

ARTICLE OPEN



Adaptive quantum error mitigation using pulse-based inverse evolutions

Ivan Henao¹, Jader P. Santos¹ and Raam Uzdin¹

Quantum Error Mitigation (QEM) enables the extraction of high-quality results from the presently-available noisy quantum computers. In this approach, the effect of the noise on observables of interest can be mitigated using multiple measurements without additional hardware overhead. Unfortunately, current QEM techniques are limited to weak noise or lack scalability. In this work, we introduce a QEM method termed 'Adaptive KIK' that adapts to the noise level of the target device, and therefore, can handle moderate-to-strong noise. The implementation of the method is experimentally simple — it does not involve any tomographic information or machine-learning stage, and the number of different quantum circuits to be implemented is independent of the size of the system. Furthermore, we have shown that it can be successfully integrated with randomized compiling for handling both incoherent as well as coherent noise. Our method handles spatially correlated and time-dependent noise which enables us to run shots over the scale of days or more despite the fact that noise and calibrations change in time. Finally, we discuss and demonstrate why our results suggest that gate calibration protocols should be revised when using QEM. We demonstrate our findings in the IBM quantum computers and through numerical simulations.

npj Quantum Information (2023)9:120; <https://doi.org/10.1038/s41534-023-00785-7>

INTRODUCTION

Quantum computers have reached a point where they outperform even the most powerful classical computers in specific tasks^{1–3}. However, these quantum devices still face considerable noise levels that need to be managed for quantum algorithms to excel in practical applications. Quantum error correction (QEC) is a prominent solution, although its implementation, particularly in complex problems such as Shor's factoring algorithm, might demand thousands of physical qubits for each encoded logical qubit^{4,5}.

A different approach, quantum error mitigation (QEM), has garnered substantial attention recently^{6–23}. Its viability has been demonstrated through experiments involving superconducting circuits^{19–21,24–29}, trapped ions³⁰, and circuit QED³¹. QEM protocols aim to estimate ideal expectation values from noisy measurements, without the resource-intensive requirements of QEC. This positions them as potential solutions for achieving quantum advantage in practical computational tasks^{19,28}. Some QEM strategies require moderate hardware overheads and can be seen as intermediary solutions between NISQ (Noisy Intermediate-Scale Quantum) computers and devices that fully exploit QEC^{11,12}. These strategies aim to virtually refine the pure final state, by utilizing extra qubits for error mitigation without actively correcting errors. The approach introduced here fits into the more common class of QEM techniques that maintain the qubit count of the original circuit.

The objective of QEM is to reduce errors in post-processing, rather than fixing them in real time. For instance, the zero-noise extrapolation (ZNE) method^{6,7} employs circuits that mimic the ideal target evolution but amplify noise by a controlled factor. The noiseless expectation values are estimated via extrapolation to the zero-noise limit, after fitting a noise scaling ansatz to the measured data. While the construction of circuits that correctly scale the noise is straightforward if the noise is time independent⁶ or if it is described by a global depolarizing channel¹³, it has been

observed that circuits designed to amplify depolarizing noise fail to achieve the intended noise scaling, when applied to more realistic noise models¹⁹. Our experimental findings also show related issues when applying such circuits to QEM in a real system. Another strategy is to simplify the actual noise appearing in multi-qubit gates such as the CNOT, CZ, Toffoli and Fredkin gates, by using randomized compiling, which renders the noise to a Pauli channel^{32,33}. A sufficiently sparse Pauli channel facilitates accurate characterization and noise amplification for ZNE²⁸. Additionally, as in other QEM methods, the performance of ZNE can be enhanced by integrating it with other error mitigation techniques³⁴, such as readout error mitigation¹⁷.

In comparison to ZNE, Probabilistic Error Cancellation (PEC) is a QEM scheme that relies on experimental characterization of the noise to effectively suppress the associated error channel^{6,8,9,20,25}. To this end, PEC uses a Monte Carlo sampling of noisy operations that on average cancel out the noise, thereby providing an unbiased estimation of the noise-free expectation value. However, this objective can only be accomplished when precise and complete tomographic details of the noise process are accessible. In practice, the success of bias suppression in PEC is limited by the scalability and accuracy of gate set tomography in realistic scenarios. Additionally, since noise characteristics evolve over time, the learning process for PEC must be carried out efficiently within a timescale that is shorter than the timescale in which the noise parameters change. A more realistic approach aims for a partial characterization of the noise, using tools like local gate set tomography⁸ or learning of a sparse noise model²⁰. The latter strategy was also employed to assist the implementation of ZNE in the experiment of ref. ²⁸. Alternatively, it is possible to learn a noise model by taking advantage of circuits that are akin to the target circuit but admit an efficient classical simulation^{9,10,16,35}. By concatenating the outcomes from the ideal (simulated) circuits with their experimental counterparts, the noise-free expectation value can be estimated through some form of data regression^{10,35}.

¹Fritz Haber Research Center for Molecular Dynamics, Institute of Chemistry, The Hebrew University of Jerusalem, Jerusalem 9190401, Israel. ✉email: raam@mail.huji.ac.il

Similar learning-based schemes have also been integrated with PEC⁹ and ZNE¹⁶.

In this work, we introduce the ‘Adaptive KIK’ method (‘KIK’ for brevity) for handling time-dependent and spatially correlated noise in QEM. This technique bears a certain (misleading) similarity to a ZNE variant known as circuit (or ‘global’³⁴) unitary folding¹³, where noise is augmented through identity operations that comprise products of the target evolution and its inverse. While both methods utilize folding to mitigate noise, they differ in the error mitigation mechanism and the way the measured data is processed. Instead of extrapolating to the zero-noise limit, we combine appropriately folded circuits to effectively construct the ‘inverse noise channel’ and approximate the ideal unitary evolution. As opposed to PEC, the implementation of the KIK method does not involve any tomographic information or noise learning subroutine. More precisely, the coefficients that weigh the folded circuits are analytically optimized according to a single experimental parameter that probes the intensity of the noise. Another distinctive aspect of KIK mitigation is a specific inversion of the target circuit for the folding procedure. This constitutes a pivotal difference with respect to circuit folding and has practical consequences, as we show experimentally. The combination of a proper inverse and coefficients adapted to the noise strength allows us to mitigate moderate-to-strong noise and significantly outperform circuit folding ZNE in experiments and simulations. Although we show that the weak noise limit of our theory has a clear connection with Richardson ZNE using circuit folding¹³, the correct inversion of the target circuit is still crucial in this limit.

Recently, important results on the fundamental limitations of QEM protocols have been obtained^{36,37}. These studies address the degradation in the statistical precision of generic QEM schemes, as noise accumulates in circuits of increasing size. In this work, instead of analyzing the degradation of statistical precision, our focus is on the accuracy of error mitigation. We obtain upper bounds for the bias between the ideal expectation value of an arbitrary observable and the value estimated using the KIK method, as a function of the accumulated noise. Our bounds show exponential suppression of the bias with respect to the number of foldings when the noise is below a certain threshold. This is in contrast with ZNE schemes which, in general, do not provide accuracy guarantees.

We test the KIK method on a ten-swap circuit and in a CNOT calibration process, using the IBM quantum computing platform. In the ten-swap experiment, we demonstrate the success of our approach for mitigating strong noise. In the calibration experiment, it is illustrated that a noise-induced bias in gate parameters leads to coherent errors. KIK-based calibration can efficiently mitigate these coherent errors by reducing the bias in the calibration measurements. Furthermore, we find that circuit folding (which uses the CNOT as its own inverse) produces erroneous and inconsistent results. Our experimental findings are enhanced by complementing the KIK method with randomized compiling and readout mitigation. We also simulate the fidelity obtained with a noisy ten-step Trotterization³⁸ of the transverse Ising model on five qubits. For unmitigated fidelities as low as 0.85, we show that KIK error mitigation produces final fidelities beyond 0.99.

RESULTS

The KIK formula for time-dependent noise

To derive our results, we adopt the Liouville-space formalism of Quantum Mechanics³⁹ (see Supplementary Note 1), in which density matrices that describe quantum states are written as vectors, and quantum operations as matrices that act on these vectors. In the following, we will employ calligraphic fonts to denote quantum operations. For example, the unitary evolution

associated with an ideal (noise-free) quantum circuit and its noisy implementation will be written as \mathcal{U} and \mathcal{K} , respectively.

In the standard representation involving superoperators and density matrices, the noisy evolution is governed by the equation

$$\frac{d}{dt}\rho = -i[H(t), \rho] + \hat{L}(t)[\rho]. \quad (1)$$

The ideal evolution is generated by the time-dependent Hamiltonian $H(t)$. On the other hand, the effect of noise is characterized by the superoperator $\hat{L}(t)$. In the following, we will refer to this superoperator as the ‘dissipator’. The equivalent of Eq. (1) in Liouville space is the equation

$$\frac{d}{dt}|\rho\rangle = (-i\mathcal{H}(t) + \mathcal{L}(t))|\rho\rangle, \quad (2)$$

where $|\rho\rangle$ is the vectorized form of ρ . Moreover, $\mathcal{H}(t)$ and $\mathcal{L}(t)$ are square matrices that represent the Hamiltonian $H(t)$ and the dissipator, respectively. We refer the reader to Supplementary Note 2 for more details.

The dynamics (2) gives rise to the noisy target evolution, which we have denoted by \mathcal{K} . As shown in Supplementary Note 3, we can write the solution to Eq. (2) as $\mathcal{K} = \mathcal{U}e^{\Omega(T)}$, where $\Omega(T) = \sum_{n=1}^{\infty} \Omega_n(T)$ is the so-called Magnus expansion⁴⁰. The time T is the total evolution time and $\Omega_n(T)$ is the n th order Magnus term corresponding to T . Here, we are specifically interested in the first Magnus term $\Omega_1(T)$, for reasons that will be clarified below. In our framework, $\Omega_1(T)$ characterizes the impact of noise and is given by

$$\Omega_1(T) = \int_0^T dt \mathcal{U}^\dagger(t) \mathcal{L}(t) \mathcal{U}(t), \quad (3)$$

where $\mathcal{U}(t)$ is the noise-free evolution at time t . In particular, $\mathcal{U} := \mathcal{U}(T)$ is the unitary associated with the noise-free target circuit.

Our basic approximation is the truncation of the Magnus series to first order. This leads to

$$\mathcal{K} \approx \mathcal{U}e^{\Omega_1(T)}. \quad (4)$$

Next, we apply the same approximation to a suitable inverse evolution \mathcal{K}_1 , such that \mathcal{K}_1 reproduces the unitary \mathcal{U}^\dagger in the absence of noise. We construct \mathcal{K}_1 through an inverse driving $\mathcal{H}_1(t)$ defined by

$$\mathcal{H}_1(t) = -\mathcal{H}(T - t). \quad (5)$$

The driving $\mathcal{H}_1(t)$ undoes the action of $\mathcal{H}(t)$, and it produces \mathcal{U}^\dagger . By using $\mathcal{H}_1(t)$, we find in Supplementary Note 3 that, to first order in the Magnus expansion, the solution to the corresponding noisy dynamics satisfies

$$\mathcal{K}_1 \approx e^{\Omega_1(T)} \mathcal{U}^\dagger. \quad (6)$$

Note that this approximation does not mean that we keep only the linear term $\Omega_1(T)$, since all the powers of $\Omega_1(T)$ are included in the exponential $e^{\Omega_1(T)}$. In Eqs. (6) and (7), we use the symbol ‘ \approx ’ to denote equality up to the first Magnus term.

The fact that $\Omega_1(T)$ is also present in the inverse evolution \mathcal{K}_1 allows us to express the error channel as $e^{\Omega_1(T)} \approx (\mathcal{K}_1 \mathcal{K})^{\frac{1}{2}}$. While $\mathcal{H}_1(t)$ is not the only alternative for generating \mathcal{U}^\dagger , it guarantees the generation of a noise channel that is identical, within our approximation, to the noise channel of \mathcal{K} . Thus, by working within the first-order truncation of the Magnus expansion, we can combine Eqs. (4) and (6) to obtain

$$\begin{aligned} \mathcal{U} &\approx \mathcal{K} e^{-\Omega_1(T)} \\ &\approx \mathcal{K} (\mathcal{K}_1 \mathcal{K})^{-\frac{1}{2}}. \end{aligned} \quad (7)$$

The ‘KIK formula’ in the second line of (7) is our main result. In the next section, we discuss the implementation of the KIK method through polynomial expansions of the operator $(\mathcal{K}_1 \mathcal{K})^{-\frac{1}{2}}$ appearing in this formula.

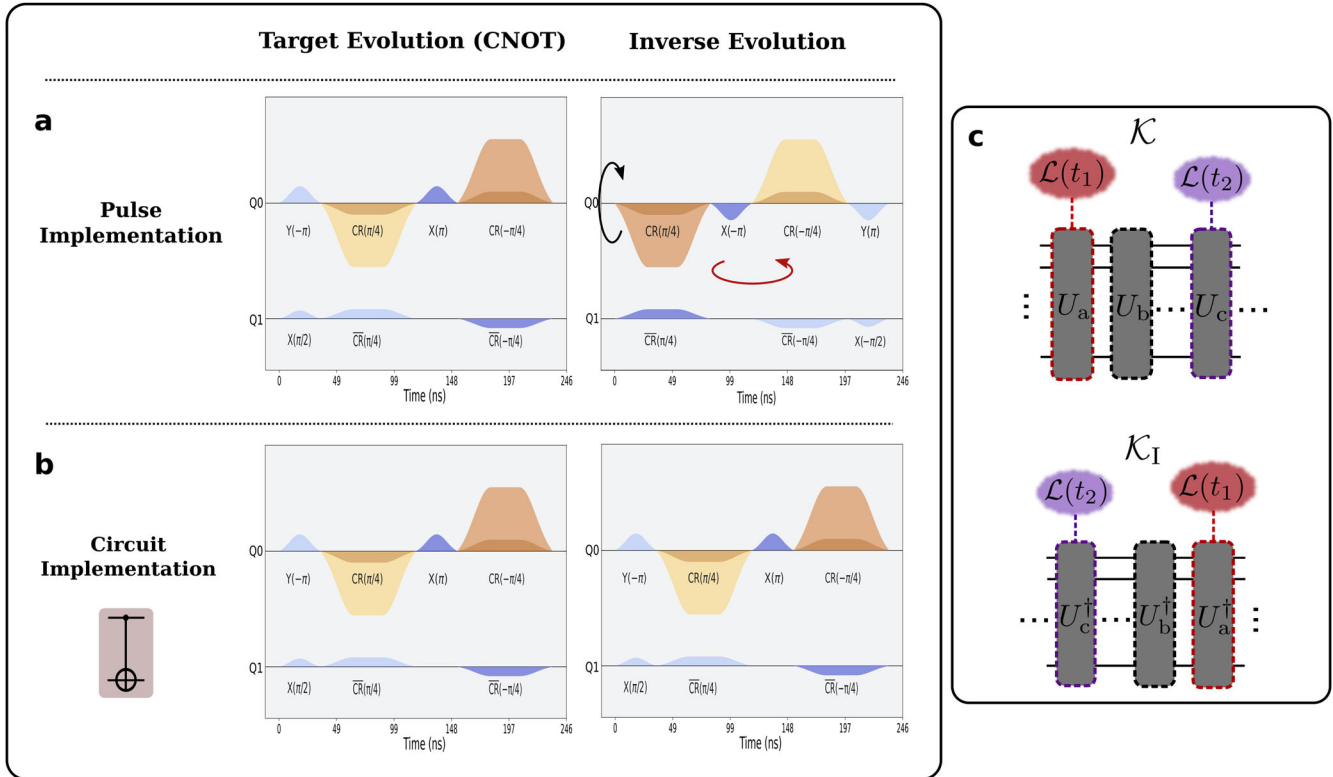


Fig. 1 Illustration of the pulse inverse used in the KIK method. **a** Quantum gates are executed via classical control signals, or pulses. The left panel shows a pulse schedule used for a CNOT gate in the IBM quantum computing platform. The pulse schedule in the right panel performs the inverse of the CNOT through the inverse driving $\mathcal{H}_I(t)$. It is constructed from the original pulse schedule $\mathcal{H}(t)$, by inverting the amplitudes of the pulses (black curved arrow) and their time ordering (red curved arrow). **b** Instead of the pulse inverse, circuit folding and other variants of unitary folding^{13,34} use the CNOT as its own inverse. Therefore, the pulse schedule for the inverse evolution is not modified. **c** Noisy implementations of \mathcal{K} and \mathcal{K}_I . We assume that during the executions of \mathcal{K} and \mathcal{K}_I temporal variations of the noise due to external factors (e.g. temperature variations) are negligible. Thus, any time dependence in $\mathcal{L}(t)$ is induced by the time dependence of $\mathcal{H}(t)$. (Top) This leads to gate dependent noise depicted by different border colors in the gates U_a , U_b , and U_c . (Bottom) Since $\mathcal{H}_I(t)$ reverses the time ordering of $\mathcal{H}(t)$, the time ordering of $\mathcal{L}(t)$ is also reversed. However, the sign of $\mathcal{L}(t)$ does not change because otherwise the inverse evolution would undo the noise.

We stress that until now the only assumption regarding the nature of the noise is that (see Supplementary Note 2)

$$\mathcal{L}_I(t) = \mathcal{L}(T - t), \quad (8)$$

where $\mathcal{L}_I(t)$ is the dissipator acting alongside $\mathcal{H}_I(t)$. This relationship follows from the form of the driving (5), and is schematically explained in Fig. 1. As detailed in Supplementary Note 2, Eq. (8) relies on the time locality of the noise. That is, on the assumption that the dissipators $\mathcal{L}(t)$ and $\mathcal{L}_I(t)$ are only determined by the current time t and not by the previous history of the evolution. Therefore, Eq. (8) may be violated or only hold approximately in the presence of pronounced non-Markovian noise.

Due to the generality of $\mathcal{L}(t)$, Eq. (7) is applicable to quantum circuits \mathcal{K} that feature time-dependent and spatially correlated noise, as well as gate-dependent errors. In Supplementary Note 3, we also discuss the scenario where noise parameters drift during the experiment, which occurs for example due to temperature variations or laser instability. We show that the impact of noise drifts can be practically eliminated in our method, if the execution order of the circuits $\mathcal{K}(\mathcal{K}_I\mathcal{K})^m$ in Eq. (9) is properly chosen. As a final remark, we note that the time independent Lindblad master equation⁴¹ is a special case of Eq. (1). Therefore, our formalism goes beyond QEM proposals based on such a master equation, like the one adopted in ref. ⁴².

QEM using the KIK formula

Since $\mathcal{K}(\mathcal{K}_I\mathcal{K})^{-\frac{1}{2}}$ is not directly implementable in a quantum device, we utilize polynomial expansions of $(\mathcal{K}_I\mathcal{K})^{-\frac{1}{2}}$ such that

$$\mathcal{U}_{\text{KIK}}^{(M)} = \sum_{m=0}^M a_m^{(M)} \mathcal{K}(\mathcal{K}_I\mathcal{K})^m. \quad (9)$$

The notation $\mathcal{U}_{\text{KIK}}^{(M)}$ represents an M th-order approximation to \mathcal{U}_{KIK} , with real coefficients $\{a_m^{(M)}\}_{m=0}^M$. In this way, we estimate the error-free expectation of an observable A as

$$\langle A \rangle_{\text{KIK}}^{(M)} = \sum_{m=0}^M a_m^{(M)} \langle A \rangle_m, \quad (10)$$

where $\langle A \rangle_m$ is the expectation value measured after executing the circuit $\mathcal{K}(\mathcal{K}_I\mathcal{K})^m$ on the initial state ρ . Before discussing the evaluation of the coefficients $a_m^{(M)}$, used in Eq. (10), it is instructive to clarify some similarities and differences between the KIK method and ZNE based on circuit folding.

The application of the KIK formula is operationally analogous to the use of circuit folding for ZNE^{13,34}. However, there are two crucial differences between these two techniques. Circuit folding is a variant of unitary folding, first introduced in ref. ¹³ as a user-friendly strategy for noise amplification in ZNE. It operates by inserting quantum gates that are logically equivalent to the identity operation, which leave the noiseless circuit unmodified. In

the case of ‘circuit folding’, identities are generated by folding the target circuit with a corresponding inverse circuit. Hence, the noise is scaled through evolutions that have the structure $\mathcal{U}(\mathcal{U}^\dagger\mathcal{U})^m$ ¹³. Notably, excluding the trivial case of a global depolarizing channel¹³, a rigorous description of how noise manifests when executing $\mathcal{U}(\mathcal{U}^\dagger\mathcal{U})^m$ was never presented, to the best of our knowledge. In this sense, circuit folding and other variants of unitary folding can be considered as a heuristic approach to QEM. Upon measuring the observable of interest on these circuits, the noiseless expectation value is estimated by combining the results corresponding to different values of m , with weights that depend on the noise scaling ansatz.

The similarity with respect to the KIK method comes from the fact that the circuits $\mathcal{K}(\mathcal{K}_i\mathcal{K})^m$ in Eq. (9) are noisy implementations of $\mathcal{U}(\mathcal{U}^\dagger\mathcal{U})^m$. However, a key difference is that in our case \mathcal{U}^\dagger is performed using the driving (5). Hereafter, we shall refer to this implementation as the ‘pulse inverse’. Conversely, unitary folding (and particularly circuit folding) relies on a circuit-based inversion, where gates that are their own inverses are executed in their original form. This is true for both foldings of single gates (or circuit layers) and for circuit foldings. A paradigmatic example would be the CNOT gate. In contrast, the driving (5) reverses the pulse schedule for each gate in the target circuit, including CNOTs and other gates that are their own inverses. This translates into a very distinct execution of \mathcal{U}^\dagger , as illustrated in Fig. 1a. Even if \mathcal{U} is just a single CNOT, we show in the section ‘Experimental results’ that properly folded circuits correspond to products between the CNOT and its pulse inverse, while circuit folding (i.e. products of the CNOT with itself) leads to erroneous results. Regarding the implementation of our method on cloud-based platforms, we are currently writing an open source Qiskit module that generates pulse-inverse circuits automatically, using only gate-level control. Consequently, users will not need to master pulse-level control to utilize our QEM technique.

Let us now discuss another major difference between our scheme and QEM protocols based on ZNE (including circuit folding). In the case of ZNE, the coefficients that weigh different noise amplification circuits are determined by the fitting of the noise scaling ansatz to experimental data. Rather than that, we ask how to choose these coefficients in such a way that $\mathcal{U}_{\text{KIK}}^{(M)}$ constitutes a good approximation to the KIK formula. This problem can be formulated in terms of the eigenvalues of the operators $(\mathcal{K}_i\mathcal{K})^{-\frac{1}{2}}$ and $\sum_{m=0}^M a_m^{(M)}(\mathcal{K}_i\mathcal{K})^m$. If λ denotes a generic eigenvalue of $\mathcal{K}_i\mathcal{K}$, our goal is to find a polynomial $\sum_{m=0}^M a_m^{(M)}\lambda^m$ that is as close as possible to $\lambda^{-1/2}$. Depending on the noise strength, we follow the two strategies presented in the following two sections. This will further clarify why our method cannot be not considered as a ZNE variant.

QEM in the weak noise regime

In the limit of weak noise, the circuit $\mathcal{K}_i\mathcal{K}$ resembles the identity operation and therefore in this case it is reasonable to approximate the function $\lambda^{-\frac{1}{2}}$ by a truncated Taylor series around $\lambda=1$. The resulting Taylor polynomial leads to the Taylor mitigation coefficients $a_m^{(M)}=a_{\text{Taylor},m}^{(M)}$, derived in Supplementary Note 4. Explicitly,

$$a_{\text{Taylor},m}^{(M)} = (-1)^m \frac{(2M+1)!!}{2^M [(2m+1)m!(M-m)!]}. \quad (11)$$

In the same supplementary note we show that $a_{\text{Taylor},m}^{(M)}$ coincide with the coefficients obtained from Richardson ZNE, by assuming that noise scales linearly with respect to m . Nevertheless, it is worth stressing that a distinctive characteristic of our approach is the pulse-based inverse \mathcal{K}_i . As proven in Supplementary Note 4, for gates that satisfy $\mathcal{U}^2 = \mathcal{I}$, using the circuit-based inverse $\mathcal{K}_i = \mathcal{K}$ introduces an additional error term that afflicts $\mathcal{U}_{\text{KIK}}^{(M)}$ (cf. Eq. (9)) for any mitigation order M . Thus, ignoring the pulse inverse

hinders QEM performance in paradigmatic gates such as the CNOT, swap, or Toffoli gate.

As a final remark, we note that circuit folding does not explicitly distinguish between noise amplification using powers of $\mathcal{K}_i\mathcal{K}$ or $\mathcal{K}\mathcal{K}_i$, as both choices reproduce the identity operation in the absence of noise. However, we show in Supplementary Note 3 that a correct application of the KIK formula involves powers of $\mathcal{K}_i\mathcal{K}$.

QEM in the strong noise regime

In this section, we present a strategy to adapt the coefficients $a_m^{(M)}$ to the noise strength, for handling moderate or strong noise. To this end, we introduce the quantity

$$\varepsilon_{\text{L2}}^{(M)} := \int_{g(\mu)}^1 \left(\sum_{m=0}^M a_m^{(M)} \lambda^m - \lambda^{-\frac{1}{2}} \right)^2 d\lambda, \quad (12)$$

where $\mu = \text{Tr}(\rho'\rho)$, ρ is the initial state, and ρ' is the state obtained by evolving ρ with the KIK cycle $\mathcal{K}_i\mathcal{K}$.

Let us elaborate on the physical meaning of $\varepsilon_{\text{L2}}^{(M)}$. For a pure state ρ , μ is the survival probability under the evolution $\mathcal{K}_i\mathcal{K}$. Note that, in this case, $\mu = 1$ if $\mathcal{K}_i\mathcal{K} = \mathcal{I}$. The lower integration limit $g(\mu)$ in Eq. (12) is a monotonically increasing function of μ , such that $0 \leq g(\mu) \leq 1$ for $0 \leq \mu \leq 1$ and $g(\mu) = 1$ if $\mu = 1$. Therefore, $g(\mu)$ serves as a proxy for the intensity of the noise affecting the circuit $\mathcal{K}_i\mathcal{K}$. More precisely, $g(\mu)$ represents an approximation to the smallest eigenvalue of $\mathcal{K}_i\mathcal{K}$, which equals 1 in the noiseless case. As the noise becomes stronger, both the smallest eigenvalue of $\mathcal{K}_i\mathcal{K}$ and $g(\mu)$ get closer to 0, which implies that the interval $[g(\mu), 1]$ is representative of the region where all the eigenvalues of $\mathcal{K}_i\mathcal{K}$ lie. Now, letting λ denote a general eigenvalue of this operator, the eigenvalues of $(\mathcal{K}_i\mathcal{K})^{-\frac{1}{2}}$ and $\sum_{m=0}^M a_m^{(M)}(\mathcal{K}_i\mathcal{K})^m$ can be written as $\lambda^{-\frac{1}{2}}$ and $\sum_{m=0}^M a_m^{(M)}\lambda^m$, respectively. Since the integrand of Eq. (12) quantifies the deviation between these quantities, $\varepsilon_{\text{L2}}^{(M)}$ represents the total error when using Eq. (9) to approximate the KIK formula (7).

Figure 2a, b illustrates the circuits involved in our adaptive approach to error mitigation. The experimental data comprise the expectation values measured on the noisy circuits $\mathcal{K}(\mathcal{K}_i\mathcal{K})^m$, shown in Fig. 2a, and the survival probability μ (Fig. 2b). In the weak noise limit, the circuit of Fig. 2b is not necessary and the $a_m^{(M)}$ become the Taylor coefficients given in Eq. (11) (which can also be obtained by setting $g(\mu) = 1$ in the adapted coefficients).

We point out that the L2 norm used to express $\varepsilon_{\text{L2}}^{(M)}$ in Eq. (12) is not the only possibility to quantify this error. However, it allows us to greatly simplify the derivation of $a_m^{(M)}$. The adaptive aspect of our method is based on the minimization of the error $\varepsilon_{\text{L2}}^{(M)}$ with respect to these coefficients, under the condition that $\mathcal{U}_{\text{KIK}}^{(M)}$ constitutes a trace-preserving map. In this way, we obtain the ‘adapted’ mitigation coefficients $a_m^{(M)} = a_{\text{Adap},m}^{(M)}$, which depend on $g(\mu)$ by virtue of Eq. (12) (for brevity, this dependence is not explicit in the notation for the adapted coefficients but it is expressed through the subscript ‘Adap’). In particular, we obtain in Supplementary Note 4 the expressions

$$a_{\text{Adap},0}^{(1)} = 1 + \frac{1}{(1+\sqrt{g})^3} + \frac{3}{2(1+\sqrt{g})^2}, \quad (13)$$

$$a_{\text{Adap},1}^{(1)} = -\frac{5+3\sqrt{g}}{2(1+\sqrt{g})^3}, \quad (14)$$

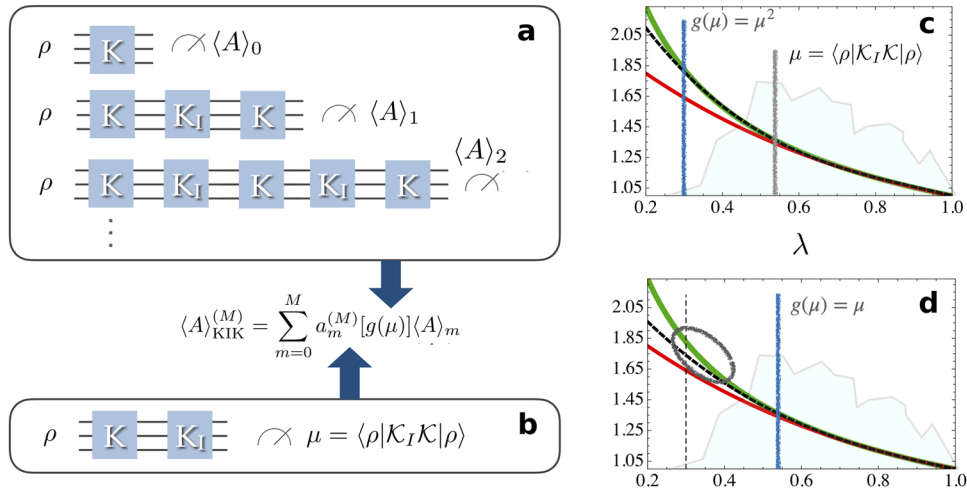


Fig. 2 Adaptive KIK error mitigation. The estimate $\langle A \rangle_{\text{KIK}}^{(M)}$ of a noiseless expectation value involves the execution of the circuits shown in (a) and (b). In particular, the survival probability μ is used to evaluate the coefficients $a_{\text{Adap},m}^{(M)}[g(\mu)]$, for adaptive error mitigation (see main text for details). The green curve in (c, d) is the plot of $\lambda^{-1/2}$ and it contains the eigenvalues of the operation that effectively suppresses the error channel $((\mathcal{K}_I \mathcal{K})^{-1/2}$ in Eq. (7)). The black dashed curves represent the polynomial approximations $\sum_{m=0}^M a_m^{(M)} \lambda^m$ that appear in the integrand of (12), for third-order mitigation ($M=3$). The better these approximations, the more accurate the corresponding error mitigation. This accuracy is related to the argument $g(\mu)$ in the optimal coefficients $a_m^{(3)} = a_{\text{Adap},m}^{(3)}[g(\mu)]$, which are obtained by minimizing (12) over the interval $[g(\mu), 1]$. Figures (c) and (d) correspond to $g(\mu) = \mu^2$ and $g(\mu) = \mu$, respectively. In (c), $\lambda^{-1/2}$ is very well approximated by $\sum_{m=0}^3 a_{\text{Adap},m}^{(3)}[\mu^2] \lambda^m$ in the interval where the eigenvalues of $(\mathcal{K}_I \mathcal{K})^{-1/2}$ are distributed (jagged line in the background). In (d), the interval $[\mu, 1]$ is too small to cover the full eigenvalue distribution and thus $\sum_{m=0}^3 a_{\text{Adap},m}^{(3)}[\mu] \lambda^m$ starts to deviate significantly from the green curve, as shown by the gray ellipse. The red curve corresponds to the Taylor polynomial $\sum_{m=0}^3 a_{\text{Adap},m}^{(3)}[1] \lambda^m$ and is the less effective approximation, as seen in both (c) and (d).

for $M=1$, and

$$a_{\text{Adap},0}^{(2)} = 1 + \frac{16}{3(1+\sqrt{g})^5} - \frac{14}{3(1+\sqrt{g})^4} + \frac{4}{(1+\sqrt{g})^2}, \quad (15)$$

$$a_{\text{Adap},1}^{(2)} = -4 \frac{10 + 8\sqrt{g} + 9g + 3g^2}{3(1+\sqrt{g})^5}, \quad (16)$$

$$a_{\text{Adap},2}^{(2)} = 2 \frac{13 + 5\sqrt{g}}{3(1+\sqrt{g})^5}, \quad (17)$$

for $M=2$. The coefficients corresponding to $M=3$ are also derived in the same supplementary note.

According to our previous remarks, we can recover the limit of weak noise by setting $g(\mu) = 1$. As expected, in this limit Eqs. (13)–(17) coincide with the coefficients $a_{\text{Tay},m}^{(M)}$ in Eq. (11) (and similarly for $M=3$, see Supplementary Note 4).

An important question is how the choice of $g(\mu)$ affects the quality of our adaptive KIK scheme. We consider functions $\{g(\mu)\} = \{1, \mu, \mu^2\}$ in the ten-swap experiment presented below, and $\{g(\mu)\} = \{1, \mu, \mu^2, \mu^{2.5}\}$ for a simulation of the transverse Ising model on five qubits, in Supplementary Note 5. In both cases, we observe that $g(\mu) = 1$ is outperformed by the functions that explicitly depend on μ . This shows that the adaptive KIK method consistently produces better results, and demonstrates the usefulness of probing the noise strength through the survival probability μ . For M sufficiently large, the adaptive scheme and the Taylor scheme produce similar results. Yet, the adaptive scheme enables to achieve substantially higher accuracies using lower mitigation orders. This is of key importance in practical applications, as low-order mitigation involves fewer circuits with lower depth (cf. Eq. (9)) and is therefore more robust to noise drifts. In addition, the approximation of keeping only the first Magnus term becomes less accurate as M increases.

The function $g(\mu) = \mu^2$ yields the best error mitigation performance, both in the ten-swap experiment and in the simulation presented in Supplementary Note 5. To understand why this happens, it is instructive to consider Fig. 2c, d. These figures show plots of $\lambda^{-1/2}$ (green solid curves), which denotes a generic eigenvalue of the noise inversion operation $(\mathcal{K}_I \mathcal{K})^{-1/2}$, and the polynomial approximations involved in third-order error mitigation (cf. Eq. (12)). The polynomials with coefficients $a_{\text{Tay},m}^{(3)}$ (Taylor mitigation) and coefficients $a_{\text{Adap},m}^{(3)}$ (adaptive mitigation) correspond to the red solid and black dashed curves, respectively. The jagged line in the background depicts a possible distribution of the eigenvalues of $(\mathcal{K}_I \mathcal{K})^{-1/2}$ (the height for a given value of λ represents the density of eigenvalues close to that value). In Fig. 2c, the adapted coefficients are evaluated at $g(\mu) = \mu^2$, and the interval $[\mu^2, 1]$ approximately covers the full region where the eigenvalues of $(\mathcal{K}_I \mathcal{K})^{-1/2}$ are contained. Thus, the associated polynomial constitutes a very good approximation to the curve $\lambda^{-1/2}$, as seen in Fig. 2c. In contrast, the black curve in Fig. 2d corresponds to coefficients $a_{\text{Adap},m}^{(3)}$ evaluated at $g(\mu) = \mu$, which leads to a poor approximation outside the interval $[\mu, 1]$ (area enclosed by the gray ellipse). This behavior sheds light on the advantage provided by $g(\mu) = \mu^2$ in our experiments and simulations. Note also that all the polynomials converge as λ tends to 1 but the Taylor polynomial (red curve) substantially separates from $\lambda^{-1/2}$ for small λ .

It is important to remark that Eq. (12) represents a measure of the distance between the polynomial (9) and the KIK formula (7), in terms of the L2 norm. In this expression, we assume that the eigenvalues λ of $\mathcal{K}_I \mathcal{K}$ are uniformly distributed across the integration interval. This is a conservative approach, given that no information besides μ is available, and in this sense it is also agnostic to the specific noise structure of $\mathcal{K}_I \mathcal{K}$. However, the

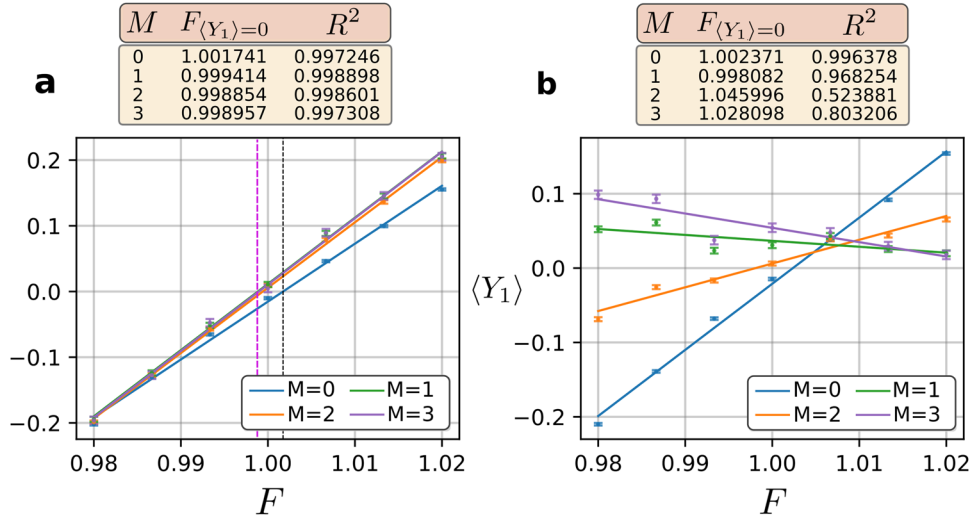


Fig. 3 Calibration curve of the pulse amplitude of a CNOT gate in the IBM processor Jakarta, using the KIK method. In (a) and (b) \mathcal{K}_1 is given by the pulse inverse and the circuit inverse $\mathcal{K}_1 = \mathcal{K}$, respectively. The initial state is $\rho = |\psi\rangle\langle\psi|$, with $|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \otimes |0\rangle$. The default amplitude is increased by the factors F shown in the x axis of the figure, and for each factor we apply Eq. (9) to evaluate the expectation value $\langle Y_1 \rangle$, where Y_1 is the y-Pauli matrix acting on the target qubit. The factor $F_{\langle Y_1 \rangle=0}$ corresponds to the ideal expectation value $\langle Y_1 \rangle = 0$ and yields the calibrated amplitude. The factors $F_{\langle Y_1 \rangle=0}$ associated with the magenta and black dashed lines are different, which indicates a shift in the amplitude obtained without KIK calibration. In (b), we see that the convergence achieved for increasing M in (a) is spoiled by the use of the circuit inverse.

evaluation of the distance $\varepsilon_{L_2}^{(M)}$ could benefit from additional knowledge about the eigenvalue distribution, which can be incorporated through a weight function $w(\lambda) \neq 1$ in the integrand of Eq. (12).

We leave the study of experimental criteria for choosing $g(\mu)$ and the potential improvements that this possibility entails for the KIK method for future work. For example, by considering higher order moments such as $\mu_2 := \langle \rho | (\mathcal{K}_1 \mathcal{K})^2 | \rho \rangle$ it is possible to devise more systematic choices of $g(\mu)$, e.g. $g(\mu) = \mu - \sqrt{\mu_2 - \mu^2}$. Yet, in the studied examples we observed no significant advantage over the simple heuristic choice $g(\mu) = \mu^2$. As for other modifications and improvements, one could also explore the use of norms other than the L2 norm employed in Eq. (12). Furthermore, the approximating polynomial can be determined in a non-integral manner. For example, by using Lagrange polynomials or a two-point Taylor expansion⁴³.

Finally, we remark that, apart from the circuits $\mathcal{K}(\mathcal{K}_1 \mathcal{K})^m$, used for the error mitigation itself, the estimation of μ only involves the additional circuit $\mathcal{K}_1 \mathcal{K}$. Therefore, our adaptive strategy is not based on any tomographic procedure or noise learning stage. Since μ is a survival probability, its variance is given by $\mu(1 - \mu)$ and has the maximum value 0.25, irrespective of the size of the system. This allows for a scalable evaluation of the coefficients for adaptive KIK mitigation. Once these coefficients are determined, the next step is the estimation of the noise-free expectation value using Eq. (10). In the section ‘Fundamental limits and measurement cost of KIK error mitigation’, we will present the corresponding measurement cost, for $1 \leq M \leq 3$, and discuss why and in what sense the KIK method is scalable.

Experimental results

In the experiments described below, the KIK mitigation of noise on the target evolution \mathcal{K} is complemented by an independent mitigation of readout errors and a simple protocol for mitigating the coherent preparation error of the initial state $\rho = |00\rangle\langle 00|$ ⁴⁴. The results of the section ‘Quantum error mitigation in a ten-swap circuit’ also include the application of randomized compiling³² to the evolutions \mathcal{K} and \mathcal{K}_1 , where circuits logically equivalent to the

corresponding ideal evolutions are randomly implemented. This is useful for turning coherent errors into incoherent noise, which can be addressed by our method. Details concerning these experimental methods can be found in Supplementary Note 6.

KIK-based gate calibration for mitigating coherent errors. A usual approach to handle coherent errors in QEM is to first transform them into incoherent errors via randomized compiling³², and then apply QEM. In this section, we discuss the application of the KIK formula to directly mitigate the coherent errors caused by a faulty calibration of a CNOT gate.

The calibration process involves measurements and adjustments of gate parameters to achieve the results that these measurements would produce in the absence of noise. Since noise affects measured expectation values, the resulting bias leads to incorrect adjustments, i.e. miscalibration. This ‘noise-induced coherent error’ effect may be small in each gate but it builds up to a substantial error in sufficiently deep circuits. Our idea is to complement the KIK error mitigation for a whole circuit, with a KIK-based calibration of the individual gates.

Figure 3 shows the results of our calibration test of a CNOT in the IBM processor Jakarta. We apply the gate on the initial state $\rho = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \otimes |0\rangle$, and measure the expectation value of the Pauli matrix Y acting on the target qubit (i.e. the qubit prepared in the state $|0\rangle$), denoted by Y_1 . We repeat this procedure for different amplitudes of the cross-resonance pulse⁴⁵, which constitutes the two-qubit interaction in the IBM CNOT implementation. Experimental details can be found in Supplementary Note 6. Each data point of Fig. 3 is obtained by applying Taylor mitigation (i.e. by applying Eq. (10) with the coefficients (11)), for $0 \leq M \leq 3$, and linear regression (least squares) is used to determine the line that best fits the experimental data. We also verify that in this case error mitigation with the adapted coefficients $a_{\text{Adap},m}^{(M)}$ does not yield a noticeable advantage. This indicates that noise is sufficiently weak, which is further supported by the quick convergence of the lines corresponding to $M \geq 1$ in Fig. 3a.

Keeping in mind that the calibrated amplitude must reproduce the ideal expectation value $\langle Y_1 \rangle = 0$, we can see from Fig. 3a that

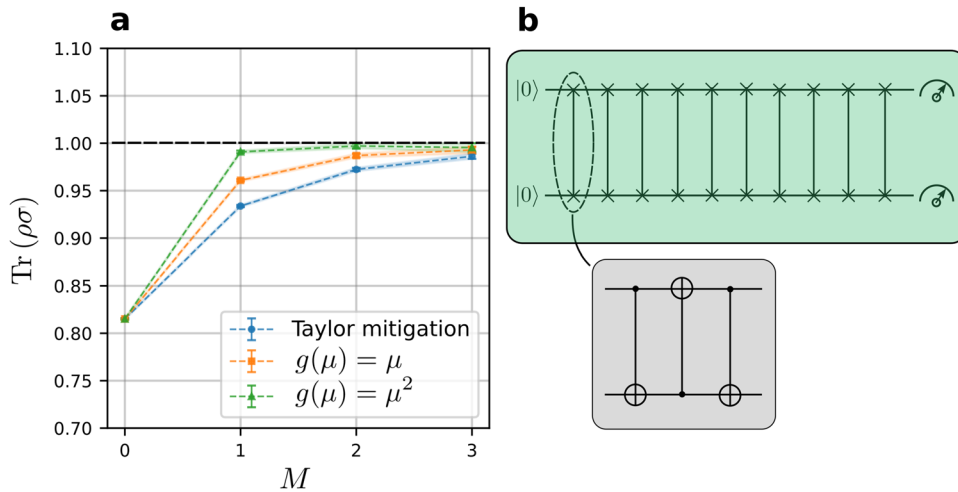


Fig. 4 Experimental QEM in the IBM processor Quito. **a** Error-mitigated survival probability for the circuit of **(b)**, as a function of the mitigation order. The ideal survival probability is 1 (dashed black line). Green and orange curves show QEM adapted to the noise intensity, and the blue curve stands for mitigation assuming weak noise (Taylor mitigation). The thickness of the lines stands for the experimental error bars. We see that Taylor mitigation is outperformed by adapted mitigation. **b** The circuit used in the experiments. Each swap is implemented as a sequence of three CNOTs.

the predicted amplitude without QEM ($M=0$) and with QEM are different. Since the CNOT is subjected to stochastic noise, without QEM the measured expectation values will be shifted and the corresponding linear regression results in a calibrated amplitude that is also shifted with respect to the correct value. This is illustrated by the separation between the black and magenta dashed lines in Fig. 3a. The magenta line represents the calibrated amplitude using KIK error mitigation, while the black one is the amplitude obtained without noise mitigation. Calibration based on the black line leads to a noise-induced coherent error. It is important to stress that the benefit of this calibration procedure would manifest when combined with QEM of the target circuit in which the CNOTs participate. The reason is that the calibrated field is consistent with gates of reduced (stochastic) noise (due to the use of QEM in the calibration process), and therefore it is not useful if the target circuit is implemented without QEM.

In Fig. 3 we also observe that a proper implementation of KIK QEM requires the pulse-based inverse \mathcal{K}_1 (Fig. 3a), performed through the driving (5), while the use of another CNOT for \mathcal{K}_1 (Fig. 3b) does not show the expected convergence as the mitigation order M increases. Note also that although a CNOT is its own inverse in the noiseless scenario, it leads to a coefficient of determination R^2 whose values show a poor linear fit. This further illustrates the importance of using the pulse inverse instead of the circuit inverse, characteristic of ZNE based on global folding. We point out that odd powers of the CNOT gate are a common choice for the application of local folding ZNE^{34,46,47}, where the goal is to amplify the noise on local sectors of the circuit rather than globally. As such, we believe that in practice this procedure would display inconsistencies similar to those observed in our CNOT experiment. More generally, we show in Supplementary Note 4 that foldings of any self-inverse gate with itself produce a residual error that is not present when the pulse inverse is applied.

Quantum error mitigation in a ten-swap circuit. In Fig. 4a, we show the results of QEM for a circuit \mathcal{K} given by a sequence of 10 swap gates. The experiments were executed in the IBM quantum processor Quito. The schematic of \mathcal{K} is illustrated in Fig. 4b.

We mitigate errors in the survival probability $\text{Tr}(\rho\sigma)$, where σ is the noisy final state that results from applying \mathcal{K} to ρ . To perform QEM, we consider the truncated expansion (9) with mitigation orders $1 \leq M \leq 3$. The blue curve in Fig. 4a corresponds to Taylor

mitigation $a_m^{(M)} = a_{\text{Taylor},m}^{(M)}$. Coefficients $a_m^{(M)} = a_{\text{Adap},m}^{(M)}$ that are adapted with functions $g(\mu) = \mu$ and $g(\mu) = \mu^2$ in Eq. (12) give rise to the orange and green curves, respectively. Furthermore, for \mathcal{K}_1 we perform the pulse inverse according to the pulse schedule described by Eq. (5).

In Fig. 4a we observe that the adapted coefficients $a_{\text{Adap},m}^{(M)}$ outperform Taylor mitigation. This shows that, beyond the limit of weak noise, QEM can be substantially improved by adapting it to the noise intensity. Within our Magnus truncation approximation, we observe that the ideal survival probability is almost fully recovered. The small residual bias is of order 10^{-3} and can be associated with small experimental imperfections (e.g. small errors in the detector calibration), or with the higher-order Magnus terms discarded in our framework. In Supplementary Note 7, we also provide a numerical example where neglecting higher-order Magnus terms leads to an eventual saturation of the QEM accuracy. However, in this example, we find that fourth-order QEM ($M=4$) yields a relative error as low as 10^{-4} , which further illustrates the accuracy achieved by the KIK formula.

Due to experimental limitations, it was not possible to implement the ten-swap circuit using CNOTs calibrated through the KIK method. Specifically, we could not guarantee that calibration circuits and error mitigation circuits would run sequentially, and without the interference of intrinsic (noncontrollable) calibrations of the processor. Moreover, this demonstration requires that all the parameters of the gate are calibrated using the KIK method, and not just the cross resonance amplitude. However, we numerically verify in Supplementary Note 6 that coherent errors vanish for a gate calibrated using KIK QEM, to the point that randomized compiling is no longer needed.

Fundamental limits and measurement cost of KIK error mitigation

Fundamental limits of KIK error mitigation. The performance of QEM protocols is often analyzed using two figures of merit. One of them is the bias between the noisy expectation value of an observable and its ideal counterpart, and the other is the statistical precision of the error-mitigated expectation value. The bias defines the QEM accuracy and is evaluated in the limit of infinite measurements. However, any experiment has a limited precision because it always involves a finite number of samples. In QEM

protocols, the estimation of ideal expectation values is usually accompanied by an increment of statistical uncertainty, which can be exponential in worst-case scenarios^{36,37}. This results in a sampling overhead for achieving a given precision, as compared to the number of samples required without using QEM.

In Supplementary Note 8, we derive the accuracy bounds

$$\epsilon_{\text{KIK}}^{(M)} \leq \sqrt{\text{Tr}(A^2) - \frac{[\text{Tr}(A)]^2}{\text{Tr}(I)}} \left| 1 - \sum_{m=0}^M a_{\text{Adap},m}^{(M)}(\mu) e^{-2(m+1/2) \int_0^T \|\mathcal{L}(t)\| dt} \right|, \text{ for } M = 1, 2, 3, \quad (18)$$

$$\leq \sqrt{\text{Tr}(A^2) - \frac{[\text{Tr}(A)]^2}{\text{Tr}(I)}} \left| 1 - \sum_{m=0}^M a_{\text{Adap},m}^{(M)}(1) e^{-2(m+1/2) \int_0^T \|\mathcal{L}(t)\| dt} \right|, \text{ for } M = 1, 2, 3, \quad (19)$$

$$\leq \frac{(2M+1)!!}{2^{M+1}(M+1)!} \sqrt{\text{Tr}(A^2) - \frac{[\text{Tr}(A)]^2}{\text{Tr}(I)}} (e^{2 \int_0^T \|\mathcal{L}(t)\| dt} - 1)^{M+1}. \quad (20)$$

These are upper bounds on the bias $\epsilon_{\text{KIK}}^{(M)}$, for an arbitrary observable A and an arbitrary initial state. We also note that the only approximation in Eqs. (18)–(20) and any of our derivations is the truncation of the Magnus expansion to its dominant term. Importantly, this does not exclude errors of moderate or strong magnitude associated with such a term. On the other hand, discarding Magnus terms beyond first order naturally leads to a saturation of accuracy. Such a saturation manifests in a residual bias that cannot be reduced by indefinitely increasing the mitigation order. Therefore, for the tighter bounds (18) and (19) we restrict ourselves to the mitigation orders used in our experiments and simulations, given by $1 \leq M \leq 3$.

On the other hand, the loosest bound (20) provides a clearer picture of how the bias associated with the first Magnus term is suppressed by increasing M . The quantity $\int_0^T \|\mathcal{L}(t)\| dt$ is the integral of the spectral norm of the dissipator $\|\mathcal{L}(t)\|$, over the total evolution time $(0, T)$. This parameter serves as a quantifier of the noise accumulated during the execution of the target evolution \mathcal{K} . Since $\frac{(2M+1)!!}{2^{M+1}(M+1)!} \leq \frac{3}{8}$, Eq. (20) implies that $\epsilon_{\text{KIK}}^{(M)}$ is exponentially suppressed if the accumulated noise is such that

$$e^{2 \int_0^T \|\mathcal{L}(t)\| dt} < 2. \quad (21)$$

In the case of noise acting locally on individual gates, $\mathcal{L}(t)$ is given by a sum of local dissipators and one can show that $\|\mathcal{L}(t)\|$ is upper bounded by the summation of all the gate errors in the circuit.

We remark that, in the NISQ era, errors escalate in quantum algorithms due to the lack of QEC. Thus, NISQ computers can perform useful computations only if the accumulated noise $\int_0^T \|\mathcal{L}(t)\| dt$ is below a certain value. Our notion of scalability is that under the constraint of moderate accumulated noise the KIK method is scale independent. In particular, when $\int_0^T \|\mathcal{L}(t)\| dt$ is sufficiently small to satisfy Eq. (21), the exponential error mitigation referred to above is applicable to circuits of any size and topology. While achieving a low accumulated noise in big circuits is technologically challenging, if this condition is met the KIK method and the resources that it requires are agnostic to the size of the circuit. Moreover, it is worth noting that Eq. (21) represents a sufficient condition for scalable error mitigation. The possibility of extending this scalability to values of $\int_0^T \|\mathcal{L}(t)\| dt$ that violate Eq. (21) depends on the tightness of the accuracy bounds (18)–(20), and constitutes an open problem.

Equations (18)–(20) are applicable to both adaptive mitigation and Taylor mitigation. In contrast, the tightest bound (18) is exclusive of adaptive mitigation. The coefficients $a_{\text{Adap},m}^{(M)}$ in this bound are evaluated at $g(\mu) = \mu$. Importantly, (18) is upper bounded by (19) and (20) for any $0 \leq \mu \leq 1$, as proven in Supplementary Note 8. According to our experiments and

simulations, we believe that even tighter bounds can be obtained for $g(\mu) = \mu^2$ or other choices of $g(\mu)$. This topic is left for future investigation.

Lastly, we stress that condition (21) does not imply that the KIK method is restricted to error mitigation for weak noise. This is related to the reiterated fact that Eqs. (18)–(20) and particularly (20) probably overestimate the actual bias between the error-mitigated expectation value and its ideal counterpart. More importantly, we have shown experimentally and numerically the substantial advantage achieved by the adaptive KIK strategy, as compared to QEM under the assumption of weak noise. This further indicates that the regime of validity of our method likely goes beyond the prediction of Eq. (20).

Measurement cost of KIK error mitigation. For the sampling overhead, we adopt the variance as the measure of statistical precision. Let $\text{Var}_0(A)$ denote the variance in the estimation of the expectation value $\langle A \rangle$, without using error mitigation, and $\text{Var}_M(A)$ the variance associated with KIK mitigation of order $M \geq 1$. The sampling overhead is defined as the increment in the number of samples needed to achieve the same precision as in the unmitigated case. Suppose that N measurements constitute the shot budget for KIK mitigation. For a given value of M , the sampling overhead is evaluated by minimizing $\text{Var}_M(A)$ over the distribution of measurements between the different circuits $\mathcal{K}(\mathcal{K}_i \mathcal{K})^m$. If N_m measurements are allocated to $\mathcal{K}(\mathcal{K}_i \mathcal{K})^m$, then

$$\text{Var}_M(A) = \sum_{m=0}^M \left(a_m^{(M)}(g) \right)^2 \frac{\text{var}_m(A)}{N_m}, \quad (22)$$

where $\text{var}_m(A)$ denotes the variance that results from measuring A on the circuit $\mathcal{K}(\mathcal{K}_i \mathcal{K})^m$.

Taking into account the constraint $\sum_{m=0}^M N_m = N$, the minimization of Eq. (22) with respect to $\{N_m\}_m$ yields $N_m = |a_m^{(M)}| N$. Of course, these values have to be approximated to the closest integer in practice. Now, we assume that $\text{var}_m(A) = \text{var}_n(A)$ for all $0 \leq m, n \leq M$. Since, for reasons previously discussed, we are interested in low mitigation orders $1 \leq M \leq 3$, $(\mathcal{K}_i \mathcal{K})^m$ does not deviate too much from the identity operation and therefore the assumption stated above is reasonable. In this way, replacing $N_m = |a_m^{(M)}| N$ into Eq. (22) yields

$$\text{Var}_M(A) = \sum_{m=0}^M |a_m^{(M)}(g)| \frac{\text{var}_0(A)}{N}. \quad (23)$$

The quantity $\frac{\text{var}_0(A)}{N}$ is the variance obtained without using error mitigation. Accordingly,

$$\gamma_M(g) = \sum_{m=0}^M |a_m^{(M)}(g)| \quad (24)$$

represents the sampling overhead. In Fig. 5, we show the sampling overheads for $1 \leq M \leq 3$, as a function of $g = g(\mu)$. As expected, larger noise strengths (corresponding to smaller values of g) lead to larger values of $\gamma_M(g)$. However, as shown in Fig. 5, these sampling overheads are quite moderate and do not represent an obstacle for scalable error mitigation. In addition, we show in Supplementary Note 3 that our method is robust to noise drifts and miscalibrations that may result from larger sampling overheads, e.g. when higher mitigation orders ($M \geq 4$) are considered.

DISCUSSION

Quantum error mitigation (QEM) is becoming a standard practice in NISQ experiments. However, QEM methods that are free from intrinsic scalability issues lack a physically rigorous formulation, or are unable to cope with significant levels of noise. The KIK method allows for scalable QEM whenever the noise accumulated in the target circuit is not too high, as implied by our upper bounds on

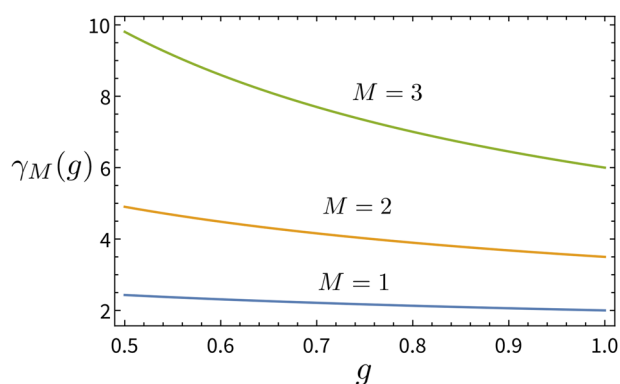


Fig. 5 Variance overhead $\gamma_M(g)$ of KIK error mitigation, for mitigation orders $1 \leq M \leq 3$. The graph shows $\gamma_M(g)$ in terms of the function $g = g(\mu)$, used to evaluate the adaptive coefficients $a_m^{(M)}(g)$ (cf. Eq. (24)). The overheads for the Taylor coefficients (low noise limit) correspond to the values at $g = 1$.

the QEM accuracy (cf. Eqs. (18)–(20)). This QEM technique is based on a master equation analysis that incorporates time-dependent and spatially correlated noise, and does not require that the noise is trace-preserving. As such, based on elementary simulations we observe that it can also mitigate leakage noise, which can take place in superconducting circuits. In the limit of weak noise, the KIK method reproduces some features of zero noise extrapolation using circuit unitary folding, and outperforms it. This is achieved thanks to the use of pulse-based inverses for the implementation of QEM circuits, and the adaptation of QEM parameters to the noise intensity for handling moderate and strong noise.

The shot overhead of our method depends only on the noise level and not on the size of the target circuit. For moderate noise, the sampling overhead for mitigation order three or lower is smaller than ten. While the KIK method can be adapted to the strength of the noise, this only requires measuring a single experimental parameter whose sampling cost is negligible and independent of the size of the system. Usually, the performance of QEM techniques may be compromised in experiments involving a large number of samples. When considering long runs, the system needs to be recalibrated multiple times, and noise parameters can undergo significant drifts. This poses challenges in the context of noise learning for QEM protocols that rely on this approach. We show in Supplementary Note 3 that our approach is resilient to drifts in the noise and calibration parameters (the latter holds if randomized compiling is applied). This enables it to be applied in calculations over runtimes of days or even weeks, including pauses for calibrations, maintenance, or execution of supporting jobs. On a similar basis, it is possible to parallelize the error mitigation task, by averaging over data collected from different quantum processors or platforms with spatially differentiated noise profiles (see Supplementary Note 3).

We have demonstrated our findings using the IBM quantum processors Quito and Jakarta. In Quito, we implemented KIK error mitigation in a circuit composed of 10 swap gates (30 sequential CNOTs). Despite the substantial noise in this setup, the tiny bias between the error-mitigated expectation value and the ideal result demonstrates that, at least in this experiment, our theoretical approximations are quite consistent with the actual noise in the system. Using the processor Jakarta, we also showed that even the calibration of a basic building block of quantum computing, such as the CNOT gate, can be affected by unmitigated noise. As a consequence, calibrated gate parameters feature erroneous values leading to coherent errors. These errors can be avoided by incorporating the KIK method in the calibration process. The integration of randomized compiling into our technique also enables the mitigation of coherent errors in the

CNOT gates. This is possible because randomized compiling transforms coherent errors into incoherent noise, which can be addressed by the KIK method.

Despite these successful demonstrations, we believe that there is room for improvement by exploring some of the possibilities mentioned in the Section ‘QEM in the strong noise regime’. We also hope that the performance shown here can be exploited for new demonstrations of quantum algorithms on NISQ devices, with the potential of achieving quantum advantage in applications of interest.

DATA AVAILABILITY

Code employed in the execution of the experiments as well as raw experimental data and data underlying figures is hosted at <https://doi.org/10.5281/zenodo.7652322>. All additional data are provided in the [supplementary information](#).

Received: 5 April 2023; Accepted: 27 October 2023;

Published online: 22 November 2023

REFERENCES

- Arute, F. et al. Quantum supremacy using a programmable superconducting processor. *Nature* **574**, 505–510 (2019).
- Zhong, H.-S. et al. Quantum computational advantage using photons. *Science* **370**, 1460–1463 (2020).
- Madsen, L. S. et al. Quantum computational advantage with a programmable photonic processor. *Nature* **606**, 75–81 (2022).
- Fowler, A. G., Mariantoni, M., Martinis, J. M. & Cleland, A. N. Surface codes: towards practical large-scale quantum computation. *Phys. Rev. A* **86**, 032324 (2012).
- O Gorman, J. & Campbell, E. T. Quantum computation with realistic magic-state factories. *Phys. Rev. A* **95**, 032338 (2017).
- Temme, K., Bravyi, S. & Gambetta, J. M. Error mitigation for short-depth quantum circuits. *Phys. Rev. Lett.* **119**, 180509 (2017).
- Li, Y. & Benjamin, S. C. Efficient variational quantum simulator incorporating active error minimization. *Phys. Rev. X* **7**, 021050 (2017).
- Endo, S., Benjamin, S. C. & Li, Y. Practical quantum error mitigation for near-future applications. *Phys. Rev. X* **8**, 031027 (2018).
- Strikis, A., Qin, D., Chen, Y., Benjamin, S. C. & Li, Y. Learning-based quantum error mitigation. *PRX Quantum* **2**, 040330 (2021).
- Czarnik, P., Arrasmith, A., Coles, P. J. & Cincio, L. Error mitigation with Clifford quantum-circuit data. *Quantum* **5**, 592 (2021).
- Koczor, B. Exponential error suppression for near-term quantum devices. *Phys. Rev. X* **11**, 031057 (2021).
- Huggins, W. J. et al. Virtual distillation for quantum error mitigation. *Phys. Rev. X* **11**, 041036 (2021).
- Giurgica-Tiron, T., Hindy, Y., LaRose, R., Mari, A. & Zeng, W. J. Digital zero noise extrapolation for quantum error mitigation. In: *2020 IEEE International Conference on Quantum Computing and Engineering (QCE)* 306–316 (IEEE, 2020).
- Cai, Z. Quantum error mitigation using symmetry expansion. *Quantum* **5**, 548 (2021).
- Mari, A., Shammah, N. & Zeng, W. J. Extending quantum probabilistic error cancellation by noise scaling. *Phys. Rev. A* **104**, 052607 (2021).
- Lowe, A. et al. Unified approach to data-driven quantum error mitigation. *Phys. Rev. Res.* **3**, 033098 (2021).
- Nation, P. D., Kang, H., Sundaresan, N. & Gambetta, J. M. Scalable mitigation of measurement errors on quantum computers. *PRX Quantum* **2**, 040326 (2021).
- Bravyi, S., Sheldon, S., Kandala, A., McKay, D. C. & Gambetta, J. M. Mitigating measurement errors in multiqubit experiments. *Phys. Rev. A* **103**, 042605 (2021).
- Kim, Y. et al. Scalable error mitigation for noisy quantum circuits produces competitive expectation values. *Nat. Phys.* **19**, 752–759 (2023).
- Van Den Berg, E., Mineev, Z. K., Kandala, A. & Temme, K. Probabilistic error cancellation with sparse pauli-Lindblad models on noisy quantum processors. *Nat. Phys.* **19**, 1116–1121 (2023).
- Ferracin, S. et al. Efficiently improving the performance of noisy quantum computers. Preprint at <https://arxiv.org/abs/2201.10672> (2022).
- Endo, S., Cai, Z., Benjamin, S. C. & Yuan, X. Hybrid quantum-classical algorithms and quantum error mitigation. *J. Phys. Soc. Jpn.* **90**, 032001 (2021).
- Cai, Z. et al. Quantum error mitigation. Preprint at <https://arxiv.org/abs/2210.00921v2> (2022).
- Kandala, A. et al. Error mitigation extends the computational reach of a noisy quantum processor. *Nature* **567**, 491–495 (2019).
- Song, C. et al. Quantum computation with universal error mitigation on a superconducting quantum processor. *Sci. Adv.* **5**, eaaw5686 (2019).

26. Quantum, G. A. et al. Hartree-Fock on a superconducting qubit quantum computer. *Science* **369**, 1084–1089 (2020).
27. Urbaneck, M. et al. Mitigating depolarizing noise on quantum computers with noise-estimation circuits. *Phys. Rev. Lett.* **127**, 270502 (2021).
28. Kim, Y. et al. Evidence for the utility of quantum computing before fault tolerance. *Nature* **618**, 500–505 (2023).
29. Shtanko, O. et al. Uncovering local integrability in quantum many-body dynamics. Preprint at <https://arxiv.org/abs/2307.07552> (2023).
30. Zhang, S. et al. Error-mitigated quantum gates exceeding physical fidelities in a trapped-ion system. *Nat. Commun.* **11**, 587 (2020).
31. Sagastizabal, R. et al. Experimental error mitigation via symmetry verification in a variational quantum eigensolver. *Phys. Rev. A* **100**, 010302 (2019).
32. Wallman, J. J. & Emerson, J. Noise tailoring for scalable quantum computation via randomized compiling. *Phys. Rev. A* **94**, 052325 (2016).
33. Hashim, A. et al. Randomized compiling for scalable quantum computing on a noisy superconducting quantum processor. *Phys. Rev. X* **11**, 041039 (2021).
34. Majumdar, R., Rivero, P., Metz, F., Hasan, A. & Wang, D. S. Best practices for quantum error mitigation with digital zero-noise extrapolation. Preprint at <https://arxiv.org/abs/2307.05203> (2023).
35. Czarnik, P., McKerns, M., Sornborger, A. T. & Cincio, L. Improving the efficiency of learning-based error mitigation. Preprint at <https://arxiv.org/abs/2204.07109> (2022).
36. Takagi, R., Endo, S., Minagawa, S. & Gu, M. Fundamental limits of quantum error mitigation. *npj Quantum Inf.* **8**, 114 (2022).
37. Quek, Y., França, D. S., Khatri, S., Meyer, J. J. & Eisert, J. Exponentially tighter bounds on limitations of quantum error mitigation. Preprint at <https://arxiv.org/abs/2210.11505> (2022).
38. Trotter, H. F. On the product of semi-groups of operators. *Proc. Am. Math. Soc.* **10**, 545–551 (1959).
39. Gyamfi, J. A. Fundamentals of quantum mechanics in Liouville space. *Eur. J. Phys.* **41**, 063002 (2020).
40. Blanes, S., Casas, F., Oteo, J.-A. & Ros, J. The Magnus expansion and some of its applications. *Phys. Rep.* **470**, 151–238 (2009).
41. Breuer, H.-P. & Petruccione, F. *The Theory of Open Quantum Systems* (Oxford University Press, 2002).
42. Sun, J. et al. Mitigating realistic noise in practical noisy intermediate-scale quantum devices. *Phys. Rev. Appl.* **15**, 034026 (2021).
43. López, J. L. & Temme, N. M. Two-point Taylor expansions of analytic functions. *Stud. Appl. Math.* **109**, 297–311 (2002).
44. Landa, H., Meiroum, D., Kanazawa, N., Fitzpatrick, M. & Wood, C. J. Experimental Bayesian estimation of quantum state preparation, measurement, and gate errors in multiqubit devices. *Phys. Rev. Res.* **4**, 013199 (2022).
45. Alexander, T. et al. Qiskit pulse: programming quantum computers through the cloud with pulses. *Quantum Sci. Technol.* **5**, 044006 (2020).
46. He, A., Nachman, B., de Jong, W. A. & Bauer, C. W. Zero-noise extrapolation for quantum-gate error mitigation with identity insertions. *Phys. Rev. A* **102**, 012426 (2020).
47. Pascuzzi, V. R., He, A., Bauer, C. W., de Jong, W. A. & Nachman, B. Computationally efficient zero-noise extrapolation for quantum-gate-error mitigation. *Phys. Rev. A* **105**, 042406 (2022).

ACKNOWLEDGEMENTS

The authors acknowledge the use of IBM Quantum services for this work. The views expressed are those of the authors, and do not reflect the official policy or position of IBM or the IBM Quantum team. Raam Uzdin is grateful for support from the Israel Science Foundation (Grant No. 2556/20).

AUTHOR CONTRIBUTIONS

R.U. conceived the method, set the theoretical framework, including most of the analytical derivations, and performed numerical simulations. J.P.S. designed and executed the experiments, and performed numerical simulations. I.H. derived some theoretical results, in particular the performance bounds. All the authors were involved in the analysis of theoretical and experimental results, and in the writing and presentation of the paper.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41534-023-00785-7>.

Correspondence and requests for materials should be addressed to Raam Uzdin.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023