

Adaptive quantum error mitigation using pulse-based inverse evolutions

Ivan Henao^{1,*}, Jader P. Santos^{1,†} and Raam Uzdin^{1‡}
¹*Fritz Haber Research Center for Molecular Dynamics, Institute of Chemistry,
The Hebrew University of Jerusalem, Jerusalem 9190401, Israel*

CONTENTS

Supplementary Note 1: Quantum mechanics in Liouville space	1
Supplementary Note 2: Dynamical description of noise for the target evolution and its inverse	2
Noise for pulse-based inverse evolution	3
Supplementary Note 3: Derivation of the KIK formula	4
Relation between $\mathcal{K}_I\mathcal{K}$ and $\mathcal{K}\mathcal{K}_I$ in the KIK formula	6
Robustness of KIK error mitigation to noise drifts and spatially varying noise profiles	7
Supplementary Note 4: Implementations of the inverse of the noise channel $\mathcal{N}_{\text{KIK}}$	11
Weak noise limit and Taylor approximation	11
Richardson ZNE using Circuit folding and linear scaling of noise	11
Error induced by the circuit inverse $\mathcal{K}_I = \mathcal{K}$ in circuits that satisfy $\mathcal{U}^2 = \mathcal{I}$	12
Coefficients for Adaptive QEM based on the KIK formula, for mitigation orders $M = 1, 2, 3$	14
Supplementary Note 5: KIK QEM applied to the transverse Ising model	15
Error mitigation	16
Supplementary Note 6: Description of the experimental procedures	18
Integration with complementary error mitigation methods	18
Statistical analysis of experimental data	20
CNOT calibration experiment	21
Ten-swap experiment	21
Simulation of a noisy calibration of a CNOT gate, using the KIK method	23
Supplementary Note 7: Numerical example to illustrate saturation of the KIK formula	26
Supplementary Note 8: Derivation of upper bounds for the performance of the KIK method	27
Lower bound on the smallest eigenvalue of $\mathcal{K}_I\mathcal{K}$	28
Sufficient condition for $\mathcal{K}_I = \mathcal{K}^\dagger$	29
First error bounds for Adaptive mitigation and Taylor mitigation	30
Alternative bound for Taylor mitigation	32
Traceless observables and second (tighter) error bounds for Adaptive error mitigation and Taylor error mitigation	34
Summary	35
References	36

SUPPLEMENTARY NOTE 1: QUANTUM MECHANICS IN LIOUVILLE SPACE

In the standard description of Quantum Mechanics, a system of dimension d is represented by a density matrix ρ of dimension $d \times d$. Moreover, a CPTP (completely positive and trace preserving) quantum operation can be expressed as

$$\rho' = \sum_i K_i \rho K_i^\dagger, \quad (1)$$

where ρ' is a density matrix and $\{K_i\}$ are Kraus operators that satisfy the completeness relation $\sum_i K_i^\dagger K_i = I$, and I is the $d \times d$ identity matrix. Observables correspond to hermitian operators A , and the associated expectation value for a system in a state ρ reads

$$\langle A \rangle = \text{Tr}(A\rho). \quad (2)$$

The Liouville space formalism is an alternative formulation that is particularly useful to simplify notation and handle quantum operations. In this framework, a density matrix is replaced by a vector $|\rho\rangle$ of dimension d^2 and a quantum operation is a matrix of dimension $d^2 \times d^2$. Using the calligraphic notation $\mathcal{O}$ for a quantum operation, the analogous of Eq. (1) in Liouville space is given by

$$|\rho'\rangle = \mathcal{O}|\rho\rangle. \quad (3)$$

Here, we adopt the approach of Ref.¹, where $|\rho\rangle$ is the column vector whose first d components correspond to the first row of ρ , the next d components correspond to the second row of ρ , and so forth. More formally, the vector representation of a $d \times d$ generic matrix B (not necessarily a density matrix) is given by $|B\rangle = (B_{11}, B_{12}, \dots, B_{1d}, B_{21}, B_{22}, \dots, B_{2d}, \dots, B_{d1}, B_{d2}, \dots, B_{dd})^T$, where B_{ij} is the ij entry of B . With this convention, in Liouville space the quantum operation (1) takes the form¹

$$|\rho'\rangle = \sum_i K_i \otimes K_i^* |\rho\rangle, \quad (4)$$

where K_i^* is the element-wise complex conjugate of K_i . For example, a unitary operation $\rho' = U\rho U^\dagger$ is written as $|\rho'\rangle = \mathcal{U}|\rho\rangle$, where $\mathcal{U} = U \otimes U^*$.

Equation (4) is obtained by following the rule to vectorize a product of three matrices B , C and D . Denoting the associated vector as $|BCD\rangle$, this rule states that¹

$$|BCD\rangle = B \otimes D^T |C\rangle, \quad (5)$$

where the superscript T denotes transposition. Setting $B = K_i$, $C = \rho$, and $D = K_i^\dagger$, Eq. (4) follows by applying (5) to (1) and using the linearity property of the vectorization.

Finally, to express the expectation value (2) in Liouville space one writes A as a row vector $\langle A|$ defined by $\langle A| = (A_{11}^*, A_{12}^*, \dots, A_{1d}^*, \dots, A_{d1}^*, A_{d2}^*, \dots, A_{dd}^*)$. In this way, the hermiticity of A leads to

$$\begin{aligned} \langle A|\rho \rangle &= \sum_{i,j} A_{ij}^* \rho_{ij} \\ &= \sum_{i,j} A_{ji} \rho_{ij} \\ &= \langle A \rangle. \end{aligned} \quad (6)$$

SUPPLEMENTARY NOTE 2: DYNAMICAL DESCRIPTION OF NOISE FOR THE TARGET EVOLUTION AND ITS INVERSE

In this section, we set the framework for the derivation of the KIK formula (Eq. (1) in the main text). For the sake of clarity and completeness, we will discuss again some topics addressed in the main text and rewrite a few equations that were already introduced. We consider a continuous-time description of the system evolution, modeled by the master equation

$$\frac{d}{dt}\rho = -i[H(t), \rho] + \hat{L}(t)[\rho]. \quad (7)$$

Here, $H(t)$ is a time-dependent Hamiltonian and $\hat{L}(t)$ is a time-dependent dissipator that accounts for the non-unitary contribution to the dynamics, which is induced by external noise. The hat symbol in $\hat{L}(t)$ is used to emphasize that it represents a superoperator in Hilbert space.

To specify the form of $\hat{L}(t)$ one could invoke a microscopic description of the dynamics, where the system is coupled to some external environment and the total system obeys the Schrodinger equation. The time-independent case is extensively studied in², and various time-dependent Markovian master equations have been derived³. The dissipator

is often given in the Lindblad form, which represents the most general dissipator for a Markovian and CPTP evolution. For our purposes, this is not necessary. For example, $\hat{L}(t)$ could incorporate leakage noise, which does not preserve probability and thus is not trace preserving.

Now, let us rewrite Eq. (7) in Liouville space. Using the linearity of the vectorization operation, we have that

$$\frac{d}{dt}|\rho\rangle = -i|[H(t), \rho]\rangle + |\hat{L}(t)[\rho]\rangle, \quad (8)$$

where $|[H(t), \rho]\rangle$ and $|\hat{L}(t)[\rho]\rangle$ are the vectors corresponding to $[H(t), \rho]$ and $\hat{L}(t)[\rho]$, respectively. By applying the rule (5) to the commutator $[H(t), \rho] = H(t)\rho - \rho H(t)$, we obtain

$$\begin{aligned} |[H(t), \rho]\rangle &= |H(t)\rho\rangle - |\rho H(t)\rangle \\ &= (H(t) \otimes I - I \otimes H(t)^T) |\rho\rangle \\ &:= \mathcal{H}(t)|\rho\rangle, \end{aligned} \quad (9)$$

where D is identified with the identity I for the product $H(t)\rho$, and with $H(t)$ for $\rho H(t)$. In both cases, C is associated with ρ . For a general dissipator $\hat{L}(t)$ we can simply write $|\hat{L}(t)[\rho]\rangle = \mathcal{L}(t)|\rho\rangle$, because ρ can always be associated with C in Eq. (5).

In this way, the Liouville-space representation of (7) reads

$$\frac{d}{dt}|\rho\rangle = (-i\mathcal{H}(t) + \mathcal{L}(t))|\rho\rangle. \quad (10)$$

We note that both $\mathcal{H}(t)$ and $\mathcal{L}(t)$ are linear operators that correspond to matrices of dimension $d^2 \times d^2$. While we do not impose any physical constraint on $\mathcal{L}(t)$, in what follows we introduce and physically justify a relationship between $\mathcal{L}(t)$ and the dissipator that affects the pulse inverse evolution. This relationship is crucial for the derivation of the KIK formula in Supplementary Note 3.

Noise for pulse-based inverse evolution. Suppose that applying the driving $\mathcal{H}(t)$ in Eq. (10) during a total time T leads to an ideal unitary evolution $\mathcal{U} = \mathcal{T}e^{-\int_0^T i\mathcal{H}(t)dt}$, where $\mathcal{T}$ is the time-ordering operator. Even in the circuit model of quantum computing, where unitary operations are composed of discrete quantum gates, each elementary gate is itself generated by a pulse schedule that can be represented as a time-dependent Hamiltonian. Hence, any quantum circuit is ultimately generated by some pulse schedule $\mathcal{H}(t)$.

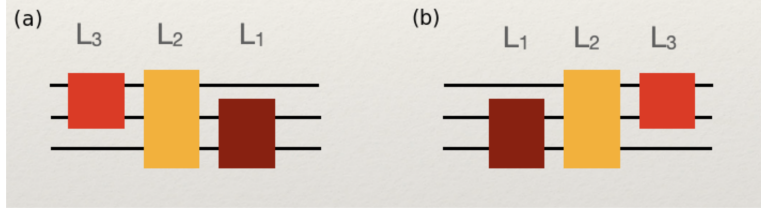
Different pulse schedules can result in the same ideal evolution $\mathcal{U}$, and naturally the same is true for its inverse $\mathcal{U}^\dagger$. However, in the presence of noise this equivalence does not hold in general. The derivation of the KIK formula relies on relating the pulse schedule for $\mathcal{U}^\dagger$ with the pulse schedule for $\mathcal{U}$ in a specific manner. Denoting the driving that generates $\mathcal{U}^\dagger$ by $\mathcal{H}_I(t)$, this relationship reads

$$\mathcal{H}_I(t) = -\mathcal{H}(T - t). \quad (11)$$

In combination with Eq. (11), the other ingredient for obtaining the KIK formula has to do with how noise comes into play for a given driving $\mathcal{H}(t)$. On the one hand, Eq. (10) describes noise that acts locally in time, i.e. that $\mathcal{L}(t)$ only depends on the current instant t and not on the previous history of the evolution. On the other hand, this time dependence can have two origins. One of them is the time-dependence of $\mathcal{H}(t)$ itself, and the other are intrinsic fluctuations of the noise that are related e.g. to changes in the environment or miscalibrations that occur during the execution of an experiment. The second possibility is discussed in detail in Supplementary Note 3. As for the influence of the driving $\mathcal{H}(t)$ on the noise, we assume that $\mathcal{L}(t)$ does not depend on the sign of $\mathcal{H}(t)$, but only on its amplitude. This reflects the fact that noise cannot be undone when running the reverse schedule described in Eq. (11). Taking into account (11), the dissipator $\mathcal{L}_I(t)$ for the “pulse inverse” $\mathcal{H}_I(t)$ should then satisfy

$$\mathcal{L}_I(t) = \mathcal{L}(T - t). \quad (12)$$

Equation (12) implies that if different gates are affected by different noise mechanisms (e.g. very fast gates may be prone to leakage noise due to non-adiabatic couplings to the levels outside the computational basis), the order in which these noise mechanisms operate is reversed when applying the pulse inverse (11), as shown in Supplementary Figure 1. This is a consequence of the time locality of both $\mathcal{L}(t)$ and $\mathcal{L}_I(t)$, and the reversed time schedule that $\mathcal{H}_I(t)$ imprints on the corresponding inverse gates. To summarize, our sole assumptions on the noise are:



Supplementary Figure 1. Time dependence of dissipators $\mathcal{L}(t)$ and $\mathcal{L}_1(t)$. (a) Illustrative noisy quantum circuit, composed of three gates that act on three qubits. The total dissipator $\mathcal{L}(t)$ is given by a sequence of dissipators $\mathcal{L}_i$, $1 \leq i \leq 3$, which affect each gate. (b) Upon reversing the drive pulse schedule (cf. Eq. (11)), the order of the gates and the corresponding dissipators is also reversed. Hence, if the sequence of dissipators in (a) is $\mathcal{L}_1\mathcal{L}_2\mathcal{L}_3$, in (b) it is given by $\mathcal{L}_3\mathcal{L}_2\mathcal{L}_1$.

1. The linearity and the time locality of the dissipator $\mathcal{L}(t)$.
2. The relationship between the dissipator $\mathcal{L}_1(t)$, which affects the pulse inverse $\mathcal{H}_1(t)$ (cf. Eq. (11)) at time t , and $\mathcal{L}(t)$. This relationship is encapsulated by Eq. (12).

We remark that $\mathcal{L}(t)$ is not restricted to have a Lindblad form or to give rise to a trace preserving map. For example, it can incorporate leakage noise, which does not conserve the total probability and therefore is not trace preserving.

Following Eq. (10), the noisy evolutions $\mathcal{K}$ and $\mathcal{K}_1$ that appear in the KIK formula are thus given by

$$\mathcal{K} = \mathcal{T}e^{\int_0^T (-i\mathcal{H}(t) + \mathcal{L}(t))dt}, \quad (13)$$

$$\mathcal{K}_1 = \mathcal{T}e^{\int_0^T (-i\mathcal{H}_1(t) + \mathcal{L}_1(t))dt}. \quad (14)$$

It is important to remark that no restrictions are imposed on the pulse $\mathcal{H}(t)$, so long as it reproduces the noise-free evolution $\mathcal{U}$. On the other hand, we will see in Supplementary Note 3 that Eqs. (11) and (12) allow us to approximate the noise channel for $\mathcal{K}$ as $(\mathcal{K}_1\mathcal{K})^{\frac{1}{2}}$, which is why the form of the inverse driving in (11) is important for our main finding.

We also note that the level of control required for implementing $\mathcal{H}_1(t)$ is very similar to that used for $\mathcal{H}(t)$. In essence, we only need to time-reverse the pulse schedule corresponding to $\mathcal{H}(t)$ and flip its sign. In this work, we use the pulse-gate capabilities of the IBM processors to implement $\mathcal{H}_1(t)$. No stretching of the pulses or any modification of their shape is involved. Therefore, the powers of $\mathcal{K}_1\mathcal{K}$ that enter the implementation of the KIK method are basically as easy to execute $\mathcal{K}$ itself.

SUPPLEMENTARY NOTE 3: DERIVATION OF THE KIK FORMULA

In this section, we derive the KIK formula

$$\mathcal{U}_{\text{KIK}} = \mathcal{K}(\mathcal{K}_1\mathcal{K})^{-\frac{1}{2}}, \quad (15)$$

where $\mathcal{U}_{\text{KIK}}$ is a first-order Magnus approximation to the ideal evolution $\mathcal{U}$ that we will clarify in what follows. Hereafter, we will refer to $\mathcal{K}$ and $\mathcal{K}_1$ as target evolution and inverse evolution, respectively. In particular, $\mathcal{K}$ is the noisy evolution over which we intend to perform error mitigation. For now, we assume that the noise-free unitary is given by $\mathcal{U} = \mathcal{T}e^{-\int_0^T i\mathcal{H}(t)dt}$, meaning that the pulse schedule $\mathcal{H}(t)$ is perfectly calibrated. Hence, the KIK formula is useful to mitigate errors caused by the dissipator $\mathcal{L}(t)$. On the other hand, we will see in Supplementary Note 6 that randomized compiling⁴ complements and enhances the error mitigation achieved with the KIK method. Accordingly, integrating randomized compiling into our QEM technique also allows for the mitigation of coherent errors, related to miscalibrations of $\mathcal{H}(t)$.

To arrive at Eq. (15) we shall proceed as follows. We consider that the driving $\mathcal{H}(t)$ acts in the time interval $(0, T)$ and the driving $\mathcal{H}_1(t)$ is applied in the interval $(T, 2T)$. Thus, the total evolution at time $t = 2T$ is $\mathcal{K}_1\mathcal{K}$. For any time $t \in (0, 2T)$ the dynamics is modeled according to

$$\frac{d}{dt}|\rho\rangle = \left(-i\tilde{\mathcal{H}}(t) + \tilde{\mathcal{L}}(t)\right)|\rho\rangle, \quad (16)$$

where

$$\begin{aligned}\tilde{\mathcal{H}}(t) &= \begin{cases} \mathcal{H}(t) & \text{for } t \in (0, T) \\ \mathcal{H}_I(t - T) & \text{for } t \in (T, 2T), \end{cases} \\ \tilde{\mathcal{L}}(t) &= \begin{cases} \mathcal{L}(t) & \text{for } t \in (0, T) \\ \mathcal{L}_I(t - T) & \text{for } t \in (T, 2T). \end{cases}\end{aligned}\quad (17)$$

Note that the action of $\mathcal{H}_I$ and $\mathcal{L}_I$ on $(T, 2T)$ requires the time shift by T as described in Eqs. (17). With this notation, the ideal evolution for $t \in (0, 2T)$ is given by $\tilde{\mathcal{U}}(t) = \mathcal{T}e^{-i\int_0^t \tilde{\mathcal{H}}(t')dt'}$, and therefore, $\tilde{\mathcal{U}}(T) = \mathcal{U}$, and $\tilde{\mathcal{U}}(2T) = \mathcal{I}$. Similarly, for the noisy evolution we have that $\tilde{\mathcal{K}}(t) = \mathcal{T}e^{\int_0^t (-i\tilde{\mathcal{H}}(t') + \tilde{\mathcal{L}}(t'))dt'}$, $\tilde{\mathcal{K}}(T) = \mathcal{K}$, and $\tilde{\mathcal{K}}(2T) = \mathcal{K}_I\mathcal{K}$.

By expressing Eq. (16) in the interaction picture we can write the evolution operator in the form $\mathcal{K} = \mathcal{U}\mathcal{N}$, where the noise channel $\mathcal{N}$ is the solution to Eq. (16) in interaction picture, at time $t = T$. Next, using the Magnus expansion⁵, we will find that $\mathcal{N}$ can be approximated by $(\mathcal{K}_I\mathcal{K})^{\frac{1}{2}}$. These are the main ingredients for the derivation of the KIK formula (15).

To define the transformed states and operators in interaction picture, we use the noiseless evolution $\tilde{\mathcal{U}}(t)$. Denoting interaction picture vectors and matrices with the subscript “int”, we have that

$$|\rho_{\text{int}}(t)\rangle = \tilde{\mathcal{U}}^\dagger(t)|\rho(t)\rangle, \quad (18)$$

$$\tilde{\mathcal{L}}_{\text{int}}(t) = \tilde{\mathcal{U}}^\dagger(t)\tilde{\mathcal{L}}(t)\tilde{\mathcal{U}}(t), \quad (19)$$

$$\frac{d}{dt}|\rho_{\text{int}}(t)\rangle = \tilde{\mathcal{L}}_{\text{int}}(t)|\rho_{\text{int}}(t)\rangle. \quad (20)$$

The solution $|\rho_{\text{int}}(t)\rangle = \tilde{\mathcal{K}}_{\text{int}}(t)|\rho_{\text{int}}(0)\rangle$ to Eq. (20) is related to the original (Schrodinger-picture) solution by $\tilde{\mathcal{K}}_{\text{int}}(t) = \tilde{\mathcal{U}}^\dagger(t)\tilde{\mathcal{K}}(t)$. Therefore,

$$\begin{aligned}\tilde{\mathcal{K}}(t) &= \tilde{\mathcal{U}}(t)\tilde{\mathcal{K}}_{\text{int}}(t) \\ &= \tilde{\mathcal{U}}(t)e^{\Omega(t)},\end{aligned}\quad (21)$$

where we express $\tilde{\mathcal{K}}_{\text{int}}(t)$ in terms of the Magnus expansion $\Omega(t) = \sum_{n=1}^{\infty} \Omega_n(t)$ ⁵. The first-order Magnus term $\Omega_1(t)$ is central to our analysis, and is given by

$$\Omega_1(t) = \int_0^t \tilde{\mathcal{L}}_{\text{int}}(t')dt'. \quad (22)$$

Regarding higher order terms $\Omega_{n \geq 2}(t)$, we only mention that they contain nested commutators that obey time ordering. For example, $\Omega_2(t) = \frac{1}{2} \int_0^t dt' \int_0^{t'} dt'' [\mathcal{L}_{\text{int}}(t'), \mathcal{L}_{\text{int}}(t'')]$.

Setting $t = T$ and $t = 2T$ in Eq. (21) leads us to the exact solutions $\mathcal{K} = \mathcal{U}e^{\Omega(T)}$ and $\mathcal{K}_I\mathcal{K} = e^{\Omega(2T)}$. If we keep only the first Magnus term in the corresponding Magnus expansions,

$$\mathcal{K} \approx \mathcal{U}e^{\Omega_1(T)}, \quad (23)$$

$$\mathcal{K}_I\mathcal{K} \approx e^{\Omega_1(2T)}. \quad (24)$$

From these expressions, our final step in the derivation of (15) is to show that

$$\Omega_1(2T) = 2\Omega_1(T) \Leftrightarrow \int_0^T \tilde{\mathcal{L}}_{\text{int}}(t')dt' = \int_T^{2T} \tilde{\mathcal{L}}_{\text{int}}(t')dt'. \quad (25)$$

Taking into account Eqs. (23) and (24), Eq. (25) implies that the noise channel $\mathcal{N} = e^{\Omega(T)}$ for the evolution $\mathcal{K}$ can be approximated by

$$\mathcal{N} \approx \mathcal{N}_{\text{KIK}} := (\mathcal{K}_I\mathcal{K})^{\frac{1}{2}}. \quad (26)$$

In this way, the KIK formula is obtained by multiplying $\mathcal{K} \approx \mathcal{U}\mathcal{N}_{\text{KIK}}$ by the inverse $\mathcal{N}_{\text{KIK}}^{-1} = (\mathcal{K}_I\mathcal{K})^{-\frac{1}{2}}$.

Before proving Eq. (25), it is instructive to write also the inverse evolution $\mathcal{K}_I$ in the first Magnus approximation. Using Eqs. (23)-(25), we have that $\mathcal{K}_I\mathcal{K} \approx \mathcal{K}_I\mathcal{U}e^{\Omega_1(T)} \approx e^{2\Omega_1(T)}$. Therefore, we can multiply the expression $\mathcal{K}_I\mathcal{U}e^{\Omega_1(T)} \approx e^{2\Omega_1(T)}$ from the right hand side by $e^{-\Omega_1(T)}\mathcal{U}^\dagger$, to obtain

$$\mathcal{K}_I \approx e^{\Omega_I(T)} \mathcal{U}^\dagger. \quad (27)$$

Let us now prove Eq. (25). First, we note that $\tilde{\mathcal{U}}(t) = \mathcal{U}^\dagger(t-T)\mathcal{U}(T) = \mathcal{U}(2T-t)$ for $t \in (T, 2T)$. In addition, for the same time interval Eqs. (12) and (17) lead to $\tilde{\mathcal{L}}(t) = \mathcal{L}(2T-t)$. Therefore,

$$\begin{aligned} \int_T^{2T} \tilde{\mathcal{L}}_{\text{int}}(t) dt &= \int_T^{2T} \tilde{\mathcal{U}}^\dagger(t) \tilde{\mathcal{L}}(t) \tilde{\mathcal{U}}(t) dt \\ &= \int_T^{2T} \mathcal{U}^\dagger(2T-t) \mathcal{L}(2T-t) \mathcal{U}(2T-t) dt \\ &= \int_0^T \mathcal{U}^\dagger(t') \mathcal{L}(t') \mathcal{U}(t') dt' \\ &= \int_0^T \tilde{\mathcal{L}}_{\text{int}}(t) dt, \end{aligned} \quad (28)$$

where the last line follows by performing the change of variable $t' = 2T - t$.

To conclude this section, we stress that Eq. (12) is key for the proof of Eq. (25). In turn, within our characterization of noise it is specifically the inverse driving (11) which provides the form taken by the dissipator (12). This shows the crucial role of using the pulse schedule (11) for the inverse evolution, rather than a different alternative that generates the ideal unitary $\mathcal{U}^\dagger$ in the absence of noise.

Relation between $\mathcal{K}_I \mathcal{K}$ and $\mathcal{K} \mathcal{K}_I$ in the KIK formula. In the following, we show that

$$\mathcal{K} \mathcal{K}_I = \mathcal{U} \mathcal{K}_I \mathcal{K} \mathcal{U}^\dagger. \quad (29)$$

Clearly, this implies that we cannot substitute $\mathcal{K}_I \mathcal{K}$ by $\mathcal{K} \mathcal{K}_I$ in the KIK formula or in the corresponding expansions. In particular, the coincidence with Richardson ZNE applying circuit unitary folding, discussed in Supplementary Note 4, is sound whenever noise amplification is performed using the correct ordering $\mathcal{K}_I \mathcal{K}$. This is different from the heuristic approach taken in Ref. ⁶, where $\mathcal{K} \mathcal{K}_I$ could be an equally valid choice because it also reproduces the identity operation in the absence of noise.

More specifically, we show that the relation (29) holds under the same approximation that leads to Eq. (15). Namely, when the Magnus expansion used to express the evolution $\mathcal{K} \mathcal{K}_I$ is also truncated to the first Magnus term. Following our noise model, this evolution is the solution to the equation

$$\frac{d}{dt} |\rho\rangle = (-i\bar{\mathcal{H}}(t) + \bar{\mathcal{L}}(t)) |\rho\rangle, \quad (30)$$

where

$$\begin{aligned} \bar{\mathcal{H}}(t) &= \begin{cases} \mathcal{H}_I(t) & \text{for } t \in (0, T) \\ \mathcal{H}(t-T) & \text{for } t \in (T, 2T), \end{cases} \\ \bar{\mathcal{L}}(t) &= \begin{cases} \mathcal{L}_I(t) & \text{for } t \in (0, T) \\ \mathcal{L}(t-T) & \text{for } t \in (T, 2T). \end{cases} \end{aligned} \quad (31)$$

In interaction picture, the first Magnus term for the solution of (30) at time $t = 2T$ reads $\int_0^{2T} \bar{\mathcal{L}}_{\text{int}}(t) dt$, where $\bar{\mathcal{L}}_{\text{int}}(t) = \bar{\mathcal{U}}^\dagger(t) \bar{\mathcal{L}}(t) \bar{\mathcal{U}}(t)$. Taking this into account, we will show that

$$\int_0^{2T} \bar{\mathcal{L}}_{\text{int}}(t) dt = \mathcal{U}(T) \int_0^{2T} \tilde{\mathcal{L}}_{\text{int}}(t) dt \mathcal{U}^\dagger(T). \quad (32)$$

Accordingly, up to first order in the Magnus expansion we have that $\mathcal{K} \mathcal{K}_I \approx e^{\int_0^{2T} \bar{\mathcal{L}}_{\text{int}}(t) dt} = \mathcal{U}(T) e^{\int_0^{2T} \tilde{\mathcal{L}}_{\text{int}}(t) dt} \mathcal{U}^\dagger(T) \approx \mathcal{U}(T) \mathcal{K}_I \mathcal{K} \mathcal{U}^\dagger(T)$, which is tantamount to Eq. (29).

To prove Eq. (32), we derive the following two alternative forms of $\int_0^T \bar{\mathcal{L}}_{\text{int}}(t) dt$:

$$\int_0^T \bar{\mathcal{L}}_{\text{int}}(t) dt = \mathcal{U}(T) \int_0^T \tilde{\mathcal{L}}_{\text{int}}(t) dt \mathcal{U}^\dagger(T), \quad (33)$$

$$\int_0^T \bar{\mathcal{L}}_{\text{int}}(t) dt = \int_T^{2T} \bar{\mathcal{L}}_{\text{int}}(t) dt. \quad (34)$$

Noting that $\bar{\mathcal{U}}(t) = \mathcal{U}(t)$ for $t \in (0, T)$, we obtain

$$\begin{aligned} \int_0^T \bar{\mathcal{L}}_{\text{int}}(t) dt &= \int_0^T \bar{\mathcal{U}}^\dagger(t) \mathcal{L}_1(t) \bar{\mathcal{U}}(t) dt \\ &= \int_0^T \mathcal{U}(t) \mathcal{L}(T-t) \mathcal{U}^\dagger(t) dt \\ &= \int_0^T \mathcal{U}(T-t') \mathcal{L}(t') \mathcal{U}^\dagger(T-t') dt' \\ &= \mathcal{U}(T) \left(\int_0^T \mathcal{U}^\dagger(t') \mathcal{L}(t') \mathcal{U}(t') dt' \right) \mathcal{U}^\dagger(T), \end{aligned} \quad (35)$$

where the change of variable $t' = T - t$ is performed in the third line, and in the last line we use the relation $\mathcal{U}(T) = \mathcal{U}(T - t') \mathcal{U}(t')$. This proves Eq. (33).

For the time interval $t \in (T, 2T)$, the evolution $\bar{\mathcal{U}}(t)$ reads $\bar{\mathcal{U}}(t) = \mathcal{U}(t - T) \mathcal{U}^\dagger(T)$. Therefore,

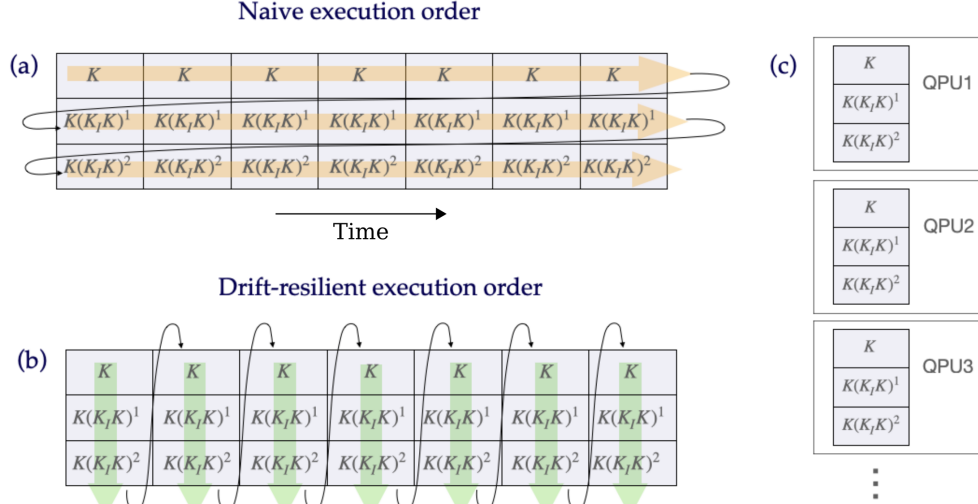
$$\begin{aligned} \int_T^{2T} \bar{\mathcal{L}}_{\text{int}}(t) dt &= \int_T^{2T} \bar{\mathcal{U}}^\dagger(t) \mathcal{L}(t - T) \bar{\mathcal{U}}(t) dt \\ &= \int_T^{2T} \mathcal{U}(T) \mathcal{U}^\dagger(t - T) \mathcal{L}(t - T) \\ &\quad \mathcal{U}(t - T) \mathcal{U}^\dagger(T) dt \\ &= \mathcal{U}(T) \left(\int_0^T \mathcal{U}^\dagger(t') \mathcal{L}(t') \mathcal{U}(t') dt' \right) \mathcal{U}^\dagger(T), \end{aligned} \quad (36)$$

which (in combination with (35)) proves Eq. (34). Equation (32) follows straightforwardly by combining Eqs. (33) and (34).

We note that the simulations studied in Ref.⁶ are based on the assumption of a global depolarizing channel that is identical for $\mathcal{K}$ and $\mathcal{K}_1$. Because global depolarizing noise commutes with any unitary $\mathcal{U}$, in this case the total noise channel for both $\mathcal{K}_1 \mathcal{K}$ and $\mathcal{K} \mathcal{K}_1$ is simply another depolarizing channel with an increased error rate. Hence, for this simple model both the KIK formula and circuit unitary folding can be applied using either $\mathcal{K}_1 \mathcal{K}$ or $\mathcal{K} \mathcal{K}_1$. However, as we have shown here, in a more realistic scenario the proper time ordering corresponding to $\mathcal{K}_1 \mathcal{K}$ is crucial for a correct application of QEM.

Robustness of KIK error mitigation to noise drifts and spatially varying noise profiles. Until now, we have approached the time dependence of $\tilde{\mathcal{L}}(t)$ (cf. Eq. (17)) as being a consequence of the time dependence associated with the pulse schedules, $\tilde{\mathcal{H}}(t)$. In this framework, any implementation of $\mathcal{K}$ or $\mathcal{K}_1$ would be affected by the same dissipators $\mathcal{L}(t)$ and $\mathcal{L}_1(t - T)$. However, it is more realistic to include the possibility of noise sources that also change in time. For example, a varying temperature or external electromagnetic field can be such that the dissipator $\mathcal{L}(t)$ acting during a given implementation of $\mathcal{K}$ differs from the dissipator $\mathcal{L}'(t)$, associated with an execution of the same evolution at a later time. In the present section we discuss a technique to collect the QEM data that minimizes the effect of noise drifts. As we shall see, this is possible by distributing the circuits for QEM into suitable sets, and separately applying the KIK formula (15) to each of these sets.

As discussed in the main text (see also Supplementary Note 4), performing QEM with the KIK method involves executing circuits of the form $\mathcal{K}(\mathcal{K}_1 \mathcal{K})^m$, for $0 \leq m \leq M$, where M is the mitigation order. Therefore, the time for running $\mathcal{K}(\mathcal{K}_1 \mathcal{K})^m$ is $(2m + 1)T$, where T is the evolution time of $\mathcal{K}$ or $\mathcal{K}_1$. In the computation of expectation values, it is necessary to implement each $\mathcal{K}(\mathcal{K}_1 \mathcal{K})^m$ a certain number of times N_m . Following standard terminology in quantum computing, a single execution of a circuit, including the preparation of the initial state $|\rho\rangle$ and the measurement of the final state, is dubbed a “shot”. Hence, N_m shots are used for each $\mathcal{K}(\mathcal{K}_1 \mathcal{K})^m$, and the “shot budget” to collect all



Supplementary Figure 2. Two strategies to collect data for second-order error mitigation, using the KIK method. The arrows indicate the order of implementation of different circuits. In (a), the implementations (shots) are separated into repetitions of each circuit $\mathcal{K}(\mathcal{K}_I \mathcal{K})^m$, for $0 \leq m \leq 2$ (horizontal groups). In (b), the shots are divided into S sets (vertical groups) that include implementations of all the circuits $\mathcal{K}(\mathcal{K}_I \mathcal{K})^m$. (a) and (b) depict circuit distributions with respect to time, for addressing noise drifts. The temporal sequence of circuit sets shown in (b) is rearranged as a “spatial distribution”, where different sets are implemented in different Quantum Processing Units (QPUs). This is useful for a parallel implementation of the KIK method.

the QEM data characteristic of the KIK method reads $N = \sum_{m=0}^M N_m$ (note that this excludes the shots invested in the estimation of the survival probability $\mu = \langle \rho | \mathcal{K}_I \mathcal{K} | \rho \rangle$, in the case of adaptive mitigation). Assuming for now that the time for preparing $|\rho\rangle$ and the time for measuring the corresponding final states are negligible with respect to T , performing N shots takes a total time

$$t_N = \sum_{m=0}^M N_m (2m+1)T. \quad (37)$$

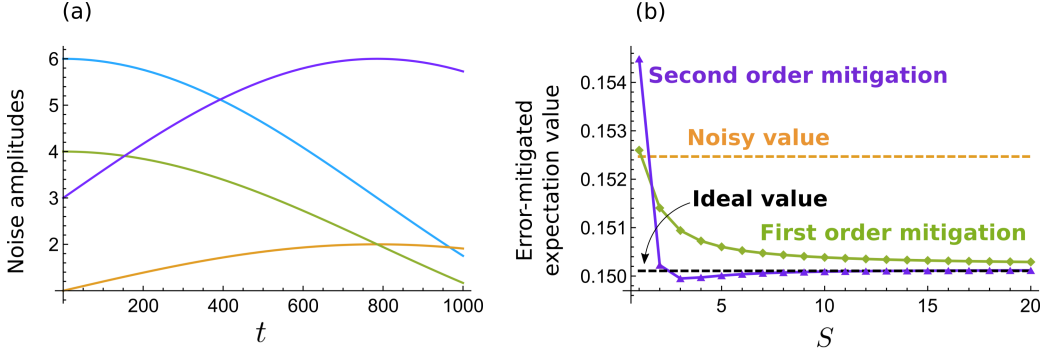
For our analysis, it is useful to extend the time domain of $\tilde{\mathcal{L}}(t)$, to account for the behavior of the noise under repetitions of the evolutions $\mathcal{K}$ and $\mathcal{K}_I$. In this way, the time for an arbitrary repetition of $\mathcal{K}$ or $\mathcal{K}_I$ can be expressed as $t + 2kT$, with $0 \leq t \leq 2T$ and k a positive integer, and stationary noise is characterized by the condition $\tilde{\mathcal{L}}(t + 2kT) = \tilde{\mathcal{L}}(t)$, where $\tilde{\mathcal{L}}(t)$ is the dissipator in Eq. (17). Conversely, noise drifts take place within the total time interval $(0, t_N)$ if $\tilde{\mathcal{L}}(t + 2kT) \neq \tilde{\mathcal{L}}(t)$ for some k .

Let us now suppose that noise unavoidably drifts in the interval $(0, t_N)$. In this scenario, the consistency of the evolutions $\mathcal{K}$ or $\mathcal{K}_I$ in different shots can break down and prevent a correct implementation of the KIK formula (15). However, we can avoid or at least alleviate this effect through a proper distribution of the shot budget N . Consider Supplementary Figure 2, where two strategies for implementing the circuits $\{\mathcal{K}(\mathcal{K}_I \mathcal{K})^m\}_{m=0}^2$ (second-order mitigation) are illustrated. In the case of Supplementary Figure 2(a), all the shots corresponding to a given N_m are sequentially implemented, i.e., N_0 shots are first performed, followed by N_1 shots, and finally by N_2 shots. On the other hand, the strategy of Supplementary Figure 2(b) relies on dividing the N shots into S sets $\{n_0, n_1, n_2\}$ of $N_S = n_0 + n_1 + n_2$ shots each, where n_m shots are employed for the circuit $\mathcal{K}(\mathcal{K}_I \mathcal{K})^m$. Therefore, all the mitigation circuits $\mathcal{K}(\mathcal{K}_I \mathcal{K})^m$ appear in each set. Without loss of generality for our argumentation, we can focus on the simple case where $n_m = N_S/3$ for $0 \leq m \leq 2$, i.e. when the shots of each set are equally distributed into the different circuits $\mathcal{K}(\mathcal{K}_I \mathcal{K})^m$.

Since each set $\{n_0, n_1, n_2\}$ contains data produced by all the circuits $\mathcal{K}(\mathcal{K}_I \mathcal{K})^m$, the KIK formula can be individually applied to these data sets. Let

$$\mathcal{U}_{\text{KIK},s} = \mathcal{K}_s (\mathcal{K}_{I,s} \mathcal{K}_s)^{-\frac{1}{2}} \quad (38)$$

denote the KIK formula corresponding to evolutions $\mathcal{K}_s$ and $\mathcal{K}_{I,s}$ that are executed in shots of the s th set $\{n_0, n_1, n_2\}^{(s)}$. If there were no noise drifts, $\mathcal{K}_s = \mathcal{K}_1 = \mathcal{K}$ and $\mathcal{K}_{I,s} = \mathcal{K}_{I,1} = \mathcal{K}_I$ for $1 \leq s \leq S$. Therefore, the two strategies depicted



Supplementary Figure 3. KIK method applied to a system with noise drifts. (a) The four time-dependent amplitudes $f_k(t)$ that appear in the dissipator (40), and characterize the drift in noise parameters. (b) First order mitigation and second order mitigation using the average formula (39), as a function of the number of sets S (see Supplementary Figure 2 and main text for details).

in Supplementary Figure 2 would lead to the same result. Nevertheless, the non-stationary character of the noise may cause that $\mathcal{K}_s$ or $\mathcal{K}_{I;s}$ change significantly when moving between different sets $\{n_0, n_1, n_2\}^{(s)}$, or even within a fixed set. The second possibility is less likely though, if the time $\sum_{m=0}^M n_m(2m+1)T$ invested in implementing each set $\{n_0, n_1, n_2\}^{(s)}$ is smaller than the characteristic time for noise drifts to be significant. In other words, if the time scale over which noise drifts occur is sufficiently large to allow a consistent execution of $\mathcal{U}_{\text{KIK},s}$. Assuming that N_S is sufficiently small (equivalently, S sufficiently large) for this to happen, for any $1 \leq s \leq S$ we can implement the formula (38) without worrying about variations in the evolutions $\mathcal{K}_s$ or $\mathcal{K}_{I;s}$. In this way, for the shot budget N we utilize the “average KIK formula”

$$\mathcal{U}_{\text{KIK,av}}^{(M)} = \frac{1}{S} \sum_{s=1}^S \mathcal{U}_{\text{KIK},s}^{(M)}. \quad (39)$$

A pertinent question when executing a computation is whether it can be parallelized and save time. The parallelization can be carried out using subsets of qubits in the same quantum processing unit (QPU), in a different QPU on the same chip, or on a different platform in a remote location. In the context of QEM, a recent proposal is to integrate this approach with virtual distillation^{7,8} techniques⁹.

The KIK method can easily be parallelized by assigning different sets to different QPUs as shown in Supplementary Figure 2(c). In this case, the index s in Eq. (38) would label different QPUs, rather than different instants of time, and the corresponding error-mitigated expectation value would also be computed using (39). The justification is the same as in the case of noise drifts discussed above.

Let us now illustrate the application of this methodology to KIK mitigation under noise drifts. In Supplementary Figure 3, we consider two qubits subjected to a dissipator

$$\tilde{\mathcal{L}}(t) = \xi \sum_{k=1}^4 f_k(t) \mathcal{L}_k, \quad (40)$$

where $\xi = 0.05$,

$$\mathcal{L}_k = A_k \otimes A_k^* - \frac{1}{2} A_k^\dagger A_k \otimes I - \frac{1}{2} I \otimes (A_k^\dagger A_k)^T, \quad (41)$$

$$A_1 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad (42)$$

$$A_2 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad (43)$$

$$A_3 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad (44)$$

$$A_4 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad (45)$$

and

$$f_1(t) = 3(1 + \cos(2t/T)), \quad (46)$$

$$f_2(t) = 1 + \sin(2t/T), \quad (47)$$

$$f_3(t) = 2(1 + \cos(2t/T)), \quad (48)$$

$$f_4(t) = 3(1 + \sin(2t/T)). \quad (49)$$

The noise amplitudes $f_k(t)$ are plotted in Supplementary Figure 3(a), for $1 \leq t \leq 1000$. Moreover, we assume a time-independent Hamiltonian (cf. Eq. (9))

$$\mathcal{H} = H \otimes I - I \otimes H^\top, \quad (50)$$

$$H = 3X \otimes X + I \otimes X, \quad (51)$$

where $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

$\tilde{\mathcal{L}}(t)$ is the Liouville-space representation of the Hilbert space dissipator that satisfies $\hat{L}(t)[\rho] = \xi \sum_{k=1}^4 f_k(t) \hat{L}_k$, which follows by applying the vectorization rule (5) to each term $\hat{L}_k[\rho] = A_k \rho A_k^\dagger - \frac{1}{2} A_k^\dagger A_k \rho - \frac{1}{2} \rho A_k^\dagger A_k$. Since the associated master equation

$$\frac{d}{dt} |\rho\rangle = (-i\mathcal{H} + \tilde{\mathcal{L}}(t)) |\rho\rangle \quad (52)$$

has GKSL (Gorini–Kossakowski–Sudarshan–Lindblad) form, it is guaranteed that its integration results in a completely positive and trace-preserving evolution. We also stress that $\tilde{\mathcal{L}}(t)$ is now defined in the total time interval $(0, t_N)$, to account for the many repetitions of the pulses $\mathcal{H}(t)$ and $\mathcal{H}_I(t - T)$ that come with the N shots.

We numerically simulate QEM using the average formula (39), by setting

$$N = 1000, \quad (53)$$

$$N_S = N/S, \quad (54)$$

$$n_1 = n_2 = n_3 = N_S/3, \quad (55)$$

and $1 \leq S \leq 20$. An execution of $\mathcal{K}$ or $\mathcal{K}_1$ is performed in a unit of time $T = 1$, for which we assume that noise is essentially time independent. In other words, a more precise description of the noise occurring during the N shots is given by the discrete dissipator

$$\tilde{\mathcal{L}}_{\text{disc}}(t) = \tilde{\mathcal{L}}(n), \text{ if } n \leq t \leq n+1, \quad (56)$$

with $\tilde{\mathcal{L}}(t)$ satisfying Eq. (40) and $0 \leq n \leq N-1$.

Supplementary Figure 3(b) shows the error-mitigated expectation value $\langle \rho \rangle_{\text{mit}} = \langle \rho | \mathcal{U}_{\text{KIK,av}}^{(M)} | \rho \rangle$, which quantifies the overlap with the initial state $\rho = I/2 \otimes |0\rangle\langle 0|$. We apply Taylor error mitigation, using the coefficients $a_{\text{Tay},m}^{(M)}$ given in Eq. (61). If $S = 1$, the standard strategy represented in Supplementary Figure 2(a) is recovered. We observe in Supplementary Figure 3(b) that in this case $\langle \rho \rangle_{\text{mit}}$ deviates drastically from the noiseless expectation value $\langle \rho \rangle_{\text{id}} = \langle \rho | \mathcal{U} | \rho \rangle$. As S increases, second-order mitigation quickly approaches $\langle \rho \rangle_{\text{id}}$ and converges to it at $S \approx 10$. We also stress that for $S \geq 10$ the quality of error mitigation is maintained, both for $M = 1$ and $M = 2$, which shows that averaging the KIK formula over more sets does not degrade the performance of the KIK method. The success of this strategy is explained because $\langle \rho \rangle_{\text{mit}} = \frac{1}{S} \sum_{s=1}^S \langle \rho | \mathcal{U}_{\text{KIK},s}^{(M)} | \rho \rangle$, and if noise is approximately stationary for each $\mathcal{U}_{\text{KIK},s}^{(M)}$ then the corresponding $\langle \rho | \mathcal{U}_{\text{KIK},s}^{(M)} | \rho \rangle$ is not affected by the action of noise drifts. However, the number of shots N_S associated with each set is not sufficiently large to achieve the accuracy shown in Supplementary Figure 3(b). This accuracy is achieved after averaging over the total number of sets. The convergence observed in Supplementary Figure 3(b) confirms that in this example increasing the number of sets enhances the QEM performance, and solves the noise drift problem.

SUPPLEMENTARY NOTE 4: IMPLEMENTATIONS OF THE INVERSE OF THE NOISE CHANNEL $\mathcal{N}_{\text{KIK}}$

The KIK formula (15) provides a compact approximation for the ideal target evolution $\mathcal{U}$. However, performing QEM with this formula also requires being able to physically implement the inverse

$$\mathcal{N}_{\text{KIK}}^{-1} = (\mathcal{K}_I \mathcal{K})^{-\frac{1}{2}}. \quad (57)$$

In this section, we compute various approximations to this inverse, given as polynomials of the KIK cycle $\mathcal{K}_I \mathcal{K}$.

Weak noise limit and Taylor approximation. First, we consider the limit of weak noise. As mentioned in the main text, in this case any eigenvalue λ of $\mathcal{K}_I \mathcal{K}$ is close to 1, and we can obtain the eigenvalues of $(\mathcal{K}_I \mathcal{K})^{-\frac{1}{2}}$ by Taylor expanding $\lambda^{-\frac{1}{2}}$ around $\lambda = 1$. Note that here we assume that noise is such that $\mathcal{K}_I \mathcal{K}$ is still diagonalizable, and therefore the eigenvalues of $(\mathcal{K}_I \mathcal{K})^{-\frac{1}{2}}$ can be obtained as $\lambda^{-\frac{1}{2}}$.

A Taylor expansion of $\lambda^{-\frac{1}{2}}$ around $\lambda = 1$ is thus equivalent to expand $(\mathcal{K}_I \mathcal{K})^{-\frac{1}{2}}$ around the identity $\mathcal{I}$. Namely,

$$\begin{aligned} (\mathcal{K}_I \mathcal{K})^{-\frac{1}{2}} &= \sum_{m=0}^{\infty} c_m (\mathcal{K}_I \mathcal{K} - \mathcal{I})^m \\ &= \sum_{m=0}^{\infty} c_m \sum_{k=0}^m \frac{m!(-1)^{m-k}}{k!(m-k)!} (\mathcal{K}_I \mathcal{K})^k, \end{aligned} \quad (58)$$

where

$$c_m = \frac{1}{m!} \frac{d^m \lambda^{-\frac{1}{2}}}{d\lambda^m} \Big|_{\lambda=1} = (-1)^m \frac{(2m-1)!!}{m!2^m}. \quad (59)$$

Since the series (58) involves infinite powers of $\mathcal{K}_I \mathcal{K}$, we must truncate it to some fixed order M for the implementation of $(\mathcal{K}_I \mathcal{K})^{-\frac{1}{2}}$. In this way,

$$\begin{aligned} (\mathcal{K}_I \mathcal{K})^{-\frac{1}{2}} &\approx \sum_{m=0}^M c_m \sum_{k=0}^m \frac{m!(-1)^{m-k}}{k!(m-k)!} (\mathcal{K}_I \mathcal{K})^k \\ &= \sum_{m=0}^M a_{\text{Tay},m}^{(M)} (\mathcal{K}_I \mathcal{K})^m, \end{aligned} \quad (60)$$

where

$$a_{\text{Tay},m}^{(M)} = (-1)^m \frac{(2M+1)!!}{2^M [(2m+1)m!(M-m)!]}. \quad (61)$$

Richardson ZNE using Circuit folding and linear scaling of noise. In the following, we show that the expansion coefficients (61) predict the result of QEM using Richardson ZNE and noise amplification through a method known as circuit folding⁶, under the assumption of a linear scaling of the noise. To put this result into context, we start by presenting the basics of Richardson ZNE and circuit folding.

The goal of ZNE is to infer the noise-free expectation value of an observable A , by measuring this observable at different levels of noise and then extrapolating to the zero-noise limit. Therefore, the application of ZNE requires assuming a certain functional dependence $\langle A \rangle(\lambda)$, between the expectation value $\langle A \rangle$ and some noise parameter λ over which the experimentalist should have control. By measuring expectation values $\langle A \rangle_k$ corresponding to different levels of noise λ_k , an experimentalist can fit the data $[\lambda_k, \langle A \rangle_k]$ to the model $\langle A \rangle(\lambda)$ and thereby estimate the noiseless expectation value as $\langle A \rangle(0)$. In the case of Richardson extrapolation, for $M+1$ data points $[\lambda_k, \langle A \rangle_k]$, $\langle A \rangle(\lambda)$ is taken as a polynomial in λ of degree M .

There exists a unique polynomial $P(\lambda)$ of degree M that intersects all the points $[\lambda_k, \langle A \rangle_k]$. This polynomial can be constructed as $P(\lambda) = \sum_{m=0}^M \langle A \rangle_m l_m(\lambda)$, where

$$l_m(\lambda) := \prod_{0 \leq k \leq M, k \neq m} \frac{\lambda - \lambda_k}{\lambda_m - \lambda_k} \quad (62)$$

is a Lagrange polynomial of degree M . Noting that $l_m(\lambda_k) = \delta_{km}$, it follows that $P(\lambda_k) = \langle A \rangle_k$ for $0 \leq k \leq M$. Therefore, the noise-free expectation value is estimated by

$$\begin{aligned} \langle A \rangle(0) &= P(0) \\ &= \sum_{m=0}^M \langle A \rangle_m l_m(0) \\ &= \sum_{m=0}^M \langle A \rangle_m \prod_{0 \leq k \leq M, k \neq m} \frac{\lambda_k}{\lambda_k - \lambda_m}. \end{aligned} \quad (63)$$

Equation (63) gives $\langle A \rangle(0)$ in terms of $\langle A \rangle_k$ and the noise strengths λ_k . One of the first techniques proposed to artificially increase the value of λ is pulse stretching¹⁰, which involves pulse control from the user. In addition, we point out that pulse stretching also assumes that the noise is time-independent. Unitary folding is an alternative that does not require this level of control. If $\mathcal{U}$ describes the target ideal evolution, unitary folding operates by adding quantum gates to $\mathcal{U}$ that in the noise-free case are just identity operations. This can be done either by using “circuit foldings” $\mathcal{U}^\dagger \mathcal{U}$, or by inserting products between gates and their own inverses. Noting that the polynomial (60) contains powers of the (noisy) implementation of $\mathcal{U}^\dagger \mathcal{U}$, we are specifically interested in the connection between this polynomial and the use of circuit folding for ZNE, rather than folding at the level of gates. In this context, the assumption of linear scaling of the noise means that each power $(\mathcal{K}_1 \mathcal{K})^k$ increases the noise characteristic of $\mathcal{K}$ by a factor of $2k$, i.e. that the noise increases proportionally to the depth of the circuit $(\mathcal{K}_1 \mathcal{K})^k$. If λ_0 corresponds to the natural noise in the target circuit $\mathcal{K}$, then, the folding $\mathcal{K}(\mathcal{K}_1 \mathcal{K})^k$ results in $\lambda_k = (2k + 1)\lambda_0$. By substituting this expression of λ_k into $\prod_{0 \leq k \leq M, k \neq m} \frac{\lambda_k}{\lambda_k - \lambda_m}$, we obtain:

$$\begin{aligned} \prod_{0 \leq k \leq M, k \neq m} \frac{\lambda_k}{\lambda_k - \lambda_m} &= \prod_{0 \leq k \leq M, k \neq m} \frac{2k + 1}{2(k - m)} \\ &= \frac{\prod_{0 \leq k \leq M, k \neq m} (2k + 1)}{2^M \prod_{0 \leq k \leq M, k \neq m} (k - m)} \\ &= \frac{(2M + 1)!! / (2m + 1)}{2^M [(-m)(1 - m) \dots (-1)][(1)(2) \dots (M - m)]} \\ &= (-1)^m \frac{(2M + 1)!!}{2^M [(2m + 1)m!(M - m)!]}, \end{aligned} \quad (64)$$

which coincides with the coefficient $a_{\text{Tay},m}^{(M)}$ [cf. Eq. (61)].

Taking into account that $\langle A \rangle_m = \langle A | \mathcal{K}(\mathcal{K}_1 \mathcal{K})^m | \rho \rangle$, it follows that the application of the KIK formula in the weak noise limit reproduces the prediction of Richardson ZNE, with circuit folding and linear scaling of noise [cf. Eqs. (63) and (64)]. As a final remark, it is worth noting that in circuit folding the realization of $\mathcal{U}^\dagger$ using circuit level of control does not involve modifying gates that are their own inverse. A fundamental example in this respect is the CNOT gate, which is the basic unit of two-qubit interactions. In contrast, the pulse inverse $\mathcal{H}_1(t)$ used in our method reverses also the schedules of the CNOT gates, to keep consistency with the pulse-based inverse $\mathcal{K}_1$ appearing in the KIK formula (15).

In what follows, we also show that using $\mathcal{K} = \mathcal{K}_1$ in the case of target circuits that are their own inverse introduces an additional error term that is absent when $\mathcal{K}_1$ is the pulse inverse. This further demonstrates the importance of the correct implementation of $\mathcal{K}_1$ for KIK QEM.

Error induced by the circuit inverse $\mathcal{K}_1 = \mathcal{K}$ in circuits that satisfy $\mathcal{U}^2 = \mathcal{I}$. Previously, we showed that the coefficients $a_{\text{Tay},m}^{(M)}$ reproduce Richardson ZNE when noise is linearly scaled through circuit folding. However, even in this limit of weak noise the KIK method provides insights that elude a naive application of circuit folding. An example of this was already given in Supplementary Note 3, by showing that, even though $\mathcal{K}\mathcal{K}_1$ and $\mathcal{K}_1\mathcal{K}$ are equivalent in the noiseless scenario, the product $\mathcal{K}_1\mathcal{K}$ is the correct choice for using the KIK formula (15). This is valid in particular for Eq. (60).

Here, we will see that ignoring the pulse inverse $\mathcal{K}_1$ also has negative consequences for QEM using Eq. (60), which for simplicity we call “Taylor mitigation”. Specifically, we consider circuits such that

$$\mathcal{U}^2 = \mathcal{I}, \quad (65)$$

which suggests the application of (60) with $\mathcal{K}_I = \mathcal{K}$. In this way, the M th order approximation to $\mathcal{U}$ (cf. Eq. (2) in the main text) reads

$$\begin{aligned}\mathcal{U}_{\text{KIK}}^{(M)} &= \sum_{m=0}^M a_{\text{Tay},m}^{(M)} \mathcal{K} (\mathcal{K}\mathcal{K})^m \\ &= \sum_{m=0}^M a_{\text{Tay},m}^{(M)} (\mathcal{K})^{2m+1} \\ &= \sum_{m=0}^M a_{\text{Tay},m}^{(M)} (\mathcal{U}e^{\Omega_1})^{2m+1},\end{aligned}\tag{66}$$

where $\mathcal{K}$ has been replaced by Eq. (23) and $\Omega_1 = \Omega_1(T)$.

Next, we approximate $\mathcal{U}e^{\Omega_1}$ by $\mathcal{U}e^{\Omega_1} \approx \mathcal{U} + \mathcal{U}\Omega_1$, and keep only terms that are linear in Ω_1 in $\mathcal{U}_{\text{KIK}}^{(M)}$. Using $\mathcal{U}^{2m+1} = \mathcal{U}$, we have that

$$(\mathcal{U} + \mathcal{U}\Omega_1)^{2m+1} \approx \mathcal{U} + \sum_{k=1}^{2m+1} \mathcal{U}^k \Omega_1 \mathcal{U}^{2m+1-k},\tag{67}$$

for $m \geq 1$. Therefore,

$$\begin{aligned}\mathcal{U}_{\text{KIK}}^{(M)} &\approx a_{\text{Tay},0}^{(M)} (\mathcal{U} + \mathcal{U}\Omega_1) + \sum_{m=1}^M a_{\text{Tay},m}^{(M)} (\mathcal{U}e^{\Omega_1})^{2m+1} \\ &\approx a_{\text{Tay},0}^{(M)} (\mathcal{U} + \mathcal{U}\Omega_1) + \sum_{m=1}^M a_{\text{Tay},m}^{(M)} \left[\mathcal{U} + \sum_{k=1}^{2m+1} \mathcal{U}^k \Omega_1 \mathcal{U}^{2m+1-k} \right] \\ &= \mathcal{U} + a_{\text{Tay},0}^{(M)} \mathcal{U}\Omega_1 + \sum_{m=1}^M a_{\text{Tay},m}^{(M)} \sum_{k=1}^{2m+1} \mathcal{U}^k \Omega_1 \mathcal{U}^{2m+1-k},\end{aligned}\tag{68}$$

where we have applied $\sum_{m=0}^M a_{\text{Tay},m}^{(M)} = 1$.

Now we divide the sum $\sum_{k=1}^{2m+1} \mathcal{U}^k \Omega_1 \mathcal{U}^{2m+1-k}$ into two sums such that one of them contains only even powers $\mathcal{U}^k$, and the other contains only odd powers $\mathcal{U}^k$. Taking into account that $\mathcal{U}^{-1} = \mathcal{U}$ and that any even power of $\mathcal{U}$ yields $\mathcal{I}$, we obtain

$$\begin{aligned}\sum_{k=1}^{2m+1} \mathcal{U}^k \Omega_1 \mathcal{U}^{2m+1-k} &= \sum_{k=1}^{m+1} \mathcal{U}^{2k-1} \Omega_1 \mathcal{U}^{2m+2-2k} + \sum_{k=1}^m \mathcal{U}^{2k} \Omega_1 \mathcal{U}^{2m+1-2k} \\ &= \sum_{k=1}^{m+1} \mathcal{U} \Omega_1 + \sum_{k=1}^m \Omega_1 \mathcal{U} \\ &= (m+1) \mathcal{U} \Omega_1 + m \Omega_1 \mathcal{U}.\end{aligned}\tag{69}$$

Substituting Eq. (69) into Eq. (68), we find that

$$\begin{aligned}\mathcal{U}_{\text{KIK}}^{(M)} &\approx \mathcal{U} + a_{\text{Tay},0}^{(M)} \mathcal{U}\Omega_1 + \sum_{m=1}^M a_{\text{Tay},m}^{(M)} [(m+1) \mathcal{U}\Omega_1 + m \Omega_1 \mathcal{U}] \\ &= \mathcal{U} + a_{\text{Tay},0}^{(M)} \mathcal{U}\Omega_1 + \sum_{m=0}^M a_{\text{Tay},m}^{(M)} [(m+1) \mathcal{U}\Omega_1 + m \Omega_1 \mathcal{U}] - a_{\text{Tay},0}^{(M)} \mathcal{U}\Omega_1 \\ &= \mathcal{U} + \sum_{m=0}^M a_{\text{Tay},m}^{(M)} [(m+1) \mathcal{U}\Omega_1 + m \Omega_1 \mathcal{U}],\end{aligned}\tag{70}$$

where in the second line we added and subtracted the term corresponding to $m = 0$ in the sum.

Finally, we use again $\sum_{m=0}^M a_{\text{Taylor},m}^{(M)} = 1$, and $\sum_{m=0}^M a_{\text{Taylor},m}^{(M)} m = -\frac{1}{2}$. In this way, for any $M \geq 1$,

$$\mathcal{U}_{\text{KIK}}^{(M)} \approx \mathcal{U} + \frac{1}{2}[\mathcal{U}, \Omega_1]. \quad (71)$$

Because Taylor mitigation corresponds to weak noise, in the limit when M tends to infinity $\mathcal{U}_{\text{KIK}}^{(M)}$ should converge to the evolution $\mathcal{U}_{\text{KIK}}$ in the KIK formula (15). By construction, this is the case if $\mathcal{K}_I$ is given by the pulse inverse. However, we see from Eq. (71) that when $\mathcal{K}_I = \mathcal{K}$ the term $\frac{1}{2}[\mathcal{U}, \Omega_1]$ remains in the approximation $\mathcal{U}_{\text{KIK}}^{(M)}$, irrespective of the mitigation order M . Note also that this term cannot be avoided in general by considering higher-order (nonlinear) contributions in Ω_1 . In this way, Eq. (71) shows that using $\mathcal{K}_I = \mathcal{K}$ instead of the pulse inverse leads to an inconsistent application of QEM, which is afflicted by an additional error term $\frac{1}{2}[\mathcal{U}, \Omega_1]$.

Finally, we remark that for global depolarizing noise the associated noise channel commutes with any unitary $\mathcal{U}$, and one can easily check that in this case the KIK formula (15) is exact when $\mathcal{K}_I = \mathcal{K}$. Thus, while the pulse-based inverse $\mathcal{K}_I$ becomes unnecessary for such a simplified noise model⁶, it is of paramount importance in practical applications under realistic noise.

Coefficients for Adaptive QEM based on the KIK formula, for mitigation orders $M = 1, 2, 3$.

To obtain the coefficients $a_{\text{Adap},m}^{(M)}$, we minimize the quantity

$$\varepsilon_{\text{L2}}^{(M)} := \int_{g(\mu)}^1 \left(\sum_{m=0}^M a_m^{(M)} \lambda^m - \lambda^{-\frac{1}{2}} \right)^2 d\lambda \quad (72)$$

with respect the first M coefficients $\left\{ a_m^{(M)} \right\}_{m=0}^{M-1}$. Therefore, we have to solve the equations

$$\frac{\partial \varepsilon_{\text{L2}}^{(M)}}{\partial a_m^{(M)}} = 0, \quad \text{for } 0 \leq m \leq M-1. \quad (73)$$

The M th coefficient is obtained by imposing the constraint $a_{\text{Adap},M}^{(M)} = 1 - \sum_{m=0}^{M-1} a_{\text{Adap},m}^{(M)}$. This is equivalent to the normalization condition $\sum_{m=0}^M a_{\text{Adap},m}^{(M)} = 1$, which ensures that the map $\mathcal{U}_{\text{KIK}}^{(M)} = \sum_{m=0}^M a_{\text{Adap},m}^{(M)} [g(\mu)] \mathcal{K} (\mathcal{K}_I \mathcal{K})^m$ is trace-preserving if $\mathcal{K}$ and $\mathcal{K}_I$ are trace-preserving. For the sake of notational simplicity, in the following we will write $g(\mu)$ as g .

By explicitly evaluating $\varepsilon_{\text{L2}}^{(M)}$ with $M = 1, 2, 3$, we find that

$$\varepsilon_{\text{L2}}^{(1)} = (1-g) \left(a_0^{(1)} \right)^2 + \frac{1}{3}(1-g^3) \left(a_1^{(1)} \right)^2 + (1-g^2) a_0^{(1)} a_1^{(1)} - 4(1-g^{\frac{1}{2}}) a_0^{(1)} - \frac{4}{3}(1-g^{\frac{3}{2}}) a_1^{(1)} - \ln(g), \quad (74)$$

$$\begin{aligned} \varepsilon_{\text{L2}}^{(2)} = & (1-g) \left(a_0^{(2)} \right)^2 + \frac{1}{3}(1-g^3) \left(a_1^{(2)} \right)^2 + \frac{1}{5}(1-g^5) \left(a_2^{(2)} \right)^2 + (1-g^2) a_0^{(2)} a_1^{(2)} + \frac{2}{3}(1-g^3) a_0^{(2)} a_2^{(2)} \\ & + \frac{1}{2}(1-g^4) a_1^{(2)} a_2^{(2)} - 4(1-g^{\frac{1}{2}}) a_0^{(2)} - \frac{4}{3}(1-g^{\frac{3}{2}}) a_1^{(2)} - \frac{4}{5}(1-g^{\frac{5}{2}}) a_2^{(2)} - \ln(g), \end{aligned} \quad (75)$$

$$\begin{aligned} \varepsilon_{\text{L2}}^{(3)} = & (1-g) \left(a_0^{(3)} \right)^2 + \frac{1}{3}(1-g^3) \left(a_1^{(3)} \right)^2 + \frac{1}{5}(1-g^5) \left(a_2^{(3)} \right)^2 + \frac{1}{7}(1-g^7) \left(a_3^{(3)} \right)^2 \\ & + (1-g^2) a_0^{(3)} a_1^{(3)} + \frac{2}{3}(1-g^3) a_0^{(3)} a_2^{(3)} + \frac{1}{2}(1-g^4) a_0^{(3)} a_3^{(3)} + \frac{1}{2}(1-g^4) a_1^{(3)} a_2^{(3)} + \frac{2}{5}(1-g^5) a_1^{(3)} a_3^{(3)} \\ & + \frac{1}{3}(1-g^6) a_2^{(3)} a_3^{(3)} - 4(1-g^{\frac{1}{2}}) a_0^{(3)} - \frac{4}{3}(1-g^{\frac{3}{2}}) a_1^{(3)} - \frac{4}{5}(1-g^{\frac{5}{2}}) a_2^{(3)} - \frac{4}{7}(1-g^{\frac{7}{2}}) a_3^{(3)} - \ln(g). \end{aligned} \quad (76)$$

By taking the partial derivatives (73) and solving the resulting linear equations, the corresponding coefficients read

$$a_{\text{Adap},0}^{(1)} = 1 + \frac{1}{(1+\sqrt{g})^3} + \frac{3}{2(1+\sqrt{g})^2}, \quad (77)$$

$$a_{\text{Adap},1}^{(1)} = -\frac{5+3\sqrt{g}}{2(1+\sqrt{g})^3}, \quad (78)$$

for $M = 1$,

$$a_{\text{Adap},0}^{(2)} = 1 + \frac{16}{3(1+\sqrt{g})^5} - \frac{14}{3(1+\sqrt{g})^4} + \frac{4}{(1+\sqrt{g})^2}, \quad (79)$$

$$a_{\text{Adap},1}^{(2)} = -4 \frac{10 + 8\sqrt{g} + 9g + 3g^{\frac{3}{2}}}{3(1+\sqrt{g})^5}, \quad (80)$$

$$a_{\text{Adap},2}^{(2)} = 2 \frac{13 + 5\sqrt{g}}{3(1+\sqrt{g})^5}, \quad (81)$$

for $M = 2$, and

$$a_{\text{Adap},0}^{(3)} = \frac{31 + 97\sqrt{g} + 276g + 300g^{\frac{3}{2}} + 270g^2 + 114g^{\frac{5}{2}} + 28g^3 + 4g^{\frac{7}{2}}}{4(1+\sqrt{g})^7}, \quad (82)$$

$$a_{\text{Adap},1}^{(3)} = -5 \frac{29 + 35\sqrt{g} + 84g + 44g^{\frac{3}{2}} + 26g^2 + 6g^{\frac{5}{2}}}{4(1+\sqrt{g})^7}, \quad (83)$$

$$a_{\text{Adap},2}^{(3)} = 3 \frac{81 + 47\sqrt{g} + 76g + 20g^{\frac{3}{2}}}{4(1+\sqrt{g})^7}, \quad (84)$$

$$a_{\text{Adap},3}^{(3)} = -5 \frac{25 + 7\sqrt{g}}{4(1+\sqrt{g})^7}, \quad (85)$$

for $M = 3$.

Let us now check that the obtained coefficients minimize $\varepsilon_{\text{L}2}^{(M)}$ in the subspace determined by the variables $\{a_m^{(M)}\}_{m=0}^{M-1}$. For $M = 1$, the second derivative $\frac{\partial^2 \varepsilon_{\text{L}2}^{(1)}}{\partial a_0^{(1)2}} = 2(1-g)$ is positive if $g \leq 1$, and therefore $a_{\text{Adap},0}^{(1)}$ minimizes $\varepsilon_{\text{L}2}^{(1)}$ with respect to $a_0^{(1)}$. To see if $a_{\text{Adap},0}^{(2)}$ and $a_{\text{Adap},1}^{(2)}$ in Eqs. (77) and (78) minimize $\varepsilon_{\text{L}2}^{(2)}$, we evaluate the determinant of the Hessian matrix of $\varepsilon_{\text{L}2}^{(2)}$ and the second partial derivative $\frac{\partial^2 \varepsilon_{\text{L}2}^{(2)}}{\partial a_0^{(2)2}}$, and check their positivity. Since $\frac{\partial^2 \varepsilon_{\text{L}2}^{(2)}}{\partial a_0^{(2)2}} = 2(1-g) \geq 0$, we only need to check the determinant of the Hessian matrix

$$H = \begin{pmatrix} \frac{\partial^2 \varepsilon_{\text{L}2}^{(2)}}{\partial a_0^{(2)2}} & \frac{\partial^2 \varepsilon_{\text{L}2}^{(2)}}{\partial a_0^{(2)} \partial a_1^{(2)}} \\ \frac{\partial^2 \varepsilon_{\text{L}2}^{(2)}}{\partial a_1^{(2)} \partial a_0^{(2)}} & \frac{\partial^2 \varepsilon_{\text{L}2}^{(2)}}{\partial a_1^{(2)2}} \end{pmatrix} = \begin{pmatrix} 2(1-g) & 1-g^2 \\ 1-g^2 & \frac{2}{3}(1-g^3) \end{pmatrix}. \quad (86)$$

Such a determinant is given by $\det(H) = \frac{4}{3} [1-g] [1-g^3] - [1-g^2]^2$, which is also positive in the interval $0 \leq g \leq 1$. Finally, for $M = 3$ we obtain the Hessian matrix

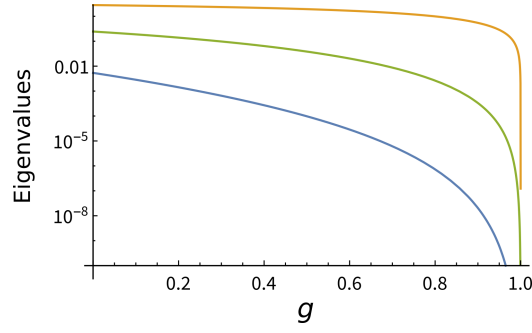
$$H = \begin{pmatrix} \frac{\partial^2 \varepsilon_{\text{L}2}^{(3)}}{\partial a_0^{(3)2}} & \frac{\partial^2 \varepsilon_{\text{L}2}^{(3)}}{\partial a_0^{(3)} \partial a_1^{(3)}} & \frac{\partial^2 \varepsilon_{\text{L}2}^{(3)}}{\partial a_0^{(3)} \partial a_2^{(3)}} \\ \frac{\partial^2 \varepsilon_{\text{L}2}^{(3)}}{\partial a_1^{(3)} \partial a_0^{(3)}} & \frac{\partial^2 \varepsilon_{\text{L}2}^{(3)}}{\partial a_1^{(3)2}} & \frac{\partial^2 \varepsilon_{\text{L}2}^{(3)}}{\partial a_1^{(3)} \partial a_2^{(3)}} \\ \frac{\partial^2 \varepsilon_{\text{L}2}^{(3)}}{\partial a_2^{(3)} \partial a_0^{(3)}} & \frac{\partial^2 \varepsilon_{\text{L}2}^{(3)}}{\partial a_2^{(3)} \partial a_1^{(3)}} & \frac{\partial^2 \varepsilon_{\text{L}2}^{(3)}}{\partial a_2^{(3)2}} \end{pmatrix} = \begin{pmatrix} 2(1-g) & 1-g^2 & \frac{2}{3}(1-g^3) \\ 1-g^2 & \frac{2}{3}(1-g^3) & \frac{1}{2}(1-g^4) \\ \frac{2}{3}(1-g^3) & \frac{1}{2}(1-g^4) & \frac{2}{5}(1-g^5) \end{pmatrix}. \quad (87)$$

The eigenvalues of this matrix are plotted in Supplementary Figure 4. Since all of them are positive in the interval $0 \leq g \leq 1$, we conclude that Eqs. (82)-(85) provide the coefficients $a_{\text{Adap},m}^{(3)}$ that minimize $\varepsilon_{\text{L}2}^{(3)}$ with respect to $\{a_m^{(M)}\}_{m=0}^2$.

SUPPLEMENTARY NOTE 5: KIK QEM APPLIED TO THE TRANSVERSE ISING MODEL

In this supplementary note, we simulate the application of the KIK method to a noisy implementation of the transverse Ising model. This model is characterized by the Hamiltonian

$$\begin{aligned} H &= g \sum_{i=1}^n X_i + J \sum_{i=1}^{n-1} Z_i \otimes Z_{i+1} \\ &:= gH_X + JH_{ZZ} \end{aligned} \quad (88)$$



Supplementary Figure 4. Eigenvalues of the Hessian matrix (87), as a function of $g = g(\mu)$.

where H_{ZZ} accounts for the magnetic interactions between nearest-neighbor spins, and the transverse magnetic fields are represented by the local term H_X . We consider $n = 5$ spins and set $g = 0.2$ and $J = 0.1$. There is no particular reason for choosing these parameter values. Similar performance was obtained with simulations using alternative values (not shown here).

For the simulation, we assume that the target evolution $U = e^{-iHT}$ is implemented in a quantum computer via Trotterization¹¹. Taking into account that the X and Z Pauli matrices do not commute, an approximation to U involving n Trotter steps has the form

$$U \approx U_n := \left(e^{-igH_X \frac{T}{n}} e^{-iJH_{ZZ} \frac{T}{n}} \right)^n, \quad (89)$$

where T is the total evolution time. Assuming $T = 10$ and 10 Trotter steps, we have that

$$\begin{aligned} U &\approx U_{10} \\ &:= \left(e^{-igH_X} e^{-iJH_{ZZ}} \right)^{10}. \end{aligned} \quad (90)$$

In what follows, I_j will denote the identity matrix of dimension $2^j \times 2^j$, with $I = I_n$. We remark that our goal is not to test the accuracy of the Trotter approximation (90), but to study the performance of the KIK method to mitigate errors in a noisy implementation of U_{10} . To this end, we model the noisy evolution $\mathcal{K}$ associated with U_{10} using

$$\mathcal{K} = \left(e^{-ig\mathcal{H}_X + \xi\mathcal{L}} e^{-iJ\mathcal{H}_{ZZ} + \xi\mathcal{L}} \right)^{10}, \quad (91)$$

$$\mathcal{H}_X = H_X \otimes I - I \otimes (H_X)^T, \quad (92)$$

$$\mathcal{H}_{ZZ} = H_{ZZ} \otimes I - I \otimes (H_{ZZ})^T, \quad (93)$$

where

$$\mathcal{L} = S \otimes S^* - \frac{1}{2} S^\dagger S \otimes I - \frac{1}{2} I \otimes (S^\dagger S)^T, \quad (94)$$

$$S = 0.5S_1 + 1.7S_2 + 0.3S_3 + 2S_4 + S_5, \quad (95)$$

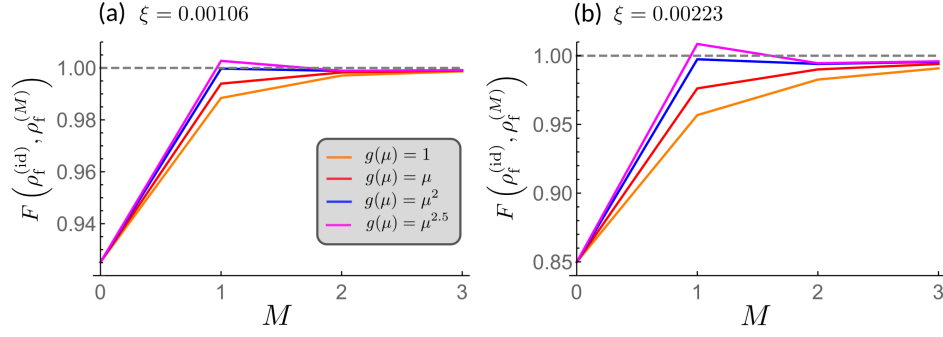
$$S_j = I_{j-1} \otimes A \otimes I_{n-j}, \text{ for } 1 \leq j \leq 5, \quad (96)$$

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}. \quad (97)$$

The initial state in our simulation is the ground state $\rho = |0\rangle\langle 0|^{\otimes 5}$.

Error mitigation. To assess the performance of the KIK QEM, we use the fidelity between the ideal final state $|\rho_f^{(\text{id})}\rangle := \mathcal{U}_{10}|\rho\rangle$ (being $\mathcal{U}_{10}$ the Liouville-space representation of U_{10}) and the error-mitigated final state $|\rho_f^{(M)}\rangle := \mathcal{U}_{\text{KIK}}^{(M)}|\rho\rangle$, which reads¹²

$$F\left(\rho_f^{(\text{id})}, \rho_f^{(M)}\right) = \left[\text{Tr} \sqrt{\sqrt{\rho_f^{(\text{id})}} \rho_f^{(M)} \sqrt{\rho_f^{(\text{id})}}} \right]^2. \quad (98)$$



Supplementary Figure 5. Error-mitigated fidelity $F(\rho_f^{(\text{id})}, \rho_f^{(M)})$, between the ideal final state $|\rho_f^{(\text{id})}\rangle = \mathcal{U}_{10}|\rho\rangle$ and the error-mitigated final state $|\rho_f^{(M)}\rangle := \mathcal{U}_{\text{KIK}}^{(M)}|\rho\rangle$. Each curve is obtained by evaluating the coefficients $a_{\text{Adap},m}^{(M)}[g(\mu)]$ in the M th-order approximation $\mathcal{U}_{\text{KIK}}^{(M)} = \sum_{m=0}^M a_{\text{Adap},m}^{(M)}[g(\mu)]\mathcal{K}(\mathcal{K}_1\mathcal{K})^m$ at the functions $g(\mu)$ indicated in Supplementary Figure 5(a). Supplementary Figures 5(a) and 5(b) stand respectively for $\xi = 0.00106$ and $\xi = 0.00223$, and both are characterized by the color code of Supplementary Figure 5(a).

Notice that the fidelity must be computed using the density matrices $\rho_f^{(\text{id})}$ and $\rho_f^{(M)}$, and not their vector representations $|\rho_f^{(\text{id})}\rangle$ and $|\rho_f^{(M)}\rangle$. We consider values of ξ given by $\xi = 0.00223$ and $\xi = 0.00106$, which lead to fidelities of unmitigated final states ($M = 0$)

$$F(\rho_f^{(\text{id})}, \rho_f^{(0)}) = 0.85, \text{ for } \xi = 0.00223, \quad (99)$$

$$F(\rho_f^{(\text{id})}, \rho_f^{(0)}) = 0.925, \text{ for } \xi = 0.00106. \quad (100)$$

QEM is performed by choosing functions $\{g(\mu)\} = \{1, \mu, \mu^2, \mu^{2.5}\}$, and $1 \leq M \leq 3$. We recall that the function $g(\mu)$ determines the lower limit of integration in $\varepsilon_{\text{L2}}^{(M)}$ (see the main text for the definition), and that for each $g(\mu)$ QEM is applied with $\mathcal{U}_{\text{KIK}}^{(M)} = \sum_{m=0}^M a_{\text{Adap},m}^{(M)}[g(\mu)]\mathcal{K}(\mathcal{K}_1\mathcal{K})^m$, using coefficients $a_{\text{Adap},m}^{(M)}[g(\mu)]$ evaluated at $g = g(\mu)$ (cf. Supplementary Note 4). In particular, $g(\mu) = 1$ corresponds to Taylor mitigation. Keeping in mind Eq. (91), the inverse $\mathcal{K}_1$ is given by

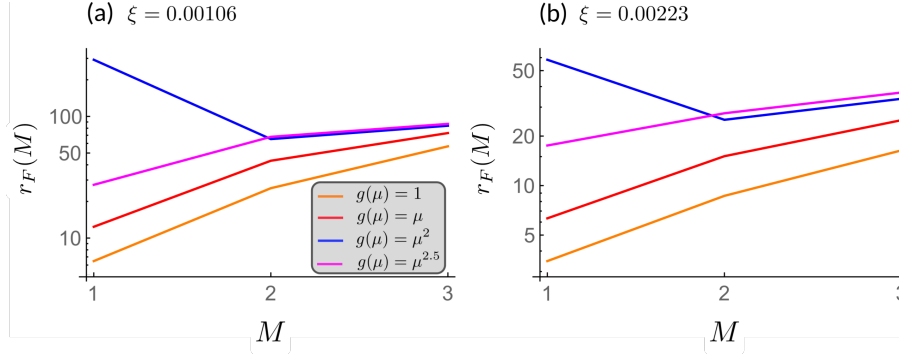
$$\mathcal{K}_1 = (e^{iJ\mathcal{H}zz + \xi\mathcal{L}} e^{ig\mathcal{H}x + \xi\mathcal{L}})^{10}. \quad (101)$$

Supplementary Figure 5 shows the fidelity $F(\rho_f^{(\text{id})}, \rho_f^{(M)})$ as a function of M , for $\xi = 0.00106$ and $\xi = 0.00223$. Overall, we observe that the best performance among the tested functions $g(\mu)$ is achieved by $g(\mu) = \mu^2$ (blue curve). In particular, $F(\rho_f^{(\text{id})}, \rho_f^{(M)})$ reaches a value extremely close to one already for $M = 1$. A clearer comparison between the different functions $g(\mu)$ is possible by looking at the ratio between the infidelity before QEM and after QEM,

$$r_F(M) = \frac{1 - F(\rho_f^{(\text{id})}, \rho_f^{(0)})}{1 - F(\rho_f^{(\text{id})}, \rho_f^{(M)})}, \quad (102)$$

which quantifies the infidelity suppression provided by the KIK method. This quantity is plotted in Supplementary Figure 6. In this figure, we see that $g(\mu) = \mu^2$ outperforms $g(\mu) = 1$ and $g(\mu) = \mu$ for all $1 \leq M \leq 3$. Although the ratio $r_F(M)$ corresponding to $g(\mu) = \mu^2$ is slightly below that associated with $g(\mu) = \mu^{2.5}$, in the case of $M = 2, 3$, $g(\mu) = \mu^2$ yields a substantially larger $r_F(M)$ if $M = 1$.

Finally, it is interesting to observe how the different functions $g(\mu)$ produce physically consistent fidelities $F(\rho_f^{(\text{id})}, \rho_f^{(M)}) \leq 1$ as M increases. In particular, we can see that the unphysical fidelity $F(\rho_f^{(\text{id})}, \rho_f^{(1)}) > 1$ corresponding to $g(\mu) = \mu^{2.5}$ is quickly fixed by going to the next mitigation order $M = 2$. The explanation for this behavior is as follows. The quality of the approximation $\sum_{m=0}^M a_{\text{Adap},m}^{(M)}[g(\mu)](\mathcal{K}_1\mathcal{K})^m$ to $(\mathcal{K}_1\mathcal{K})^{-1/2}$ is determined by how well the polynomial $\sum_{m=0}^M a_{\text{Adap},m}^{(M)}[g(\mu)]\lambda^m$ approximates the function $\lambda^{-1/2}$, where λ is an eigenvalue of $\mathcal{K}_1\mathcal{K}$. If $g(\mu)$ is too small, as compared to the smallest eigenvalue of $(\mathcal{K}_1\mathcal{K})^{-1/2}$, a polynomial $\sum_{m=0}^M a_{\text{Adap},m}^{(M)}[g(\mu)](\mathcal{K}_1\mathcal{K})^m$



Supplementary Figure 6. Enhancement ratio (102), which quantifies the infidelity suppression obtained by the KIK method. The color code in Supplementary Figure 6(b) is the same as in 6(a).

with low M may be a rough approximation to $(\mathcal{K}_I \mathcal{K})^{-1/2}$. This leads to undesired effects such as the aforementioned fidelity $F(\rho_f^{(\text{id})}, \rho_f^{(1)}) > 1$ (note that in our example the function $g(\mu) = \mu^{2.5}$ yields the smallest $g(\mu)$ for any value of μ). However, by increasing M we can always improve the approximation in the whole interval $(g(\mu), 1)$, which by assumption contains all the eigenvalues of $(\mathcal{K}_I \mathcal{K})^{-1/2}$. This would explain not only the recovery of a physical fidelity but also that for $M = 2$ and $M = 3$ the maximum values of this quantity correspond to $g(\mu) = \mu^{2.5}$.

On the contrary, when $g(\mu)$ is larger than the smallest eigenvalue of $(\mathcal{K}_I \mathcal{K})^{-1/2}$, the interval $(g(\mu), 1)$ does not contain all the eigenvalues of $(\mathcal{K}_I \mathcal{K})^{-1/2}$ and increasing M does not necessarily improve the QEM. This is likely the reason that $g(\mu) = \mu^2$ yields fidelities smaller for $M = 2$ and $M = 3$, as compared to $M = 1$.

SUPPLEMENTARY NOTE 6: DESCRIPTION OF THE EXPERIMENTAL PROCEDURES

Here, we provide additional details and complementary information concerning the experiments described in the main text. In the first subsection we describe techniques to tackle different classes of errors, which complement the KIK error mitigation. Next, we explain in detail the steps to estimate error-mitigated expectation values, as well as the corresponding uncertainties. Then, information specific to the CNOT calibration experiment and the ten-swap experiment is given in the following two subsections. We conclude in the last subsection with a simulated calibration of the CNOT gate. This simulation shows results very similar to those of the CNOT calibration experiment, and provides additional evidence of the usefulness of KIK-based calibration.

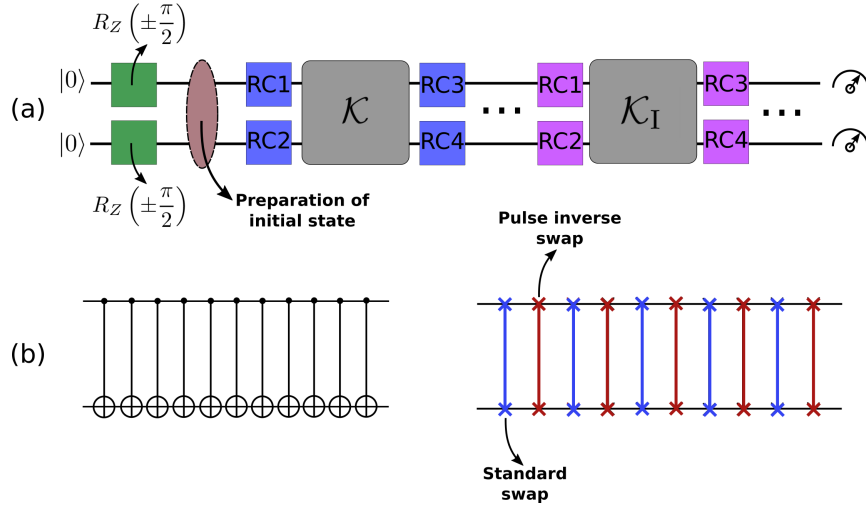
Integration with complementary error mitigation methods. We consider three complementary techniques for addressing errors that can affect an experiment in the three stages of its execution. Namely, the preparation of the initial state, the evolution, and the measurement.

In the IBM quantum processors the preparation of the single-qubit (default) computational state $|0\rangle$ may possess a small deviation angle $\delta\theta$ ¹³. As a result, the prepared state is given by $|\psi(\delta\theta)\rangle = \cos(\frac{\delta\theta}{2})|0\rangle + e^{i\varphi}\sin(\frac{\delta\theta}{2})|1\rangle$. To cope with this error, we apply rotations $R_Z(\pm\pi/2) = e^{\mp i\frac{\pi}{4}Z}$ on each qubit, in such a way that $R_Z(\pi/2)$ and $R_Z(-\pi/2)$ are equally distributed among the number of shots. This produces a mixed state

$$\varrho = \cos^2\left(\frac{\delta\theta}{2}\right)|0\rangle\langle 0| + \sin^2\left(\frac{\delta\theta}{2}\right)|1\rangle\langle 1|, \quad (103)$$

whose population $\text{Tr}(|0\rangle\langle 0|\varrho) = \cos^2(\frac{\delta\theta}{2})$ coincides with $|\langle 0|\psi(\delta\theta)\rangle|^2$. However, the coherent error associated with the rotation angles $\delta\theta$ and φ is eliminated through this procedure. Since both experiments involve circuits acting on two qubits, there are four possible rotation operations characterized by the pairs of angles $\{(\pm\pi/2, \pm\pi/2)\}$. This is illustrated by the green squares in Supplementary Figure 7(a).

Coherent errors present during the evolution stage of a quantum circuit can be specially harmful for quantum computing¹⁴. Randomized compiling (RC)⁴ is a standard method that allows to transform these errors into stochastic noise, which is amenable to the application of QEM using the KIK method. The essential idea of RC is to replace the target circuit by an average over compiled realizations that are logically equivalent to the original evolution. Each RC realization is obtained by randomly replacing and/or adding some gates that leave the noise-free target circuit



Supplementary Figure 7. Schematic of the QEM circuits used in the experiments. (a) General form of a KIK circuit $\mathcal{K}(\mathcal{K}_I\mathcal{K})^m$. For both experiments, initial rotations $R_Z(\pm\pi/2)$ (green squares) are applied on each qubit to mitigate a potential coherent error affecting the state $|00\rangle$. After that, the initial state for the CNOT calibration experiment is prepared (this preparation is not part of the ten-swap experiment), and the evolutions $\mathcal{K}$ and $\mathcal{K}_I$ are interleaved by RC operations (only in the case of the ten-swap experiment). Blue and magenta squares distinguish potentially different RC realizations, chosen from Supplementary Table 1. (b) The evolution $\mathcal{K}$ used for the CNOT calibration experiment (left), and for the ten-swap experiment (right).

RC1	RC2	RC1	RC2	RC1	RC2	RC1	RC2
I	I	X	I	Y	I	Z	I
I	X	X	X	Y	X	Z	X
I	Y	X	Y	Y	Y	Z	Y
I	Z	X	Z	Y	Z	Z	Z

Supplementary Table 1. Possible combinations of RC operations used for the evolutions $\mathcal{K}$ and $\mathcal{K}_I$ in the ten-swap experiment. Here, RC1 and RC2 stand for Pauli gates that are performed as indicated in Supplementary Figure 7(a). Since in this circuit the ideal unitary associated with $\mathcal{K}$ and $\mathcal{K}_I$ is the identity, the gate RC3 must coincide with RC1, and the gate RC4 must coincide with RC2.

invariant. However, in the presence of coherent errors, different RC realizations produce different evolutions, and coherent errors characteristic of single realizations can be averaged out into stochastic noise.

In our method, the implementation $\mathcal{U}_{\text{KIK}}^{(M)}$ of the KIK formula contains both the target evolution $\mathcal{K}$ and its inverse $\mathcal{K}_I$. Therefore, for each KIK power $\mathcal{K}(\mathcal{K}_I\mathcal{K})^m$ we independently applied RC on $\mathcal{K}$ and $\mathcal{K}_I$, as indicated in Supplementary Figure 7(a). This is done by using the Pauli gates shown in Supplementary Table 1, which are performed before and after the execution of $\mathcal{K}$ and $\mathcal{K}_I$ in the ten-swap experiment. A sequence of ten swap gates is logically equivalent to the identity operation, and one can straightforwardly check that each RC realization in Supplementary Table 1 reproduces this operation. Furthermore, we remark that an independent application of RC on $\mathcal{K}$ and $\mathcal{K}_I$ is important to perform an effective randomization of the coherent errors affecting both evolutions. This ensures that the first Magnus terms appearing in the evolutions $\mathcal{K}$ and $\mathcal{K}_I\mathcal{K}$ (cf. Eqs. (23) and (24)) are still related through Eq. (25), and thus validates the KIK formula with the randomized versions of $\mathcal{K}$ and $\mathcal{K}_I$. Along the same line, we avoid the simplification of RC layers in any sequence $\mathcal{K}(\mathcal{K}_I\mathcal{K})^m$. This means that consecutive single-qubit gates used for RC are not merged into a single gate. Otherwise, the independence of the randomization would be affected by preventing that the gates of the original pair act separately on $\mathcal{K}$ and $\mathcal{K}_I$.

On the other hand, for the CNOT calibration experiment we omitted the use of RC. The reason is that, in this experiment, we explored an alternative mitigation of coherent errors affecting the calibrated gate, which operates differently from RC. More specifically, a KIK-based mitigation of stochastic noise in the calibration process can improve the quality of the calibration, which translates into a calibrated gate with reduced coherent errors.

A final ingredient for our QEM experiments is the mitigation of readout errors. For a circuit executed on N qubits, the measurement process outputs counts of the states $|i_1 i_2 \dots i_N\rangle$ ($i_j \in \{0, 1\}$ for $1 \leq j \leq N$) in the computational basis. A readout error occurs when the state registered upon the measurement is incorrect. For example, for a single

qubit whose final state is $|1\rangle$, some counts may erroneously register the state as being $|0\rangle$.

Let us denote a general N -bit string by $\mathbf{k} = (k_1 k_2 \dots k_N)$, and the corresponding computational state by $|\mathbf{k}\rangle$. When a quantum circuit produces a final state σ , readout errors cause wrong estimates of the probabilities $p_{\mathbf{k}} = \text{Tr}(|\mathbf{k}\rangle\langle\mathbf{k}|\sigma)$. A simple way to relate the ideal probabilities $p_{\mathbf{k}}$ and the erroneous ones is by considering the probability distributions associated with each computational state. If $p(\mathbf{l}|\mathbf{k})$ denotes the conditional probability to register the state $|\mathbf{l}\rangle$, given that the actual state is $|\mathbf{k}\rangle$, for the final state σ the probability to measure $|\mathbf{l}\rangle$ in the presence of readout errors is given by

$$q_{\mathbf{l}} = \sum_{\mathbf{k}} p(\mathbf{l}|\mathbf{k}) p_{\mathbf{k}}. \quad (104)$$

Thus, the information about the readout errors is contained in a $2^N \times 2^N$ matrix $\mathbf{M}$ with entries $p(\mathbf{l}|\mathbf{k})$, whose columns and rows are associated with $\mathbf{k}$ and $\mathbf{l}$, respectively. To counteract the effect of the readout errors, we can invert the “measurement matrix” $\mathbf{M}$. In this way, the vector of error-free probabilities can be obtained by applying the inverse $\mathbf{M}^{-1}$ to the vector of measured probabilities.

Although the procedure described above is not scalable in N , because the size of $\mathbf{M}$ is exponential in the number of qubits, our experiments involve two qubits and admit an efficient estimation of the $p(\mathbf{l}|\mathbf{k})$. To this end, we prepared the four computational states $\{|ij\rangle\}_{i,j=0}^1$ using circuits with the appropriate X gates. For example, $|01\rangle$ is prepared by applying X on the second qubit. The measurement matrices were experimentally determined and then inverted for readout error mitigation in both the CNOT calibration experiment and the ten-swap experiment. The number of shots invested in the estimation of the distributions $\{p(\mathbf{l}|\mathbf{k})\}_{\mathbf{l}}$ is given in Supplementary Table 2.

As a final comment, it is important to mention that the study of methods to efficiently cope with readout errors in circuits containing a large number of qubits is an active area of research. For example, in many cases the noisy probability distribution $q_{\mathbf{l}}$ is mostly concentrated around the ideal distribution $p_{\mathbf{k}}$. This allows to construct a measurement matrix of reduced dimension that is defined over the subspace of bit strings with $q_{\mathbf{l}} \neq 0$ ¹⁵. In particular, the maximum size of such a matrix is determined by the maximum number of shots used to sample $q_{\mathbf{l}}$, and is independent of the number of qubits.

Statistical analysis of experimental data. The experimental estimates of different probability distributions are computed as frequencies over the number of shots. Ultimately, the goal is to estimate probability distributions $\{p_{\mathbf{k}}^{(m,i)}\}_{\mathbf{k}}$ for each m and i , where $p_{\mathbf{k}}^{(m,i)}$ denotes the probability to measure $|\mathbf{k}\rangle$ when the circuit $\mathcal{K}(\mathcal{K}_I \mathcal{K})^m$ is implemented in combination with the i th RC realization. These probabilities are the key element for the computation of expectation values and the application of QEM.

In the case of the ten-swap experiment, we applied 16 RC realizations of each circuit $\mathcal{K}(\mathcal{K}_I \mathcal{K})^m$. Hence, for any m the expectation value $\text{Tr}(O\sigma)$ of an observable O is estimated as

$$\langle O \rangle_m = \frac{1}{N_{\text{RC}}} \sum_{i=1}^{N_{\text{RC}}} \langle O \rangle_{m,i}, \quad (105)$$

$$\langle O \rangle_{m,i} = \sum_{\mathbf{k}} p_{\mathbf{k}}^{(m,i)} O_{\mathbf{k}}, \quad (106)$$

where $N_{\text{RC}} = 16$. Importantly, for observables that are not diagonal in the computational basis, the final state σ in $\text{Tr}(O\sigma)$ must contain a rotation that performs the corresponding change of basis. For example, to measure the observable $O = Y_1$ on qubit 1, for the CNOT calibration experiment, before the measurement we applied a rotation $R_{X_1}(\pi/2) = e^{-i\frac{\pi}{4}X_1}$ on this qubit.

Since different RC realizations correspond to independent experiments, the variance for $\langle O \rangle_m$ reads

$$\begin{aligned} \text{Var}(\langle O \rangle_m) &= \frac{1}{N_{\text{RC}}^2} \sum_{i=1}^{N_{\text{RC}}} \text{Var}(\langle O \rangle_{m,i}) \\ &= \frac{1}{N_{\text{RC}}^2} \sum_{i=1}^{N_{\text{RC}}} \left[\frac{\sum_{\mathbf{k}} p_{\mathbf{k}}^{(m,i)} O_{\mathbf{k}}^2 - \langle O \rangle_{m,i}^2}{N_i} \right], \end{aligned} \quad (107)$$

where N_i is the number of shots used in the i th RC realization. In the CNOT calibration experiment RC was not implemented, as mentioned before. However, Eqs. (105)-(107) are still applicable in this case, with $N_{\text{RC}} = 8$ being the number of circuits associated with different initial rotations, and the index i labeling any of these circuits.

The KIK method provides the error-mitigated expectation value

$$\langle O \rangle_M = \sum_{m=0}^M a_m^{(M)} \langle O \rangle_m, \quad (108)$$

with the coefficients $a_m^{(M)}$ chosen depending on the QEM strategy. Namely, Taylor mitigation, or adaptive mitigation. The independence of the experiments associated with different circuits $\mathcal{K}(\mathcal{K}_I \mathcal{K})^m$ leads to the variance

$$\text{Var}(\langle O \rangle_M) = \sum_{m=0}^M \left(a_m^{(M)} \right)^2 \text{Var}(\langle O \rangle_m). \quad (109)$$

The error bars in the plots of the main text and the plots of Supplementary Figures 8 and 9 correspond to one standard deviation. That is, half of each error bar is given by $\sqrt{\text{Var}(\langle O \rangle_M)}$.

CNOT calibration experiment. The calibration experiment was performed on the IBM processor Jakarta, using the qubits labeled by 0 and 1. The qubit 0 was employed as control for the CNOT gate and the qubit 1 as target. In the IBM processors, the two-qubit interaction used to generate the CNOT gate is effectively implemented via the so called cross-resonance interaction¹⁶ $H_{\text{CR}} = Z \otimes X$, where Z and X are the Pauli matrices $Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ and

$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ acting on the control qubit and target qubit, respectively. The CNOT thus involves a $\pi/2$ rotation with the Hamiltonian H_{CR} . This operation is performed in the quantum processor by applying a microwave pulse characterized by various calibrated parameters such as amplitude and duration. However, the values obtained for these parameters may be affected by systematic errors, due to noise present in the measurements used for calibration.

We focus on the calibration of the pulse amplitude, using the KIK method to mitigate the effect of noise. As explained in the main text, we prepare the initial state $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \otimes |0\rangle$ and measure the observable $Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$ on the target qubit for different pulse amplitudes. This choice of initial state and observable is convenient because it produces a calibration curve that is very well described by a straight line. We measured the expectation value $\langle Y_1 \rangle$ for the amplitude factors $F \in \{0.98, 0.9866, 0.9933, 1, 1.0066, 1.0133, 1.02\}$, where $F = 1$ corresponds to the default amplitude used by IBM, and fitted a straight line to the resulting data points. The calibrated amplitude factor is extracted from the intersection of the fitted line with the x axis, which corresponds to $\langle Y_1 \rangle = 0$. The range of amplitudes F was chosen to be sufficiently narrow, so that one could observe the linear behavior previously described.

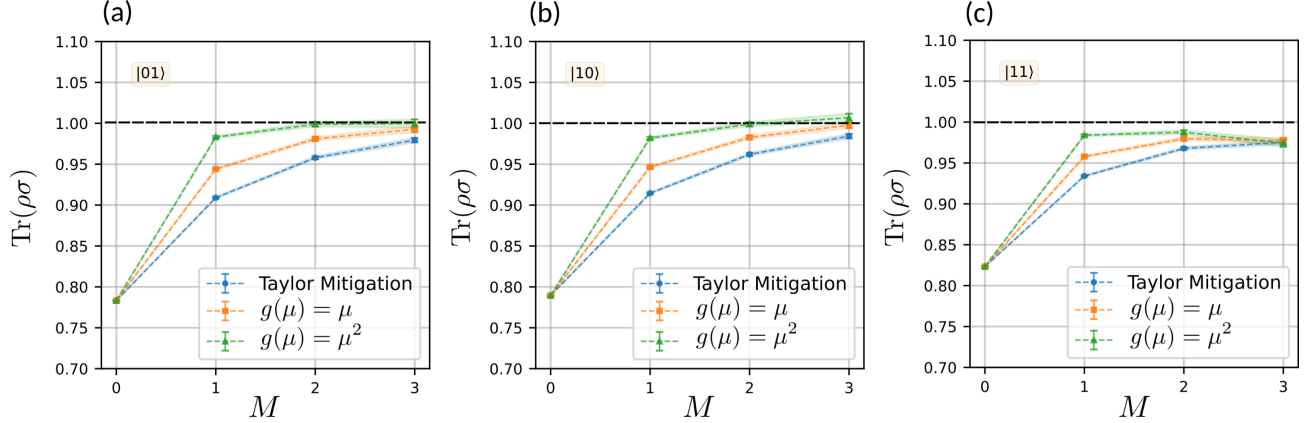
To increase the precision of the calibration, the experiment was performed using a sequence of 11 CNOT gates. Ideally, this circuit is equivalent to a single CNOT, and the repetition of CNOTs has the effect of amplifying the variations of $\langle Y_1 \rangle$ associated with different values of F . Accordingly, the inverse $\mathcal{K}_I$ for the left circuit in Supplementary Figure 7(b) consists of a sequence of gates such that each of them is the pulse inverse of a single CNOT.

In the processor Jakarta, the maximum number of shots per circuit is 32000. For the mitigation of readout errors, we repeated four times the circuits employed in the preparation of each computational state, which yields a total of $4 \times 32000 = 128000$ shots used to estimate each probability distribution $\{p(\mathbf{l}|\mathbf{k})\}_{\mathbf{l}}$. The circuits $\mathcal{K}(\mathcal{K}_I \mathcal{K})^m$, used for QEM, were preceded by any of the four rotation operations used to mitigate the preparation coherent error. These rotations were executed before the preparation of the initial state, as seen in Supplementary Figure 7(a). We repeated two times the circuit corresponding to any rotation. Hence, $8 \times 32000 = 256000$ shots were employed to measure the expectation value of Y_1 on each $\mathcal{K}(\mathcal{K}_I \mathcal{K})^m$, for each amplitude F . Since we applied KIK mitigation up to order 3, the value of m varies between 0 and 3. These experimental details are summarized in Supplementary Table 2.

Ten-swap experiment. In this experiment, the error-mitigated quantity was the probability for the system to remain in the initial state $|00\rangle$, after applying a sequence of ten swap gates. The IBM processor employed was Quito. In practice, three CNOT gates were used to implement each swap, as shown in Fig. 4(b) of the main text. Apart from the application of RC, detailed below, we reduced the impact of coherent errors by alternating a normal swap with a pulse-based inverse swap (see Supplementary Figure 7(b)). This strategy has been previously applied to gates that are their own inverse, with the purpose of mitigating coherent errors such as overrotations¹⁷. Since a swap

	shots per KIK circuit	shots for readout mitigation	RC operations
CNOT calibration experiment	256000	512000	n/a
ten-swap experiment	320000	240000	16

Supplementary Table 2. Summary of experimental resources. The ‘shots per KIK circuit’ are the number of shots associated with each circuit $\mathcal{K}(\mathcal{K}_1\mathcal{K})^m$, for $0 \leq m \leq 3$. For both experiments, these shots include the initial rotations, and also RC operations in the case of the ten-swap experiment. The ‘shots for readout mitigation’ are the total number of shots used to compute $\mathbf{M}^{-1}$.



Supplementary Figure 8. Experimental QEM in the IBM processor Quito. Error-mitigated survival probability for the ten-swap experiment, for initial states $|01\rangle$ (a), $|10\rangle$ (b), and $|11\rangle$ (c). The ideal survival probability is 1 (dashed black line). Green and orange curves show adaptive QEM, and the blue curve stands for mitigation assuming weak noise (Taylor mitigation). The thin colored areas connect small experimental error bars. Here, we see again that Taylor mitigation is outperformed by adaptive mitigation.

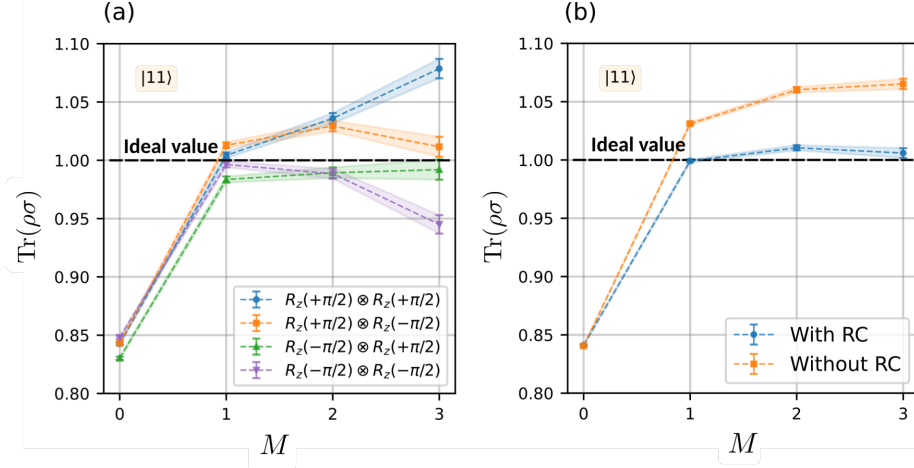
gate is its own inverse, our ideal target circuit is not modified by interleaving standard swaps with their pulse-inverse counterparts.

Let us now describe in more detail the distribution of the different circuits for this experiment. As in the previous case, any of the circuits $\mathcal{K}(\mathcal{K}_1\mathcal{K})^m$ was preceded by one of the four rotation operations applied on the state $|00\rangle$. For $0 \leq m \leq 3$, a circuit $\mathcal{K}(\mathcal{K}_1\mathcal{K})^m$ is accompanied by a rotation operation and a RC operation chosen at random from Supplementary Table 1. We repeated four times each possible rotation, which results in a total of 16 circuits per each value of m , involving 16 (not necessarily different) RC realizations. Keeping in mind that the maximum number of shots for the IBM Quito device is 20000, the total number of shots used for each circuit $\mathcal{K}(\mathcal{K}_1\mathcal{K})^m$ was $16 \times 20000 = 320000$. For readout mitigation, the circuits that prepare the computational states were repeated three times. Therefore, $3 \times 20000 = 60000$ shots were employed to estimate each probability distribution $\{p(\mathbf{l}|\mathbf{k})\}_{\mathbf{l}}$.

We also include here the error mitigation curves for the initial states $|01\rangle$, $|10\rangle$, and $|11\rangle$, which complement the curves shown in Fig. 4(a) of the main text. These plots are given in Supplementary Figure 8. As in the case of the state $|00\rangle$, we can see that adaptive mitigation with the function $g(\mu) = \mu^2$ leads to the best results. We also remark that the slightly worse results corresponding to the state $|11\rangle$ may be due to several experimental factors. Namely, imperfect averaging of the initial state (see Eq. (103)), drifts in gate parameters that characterize the single-qubit gates used for RC, and drifts in the measurement matrix $\mathbf{M}$. However, for the present experiment we do not have sufficient information for determining the most dominant factor.

In order to study the effect of coherent errors in the ten-swap circuit, we executed an additional experiment on the IBM processor Quito. As in the case of Supplementary Figure 8(c), the quantity measured was the survival probability for the system to remain in the initial state $|11\rangle$. The results presented in Supplementary Figure 9 were obtained by applying adaptive mitigation with $g(\mu) = \mu^2$. Coherent errors are manifested in Supplementary Figure 9(a) by the significant separation between different QEM curves, which correspond to different initial rotations. Furthermore, Supplementary Figure 9(b) shows how the application of RC (blue curve) produces substantially more accurate QEM, for all $1 \leq M \leq 3$. It is also worth stressing that in the absence of KIK QEM the enhancement provided by RC disappears, as evidenced by the matching of the blue and orange curves at $M = 0$.

Each curve in Supplementary Figure 9(a) shows the survival probability for a given rotation operation, obtained



Supplementary Figure 9. Effect of coherent errors on the error-mitigated survival probability for a ten-swap experiment. The initial state is $|11\rangle$ and the ideal survival probability is 1 (dashed black lines). The thin colored areas connect small experimental error bars. All the plots show adaptive mitigation using $g(\mu) = \mu^2$. (a) Error-mitigated survival probability for different initial rotations. By comparing these results with the blue curve in Supplementary Figure 9(b), it becomes clear that averaging over the initial rotations substantially mitigates the preparation coherent error and enhances the QEM. (b) Error-mitigated survival probabilities obtained with RC (blue curve) and without RC (orange curve). The QEM accuracy is significantly increased by applying RC.

by averaging over the associated RC realizations. Thus, if use the subscript r to label any of the four possible initial rotations, and i_r to label RC realizations that accompany the rotation r , for M th order mitigation the corresponding survival probability is given by

$$\langle O \rangle_{M,r} = \sum_{m=0}^M a_{\text{Adap},m}^{(M)} (\mu^2) \langle O \rangle_{m,r}, \quad (110)$$

where $O = \rho = |00\rangle\langle 00|$ and

$$\begin{aligned} \langle O \rangle_{m,r} &= \frac{1}{4} \sum_{i_r=1}^4 \langle O \rangle_{m,i_r} \\ &= \frac{1}{4} \sum_{i_r=1}^4 \sum_{\mathbf{k}} p_{\mathbf{k}}^{(m,i_r)} O_{\mathbf{k}}. \end{aligned} \quad (111)$$

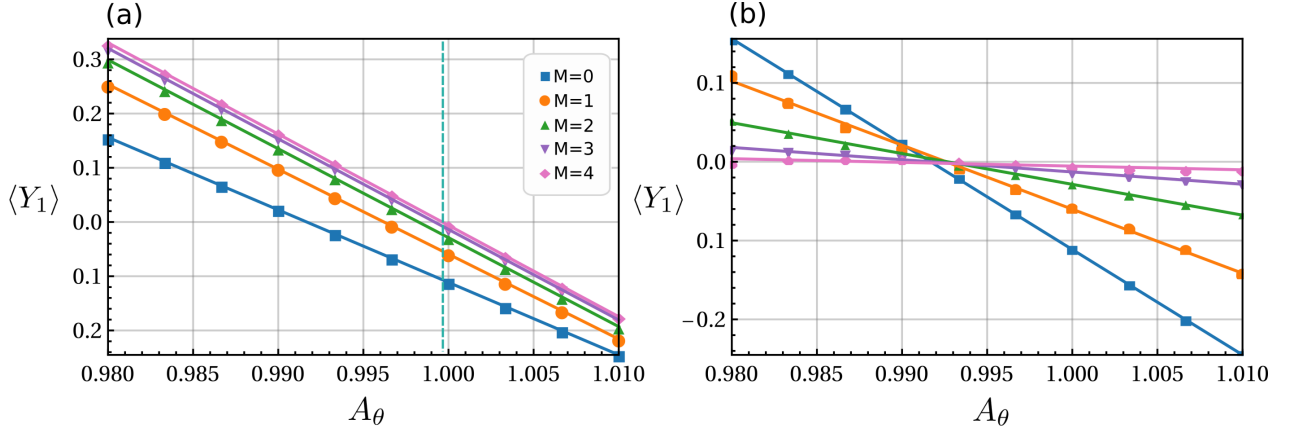
Once again, the calculation of the corresponding variances takes into account the independence of different RC realizations for a given rotation.

As compared to Supplementary Figure 8(c), the blue curve in Supplementary Figure 9(b) features a performance more consistent with that observed in Supplementary Figures 8(a) and 8(b). This provides strong evidence that the modest performance observed in Supplementary Figure 8(c) is not intrinsic to the state $|11\rangle$, but possibly related to experimental factors already mentioned. We also remark that the orange curve is obtained by omitting the application of RC in any of the four repetitions of each initial rotation operation. The construction of both curves in Supplementary Figure 9(b) involves the same number of shots per KIK circuit, as per Supplementary Table 2.

Simulation of a noisy calibration of a CNOT gate, using the KIK method. To complement the CNOT calibration experiment, we present now a simulation of the same calibration process. The ideal CNOT gate is simulated through the unitary evolution

$$U = R_Z^{(0)}(-\pi/2) e^{-i\frac{\theta}{2} H_{\text{CR}}} R_X^{(1)}(-\pi/2), \quad (112)$$

with the cross-resonance interaction $H_{\text{CR}} = Z \otimes X$, and the rotations $R_Z^{(0)}(-\pi/2) = e^{i\frac{\pi}{4}Z}$ and $R_X^{(1)}(-\pi/2) = e^{i\frac{\pi}{4}X}$ acting on the target qubit and the control qubit, respectively. The angle θ depends on the strength and the duration



Supplementary Figure 10. Simulation of the calibration of a noisy CNOT gate, for noise strength $\xi = 2/100$. The curves show the expectation value of $Y_1 = I \otimes Y$ as a function of the angle amplitude A_θ , which determines the angle $\theta = A_\theta \frac{\pi}{2}$ in the cross-resonance interaction H_{CR} (cf. Eq. (112)). The calibrated amplitude corresponds to $\langle Y_1 \rangle = 0$ (cyan dashed line). Different curves stand for different mitigation orders $0 \leq M \leq 4$. Figures (a) and (b) correspond to the application of the pulse inverse and the circuit inverse $\mathcal{K}_I = \mathcal{K}$, respectively.

of the pulse used to generate the cross-resonance interaction. An ideal (noise-free) CNOT gate corresponds to the angle $\theta = \pi/2$.

For the effect of noise, we consider the dissipator

$$\mathcal{L} = \xi \sum_{i=1}^2 \gamma_i \left[A_i \otimes A_i^* - \frac{1}{2} A_i^\dagger A_i \otimes I - \frac{1}{2} I \otimes (A_i^\dagger A_i)^T \right], \quad (113)$$

$$A_1 = Z, \quad (114)$$

$$A_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad (115)$$

where $\gamma_1 = 1$ and $\gamma_2 = 1/10$. In Supplementary Figures 10 and 11 we show the calibration curves for $\xi = 2/100$ (Supplementary Figure 10) and $\xi = 1/100$ (Supplementary Figure 11). Since the action of noise produces an imperfect CNOT gate, the angle θ obtained under a noisy calibration is in general different from $\pi/2$. We express this angle as $\theta = A_\theta \frac{\pi}{2}$, and take the “angle amplitude” A_θ as the parameter for calibration.

As in the experimental case, the initial state and observable used for calibration are $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \otimes |0\rangle$ and $Y_1 = I \otimes Y$, respectively. Furthermore, the target circuit consists of a sequence of 11 CNOTs, in order to increase the sensitivity of $\langle Y_1 \rangle$ to variations of A_θ . Therefore, the noisy target circuit and its inverse are given by

$$\mathcal{K} = \left(e^{-i \frac{\pi}{2} A_\theta \mathcal{H}_{CR} + \mathcal{L}} \right)^{11}, \quad (116)$$

$$\mathcal{K}_I = \left(e^{i \frac{\pi}{2} A_\theta \mathcal{H}_{CR} + \mathcal{L}} \right)^{11}, \quad (117)$$

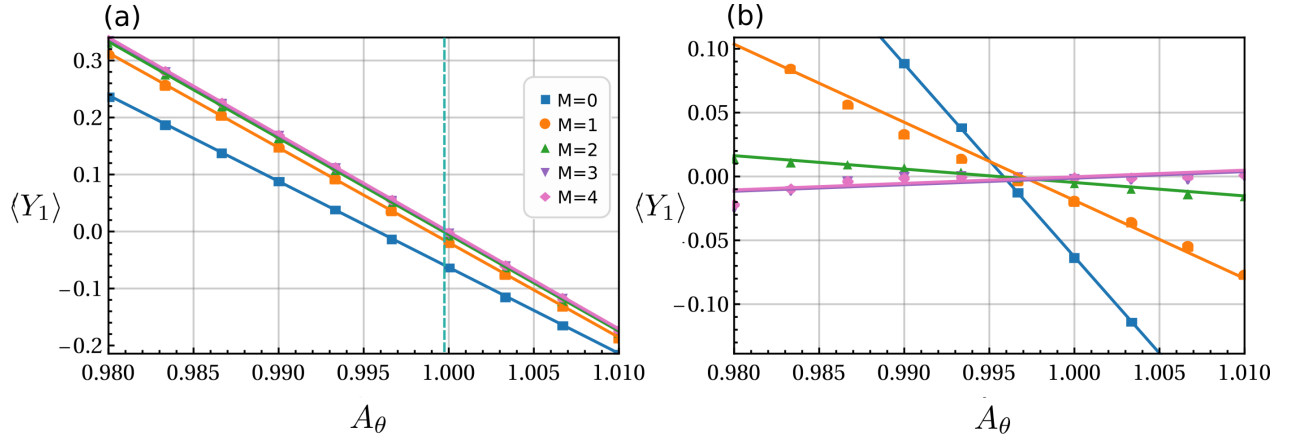
where $\mathcal{H}_{CR}$ is the Liouville space representation of H_{CR} .

The amplitudes A_θ obtained for error mitigation orders $0 \leq M \leq 4$ are given in Supplementary Table 3. Due to the small values of the noise strength ξ , we apply Taylor mitigation with the coefficients $a_{\text{Taylor}, m}^{(M)}$. In Supplementary Table 3, we can observe that A_θ tends to one as M increases. This reflects the fact that higher mitigation orders produce an evolution $\mathcal{U}_{\text{KIK}}^{(M)}$ closer to the ideal CNOT gate, and therefore the resulting calibrated angle θ also approaches its noise-free value $\theta = \pi/2$. On the other hand, for the noisy calibration corresponding to $M = 0$, the associated angle presents the largest deviation from $\theta = \pi/2$. As a consequence, such a biased angle gives rise to a coherent error.

To test the calibrated gate, we simulate the average gate fidelity after applying M th order Taylor mitigation to the noisy evolution

$$\mathcal{K}_{\text{CNOT}} = e^{-i \frac{\theta_M}{2} \mathcal{H}_{CR} + \mathcal{L}}, \quad (118)$$

where θ_M is the angle calibrated via M th order mitigation. In other words, this strategy consists of using the same order of mitigation for the calibration of the CNOT and for the evaluation of its average fidelity.



Supplementary Figure 11. Simulation of the calibration of a noisy CNOT gate, for noise strength $\xi = 1/100$. The curves show the expectation value of $Y_1 = I \otimes Y$ as a function of the angle amplitude A_θ , which determines the angle $\theta = A_\theta \frac{\pi}{2}$ in the cross resonance interaction H_{CR} (cf. Eq. (112)). The calibrated amplitude corresponds to $\langle Y_1 \rangle = 0$ (cyan dashed line). Different curves stand for different mitigation orders $0 \leq M \leq 4$. Figures (a) and (b) correspond to the application of the pulse inverse and the circuit inverse $\mathcal{K}_1 = \mathcal{K}$, respectively.

M (mitigation order)	0	1	2	3	4
A_θ for $\xi = 2/100$	0.991671	0.996210	0.998235	0.999181	0.999631
A_θ for $\xi = 1/100$	0.995830	0.998836	0.999673	0.999909	0.999977

Supplementary Table 3. Angle amplitudes A_θ obtained with KIK-based calibration of a CNOT gate. A_θ determines the angle $\theta = A_\theta \frac{\pi}{2}$ in Eq. (112), and $A_\theta = 1$ for a noise-free CNOT. ξ gives the noise strength in the dissipator (113).

The average gate fidelity between a quantum channel Λ and a unitary evolution U is defined by

$$F(\Lambda, U) = \int d\psi \langle \psi | U^\dagger \Lambda(|\psi\rangle\langle\psi|) U | \psi \rangle, \quad (119)$$

where the integral is taken over all the pure states $|\psi\rangle$ with respect to the Haar measure. Here, $\Lambda(|\psi\rangle\langle\psi|)$ denotes the state obtained by applying Λ on $|\psi\rangle\langle\psi|$. For quantum channels acting on N qubits, the fidelity (119) can be computed by using the Pauli transfer matrices of Λ and U . The Pauli transfer matrix R_Λ for a general quantum channel Λ has elements¹⁸

$$(R_\Lambda)_{ij} = \frac{1}{d} \text{Tr} [P_i \Lambda(P_j)], \quad (120)$$

where $P_i, P_j \in \{I, X, Y, Z\}^{\otimes N}$ are Pauli operators and $d = 2^N$ is the dimension of the Hilbert space.

In this way, the fidelities F_M in Supplementary Tables 4 and 5 are calculated as¹⁸

$$F(\Lambda, U) = \frac{\text{Tr}(R_\Lambda^{-1} R_U) + d}{d(d+1)}, \quad (121)$$

with $U = e^{-i\frac{\theta}{2} H_{\text{CR}}}$ and the channel Λ corresponding to M th order mitigation,

$$\mathcal{U}_{\text{KIK}}^{(M)} = \sum_{m=0}^M a_{\text{Tay}, m}^{(M)} \mathcal{K}_{\text{CNOT}} (\mathcal{K}_{\text{I,CNOT}} \mathcal{K}_{\text{CNOT}})^m, \quad (122)$$

$$\mathcal{K}_{\text{I,CNOT}} = e^{i\frac{\theta M}{2} \mathcal{H}_{\text{CR}} + \mathcal{L}}. \quad (123)$$

Using the Liouville space formalism, the matrix elements of R_Λ and R_U are given by

$$(R_\Lambda)_{ij} = \frac{1}{d} \langle P_i | \mathcal{U}_{\text{KIK}}^{(M)} | P_j \rangle, \quad (124)$$

$$(R_U)_{ij} = \frac{1}{d} \langle P_i | \mathcal{U} | P_j \rangle, \quad (125)$$

M (mitigation order)	F_0 (noisy calibration)	F_M (KIK-based calibration)	F_0 with RC
0	0.967325	0.967325	0.967325
1	0.997270	0.997298	0.997388
2	0.999692	0.999725	0.999743
3	0.999934	0.999968	0.999972
4	0.999961	0.999996	0.999997

Supplementary Table 4. KIK-mitigated gate fidelity (121) (simulation) for a CNOT gate affected by noise (113), with noise strength $\xi = 2/100$. F_0 is the fidelity obtained when the gate is calibrated without KIK error mitigation. F_M is the fidelity when M th order mitigation is performed alongside M th order KIK calibration. “ F_0 with RC” refers to the fidelity obtained without KIK-based calibration, and by combining KIK mitigation with RC. In this case the fidelity is the average over fidelities associated with all the 16 RC operations described in Supplementary Table 1.

M (mitigation order)	F_0 (noisy calibration)	F_M (KIK-based calibration)	F_0 with RC
0	0.983434	0.983434	0.983434
1	0.999288	0.999296	0.999320
2	0.999954	0.999963	0.999965
3	0.999989	0.999998	0.999998
4	0.999991	1.000000	1.000000

Supplementary Table 5. KIK-mitigated gate fidelity (121) (simulation) for a CNOT gate affected by noise (113), with noise strength $\xi = 1/100$. The description of each column is as in Supplementary Table 4.

where $\mathcal{U} = U \otimes U^*$ and $|P_j\rangle$ is the vector representation of P_j . On the other hand, F_0 is computed by replacing $\mathcal{K}_{\text{CNOT}}$ and $\mathcal{K}_{\text{I,CNOT}}$ in Eq. (122), by $e^{-i\frac{\theta_0}{2}\mathcal{H}_{\text{CR}}+\mathcal{L}}$ and $e^{i\frac{\theta_0}{2}\mathcal{H}_{\text{CR}}+\mathcal{L}}$, respectively.

We choose Taylor mitigation because the corresponding coefficients do not depend on the initial state. In contrast, in the case of adaptive mitigation the quantity $\mu = \langle \rho | \mathcal{K}_{\text{I,CNOT}} \mathcal{K}_{\text{CNOT}} | \rho \rangle$ depends on the initial state ρ , which makes it difficult to take into account when evaluating the gate fidelity.

Supplementary Tables 4 and 5 show the simulated fidelities for $\xi = 2/100$ and $\xi = 1/100$, respectively. The second column in these tables gives the error-mitigated fidelities without a KIK-based calibration, i.e. when $\theta_M = \theta_0$. The third column contains the values simulated by following the M -order mitigation strategy previously described. Finally, fidelities obtained without KIK-based calibration and RC integrated into the KIK mitigation are presented in the fourth column. Both tables show that KIK-based calibration and RC yield similar fidelities, which surpass the values obtained when none of these methods is employed. The effect of these two approaches is different though. On the one hand, RC mitigates the coherent error that characterizes the biased angle θ_0 by postprocessing measurement data. On the other hand, in the KIK-based calibration this coherent error is reduced at a physical level.

SUPPLEMENTARY NOTE 7: NUMERICAL EXAMPLE TO ILLUSTRATE SATURATION OF THE KIK FORMULA

In this section, we consider a numerical example where the accuracy of QEM using the KIK method saturates at approximately the mitigation order $M = 4$. We quantify the accuracy by the relative error

$$\frac{\varepsilon_{\text{KIK}}^{(M)}}{|\langle A | \mathcal{U} | \rho \rangle|} = \frac{|\langle A | \mathcal{U} | \rho \rangle - \langle A | \mathcal{U}_{\text{KIK}}^{(M)} | \rho \rangle|}{|\langle A | \mathcal{U} | \rho \rangle|}, \quad (126)$$

where $\varepsilon_{\text{KIK}}^{(M)} = |\langle A | \mathcal{U} | \rho \rangle - \langle A | \mathcal{U}_{\text{KIK}}^{(M)} | \rho \rangle|$ is the difference between the ideal expectation value $\langle A | \mathcal{U} | \rho \rangle$, and the error-mitigated expectation value $\langle A | \mathcal{U}_{\text{KIK}}^{(M)} | \rho \rangle$, for the observable A and the initial state ρ . In our example, the error (126) first decreases as a function of M , and then attains a value approximately constant for $M \geq 4$. This is shown in Supplementary Figure 12. We apply QEM with Taylor mitigation, characterized by $\mathcal{U}_{\text{KIK}}^{(M)} = \sum_{m=0}^M a_{\text{Taylor},m}^{(M)} \mathcal{K}(\mathcal{K}_{\text{I}}\mathcal{K})^m$, and the coefficients $a_{\text{Taylor},m}^{(M)}$ in Eq. (61).

Our example refers to a four-qubit system characterized by the time-independent Hamiltonian

$$H = X \otimes X \otimes I \otimes I + I \otimes X \otimes X \otimes I + I \otimes I \otimes X \otimes X, \quad (127)$$

$$\mathcal{H} = H \otimes I - I \otimes H^T, \quad (128)$$

where $X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ and $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Here, $\mathcal{H}$ is the Liouville space representation of H , taking into account Eq. (5). Each qubit is subjected to spontaneous emission, which can be described via a GKLS (Gorini-Kossakowski-Sudarshan-Lindblad) master equation with dissipator $\hat{L}$ such that

$$\hat{L}[\rho] = \sum_{k=1}^4 \xi_k \left(A_k \rho A_k^\dagger - \frac{1}{2} A_k^\dagger A_k \rho - \frac{1}{2} \rho A_k^\dagger A_k \right), \quad (129)$$

$$A_1 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \otimes I \otimes I \otimes I, \quad (130)$$

$$A_2 = I \otimes \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \otimes I \otimes I, \quad (131)$$

$$A_3 = I \otimes I \otimes \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \otimes I, \quad (132)$$

$$A_4 = I \otimes I \otimes I \otimes \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}. \quad (133)$$

We assume that all the qubits are affected by the same relaxation rate $\xi = \xi_k = 0.02$, for $1 \leq k \leq 4$. Using the vectorization rule (cf. Eq. (5)), in Liouville space the dissipator (129) takes the form

$$\mathcal{L} = \xi \sum_{k=1}^4 \left(A_k \otimes A_k^* - \frac{1}{2} A_k^\dagger A_k \otimes I - \frac{1}{2} I \otimes A_k^\dagger A_k \right), \quad (134)$$

where A_k^* denotes the complex conjugate of A_k .

Assuming a time unit $T = 1$ for the evolution time, we have that

$$\mathcal{U} = e^{-i\mathcal{H}}, \quad (135)$$

$$\mathcal{K} = e^{-i\mathcal{H} + \mathcal{L}}, \quad (136)$$

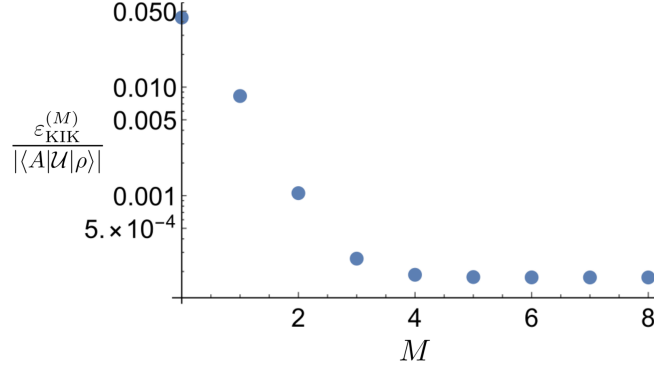
$$\mathcal{K}_1 = e^{i\mathcal{H} + \mathcal{L}}. \quad (137)$$

The total system starts in the ground state $\rho = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}^{\otimes 4}$, whose vector form is given by $|\rho\rangle = (1, 0, 0, \dots, 0)^T$. The observable A is the projector onto the ground state, and thus $\langle A | = (1, 0, 0, \dots, 0)$.

The saturation of the relative error can be understood as being a consequence of the approximations involved in the derivation of the KIK formula. More specifically, it is related to the fact that in the derivation of Eq. (26) we have discarded higher-order Magnus terms $\Omega_{n \geq 2}$. Nevertheless, it is worth noting that, according to Supplementary Figure 12, a relative error as small as $\sim 10^{-4}$ can already be achieved with a mitigation order $M = 3$. When the initial error is smaller (i.e. when ξ is smaller), the improvement is even more substantial.

SUPPLEMENTARY NOTE 8: DERIVATION OF UPPER BOUNDS FOR THE PERFORMANCE OF THE KIK METHOD

A complete assessment of a QEM technique requires establishing bounds for the remaining error in the estimation of expectation values. Furthermore, these bounds should be applicable to general observables A and circuits $\mathcal{K}$ of arbitrary size. The goal of the present section is to derive such bounds, in the case of QEM using the KIK method. In the first two subsections we will state some technical results and physically realistic assumptions, which will be



Supplementary Figure 12. Log scale plot of the relative error (126), for the QEM example described in the present supplementary note.

used for the derivation of our bounds in the last two subsections.

Lower bound on the smallest eigenvalue of $\mathcal{K}_I \mathcal{K}$. Assuming that $\mathcal{K}_I \mathcal{K}$ is diagonalizable, i.e. $\mathcal{K}_I \mathcal{K} = \sum_k \lambda_k |k\rangle\langle k|$, where $|k\rangle$ and $\langle k|$ denote respectively right and left eigenvectors (note that in general we may have $\langle k| \neq |k\rangle^\dagger$), we will obtain the bound

$$\min_k \lambda_k \geq e^{-2 \int_0^T \|\mathcal{L}(t)\| dt}. \quad (138)$$

We recall that $\mathcal{L}(t)$ is the dissipator introduced in Eq. (10), and T is the evolution time for $\mathcal{K}$. The norm $\|*\|$ in Eq (138) is the spectral norm.

First, we consider the equivalent expression of Eq. (20),

$$\frac{d}{dt} \tilde{\mathcal{K}}_{\text{int}}(t) = \tilde{\mathcal{L}}_{\text{int}}(t) \tilde{\mathcal{K}}_{\text{int}}(t), \quad (139)$$

with the interaction-picture evolution $\tilde{\mathcal{K}}_{\text{int}}(t) = \tilde{\mathcal{U}}^\dagger(t) \mathcal{K}(t)$, and dissipator $\tilde{\mathcal{L}}_{\text{int}}(t) = \tilde{\mathcal{U}}^\dagger(t) \mathcal{L}(t) \tilde{\mathcal{U}}(t)$. We point out that, in this case, we consider only the time interval $0 \leq t \leq T$. The inverse evolution $\tilde{\mathcal{K}}_{\text{int}}^{-1}(t) = \tilde{\mathcal{K}}^{-1}(t) \tilde{\mathcal{U}}(t)$ satisfies the equation

$$\frac{d}{dt} \tilde{\mathcal{K}}_{\text{int}}^{-1}(t) = -\tilde{\mathcal{L}}_{\text{int}}(T-t) \tilde{\mathcal{K}}_{\text{int}}^{-1}(t), \quad (140)$$

which will be used in the derivation of (138). Here, it is important to remark that $\tilde{\mathcal{K}}_{\text{int}}^{-1}(t)$ is the mathematical inverse of $\tilde{\mathcal{K}}_{\text{int}}(t)$ and must not be confused with $\mathcal{K}_I$.

By taking the spectral norm at both sides of Eq. (140), we get

$$\begin{aligned} \left\| \frac{d}{dt} \tilde{\mathcal{K}}_{\text{int}}^{-1}(t) \right\| &= \left\| \tilde{\mathcal{L}}_{\text{int}}(T-t) \tilde{\mathcal{K}}_{\text{int}}^{-1}(t) \right\| \\ &\leq \left\| \tilde{\mathcal{L}}(T-t) \right\| \left\| \tilde{\mathcal{K}}^{-1}(t) \right\|, \end{aligned} \quad (141)$$

where the inequality follows from the submultiplicativity and unitary invariance of the spectral norm. On the other hand, we can apply the reverse triangle inequality $\|x - y\| \geq ||x| - |y||$ to obtain

$$\begin{aligned} \left\| \frac{d}{dt} \tilde{\mathcal{K}}_{\text{int}}^{-1}(t) \right\| &= \lim_{\Delta t \rightarrow 0} \left\| \frac{\tilde{\mathcal{K}}_{\text{int}}^{-1}(t + \Delta t) - \tilde{\mathcal{K}}_{\text{int}}^{-1}(t)}{\Delta t} \right\| \\ &\geq \lim_{\Delta t \rightarrow 0} \left| \frac{\left\| \tilde{\mathcal{K}}_{\text{int}}^{-1}(t + \Delta t) \right\| - \left\| \tilde{\mathcal{K}}_{\text{int}}^{-1}(t) \right\|}{\Delta t} \right| \\ &= \frac{d}{dt} \left\| \tilde{\mathcal{K}}_{\text{int}}^{-1}(t) \right\| = \frac{d}{dt} \left\| \tilde{\mathcal{K}}^{-1}(t) \right\|. \end{aligned} \quad (142)$$

By combining this result with Eq. (141), we have the inequality

$$\frac{d}{dt} \left\| \tilde{\mathcal{K}}^{-1}(t) \right\| \leq \left\| \tilde{\mathcal{L}}(T-t) \right\| \left\| \tilde{\mathcal{K}}^{-1}(t) \right\|, \quad (143)$$

which upon integration yields

$$\left\| \tilde{\mathcal{K}}^{-1}(T) \right\| \leq e^{\int_0^T \left\| \tilde{\mathcal{L}}(T-t) \right\| dt}. \quad (144)$$

Taking into account that $\tilde{\mathcal{K}}(T) = \mathcal{K}$ and $\tilde{\mathcal{L}}(T-t) = \mathcal{L}(T-t)$ for $0 \leq t \leq T$ (cf. Eq. (17)), the change of variable $t' = T - t$ straightforwardly leads to

$$\left\| \mathcal{K}^{-1} \right\| \leq e^{\int_0^T \left\| \mathcal{L}(t) \right\| dt}. \quad (145)$$

Consider now the singular value decomposition (SVD) $\mathcal{K} = \mathcal{V}\mathcal{S}\mathcal{W}$, where $\mathcal{S}$ is a diagonal matrix with the singular values of $\mathcal{K}$ and $\mathcal{V}$ and $\mathcal{W}$ are unitary matrices. Then, the SVD of $\mathcal{K}^{-1}$ reads $\mathcal{W}^\dagger \mathcal{S}^{-1} \mathcal{V}^\dagger$, and the spectral norm (maximum singular value) of $\mathcal{K}^{-1}$ is given by $\left\| \mathcal{K}^{-1} \right\| = 1/s_{\min}$, where $s_{\min}$ is the minimum singular value of $\mathcal{K}$. Accordingly, we can rewrite (145) as

$$s_{\min} \geq e^{-\int_0^T \left\| \mathcal{L}(t) \right\| dt}. \quad (146)$$

The singular values of $\mathcal{K}$ are the eigenvalues of $(\mathcal{K}^\dagger \mathcal{K})^{\frac{1}{2}}$, where $\mathcal{K}^\dagger$ is the Hermitian conjugate of $\mathcal{K}$. If the evolution $\mathcal{K}_I$ coincides with $\mathcal{K}^\dagger$, we have that $s_{\min}$ is also the minimum eigenvalue of $(\mathcal{K}_I \mathcal{K})^{\frac{1}{2}}$, or, equivalently, the minimum eigenvalue of $\mathcal{K}_I \mathcal{K}$ satisfies the inequality (138). Hence, our last step in proving this inequality is to show that $\mathcal{K}_I = \mathcal{K}^\dagger$, under sound physical conditions. This is done in the next subsection.

Sufficient condition for $\mathcal{K}_I = \mathcal{K}^\dagger$. Suppose that the dissipator $\mathcal{L}(t)$ is Hermitian, i.e. $\mathcal{L}^\dagger(t) = \mathcal{L}(t)$ for $0 \leq t \leq T$. Then, from the definition of the first-order Magnus term (22) it also follows that

$$\Omega_1^\dagger(T) = \Omega_1(T). \quad (147)$$

If we can prove that, up to first order in the Magnus expansion (where $\Omega(T)$ is approximated by $\Omega_1(T)$),

$$\mathcal{K}_I \approx e^{\Omega_1(T)} \mathcal{U}^\dagger, \quad (148)$$

then, Eqs. (23) and (148) imply that $\mathcal{K}_I = \mathcal{K}^\dagger$ within this approximation.

Before discussing under which circumstances the property $\mathcal{L}^\dagger(t) = \mathcal{L}(t)$ is satisfied, let us first derive Eq. (148). We start by writing the formal solution to the inverse evolution $\mathcal{K}_I$. Denoting the corresponding Magnus expansion by Ω_I , and assuming for simplicity that $\mathcal{L}_I(t)$ acts on the time interval $(0, T)$ (as opposed to the interval $(T, 2T)$, used in the derivation of the KIK formula), Eq. (23) leads to

$$\begin{aligned} \mathcal{K}_I &= \mathcal{U}^\dagger e^{\Omega_I(T)} \\ &\approx \mathcal{U}^\dagger e^{\Omega_{I,1}(T)}, \end{aligned} \quad (149)$$

with the first-order Magnus term

$$\begin{aligned} \Omega_{I,1}(T) &= \int_0^T \mathcal{L}_{I,\text{int}}(t) dt \\ &= \int_0^T \mathcal{U}(t) \mathcal{L}(T-t) \mathcal{U}^\dagger(t) dt. \end{aligned} \quad (150)$$

Writing $\mathcal{U}(t)$ as $\mathcal{U}(t) = \mathcal{U}(T) \mathcal{U}^\dagger(T-t)$, and performing the change of variable $t' = T - t$ in the integral, we obtain

$$\begin{aligned} \Omega_{I,1}(T) &= -\mathcal{U}(T) \left(\int_T^0 \mathcal{U}^\dagger(t') \mathcal{L}(t') \mathcal{U}(t') dt' \right) \mathcal{U}^\dagger(T) \\ &= \mathcal{U} \left(\int_0^T \mathcal{U}^\dagger(t') \mathcal{L}(t') \mathcal{U}(t') dt' \right) \mathcal{U}^\dagger \\ &= \mathcal{U} \Omega_1(T) \mathcal{U}^\dagger. \end{aligned} \quad (151)$$

Note that we also used the simplified notation $\mathcal{U}(T) = \mathcal{U}$. In this way, we conclude that

$$\begin{aligned}\mathcal{K}_I &\approx \mathcal{U}^\dagger e^{\mathcal{U}\Omega_1(T)\mathcal{U}^\dagger} \\ &= e^{\Omega_1(T)}\mathcal{U}^\dagger.\end{aligned}\tag{152}$$

Thus, the hermiticity of $\mathcal{L}(t)$ leads to $\mathcal{K}_I = \mathcal{K}_I^\dagger$, within our approximation.

Now, let us come back to the subject of the Hermiticity of $\mathcal{L}(t)$. A sufficient condition to have this property is that

$$\mathcal{L}(t) = \sum_{\vec{k}} \alpha_{\vec{k}}(t) \left(P_{\vec{k}} \otimes P_{\vec{k}}^T - \mathcal{I} \right), \tag{153}$$

where $P_{\vec{k}} = \sigma_{\mathbf{k}_1}^{(1)} \otimes \sigma_{\mathbf{k}_2}^{(2)} \otimes \dots \otimes \sigma_{\mathbf{k}_n}^{(n)} \otimes \dots \otimes \sigma_{\mathbf{k}_N}^{(N)}$ is a Pauli operator acting on N qubits, $\sigma_{\mathbf{k}_n}^{(n)} \in \{I, X, Y, Z\}$ is a Pauli matrix acting on the n th qubit, and the time-dependent coefficients $\alpha_{\vec{k}}(t)$ are real. We note that a time-independent version of this dissipator was recently considered in Ref.¹⁹. For a given state ρ , we can check by direct application of the identity (5) that Eq. (153) is the Liouville space representation of the superoperator $\hat{L}(t)$ such that

$$\hat{L}(t)[\rho] = \sum_{\vec{k}} \alpha_{\vec{k}}(t) (P_{\vec{k}} \rho P_{\vec{k}} - \rho). \tag{154}$$

The noise model of Ref.¹⁹ relies on a time-independent $\hat{L}(t) = \hat{L}$, characterized by time-independent coefficients $\alpha_{\vec{k}}(t) = \alpha_{\vec{k}}$. In addition, $\alpha_{\vec{k}} \neq 0$ only if the index $\vec{k} = \{\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_N\}$ contains at most two components $\mathbf{k}_n$ different from 0, which limits correlated errors to occur at most between pairs of qubits. On the other hand, the more general dissipator (153) guarantees the condition $\mathcal{K}_I = \mathcal{K}_I^\dagger$. We remark that the time-dependent coefficients $\alpha_{\vec{k}}(t)$ in Eq. (153) can also be different from zero for arbitrary indices $\vec{k}$, thereby allowing for correlated errors between any number of qubits in the system.

The Pauli channel (153) arises naturally when executing RC for the purpose of mitigating coherent errors. In others words, by applying RC on a general dissipator, which can contain non-diagonal components in the Pauli operator representation, a dissipator of the form (153) is obtained. Since we exploit RC in our experiments, the effective noise is described by a Pauli channel, which according to the above results implies $\mathcal{K}_I = \mathcal{K}_I^\dagger$.

First error bounds for Adaptive mitigation and Taylor mitigation. To begin this subsection, we recall that the error-mitigated evolution $\mathcal{U}_{\text{KIK}} = \mathcal{K} (\mathcal{K}_I \mathcal{K})^{-\frac{1}{2}}$ is implemented through M -order approximations of the form $\mathcal{U}_{\text{KIK}}^{(M)} = \sum_{m=0}^M a_m^{(M)} \mathcal{K} (\mathcal{K}_I \mathcal{K})^m$. Our goal is to obtain an upper bound on the error

$$\varepsilon_{\text{KIK}}^{(M)} = \left| \langle A | \mathcal{U} | \rho \rangle - \langle A | \mathcal{U}_{\text{KIK}}^{(M)} | \rho \rangle \right|, \tag{155}$$

which quantifies how much the ideal expectation value $\langle A | \mathcal{U} | \rho \rangle$ deviates from the error-mitigated expectation value $\langle A | \mathcal{U}_{\text{KIK}}^{(M)} | \rho \rangle$. Here, A is an arbitrary observable and ρ is an arbitrary initial state.

We will start by deriving a bound for the error associated with adaptive mitigation, with the coefficients $a_{\text{Adap},m}^{(M)}[g(\mu)]$ evaluated at $g(\mu) = \mu$, and $1 \leq M \leq 3$. The possibility of obtaining tighter bounds with different choices of $g(\mu)$ is left as an open problem. Later on, we will derive another bound that, despite being looser, has the advantage of being independent of μ , and is also valid for both adaptive mitigation with $g(\mu) = \mu$ and Taylor mitigation.

Using our conventional approximations $\mathcal{K} \approx \mathcal{U} e^{\Omega_1}$ and $\mathcal{K}_I \mathcal{K} \approx e^{2\Omega_1}$, we can write $\mathcal{U}_{\text{KIK}}^{(M)}$ as

$$\begin{aligned}\mathcal{U}_{\text{KIK}}^{(M)} &\approx \mathcal{U} \sum_{m=0}^M a_{\text{Adap},m}^{(M)}(\mu) e^{(2m+1)\Omega_1} \\ &\approx \mathcal{U} \sum_{m=0}^M a_{\text{Adap},m}^{(M)}(\mu) (\mathcal{K}_I \mathcal{K})^{m+1/2}.\end{aligned}\tag{156}$$

Therefore,

$$\begin{aligned}
\varepsilon_{\text{KIK}}^{(M)} &= \left| \langle A | \left(\mathcal{U} - \mathcal{U}_{\text{KIK}}^{(M)} \right) | \rho \rangle \right| \\
&\leq \sqrt{\langle A | A \rangle} \sqrt{\langle \rho | \rho \rangle} \left\| \mathcal{U} \left(\mathcal{I} - \sum_{m=0}^M a_{\text{Adap},m}^{(M)}(\mu) (\mathcal{K}_I \mathcal{K})^{m+1/2} \right) \right\| \\
&= \sqrt{\langle A | A \rangle} \left\| \left(\mathcal{I} - \sum_{m=0}^M a_{\text{Adap},m}^{(M)}(\mu) (\mathcal{K}_I \mathcal{K})^{m+1/2} \right) \right\|,
\end{aligned} \tag{157}$$

where the first inequality follows from the definition (156) and the definition of the spectral norm $\|\cdot\|$. In the last line, we use the unitary invariance of this norm. Moreover, we assume an initial pure state ρ , which implies $\langle \rho | \rho \rangle = 1$. Assuming as before that $\mathcal{L}(t)$ satisfies Eq. (153), and hence $\mathcal{K}_I = \mathcal{K}^\dagger$, we have that the operator $\mathcal{K}_I \mathcal{K}$ is Hermitian. Since this implies that $\mathcal{I} - \sum_{m=0}^M a_{\text{Adap},m}^{(M)}(\mu) (\mathcal{K}_I \mathcal{K})^{m+1/2}$ is also Hermitian, the corresponding spectral norm is simply the absolute value of its maximum eigenvalue. In this way, Eq. (157) leads to

$$\varepsilon_{\text{KIK}}^{(M)} \leq \sqrt{\langle A | A \rangle} \max_k \left| 1 - \sum_{m=0}^M a_{\text{Adap},m}^{(M)}(\mu) (\lambda_k)^{m+1/2} \right|, \tag{158}$$

where we have again denoted the eigenvalues of $\mathcal{K}_I \mathcal{K} = \sum_k \lambda_k |k\rangle\langle k|$ by $\{\lambda_k\}$. Since these eigenvalues may be difficult to evaluate both experimentally and theoretically, we establish another upper bound to $\varepsilon_{\text{KIK}}^{(M)}$ that depends on the single quantity $\int_0^T \|\mathcal{L}(t)\| dt$, instead of all the eigenvalues λ_k .

We start by looking at the behavior of the function $f_M(\mu, \lambda) := \left| 1 - \sum_{m=0}^M a_{\text{Adap},m}^{(M)}(\mu) (\lambda)^{m+1/2} \right|$ in the plots of Supplementary Figure 13. This will allow us to derive the bound

$$\varepsilon_{\text{KIK}}^{(M)} \leq \sqrt{\langle A | A \rangle} \left| 1 - \sum_{m=0}^M a_{\text{Adap},m}^{(M)}(\mu) \left(e^{-2 \int_0^T \|\mathcal{L}(t)\| dt} \right)^{m+1/2} \right|, \tag{159}$$

for $M = 1, 2, 3$. To this end, we note that the right hand side of Eq. (158) can be written as $\sqrt{\langle A | A \rangle} \max_k f_M(\mu, \lambda_k)$. Thus, we will obtain the bound (159) by proving the inequality $\max_k f_M(\mu, \lambda_k) \leq f_M(\mu, e^{-2 \int_0^T \|\mathcal{L}(t)\| dt})$, for $M = 1, 2, 3$. First, we will show that this inequality holds when the eigenvalue that maximizes $f_M(\mu, \lambda_k)$ is smaller than μ , and then the proof will be extended to the complementary case (i.e. when the aforementioned eigenvalue is larger than μ).

For the first case we resort to the plots (a), (b) and (c) in Supplementary Figure 13. These plots show that if $\lambda \leq \mu$, the function $f_M(\mu, \lambda)$ is monotonically decreasing with respect to λ . Therefore, the inequality $\min_k \lambda_k \geq e^{-2 \int_0^T \|\mathcal{L}(t)\| dt}$ (cf. Eq. (138)) implies that

$$f_M(\mu, \lambda_k) \leq f_M(\mu, \min_k \lambda_k) \leq f_M\left(\mu, e^{-2 \int_0^T \|\mathcal{L}(t)\| dt}\right), \tag{160}$$

for any eigenvalue λ_k such that $\lambda_k \leq \mu$.

If $\lambda \geq \mu$, the plots (d), (e) and (f) in Supplementary Figure 13 show that $f_M(\mu, \lambda) \leq f_M(\mu, \mu)$. Since

$$\mu = \langle \rho | \mathcal{K}_I \mathcal{K} | \rho \rangle = \sum_k \lambda_k |\langle k | \rho \rangle|^2 \geq \min_k \lambda_k, \tag{161}$$

the monotonicity observed in Supplementary Figures 13(a)-13(c) leads to $f_M(\mu, \lambda) \leq f_M(\mu, \mu) \leq f_M(\mu, \min_k \lambda_k)$, where the rightmost inequality is a consequence of (161) and the leftmost inequality follows by hypothesis ($\lambda \geq \mu$). Therefore, the inequalities (160) also hold when $\lambda_k \geq \mu$. This implies that $\max_k f_M(\mu, \lambda_k) \leq f_M\left(\mu, e^{-2 \int_0^T \|\mathcal{L}(t)\| dt}\right)$, because (160) is valid for any eigenvalue λ_k .

Next, we derive another bound that is independent of both λ_k and μ . In this case, we will use the monotonicity of $f_M(\mu, \lambda)$ with respect to μ , for $\mu \geq \lambda$ (see Supplementary Figures 13(a)-13(c)). By combining $\min_k \lambda_k \geq$

$e^{-2 \int_0^T \|\mathcal{L}(t)\| dt}$ (Eq. (138)) with Eq. (161), we have that $\mu, 1 \geq e^{-2 \int_0^T \|\mathcal{L}(t)\| dt}$ and therefore $f_M \left(\mu, e^{-2 \int_0^T \|\mathcal{L}(t)\| dt} \right) \leq f_M \left(1, e^{-2 \int_0^T \|\mathcal{L}(t)\| dt} \right)$. By combining this result with Eq. (159), we obtain the bound

$$\begin{aligned} \varepsilon_{\text{KIK}}^{(M)} &\leq \sqrt{\langle A|A \rangle} f_M \left(\mu, e^{-2 \int_0^T \|\mathcal{L}(t)\| dt} \right) \\ &\leq \sqrt{\langle A|A \rangle} f_M \left(1, e^{-2 \int_0^T \|\mathcal{L}(t)\| dt} \right) \\ &= \sqrt{\langle A|A \rangle} \left| 1 - \sum_{m=0}^M a_{\text{Adap},m}^{(M)}(1) \left(e^{-2 \int_0^T \|\mathcal{L}(t)\| dt} \right)^{m+1/2} \right|, \end{aligned} \quad (162)$$

for $M = 1, 2, 3$. Although this bound is looser than (159), it only depends on $\int_0^T \|\mathcal{L}(t)\| dt$. Since this quantity is the integral of the spectral norm of the dissipator $\mathcal{L}(t)$, over the time T consumed by the evolution $\mathcal{K}$, it can be seen as a quantifier of the total error rate in our approach. Hence, the bound (162) has the advantage of being given only in terms of this error rate.

Let us see now that this bound is also applicable to Taylor mitigation. For M arbitrary, we have that

$$\begin{aligned} \mathcal{U}_{\text{KIK}}^{(M)} &\approx \mathcal{U} \sum_{m=0}^M a_{\text{Tay},m}^{(M)} (\mathcal{K}_I \mathcal{K})^{m+1/2}, \\ \varepsilon_{\text{KIK}}^{(M)} &= \left| \langle A | \left(\mathcal{U} - \mathcal{U}_{\text{KIK}}^{(M)} \right) | \rho \rangle \right| \\ &\leq \sqrt{\langle A|A \rangle} \left\| \mathcal{U} \left(\mathcal{I} - \sum_{m=0}^M a_{\text{Tay},m}^{(M)} (\mathcal{K}_I \mathcal{K})^{m+1/2} \right) \right\|. \end{aligned} \quad (163)$$

If $1 \leq M \leq 3$, a direct calculation allows us to corroborate that (cf. Eqs. (61) and (77)-(85)) $a_{\text{Tay},m}^{(M)} = a_{\text{Adap},m}^{(M)}(1)$. Therefore, in this case we can rewrite (163) as

$$\varepsilon_{\text{KIK}}^{(M)} \leq \sqrt{\langle A|A \rangle} \left\| \mathcal{U} \left(\mathcal{I} - \sum_{m=0}^M a_{\text{Adap},m}^{(M)}(1) (\mathcal{K}_I \mathcal{K})^{m+1/2} \right) \right\|. \quad (164)$$

Finally, it is not difficult to check that $f_M(1, \lambda) = \left| 1 - \sum_{m=0}^M a_{\text{Adap},m}^{(M)}(1) (\lambda)^{m+1/2} \right|$ is monotonically decreasing in λ , which leads to the inequalities

$$f_M \left(1, e^{-2 \int_0^T \|\mathcal{L}(t)\| dt} \right) \geq f_M(1, \min_k \lambda_k) \geq f_M(1, \lambda_k). \quad (165)$$

From this result and Eq. (164), it follows that the bound (162) is also valid for Taylor mitigation.

Alternative bound for Taylor mitigation. Now, we derive another bound on $\varepsilon_{\text{KIK}}^{(M)}$, applicable to Taylor error mitigation. While this bound is looser than Eq. (162), it holds for any $M \geq 1$ and not only for $1 \leq M \leq 3$. To this end, we apply the Taylor remainder

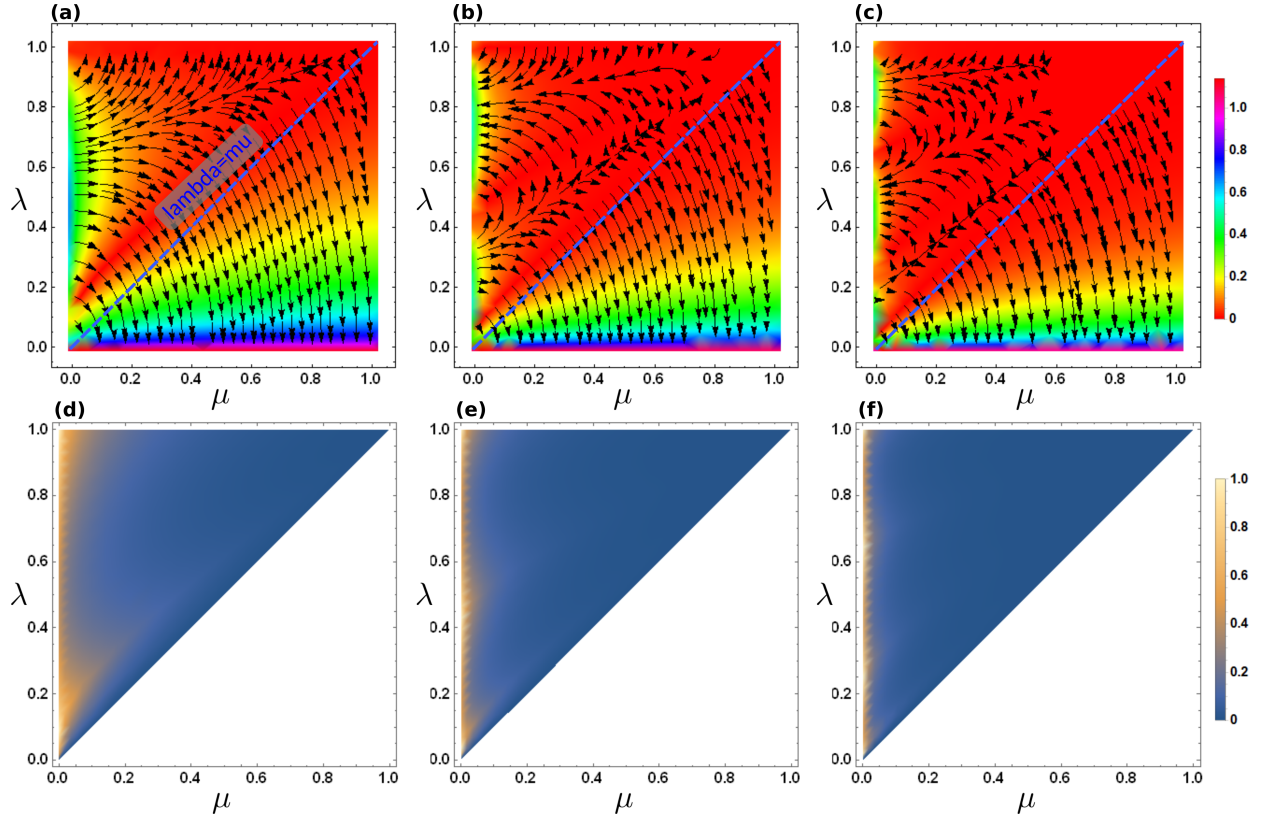
$$R_M(x) = \int_a^x \frac{1}{M!} \left(\frac{d^{M+1} f}{dx^{M+1}} \right)_{x=t} (x-t)^M dt, \quad (166)$$

which gives the error $R_M(x) = f(x) - P_M(x)$ when approximating a function $f(x)$ with the M -degree Taylor polynomial

$$P_M(x) = \sum_{m=0}^M \frac{1}{m!} \left(\frac{d^m f}{dx^m} \right)_{x=a} (x-a)^m. \quad (167)$$

In Taylor mitigation, we approximate the eigenvalues $f(\lambda) = \lambda^{-\frac{1}{2}}$ of $(\mathcal{K}_I \mathcal{K})^{-\frac{1}{2}}$, using the Taylor polynomial $P_M(\lambda) = \sum_{m=0}^M \frac{1}{m!} \left(\frac{(2m-1)!!}{2^m} \lambda^{-m-\frac{1}{2}} \right)_{\lambda=1} (\lambda-1)^m$. The corresponding Taylor remainder is

$$R_M(\lambda) = \int_1^\lambda \frac{1}{M!} \frac{(2M+1)!!}{2^{M+1}} t^{-M-\frac{3}{2}} (\lambda-t)^M dt. \quad (168)$$



Supplementary Figure 13. Plots used in the derivation of the bounds (159) and (162). (a), (b), and (c) are color density plots of $f_M(\mu, \lambda) := \left| 1 - \sum_{m=0}^M a_{\text{Adap},m}^{(M)}(\mu) (\lambda)^{m+1/2} \right|$, for $M = 1$, $M = 2$, and $M = 3$, respectively. The streamlines depict the gradient of $f_M(\mu, \lambda)$. In particular, we can see that for $\lambda \leq \mu$ the functions $f_M(\mu, \lambda)$ are monotonically decreasing with respect to λ , and monotonically increasing with respect to μ . The plots (d), (e) and (f) are color density plots of $f_M(\mu, \mu) - f_M(\mu, \lambda)$ (with $1 \leq M \leq 3$ increasing from left to right), for $\lambda \geq \mu$, and show that in this interval $f_M(\mu, \mu) \geq f_M(\mu, \lambda)$.

The eigenvalues of the Hermitian operator $\mathcal{K}_1 \mathcal{K} = \mathcal{K}^\dagger \mathcal{K}$ must satisfy $\lambda \leq 1$. Otherwise, many applications of the evolution $\mathcal{K}_1 \mathcal{K}$ would lead to a non-physical operation, characterized by divergent eigenvalues. Taking this into account, the absolute value of $R_M(\lambda)$ is upper bounded by

$$\begin{aligned}
 |R_M(\lambda)| &\leq \frac{1}{M!} \frac{(2M+1)!!}{2^{M+1}} \lambda^{-M-\frac{3}{2}} \int_{\lambda}^1 |(\lambda-t)^M| dt \\
 &= \frac{1}{M!} \frac{(2M+1)!!}{2^{M+1}} \lambda^{-M-\frac{3}{2}} \int_{\lambda}^1 (t-\lambda)^M dt \\
 &= \frac{(2M+1)!!}{2^{M+1}(M+1)!} \lambda^{-M-\frac{3}{2}} (1-\lambda)^{M+1}.
 \end{aligned} \tag{169}$$

Using the diagonal form $\mathcal{K}_1 \mathcal{K} = \sum_k \lambda_k |k\rangle \langle k|$, we can express the approximation $\sum_{m=0}^M a_{\text{Tay},m}^{(M)} (\mathcal{K}_1 \mathcal{K})^m$ to the inverse $(\mathcal{K}_1 \mathcal{K})^{-\frac{1}{2}}$ as

$$\sum_{m=0}^M a_{\text{Tay},m}^{(M)} (\mathcal{K}_1 \mathcal{K})^m = (\mathcal{K}_1 \mathcal{K})^{-\frac{1}{2}} + \sum_k (\pm |R_M(\lambda_k)|) |k\rangle \langle k|. \tag{170}$$

Therefore,

$$\begin{aligned}
\mathcal{U}_{\text{KIK}}^{(M)} &= \mathcal{K} \left[\sum_{m=0}^M a_{\text{Tay},m}^{(M)} (\mathcal{K}_I \mathcal{K})^m \right] \\
&= \mathcal{U} (\mathcal{K}_I \mathcal{K})^{\frac{1}{2}} \left[(\mathcal{K}_I \mathcal{K})^{-\frac{1}{2}} + \sum_k (\pm |R_M(\lambda_k)|) |k\rangle\langle k| \right] \\
&= \mathcal{U} + \sum_k \left(\pm \lambda_k^{\frac{1}{2}} |R_M(\lambda_k)| \right) |k\rangle\langle k|,
\end{aligned} \tag{171}$$

where we write $\mathcal{K}$ as $\mathcal{K} = \mathcal{U} (\mathcal{K}_I \mathcal{K})^{\frac{1}{2}}$ in the second line, and $(\mathcal{K}_I \mathcal{K})^{\frac{1}{2}}$ as $(\mathcal{K}_I \mathcal{K})^{\frac{1}{2}} = \sum_k \lambda_k^{\frac{1}{2}} |k\rangle\langle k|$ in the third line. Accordingly, for Taylor mitigation we have that

$$\begin{aligned}
\varepsilon_{\text{KIK}}^{(M)} &= \left| \langle A | \left(\mathcal{U} - \mathcal{U}_{\text{KIK}}^{(M)} \right) | \rho \rangle \right| \\
&\leq \frac{(2M+1)!!}{2^{M+1}(M+1)!} \sqrt{\langle A | A \rangle} \left\| \sum_k \left(\frac{1}{\lambda_k} - 1 \right)^{M+1} |k\rangle\langle k| \right\| \\
&= \frac{(2M+1)!!}{2^{M+1}(M+1)!} \sqrt{\langle A | A \rangle} \left(\frac{1}{\min_k \lambda_k} - 1 \right)^{M+1} \\
&\leq \frac{(2M+1)!!}{2^{M+1}(M+1)!} \sqrt{\langle A | A \rangle} \left(e^{2 \int_0^T \|\mathcal{L}(t)\| dt} - 1 \right)^{M+1},
\end{aligned} \tag{172}$$

where we assume again $\langle \rho | \rho \rangle = 1$, and Eqs. (169) and (138) are respectively applied in the second line and the last line.

In the following section, we will take advantage of an important property of the coefficients $a_{\text{Adap},m}^{(M)}(\mu)$ and $a_{\text{Tay},m}^{(M)}$, in order to further tighten the bounds (159), (162) and (172).

Traceless observables and second (tighter) error bounds for Adaptive error mitigation and Taylor error mitigation. Given an arbitrary observable A , we already know that the error-mitigated expectation value in the case of M th order mitigation is given by

$$\langle A \rangle_{\text{mit}} = \left\langle A \left| \sum_{m=0}^M a_m^{(M)} \mathcal{K} (\mathcal{K}_I \mathcal{K})^m \right| \rho \right\rangle, \tag{173}$$

where $a_m^{(M)} = a_{\text{Adap},m}^{(M)}(\mu)$ for adaptive mitigation and $a_m^{(M)} = a_{\text{Tay},m}^{(M)}$ for Taylor mitigation.

Now, let us see that the error $\varepsilon_{\text{KIK}}^{(M)}$ is invariant under a transformation $A \rightarrow A + bI$, where I is the identity operator and b is a real number. This shifts the trace of A by the value $b\text{Tr}(I)$. Letting $\varepsilon_{\text{KIK}}^{(M)}(b)$ denote the error corresponding to the observable $A + bI$, we have that

$$\begin{aligned}
\varepsilon_{\text{KIK}}^{(M)}(b) &= \left| \left\langle A + bI \left| \sum_{m=0}^M a_m^{(M)} \mathcal{K} (\mathcal{K}_I \mathcal{K})^m \right| \rho \right\rangle - \langle A + bI | \mathcal{U} | \rho \rangle \right| \\
&= \left| \sum_{m=0}^M a_m^{(M)} \langle A | \mathcal{K} (\mathcal{K}_I \mathcal{K})^m | \rho \rangle - \langle A | \mathcal{U} | \rho \rangle + b \sum_{m=0}^M a_m^{(M)} - b \right|,
\end{aligned} \tag{174}$$

where in the second line we use the fact that $\langle I | \mathcal{K} (\mathcal{K}_I \mathcal{K})^m | \rho \rangle = \text{Tr}(\rho^{(m)}) = 1$, for a trace-preserving evolution $\mathcal{K} (\mathcal{K}_I \mathcal{K})^m$. Here, $\rho^{(m)}$ is the state resulting from applying the circuit $\mathcal{K} (\mathcal{K}_I \mathcal{K})^m$ on ρ .

In the case of adaptive mitigation, we have that, for all M , $\sum_{m=0}^M a_m^{(M)} = \sum_{m=0}^M a_{\text{Adap},m}^{(M)}(\mu) = 1$ by construction. Since in the limit of zero noise the Taylor approximation converges to $(\mathcal{K}_I \mathcal{K})^{-1} = \mathcal{I}$, it also follows from Eq. (60) that

$$\sum_{m=0}^M a_{\text{Tay},m}^{(M)} = 1, \text{ for all } M. \tag{175}$$

Therefore, Eq. (174) is equivalent to (both for adaptive mitigation and Taylor mitigation)

$$\varepsilon_{\text{KIK}}^{(M)}(b) = \varepsilon_{\text{KIK}}^{(M)}(0), \quad (176)$$

for all b real. This result implies that the error $\varepsilon_{\text{KIK}}^{(M)}(0)$, associated with the actual observable A , can be evaluated using instead the shifted observable $A + bI$. By combining this property with Eqs. (159), (162), and (172), we can obtain the families of b -dependent bounds

$$\varepsilon_{\text{KIK}}^{(M)}(0) \leq \sqrt{\text{Tr}[(A + bI)^2]} \left| 1 - \sum_{m=0}^M a_{\text{Adap},m}^{(M)}(\mu) \left(e^{-2 \int_0^T \|\mathcal{L}(t)\| dt} \right)^{m+1/2} \right|, \quad (177)$$

$$\varepsilon_{\text{KIK}}^{(M)}(0) \leq \frac{(2M+1)!!}{2^{M+1}(M+1)!} \sqrt{\text{Tr}[(A + bI)^2]} \left(e^{2 \int_0^T \|\mathcal{L}(t)\| dt} - 1 \right)^{M+1}, \quad (178)$$

where we have written $\langle A + bI | A + bI \rangle$ as $\text{Tr}[(A + bI)^2]$. In the case of Eq. (177), values of $0 \leq \mu < 1$ yield bounds for adaptive mitigation, as per Eq. (159), and $\mu = 1$ yields bounds applicable to adaptive mitigation and Taylor mitigation, cf. Eq. (162). On the other hand, Eq. (178) is a consequence of the Taylor-mitigation bound (172).

The tightest bounds in (177) and (178) are obtained by minimizing $\text{Tr}[(A + bI)^2] = \text{Tr}(A^2) + 2b\text{Tr}(A) + b^2\text{Tr}(I)$, with respect to b . A simple calculation shows that the minimum is attained by b such that the observable $A + bI$ is traceless, i.e. $\text{Tr}(A + bI) = 0$. The resulting expression is

$$b = -\frac{\text{Tr}(A)}{\text{Tr}(I)}. \quad (179)$$

By substituting this result into Eqs. (177) and (178), we obtain the optimal bounds

$$\varepsilon_{\text{KIK}}^{(M)}(0) \leq \sqrt{\text{Tr}(A^2) - \frac{[\text{Tr}(A)]^2}{\text{Tr}(I)}} \left| 1 - \sum_{m=0}^M a_{\text{Adap},m}^{(M)}(\mu) \left(e^{-2 \int_0^T \|\mathcal{L}(t)\| dt} \right)^{m+1/2} \right|, \quad (180)$$

$$\varepsilon_{\text{KIK}}^{(M)}(0) \leq \frac{(2M+1)!!}{2^{M+1}(M+1)!} \sqrt{\text{Tr}(A^2) - \frac{[\text{Tr}(A)]^2}{\text{Tr}(I)}} \left(e^{2 \int_0^T \|\mathcal{L}(t)\| dt} - 1 \right)^{M+1}. \quad (181)$$

Summary. To conclude this supplementary note, we merge the bounds previously derived into a single expression. That is,

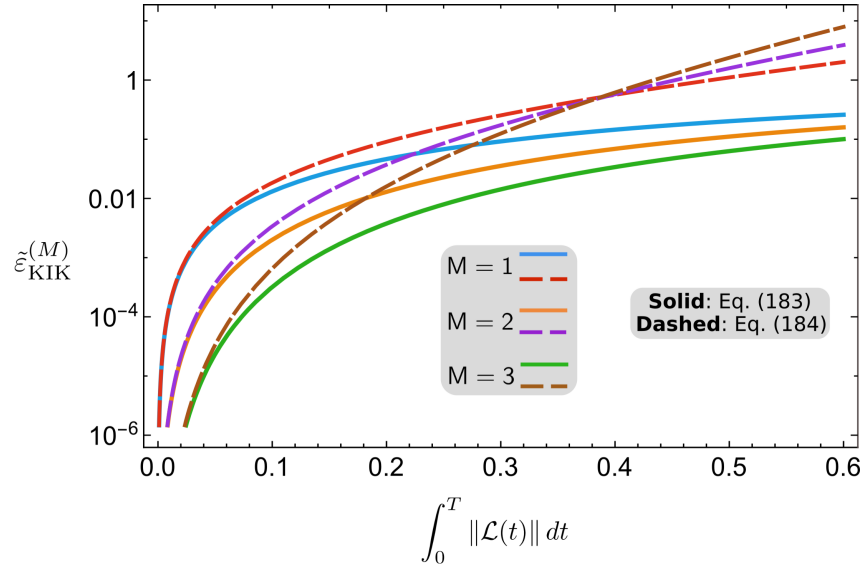
$$\varepsilon_{\text{KIK}}^{(M)} \leq \sqrt{\text{Tr}(A^2) - \frac{[\text{Tr}(A)]^2}{\text{Tr}(I)}} \left| 1 - \sum_{m=0}^M a_{\text{Adap},m}^{(M)}(\mu) e^{-2(m+1/2) \int_0^T \|\mathcal{L}(t)\| dt} \right|, \text{ for } M = 1, 2, 3, \quad (182)$$

$$\leq \sqrt{\text{Tr}(A^2) - \frac{[\text{Tr}(A)]^2}{\text{Tr}(I)}} \left| 1 - \sum_{m=0}^M a_{\text{Adap},m}^{(M)}(1) e^{-2(m+1/2) \int_0^T \|\mathcal{L}(t)\| dt} \right|, \text{ for } M = 1, 2, 3, \quad (183)$$

$$\leq \frac{(2M+1)!!}{2^{M+1}(M+1)!} \sqrt{\text{Tr}(A^2) - \frac{[\text{Tr}(A)]^2}{\text{Tr}(I)}} \left(e^{2 \int_0^T \|\mathcal{L}(t)\| dt} - 1 \right)^{M+1}, \quad (184)$$

where Eq. (182) is equivalent to Eq. (180), and (183) follows by applying the optimization strategy of the previous section to Eq. (162). Equation (184) corresponds to Eq. (181), and is the looser bound according to Supplementary Figure 14. We also recall that the bound (183) is valid for both adaptive mitigation and Taylor mitigation, in addition to being independent of μ . On the other hand, the bound (184) is valid for Taylor mitigation and for all $M \geq 1$. From this bound we can also see that the error $\varepsilon_{\text{KIK}}^{(M)}$ is exponentially decreasing in M if $\int_0^T \|\mathcal{L}(t)\| dt < \frac{1}{2} \ln(2)$, since the prefactor $\frac{(2M+1)!!}{2^{M+1}(M+1)!}$ satisfies $\frac{(2M+1)!!}{2^{M+1}(M+1)!} \leq \frac{3}{8}$.

We also stress that, once coherent noise is converted into incoherent noise, via RC, the quantity $\int_0^T \|\mathcal{L}(t)\| dt$ accounts for the total error rate affecting the evolution $\mathcal{K}$, irrespective of the depth or the width of the corresponding circuit. Therefore, the bounds (182)-(184) are meaningful to assess the performance of the KIK method applied to circuits $\mathcal{K}$ of arbitrary size. It is also important to note that generic QEM methods can be useful so long as the total error rate is not excessively high²⁰. In the context of the KIK method, we can have scalable QEM if $\int_0^T \|\mathcal{L}(t)\| dt$ is kept below a fixed value for increasingly large circuits, and if for this value the bound (182) is sufficiently small.



Supplementary Figure 14. Upper bounds (183) and (184) on $\tilde{\varepsilon}_{\text{KIK}}^{(M)} := \frac{\varepsilon_{\text{KIK}}^{(M)}}{\sqrt{\text{Tr}(A^2) - \frac{[\text{Tr}(A)]^2}{\text{Tr}(I)}}}$, for mitigation orders $M = 1, 2, 3$.

* ivhenao@gmail.com

† jader.pereira.santos@gmail.com

‡ raam@mail.huji.ac.il

- [1] Gyamfi, J. A. Fundamentals of quantum mechanics in liouville space. *Eur. J. Phys.* **41**, 063002 (2020).
- [2] Breuer, H.-P. & Petruccione, F. *The theory of open quantum systems* (Oxford University Press, 2002).
- [3] Dann, R., Levy, A. & Kosloff, R. Time-dependent markovian quantum master equation. *Phys. Rev. A* **98**, 052129 (2018).
- [4] Wallman, J. J. & Emerson, J. Noise tailoring for scalable quantum computation via randomized compiling. *Phys. Rev. A* **94**, 052325 (2016).
- [5] Blanes, S., Casas, F., Oteo, J.-A. & Ros, J. The magnus expansion and some of its applications. *Phys. Rep.* **470**, 151–238 (2009).
- [6] 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).
- [7] Koczor, B. Exponential error suppression for near-term quantum devices. *Phys. Rev. X* **11**, 031057 (2021).
- [8] Huggins, W. J. *et al.* Virtual distillation for quantum error mitigation. *Phys. Rev. X* **11**, 041036 (2021).
- [9] Jnane, H., Undseth, B., Cai, Z., Benjamin, S. C. & Koczor, B. Multicore quantum computing. *Phys. Rev. Applied* **18**, 044064 (2022).
- [10] Temme, K., Bravyi, S. & Gambetta, J. M. Error mitigation for short-depth quantum circuits. *Phys. Rev. Lett.* **119**, 180509 (2017).
- [11] Trotter, H. F. On the product of semi-groups of operators. *Proc. Amer. Math. Soc.* **10**, 545–551 (1959).
- [12] Nielsen, M. A. & Chuang, I. *Quantum computation and quantum information* (American Association of Physics Teachers, 2002).
- [13] Landa, H., Meirom, D., Kanazawa, N., Fitzpatrick, M. & Wood, C. J. Experimental bayesian estimation of quantum state preparation, measurement, and gate errors in multiqubit devices. *Phys. Rev. Research* **4**, 013199 (2022).
- [14] Wallman, J., Granade, C., Harper, R. & Flammia, S. T. Estimating the coherence of noise. *New J. Phys.* **17**, 113020 (2015).
- [15] Nation, P. D., Kang, H., Sundaresan, N. & Gambetta, J. M. Scalable mitigation of measurement errors on quantum computers. *PRX Quantum* **2**, 040326 (2021).
- [16] Alexander, T. *et al.* Qiskit pulse: programming quantum computers through the cloud with pulses. *Quantum Sci. Technol.* **5**, 044006 (2020).
- [17] Zhang, B. *et al.* Hidden inverses: Coherent error cancellation at the circuit level. *Phys. Rev. Applied* **17**, 034074 (2022).
- [18] Greenbaum, D. Introduction to quantum gate set tomography. *Preprint at <https://arxiv.org/abs/1509.02921>* (2015).
- [19] 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**, 1–2 (2023).
- [20] Endo, S., Cai, Z., Benjamin, S. C. & Yuan, X. Hybrid quantum-classical algorithms and quantum error mitigation. *J.*

