practice, this implies that noise scaling can always be meaningfully defined for any quasi-probability representation of an ideal gate.

Let us consider the representation of an ideal gate \mathcal{G} as a linear combination of noisy operations \mathcal{O}_α weighted by a quasi-probability distribution η_α . The explicit formula is given in Eq. (3) where, for simplicity, we now drop the gate index i . We now split the domain of the quasi-distribution η_α in two parts corresponding to the regions where η_α is positive and negative respectively:

$$D^{(\pm)} = \{\alpha, \text{ such that } \eta_\alpha \gtrless 0\}.\quad (38)$$

In a similar way we can define two completely positive and trace-preserving channels associated to the positive and negative regions respectively (see *e.g.* Ref. [20] or, in the context of magic states, Ref. [27]):

$$\Phi^{(\pm)} = \frac{1}{\gamma^{(\pm)}} \sum_{D^{(\pm)}} |\eta_\alpha| \mathcal{O}_\alpha, \quad \gamma^{(\pm)} = \sum_{D^{(\pm)}} |\eta_\alpha|,\quad (39)$$

where $\gamma^{(\pm)}$ are to the positive and negative volumes of the quasi-distribution, such that the total one-norm is $\gamma = \gamma^{(+)} + \gamma^{(-)}$, while the normalization implies that $\gamma^{(+)} - \gamma^{(-)} = 1$.

Given the previous definitions we can reduce the multi-term linear combination in Eq. (3) to a linear combinations of just two channels:

$$\mathcal{G} = \gamma^{(+)} \Phi^{(+)} - \gamma^{(-)} \Phi^{(-)},\quad (40)$$

where both channels $\Phi^{(\pm)}$ can be applied on the noisy hardware since they are a convex combination of implementable operations. What is the physical meaning of these channels? As we are going to show, Φ^+ can be considered as a noisy approximation of \mathcal{G} , while Φ^- could be considered an error term that needs to be subtracted from Φ^+ in order to recover the ideal gate \mathcal{G} .

This intuition suggests the following canonical noise model associated to the ideal gate \mathcal{G} :

$$\Lambda_p = (1-p) \mathcal{G} + p \Phi^{(-)}, \quad p \in [0, 1],\quad (41)$$

corresponding to a channel in which with probability p the error operation $\Phi^{(-)}$ occurs, while with probability the $(1-p)$ the ideal gate is applied without any errors. It is easy to check that, depending the single parameter p , the channel Λ_p interpolates between the three channels that appear in Eq. (40): for $p = 0$ we obtain the ideal gate \mathcal{G} , for $p = \gamma^{(-)}/\gamma^{(+)}$ we obtain $\Phi^{(+)}$ which therefore could be considered as a noisy (implementable) approximation of the ideal gate, while for $p = 1$ we get $\Phi^{(-)}$ which can be considered as the maximum noise limit.

If we identify $\tilde{p} = \gamma^{(-)}/\gamma^{(+)}$ as the noise level of the hardware, we directly obtain a natural definition of a noise scaled gate:

$$\mathcal{G}^{(\lambda)} = \Lambda_{\lambda\tilde{p}} = (1 - \lambda\tilde{p})\mathcal{G} + \lambda\tilde{p}\Phi^{(-)}, \quad \tilde{p} = \frac{\gamma^{(-)}}{\gamma^{(+)}} \quad (42)$$

where $0 \leq \lambda \leq \lambda_{\max} = \gamma^{(+)}/\gamma^{(-)}$. By construction, this type of noise reduces the negativity of the quasi-probability representation. Indeed, replacing Eq. (40) into Eq. (42), we obtain the following representation for the noise scaled gate in terms of implementable operations:

$$\mathcal{G}^{(\lambda)} = (\gamma^{(+)} - \lambda\gamma^{(-)})\Phi^{(+)} - (1 - \lambda)\gamma^{(-)}\Phi^{(-)}, \quad (43)$$

whose sampling cost (one-norm) decreases linearly with respect to the noise scale factor λ :

$$\gamma^{(\lambda)} = \begin{cases} \gamma - \lambda(\gamma - 1), & \lambda \in [0, 1], \\ 1, & \lambda \in [1, \lambda_{\max} = \frac{\gamma+1}{\gamma-1}]. \end{cases} \quad (44)$$

Note: For the optimal PEC representation of a gate in the presence of depolarizing noise given in Eq. (35) (evaluated at $\lambda = 0$), the canonical noise channel defined in Eq. (41) is equal to the actual physical depolarizing channel acting on the system. In other words, the noise-scaled quasi-probability decomposition that was manually computed in Eqs. (36) is a particular case of Eq. (43). The advantage of Eq. (43) is that it is well defined even for an arbitrary PEC representation, *e.g.*, one based on the experimental tomography of a noisy gate set.

VII. MINIMUM SAMPLING COST OF EXACT NEPEC REPRESENTATIONS

In the previous sections we presented several techniques in which ideal gates are expanded in a noise-extended basis with *approximate* representations, *i.e.*, with a non-zero bias. In this section instead we focus on *exact* gate representations and we study the possibility of reducing the optimal sampling cost γ^{opt} introduced in Eq. (5) by using NEPEC instead of PEC. Our analysis is inspired by the resource theory of error mitigation proposed in Ref. [20] and we follow a similar (but not identical) notation. Let \mathcal{I} be the set of implementable operations used in PEC. Then, by noise scaling, we obtain the extended set of implementable operations $\tilde{\mathcal{I}}$ introduced in Eq. (20), such that any ideal gate \mathcal{G}_i of a circuit can be expanded in this extended basis as described in Eq. (24). The associated optimal sampling cost is:

$$\tilde{\gamma}_i^{\text{opt}} = \min_{\substack{\{\eta_{i,\alpha,\lambda}\} \\ \{\mathcal{O}_\alpha^{(\lambda)}\}}} \left[\sum_{\alpha,\lambda} |\eta_{i,\alpha,\lambda}| \right] \text{ s.t. Eq. (24) holds } \leq \gamma_i^{\text{opt}}, \quad (45)$$

where the last inequality sign is a consequence of the fact that, since $\mathcal{I} \subset \tilde{\mathcal{I}}$, the minimization in Eq. (45) is over a larger landscape w.r.t. Eq. (5) and so, in principle, a smaller value can be reached. But is this inequality strict? We will show that this depends on the particular noise model and on the specific gate set \mathcal{I} .

A. A no-go test

Hypothesis 1: *Let us consider the particular setting in which each noise-scaled operation in $\tilde{\mathcal{I}}$ can be represented as a convex combination of operations which are in the non-scaled set \mathcal{I} , *i.e.*:*

$$\forall \mathcal{O}^{(\lambda)} \in \tilde{\mathcal{I}}, \quad \mathcal{O}^{(\lambda)} = \sum_{\alpha} \mu_{\lambda}(\alpha) \mathcal{O}_{\alpha}, \quad \mathcal{O}_{\alpha} \in \mathcal{I}, \quad (46)$$

where $\{\mu_{\lambda}(\alpha) \geq 0\}$ is a positive probability distribution with respect to α .

In practice, if Hypothesis 1 holds, instead of actually scaling the hardware noise one could effectively obtain the same result by just probabilistically drawing an operation in \mathcal{I} according to the probability distribution $\mu_{\lambda}(\alpha)$. It is reasonable to expect that, in this case, there cannot be any advantage with respect to the sampling cost and so the inequality in (45) saturates to a trivial equality.

Indeed, by replacing Eq. (46) into Eq. (24), we see that for each noise-extended representation of a gate \mathcal{G}_i there exists an equivalent representation which does not require noise scaling and has the same one-norm parameter γ_i . The corresponding quasi-distribution is:

$$\eta_{i,\alpha'} = \sum_{\alpha,\lambda} \eta_{i,\alpha,\lambda} \mu_{i,\alpha,\lambda}(\alpha'), \quad (47)$$

whose one-norm is the same as the NEPEC one-norm, since

$$\sum_{\alpha'} |\eta_{i,\alpha'}| = \sum_{\alpha,\lambda} |\eta_{i,\alpha,\lambda}| \sum_{\alpha'} |\mu_{i,\alpha,\lambda}(\alpha')| = \sum_{\alpha,\lambda} |\eta_{i,\alpha,\lambda}| = \gamma_i. \quad (48)$$

Therefore we can conclude that, if Hypothesis 1 holds, noise-scaling cannot reduce the sampling cost, *i.e.*, $\tilde{\gamma}^{\text{opt}} = \gamma^{\text{opt}}$.

A relevant example in which this no-go result applies is when the set of implementable operations is $\mathcal{I} = \{\mathcal{E}_p \circ \mathcal{G}, \forall \text{ unitary } \mathcal{G}\}$ and \mathcal{D}_p is a depolarizing channel acting on k qubits (see *e.g.* [20] for a detailed analysis of this error mitigation scenario). The corresponding noise-extended set would be $\tilde{\mathcal{I}} = \{\mathcal{D}_{\lambda p} \circ \mathcal{G}, \forall \text{ unitary } \mathcal{G}, \lambda \in [1, \lambda_{\max} = (1 - 4^{-k})p^{-1}]\}$, where the explicit expression for $\mathcal{D}_{\lambda p}$ was already given in Eq. (22). For $\lambda \in [1, \lambda_{\max}]$,

$\mathcal{D}_{\lambda p} = \mathcal{D}_p \circ \mathcal{D}_{p'(\lambda)}$ for some $p'(\lambda) \in [0, 1]$, and so we have:

$$\begin{aligned} \mathcal{G}^{(\lambda)} &= \mathcal{D}_{\lambda p} \circ \mathcal{G} = \mathcal{D}_p \circ \mathcal{D}_{p'(\lambda)} \circ \mathcal{G} \\ &= [1 - p'(\lambda)]\mathcal{D}_p \circ \mathcal{G} + p'(\lambda) \sum_{\mathcal{P} \in \mathcal{P}_k} \frac{\mathcal{D}_p \circ \mathcal{P} \circ \mathcal{G}}{4^k - 1}, \end{aligned} \quad (49)$$

where the last sum is over all the Pauli strings different from the identity as described in Eq. (22). Since $\mathcal{P} \circ \mathcal{G}$ is a unitary, $\mathcal{D}_p \circ \mathcal{P} \circ \mathcal{G} \in \mathcal{I}$ and so Eq. (49) corresponds to a convex combination of operations which are in \mathcal{I} . This means that Hypothesis 1 holds and so, in this case, noise scaling cannot help in reducing the mitigation cost of an exact representation.

B. Example: NEPEC representations with amplitude damping noise

An example in which $\tilde{\gamma}^{\text{opt}} < \gamma^{\text{opt}}$, is given by the following noisy gate set: $\mathcal{I} = \{\mathcal{A}_p \circ \mathcal{G}, \forall \text{ single-qubit unitary } \mathcal{G}\}$ where \mathcal{A}_p is a single-qubit amplitude damping channel. It is easy to check that, since all the \mathcal{G} that appear in the definition of \mathcal{I} are unitary, it is impossible to obtain a valid PEC representation with this gate set [7]. Indeed any linear combination of the noisy elements of \mathcal{I} , if applied to the maximally mixed state, will always simplify to the non-unital channel \mathcal{A}_p and so it can never represent the action of an ideal unitary gate. For this reason, the operation $\text{RESET} = |0\rangle\langle 0| + |0\rangle\langle 1|$ was used in previous works [7, 8, 20] to make PEC feasible in the presence of amplitude damping noise.

In this proof-of-principle example we assume that, for some reason, we cannot apply the RESET gate (e.g. because its physical implementation is too noisy or too slow). In this case, the mitigation norm γ^{opt} is infinite, in the sense that the minimization problem in Eq. (5) is unfeasible without noise scaling. On the other hand, if we extend the gate set \mathcal{I} by noise scaling, the minimization problem in Eq. (5) becomes feasible and we obtain a finite mitigation cost $\tilde{\gamma}^{\text{opt}} < \gamma^{\text{opt}} = \infty$. This can be formally deduced by simply observing that $\mathcal{A}_{\lambda p}$, in the maximum noise limit $\lambda \rightarrow p^{-1}$, tends to the RESET gate and so we can recover the results of Refs. [7, 8, 20] in this limit. Moreover we numerically find that even using intermediate values of noise scaling $1 < \lambda < p^{-1}$, the representation problem becomes feasible without a RESET gate at the cost of obtaining a larger (but finite) one-norm. Unfortunately we also report that, when using noise scaling on an initial gate set \mathcal{I} which already includes the RESET gate, we could not obtain any further reduction of the sampling cost.

Inspired by the gate extrapolation technique proposed in Sec. V, we may also ask if it is possible to obtain a representation in the form of Eq. (30), *i.e.*, only using a linear combination of the same gate $\mathcal{G}_i^{(\lambda)}$ applied

at different noise scale factors. Differently from the depolarizing super-operator \mathcal{D}_p which is linear in p , the amplitude damping super-operator \mathcal{A}_p is non-linear in p . This is not a problem if we aim for an approximate (biased) representation, but obtaining an exact (unbiased) representation is less trivial. Using the change of variable $p' = 1 - \sqrt{1 - p}$, \mathcal{A}_p becomes a quadratic function of p' . This implies that by scaling p' with 3 scale factors $\lambda \in S = [\lambda_1, \lambda_2, \lambda_3]$ we can always extrapolate to the exact zero-noise limit using the Richardson coefficients defined in Eq. (31). Minimizing the one-norm of the quasi-probability $\{\eta_{i,\lambda}\}$ with respect to S and imposing $\lambda_j \in [0, 1/p']$, one can obtain the optimal coefficients $S = [1, (1 + 1/p')/2, 1/p']$. The corresponding optimized one norm is:

$$\gamma_i = \frac{1 + 6p' + p'^2}{(1 - p')^2} = \frac{9 - 8\sqrt{1 - p} - p}{1 - p}. \quad (50)$$

For small p , the equation above scales as $1 + 4p + O(p^2)$. For comparison, we note that this is strictly larger than the sampling cost derived in Refs. [20] which scales as $1 + p + O(p^2)$. This is not surprising since Eq. (50) corresponds to a very restricted noisy basis (just a single noise-scaled gate).

We conclude noticing that, while the change of variable $p \rightarrow p'$ was useful for the theoretical derivation of Eq. (50) and for proving the existence of an exact quadratic extrapolation, it is not necessary to apply it in a practical scenario. In a real use case, one can directly scale the noise level p with three different noise scale factors and just solve for the coefficients $\{\eta_{i,\lambda} | \lambda \in S\}$ in Eq. (30). The previous theoretical analysis ensures that, in the presence of amplitude damping noise, three different scale factors are enough to obtain an exact solution of Eq. (30).

VIII. CONCLUSIONS

We propose a general error mitigation framework—NEPEC—in which a given ideal quantum circuit is represented in terms of a quasi-probability distribution over different circuits evaluated at different noise levels. This approach generalizes existing techniques (PEC and ZNE) and, depending on the choice of the quasi-probability distribution, gives rise to different practical implementations as summarized in Table I.

A promising implementation of NEPEC is the possibility of defining approximate quasi-probability representations of individual gates (or layers) via local extrapolation and therefore without the need of performing gate set tomography. This fact can be an important practical advantage compared to standard PEC and, by construction, it bypasses noise characterization errors. For example, in the simulation reported in Fig. 2, we have shown

the robustness of this method against noise calibration errors.

Other specific implementations of the NEPEC approach are probabilistic error reduction (PER) and virtual ZNE. In PER, noise is reduced at intermediate scale factors $\lambda \in [0, 1]$ with the advantage of requiring a smaller sampling cost compared to PEC ($\lambda = 0$). Moreover, PER can also be considered as a virtual noise scaling method, and ZNE can be applied as a second post-processing step as shown in Fig. 3 (see also [19]). As a by-product of our theoretical analysis, we also identified a canonical noise channel that can be associated to any quasi-probability representation of a gate (Sec. VID). This canonical channel is useful to apply PER with any noisy gate set and might be of independent interest beyond the scope of this work.

Finally we investigated if, by extending the basis of implementable operations via noise-scaling, it is possible to obtain exact gate representations with a smaller one-norm (sampling cost) compared to standard PEC representations. We found that there is a large class of situations (see Hypothesis 1 in Sec. VII) in which this is impossible, however, we also gave an explicit example in which NEPEC provides an advantage (Sec. VIIB).

We hope that this work can stimulate further theoretical and experimental research lines. Stacking and hybridizing different quantum error mitigation techniques seems a promising strategy [13, 14, 19, 28]. The

freedom of sampling over different circuits and different noise levels opens up a large variety of possible error mitigation techniques. In this work we explicitly proposed only some of them (see Table I), but new and perhaps better techniques based on the NEPEC framework could be found in the future. Moreover, the experimental implementation of NEPEC with real quantum processors remains an open and important research task which could be addressed in the near-future. In this regard, we comment that an experimental implementation of NEPEC is feasible with current technology. Indeed, the main technical requirements (sampling and noise scaling) have been already achieved and experimentally demonstrated in the context of PEC [12] and ZNE [11].

ACKNOWLEDGEMENTS

We thank Ryan LaRose, Sarah Kaiser, Daniel Strano and Tudor Giurgica-Tiron for discussions and feedback on this work. This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Accelerated Research in Quantum Computing under Award Number de-sc0020266.

-
- [1] P. W. Shor, *Scheme for reducing decoherence in quantum computer memory*, *Phys. Rev. A* **52**, R2493 (1995).
- [2] A. M. Steane, *Error correcting codes in quantum theory*, *Phys. Rev. Lett.* **77**, 793 (1996).
- [3] A. R. Calderbank and P. W. Shor, *Good quantum error-correcting codes exist*, *Phys. Rev. A* **54**, 1098 (1996).
- [4] D. Aharonov and M. Ben-Or, *Fault-tolerant quantum computation with constant error rate*, *SIAM J. Comp.* (2008).
- [5] A. Y. Kitaev, *Quantum computations: algorithms and error correction*, *Russ. Math. Surv.* **52**, 53 (1997).
- [6] J. Preskill, *Quantum Computing in the NISQ era and beyond*, *Quantum* **2**, 79 (2018).
- [7] K. Temme, S. Bravyi, and J. M. Gambetta, *Error mitigation for short-depth quantum circuits*, *Phys. Rev. Lett.* **119**, 180509 (2017).
- [8] S. Endo, S. C. Benjamin, and Y. Li, *Practical quantum error mitigation for near-future applications*, *Phys. Rev. X* **8**, 031027 (2018).
- [9] S. Endo, Z. Cai, S. C. Benjamin, and X. Yuan, *Hybrid quantum-classical algorithms and quantum error mitigation*, *J. Phys. Soc. Jap.* **90**, 032001 (2021).
- [10] Y. Li and S. C. Benjamin, *Efficient variational quantum simulator incorporating active error minimization*, *Phys. Rev. X* **7**, 021050 (2017).
- [11] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, *Error mitigation extends the computational reach of a noisy quantum processor*, *Nature* **567**, 491 (2019).
- [12] S. Zhang, Y. Lu, K. Zhang, W. Chen, Y. Li, J.-N. Zhang, and K. Kim, *Error-mitigated quantum gates exceeding physical fidelities in a trapped-ion system*, *Nature Commun.* **11**, 1 (2020).
- [13] A. Lowe, M. H. Gordon, P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio, *Unified approach to data-driven quantum error mitigation*, *arXiv preprint arXiv:2011.01157* (2020).
- [14] D. Bultrini, M. H. Gordon, P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio, *Unifying and benchmarking state-of-the-art quantum error mitigation techniques*, *arXiv preprint arXiv:2107.13470* (2021).
- [15] P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio, *Error mitigation with Clifford quantum-circuit data*, *arXiv preprint arXiv:2005.10189* (2020).
- [16] B. Koczor, *Exponential error suppression for near-term quantum devices*, *arXiv preprint arXiv:2011.05942* (2020).
- [17] W. J. Huggins, S. McArdle, T. E. O'Brien, J. Lee, N. C. Rubin, S. Boixo, K. B. Whaley, R. Babbush, and J. R. McClean, *Virtual distillation for quantum error mitigation*, *arXiv preprint arXiv:2011.07064* (2020).
- [18] C. Piveteau, D. Sutter, and S. Woerner, *Quasiprobability decompositions with reduced sampling overhead*, *arXiv preprint arXiv:2101.09290* (2021).
- [19] Z. Cai, *Multi-exponential error extrapolation and combining error mitigation techniques for NISQ applications*, *npj Quant. Inf.* **7**, 80 (2021).

- [20] R. Takagi, *Optimal resource cost for error mitigation*, [arXiv preprint arXiv:2006.12509](#) (2020).
- [21] H. Pashayan, J. J. Wallman, and S. D. Bartlett, *Estimating outcome probabilities of quantum circuits using quasiprobabilities*, *Phys. Rev. Lett.* **115**, 070501 (2015).
- [22] T. Giurgica-Tiron, Y. Hindy, R. LaRose, A. Mari, and W. J. Zeng, *Digital zero noise extrapolation for quantum error mitigation*, *2020 IEEE Int. Conf. Quant. Comp. Eng.*, 306 (2020).
- [23] E. F. Dumitrescu, A. J. McCaskey, G. Hagen, G. R. Jansen, T. D. Morris, T. Papenbrock, R. C. Pooser, D. J. Dean, and P. Lougovski, *Cloud quantum computing of an atomic nucleus*, *Phys. Rev. Lett.* **120**, 210501 (2018).
- [24] A. He, B. Nachman, W. A. de Jong, and C. W. Bauer, *Zero-noise extrapolation for quantum-gate error mitigation with identity insertions*, *Phys. Rev. A* **102**, 012426 (2020).
- [25] This is valid for a basic polynomial regression. For more advanced inference algorithms, e.g. those taking into account different uncertainties for each data point, the extrapolation coefficients can be data-dependent.
- [26] R. LaRose, A. Mari, S. Kaiser, P. J. Karalekas, A. A. Alves, P. Czarnik, M. E. Mandouh, M. H. Gordon, Y. Hindy, A. Robertson, P. Thakre, N. Shammah, and W. J. Zeng, *Mitiq: A software package for error mitigation on noisy quantum computers*, [arXiv preprint arXiv:2009.04417](#) (2020).
- [27] J. R. Seddon and E. T. Campbell, *Quantifying magic for multi-qubit operations*, *Proc. Roy. Soc. A* **475**, 20190251 (2019).
- [28] J. Sun, X. Yuan, T. Tsunoda, V. Vedral, S. C. Benjamin, and S. Endo, *Mitigating realistic noise in practical noisy intermediate-scale quantum devices*, *Phys. Rev. Applied* **15**, 034026 (2021).