Mitigating crosstalk errors by randomized compiling: Simulation of the BCS model on a superconducting quantum computer

Hugo Perrin ®,[1,*,†] Thibault Scoquart ®,[1,*,‡] Alexander Shnirman,[1,2] Jörg Schmalian,[1,2] and Kyrylo Snizhko[3]

[1]*Karlsruhe Institute of Technology, Institut für Theorie der Kondensierten Materie, TKM, 76049 Karlsruhe, Germany*
[2]*Karlsruher Institut für Technologie, Institut für Quantenmaterialien und Technologien, IQMT, 76021 Karlsruhe, Germany*
[3]*Univ. Grenoble Alpes, CEA, Grenoble INP, IRIG, PHELIQS, 38000 Grenoble, France*

We develop and apply an extension of the randomized compiling (RC) protocol that includes a special treatment of neighboring qubits and dramatically reduces crosstalk effects caused by the application of faulty gates on superconducting qubits in IBMQ quantum computers (`ibm_lagos` and `ibmq_ehningen`). Crosstalk errors, stemming from controlled-NOT (CNOT) two-qubit gates, are a crucial source of errors on numerous quantum computing platforms. For the IBMQ machines, their effect on the performance of a given quantum computation is often overlooked. Our RC protocol turns coherent noise due to crosstalk into a depolarizing noise channel that can then be treated using established error mitigation schemes, such as noise estimation circuits. We apply our approach to the quantum simulation of the nonequilibrium dynamics of the Bardeen-Cooper-Schrieffer (BCS) Hamiltonian for superconductivity, a particularly challenging model to simulate on quantum hardware because of the long-range interaction of Cooper pairs. With 135 CNOT gates, we work in a regime where crosstalk, as opposed to either Trotterization or qubit decoherence, dominates the error. Our twirling of neighboring qubits is shown to dramatically improve the noise estimation protocol without the need to add new qubits or circuits and allows for a quantitative simulation of the BCS model.

I. INTRODUCTION

Despite the long-lasting quest for a noiseless and scalable quantum computer (QC) that could achieve quantum advantage [1], currently available platforms require the development and improvement of the so-called noisy intermediate-scale quantum (NISQ) computer [2]. In these devices, qubit operations are subject to noise, which produces errors at a level that remains beyond the scope of fault-tolerant quantum error correction codes [3–5]. Nevertheless, several implementations of digital quantum computers are now accessible, like the IBMQ digital quantum computers [6], which can implement a universal set of quantum operations on arrays of superconducting transmon qubits (available devices containing up to 128 qubits). In order to make the most of these already existing machines, numerous error mitigation solutions have been proposed over the past decade [7–12], with the ultimate goal of pushing the efficiency and precision limit of NISQ computers until they surpass the performance of classical computers on a set of well-known quantum algorithms: quantum simulation, quantum optimization, and

quantum Fourier-transform-based algorithms (for detailed reviews on quantum algorithms see Refs. [13,14]). While the existing quantum hardware is quickly approaching the number of qubits theoretically required to do so, noise-induced errors remain very significant on NISQ computers [15,16], and practical solutions on how to reduce these errors to a minimum are lacking.

On IBMQ devices, quantum gate implementations produce errors that can mainly be attributed to the two-qubit [controlled-NOT (CNOT)] gate [17], with an announced error rate of 1%, obtained by Clifford randomized benchmarking [18,19]. However, little is known about the structure of the quantum noise channels acting on the qubits [20–22]. The noise is known to be a mixture of random, incoherent noise channels, experimentally linked to imperfect gate application and qubit decoherence, as well as coherent noise channels [23], induced by the weak residual couplings between a qubit and its environment (neighboring qubits, readout resonator, and residual electromagnetic fields, for instance). Coherent noise may be correlated, both in time and in space, and lead to systematic deviations from the perfect output. One of the identified origins of coherent noise on a given qubit is linked to unwanted resonances while applying radio-frequency (RF) pulses to implement a quantum gate on neighboring qubits. This effect, dubbed crosstalk [24,25], is known to be important in superconducting qubit architectures [26,27]. Crosstalk severely limits the performance of these NISQ implementations, in spite of considerable experimental efforts [17,28,29] and software mitigation approaches [30–32]. Finally, we emphasize that

---

*These authors contributed equally to this work.

†hugo.perrin@kit.edu

‡thibault.scoquart@kit.edu

previous experiments have shown that the effective noise acting on the transmon qubits of IBMQ QCs is very unstable in time, and may change drastically over the scale of a few hours or days [33], which limits the possibilities for benchmarking the noise channels and reusing this information.

Error mitigation schemes are, therefore, essential in order to make NISQ a reliable platform for quantum computing. Over the past decade, several of those techniques have been proposed with the aim of turning noisy outputs into practical and reliable results. They can be classified into two categories: noise-parameter-dependent and noise-parameter-free error mitigation protocols. The former, which includes, e.g., probabilistic error cancellation (PEC) [11,12], relies on a precise characterization of the noise parameters, which requires a substantial initial overhead [34–37]. Noise-parameter-free error mitigation schemes, on the other hand, usually assume a model of noise but do not require any characterization of their parameters. Therefore, they are free from this initial overhead but fail when noise model assumptions are not met. The most prominent example is known as zero-noise extrapolation (ZNE), which consists of performing several experiments with increasing noise strength (which can be achieved both with experimental or algorithmic approaches), and extrapolating the results to the zero-noise limit [11,12,38–40]. Another example is the so-called noise estimation circuit (NEC) technique, introduced in Ref. [41]. A common feature of both parameter-dependent and parameter-free methods is an exponential scaling of the required number of circuit runs with the number of noisy gates [42].

The NEC method was originally designed for noisy qubits affected by a very simple one-parameter noise channel called depolarizing noise. This assumption is unrealistic for NISQ, and particularly for IBMQ quantum computers, as discussed above. Fortunately, the noise structure can be simplified using a procedure called randomized compiling (RC) [43–45]. RC consists of averaging a given circuit output over different randomized versions of the original circuit in a way that effectively maps any noise channel to a simpler Pauli noise channel [46], bringing it closer to the depolarizing channel required to apply the NEC method.

RC is typically performed only on qubits directly affected by CNOT gates [47–50], therefore ignoring crosstalk errors. These are typically assumed to be insignificant, yet turn out to be harmful to our simulation results in practice. Methods for mitigating crosstalk via RC have been previously proposed [44] and demonstrated to be effective [45]. In this article, we propose an extension of the twirling gate set [see Fig. 1(a) or Fig. 5], which further maps crosstalk errors onto a depolarizing noise channel. We then demonstrate that using this technique drastically improves the performance of the NEC mitigation scheme implementation on IBMQ QCs. We expect that similar improvements can be achieved for other error mitigation schemes that assume a depolarizing noise model. This is relevant for all QC architectures affected by crosstalk. In addition, our results confirm that crosstalk errors on IBMQ devices cannot be neglected (we find that they lead to an effective error rate per CNOT gate of up to 5%) and need to be mitigated.

To benchmark the RC and NEC methods used in this article, we perform a quantum simulation of the BCS

Hamiltonian, the historical model describing superconductivity through the emergence of electronic Cooper pairs [51]. Starting from an initial quenched state, we simulate the out-of-equilibrium dynamics of the system using a simple Trotter algorithm. The evolution is then monitored by measuring time-dependent expectation values of Pauli string observables. Specific features make this model very suitable to benchmark the performance of QCs. First, due to the hard-core boson statistics of the Cooper pairs, the model has a natural description in terms of Pauli matrices. Second, the BCS model is integrable, so that the few-particle dynamics can be completely solved efficiently on a classical computer [52–56]. Finally, the BCS interaction term contains an all-to-all coupling between spins, which makes it challenging to implement on a low-connectivity platform like IBMQ QC (which has a quasilinear layout of qubits). Previously, the BCS model has been implemented on IBMQ with the purpose of finding its ground-state energies using variational quantum algorithms [57–59]. Also, an implementation of the BCS evolution operator for a nitrogen-vacancy-center-based QC with a star-shaped connectivity was proposed recently [60].

Our key results are summarized in Fig. 1. Specifically, Fig. 1(b) shows that crosstalk randomized compiling dramatically improves the accuracy of the quantum simulation without the need for adding new qubits or circuits (compared to the standard randomized compiling). While Fig. 1(b) presents the results averaged over multiple observables and simulation times, Figs. 1(c) and 1(d) demonstrate how the results of applying different techniques compare for one specific observable as a function of the simulated evolution time (i.e., number of Trotter steps applied). It is clear that the contribution of crosstalk is not negligible.

This article is organized as follows: In Sec. II, we Trotterize the BCS time evolution, and show how to map the corresponding unitary operator to a quantum circuit. We run naive simulations of the model on IBMQ QCs, illustrating the dramatic effect of the noise on the results. In Sec. III, we present our improved RC procedure, and perform a comparative analysis between standard RC and crosstalk RC, by applying them to our BCS simulations. Section IV is devoted to our implementation of the NEC mitigation scheme, and readout error correction. Finally, we provide in Sec. V an in-depth comparison of various levels of error mitigation—from no mitigation to NEC assisted with crosstalk RC. We discuss the observed discrepancies between classical simulations assuming a simple depolarizing noise channel and the results from real QC.

II. QUANTUM SIMULATION OF THE BCS MODEL ON A SUPERCONDUCTING QC

Throughout this work, we model a quench in the BCS model with three Cooper pairs in order to gauge the efficiency of both RC techniques and their combination with the NEC technique. The dynamical evolution of various local observables is the quantitative outcome of these simulations. In this section, we show how the BCS quench can be simulated on a quantum computer and present the results of a basic simulation—without any error mitigation—setting a reference for the performance of more advanced methods. We
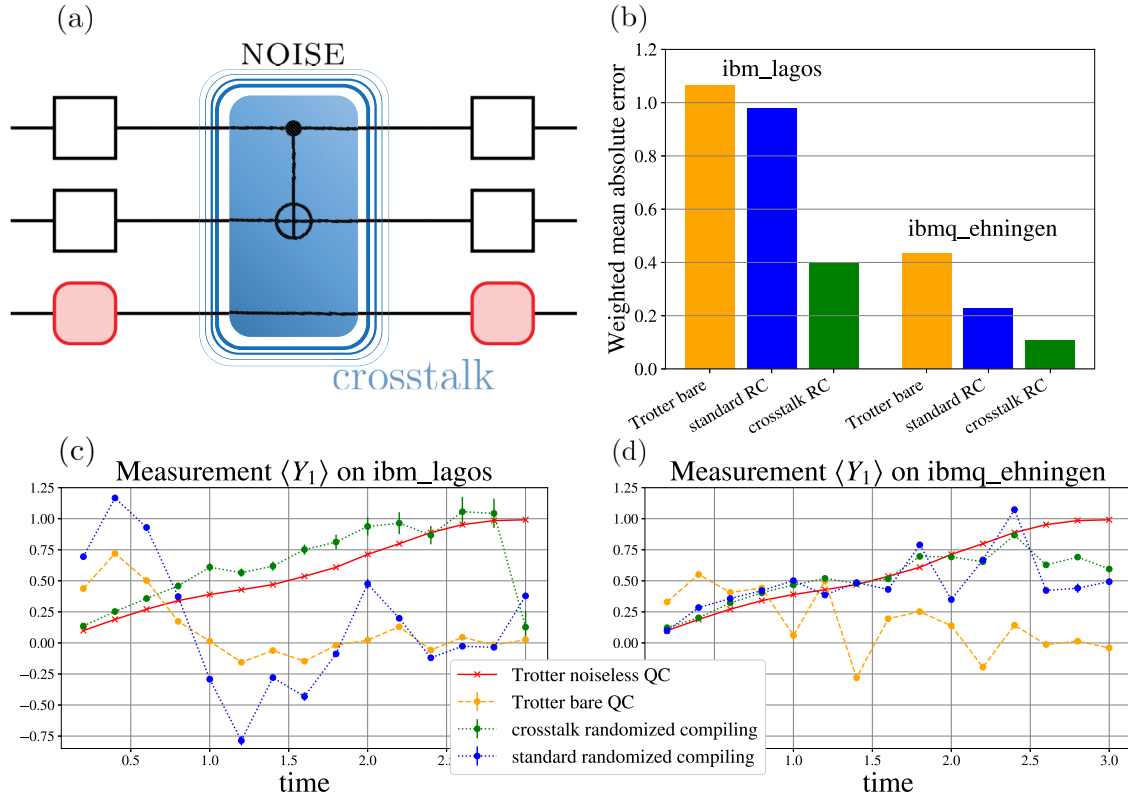
FIG. 1. Illustration of the importance of crosstalk mitigation. (a) Pictorial representation of noise occurring during the application of a CNOT gate (blue cloud). Crosstalk may extend the effect of this noise to the neighboring qubits. Gates applied before and after the CNOT gates depict the RC protocol (white gates are used for the standard RC procedure; adding the red gates constitutes the crosstalk RC). A more detailed description is given in Sec. III. (b) Figure of merit for the simulation accuracy when using different error mitigation techniques. Clearly, the effect of mitigating crosstalk is not negligible. The figure of merit is defined as $\epsilon = S^{-1} \sum_j |i_j| |(n_j - i_j)|$, where j enumerates all simulated evolution durations and observables measured, i_j is the ideal noiseless expectation value for the respective observable, n_j is the value measured on the noisy QC, and $S = \sum_j |i_j|$ represents the normalization factor, as $|i_j|$ weights prioritize the contributions of observables with expectation value away from zero. [(c), (d)] Evolution of the expectation value Y of the second qubit out of three under the three-body BCS Hamiltonian [see Eq. (1)] with energy levels $\{-1, 0, +1\}$ and coupling constant $g = 0.5$, respectively, on both IBMQ devices. The red curve is the noiseless Trotterized dynamics, and the dashed orange curve represents the simulation on a real, noisy QC. The dashed blue (green) curve shows the simulated evolution when using a number of error mitigation protocols: readout-error correction, standard randomized compiling (crosstalk randomized compiling), and NEC. While not allowing for perfect agreement with the ideal red curve, crosstalk randomized compiling dramatically improves the accuracy of the quantum simulation.

execute the quantum code on three linearly connected qubits of `ibmq_lagos` and of `ibmq_ehningen` (cf. Fig. 2). Throughout the paper, unless stated otherwise, results for `ibmq_lagos` are presented. Results for `ibmq_ehningen` are gathered in Appendix C.
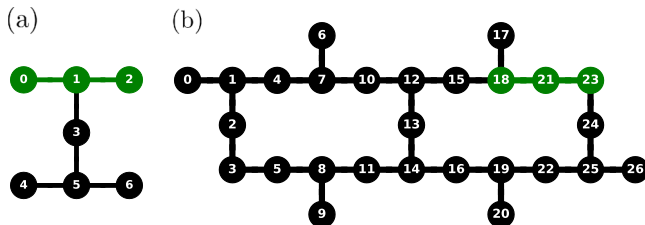

FIG. 2. Coupling map of (a) `ibm_lagos` and (b) `ibmq_ehningen` machines. Links represent the physical junctions on which two-qubit gates can be realized. The green qubits are used for the simulation and black qubits are the idle qubits.

A. Problem definition

The BCS Hamiltonian can be represented in the spin language via

$$\mathcal{H}_{\text{BCS}} = -\sum_{j=0}^{L-1} \left(\epsilon_j - \frac{g}{2} \right) \sigma_j^z - \frac{g}{2} \sum_{\substack{i,j=0 \\ i<j}}^{l-1} \left(\sigma_i^x \sigma_j^x + \sigma_i^y \sigma_j^y \right) \quad (1)$$

(see Appendix A), where $L = 3$ is the number of energy levels and σ_j^a, $a \in \{x, y, z\}$, refers to the operator such that the Pauli matrix σ^a is applied on the energy level j and the identity operator on the other energy levels. The eigenstates $|0\rangle$ and $|1\rangle$ of σ_j^z denote respectively the absence or presence of a Cooper pair on the energy level j. The system is initialized in state $|\psi_0\rangle$, corresponding to the mean-field solution of the above Hamiltonian (cf. Appendix A). The aim is to simulate the evolution of the system up to time t, described by the evolution operator $U(t) = e^{-i\mathcal{H}_{\text{BCS}}t}$, and measure expectation values of observables $\langle\psi_0| \mathcal{O} |\psi_0\rangle$, where $|\psi(t)\rangle = U(t) |\psi_0\rangle$,
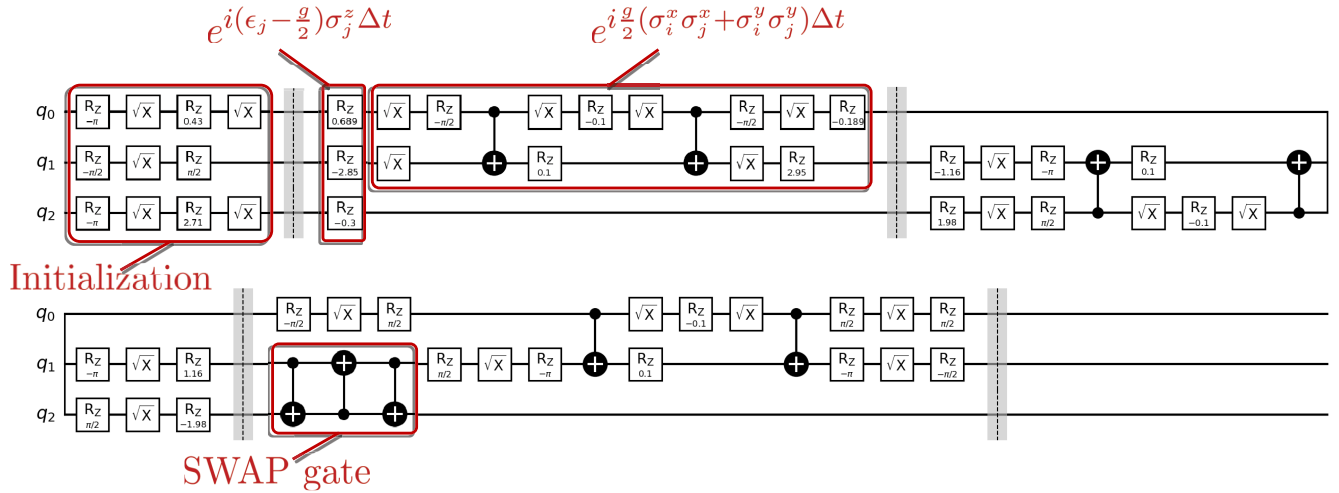
FIG. 3. Transpiled quantum circuit for a single Trotter step [Eq. (3)], preceded by the initial state preparation. The on-site energy and the interaction between the first two qubits are highlighted in red. The highlighted SWAP gate is required to ensure the interaction between the first and last qubits on a linear layout (see main text). Quantum circuits for the evolution up to arbitrary time $t = r\Delta t$ are built by concatenation of this elementary circuit r times (without the initialization), and thus contain $9r$ CNOT gates.

and the observables $\mathcal{O}$ are products of Pauli operators on the qubits.

B. Trotter algorithm for the unitary evolution

First, we need to implement the unitary evolution operator $U(t) = e^{-i\mathcal{H}_{BCS}t}$ on the quantum computer, i.e., to break down $U(t)$ into a quantum circuit comprising one- and two-qubit unitary gates that can be directly applied on the QC (native gates). This operation, dubbed transpilation, is a difficult problem that usually cannot be tackled exactly, even for simple integrable systems. To circumvent this issue, we approximate the exact unitary evolution operator using the Trotter algorithm. We slice the full time evolution into r small time steps of length $\Delta t = t/r$:

$$e^{-i\mathcal{H}t} = (e^{-i\mathcal{H}\Delta t})^r. \quad (2)$$

For each time step, we use the Suzuki-Trotter formula [61] to expand the complex exponential at first order in Δt. For the BCS Hamiltonian this yields

$$e^{-i\mathcal{H}_{BCS}\Delta t} \simeq \prod_{j=0}^{L-1} e^{i(\epsilon_j - \frac{g}{2})\sigma_j^z \Delta t} \prod_{0 \leqslant i < j \leqslant L-1} e^{i\frac{g}{2}(\sigma_i^x \sigma_j^x + \sigma_i^y \sigma_j^y)\Delta t}. \quad (3)$$

For each Trotter step, the error accumulated by this approximation has an upper bound of order $\mathcal{O}(\Delta t^2)$. This error scales as $\mathcal{O}(t\Delta t)$ for the whole dynamics (see, e.g., Refs. [62,63]). In our particular case, however, this upper bound is not reached, and the first-order Trotter algorithm remains a good approximation of the exact dynamics over the time range we consider. Thus, in what follows we compare the results obtained on a quantum computer to the result of perfect Trotter evolution, and not to the exact result of continuous evolution.

The exponential terms in Eq. (3) can now be straightforwardly transpiled into a quantum circuit comprising the native one-qubit and two-qubit gates of the IBMQ machines. More details about this procedure can be found in Appendix B, and the resulting quantum circuit is displayed in Fig. 3, where the

two first exponential factors in Eq. (3) are highlighted with red boxes as an example. The rightmost one, in particular, implements the interactions between qubits q_0 and q_1. It is then simply repeated for the interaction q_1-q_2. For the last interaction, however, a SWAP gate (composed of a series of three alternating CNOT gates) is required to account for the linear layout of the quantum computers. This gate exchanges the logical information contained in physical qubits q_1 and q_2, thus allowing to implement the interaction q_0-q_2 using physical qubit q_1 as an intermediary. Finally, we obtain the equivalent quantum circuit representation of an elementary Trotter step from Fig. 3, which requires nine CNOT gates in total. To simulate the dynamics up to an arbitrary time t, we simply concatenate $r = t/\Delta t$ copies of this circuit [64].

C. Parameters

We choose the following model parameters: on-site energies $\epsilon_j \in \{-1, 0, +1\}$ and a coupling constant $g = 0.5$. We then perform the Trotter evolution up to time $T = 3.0$ with a time step $\Delta t = 0.2$. Unless stated otherwise, quantum simulations shown throughout the article are performed on qubits q_0, q_1, and q_2 of the `ibmq_lagos` machine. We choose q_0 as the first qubit, with energy $\epsilon_0 = -1$, q_1 as the second qubit, with energy $\epsilon_1 = 0$, and q_2 as the third qubit, with energy $\epsilon_2 = +1$.

D. Observables

After performing the evolution, we monitor the dynamics of the system by evaluating the expectation values of different observables, which are tensor products of Pauli matrices. These observables may bear little physical interest, but their expectation values are straightforwardly estimated from simple measurements, and their exact evolution is easily obtained classically. We made two different experiments: in the first one we select observables from the set $\mathcal{S}_1 \equiv \{Z_0, Z_1, Z_2, Z_0Z_1, Z_1Z_2, Z_0Z_2, Z_0Z_1Z_2\}$, and in the second one from the set $\mathcal{S}_2 \equiv \{X_0, Y_1, Z_2, X_0Y_1, Y_1Z_2, X_0Z_2, X_0Y_1Z_2\}$

(capital letters correspond to the direction of the Pauli observable, and subscripts refer to the qubit they are measured on). We therefore have access to a total of 7+7 different observables. All the measurement outputs are sampled over 32 000 repetitions.

E. Raw results

Despite the small number of qubits, the three-qubit BCS circuits are challenging to simulate on a noisy QC due to their depth. Their execution is subject to significant experimental noise, mainly originating from imperfect (CNOT) gate applications (IBM estimates the CNOT gate's error rate at $\sim 10^{-2}$ while the single-qubit gate's error is of the order of 10^{-4}).

We have simulated the dynamics with circuits containing up to 15 Trotter steps with 9 CNOT gates per step, i.e., with a maximum of 135 CNOT gates. We plot in Fig. 4 the time evolution for 3 of the 14 available observables (the other observables are plotted in Appendix C): the first qubit's x component $\langle X_0 \rangle$ [see Fig. 4(a)], the third qubit's z component $\langle Z_2 \rangle$ [see Fig. 4(b)] and their joint evolution $\langle X_0 Z_2 \rangle$ [see Fig. 4(c)]. We compare their evolution with the exact BCS dynamics (black solid line), the Trotterized dynamics obtained on a noiseless classical computer (red solid line) and on a real, noisy, IBM QC (`ibm_lagos`) (orange dotted line). We first emphasize that the first-order Trotterized dynamics is indeed a good approximation of the exact evolution here, so we can expect the algorithmic Trotter error to be negligible compared to errors stemming from the noise acting on the QC.

When the Trotter circuits are run on a real, noisy QC, we observe that the measured expectation values of all observables decay with *a priori* unpredictable fluctuations. Despite the fact that the CNOT error rate is around 1%, the effect of noise is already dramatic after a few Trotter steps, and no interesting features of the dynamics can be extracted from these results. This illustrates the need for error mitigation strategies to fight the effects of the noise, which we can expect to be even more dramatic as the circuit depth increases. For the sake of comparison, simulation of this system as well as the following error mitigation protocols have also been performed on qubits q_{18}, q_{21}, and q_{23} of the `ibmq_ehningen` device measured respectively in the x, y, and z directions. The results are qualitatively similar and are shown in Fig. 13 in Appendix C.

Small systems like the one we consider, whose full dynamics is easy to simulate on a classical computer, constitute a good platform to benchmark different noise-tailoring and quantum error mitigation protocols. In what follows, we implement a few of these protocols and assess their efficiency to improve the results of our quantum simulation.

III. SIMPLIFYING THE NOISE STRUCTURE: CROSSTALK RANDOMIZED COMPILING

Errors in QCs have multiple experimental origins. They can originate, for instance, from imperfect measurement during a qubit state readout. They can also stem from the experimentally imperfect application of certain quantum gates, in particular CNOT gates. In addition, qubit decoherence constitutes another source of errors, which sets an upper
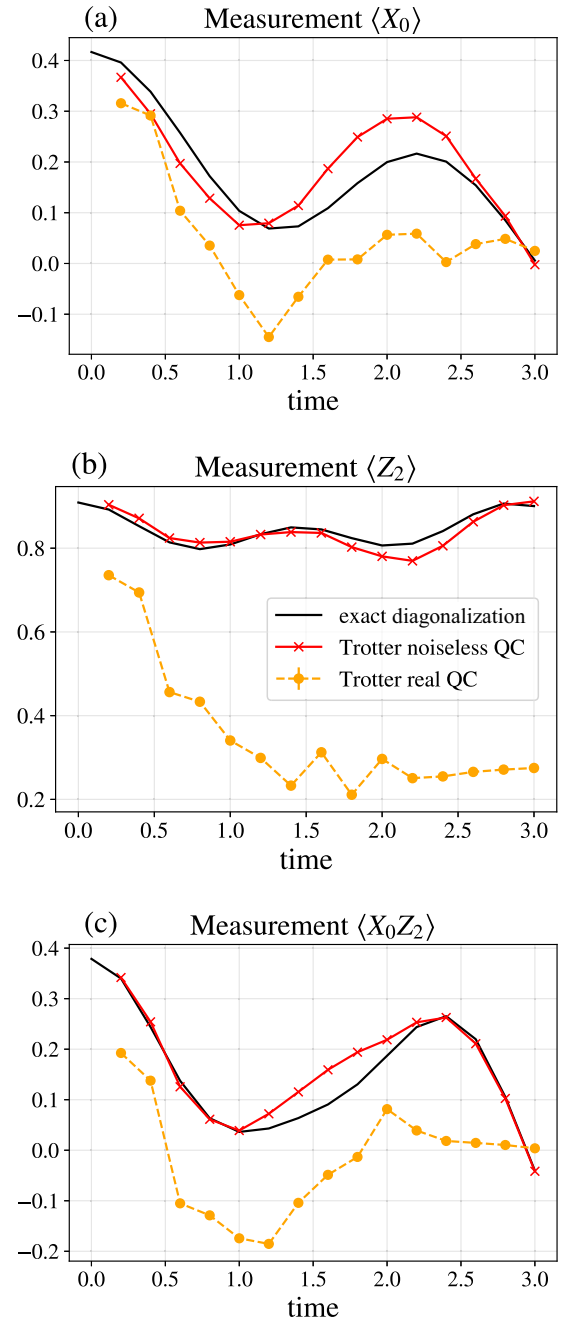

FIG. 4. Time evolution of the first qubit's x component $\langle X_0 \rangle$ (top panel), the third qubit's z component $\langle Z_2 \rangle$ (middle panel), and their joint evolution $\langle X_0 Z_2 \rangle$ (bottom panel) under the exact BCS Hamiltonian evolution (black), the Trotter noiseless circuits (red), and the Trotter circuit on a real noisy QC (orange). Error bars due to shot noise are smaller than the plot marker size $\sigma_s \sim 1/\sqrt{N_s} = 5 \times 10^{-3}$, where $N_s = 32\,000$ is the number of shots.

bound on the size of the circuits that can be run. On superconducting chips of IBM, the decoherence times T_1 and T_2 are typically of the order of ~ 100 μs while the application of the CNOT gates is ~ 300 ns [17,65] (data taken from IBMQ backend properties). Because we apply no more than 135 CNOT gates, we consider decoherence of qubits negligible compared to the noise introduced by CNOT gates.

To address errors arising from noisy CNOT gates, understanding and simplifying their noise structure is crucial. Randomized compiling was introduced in Ref. [44] as a method for converting any coherent noise channel into a Pauli (incoherent) noise channel. The original proposal recommended applying the twirling protocol to all qubits, but in practice this approach is often performed only on the active qubits of the CNOT gates [47–50], thereby overlooking spillover crosstalk effects. In the present section, we discuss the importance of applying RC to neighboring qubits as well, and improve the procedure to simplify crosstalk noise even further. We start, in Sec. III A, by presenting in detail the standard RC protocol on active qubits. Then in Sec. III B, we propose an extension of the quantum gate twirling set on idle qubits, which turns any coherent noise channel into a depolarizing (incoherent and unbiased) noise channel. By implementing the crosstalk RC protocol in our BCS simulation, we demonstrate its efficiency on IBMQ devices. Unlike for the standard RC protocol, we show that final results can be faithfully reproduced with classical simulations of noisy quantum circuits assuming a simple depolarizing noise channel.

A. Standard randomized compiling protocol

Before mitigating errors introduced by the application of CNOT gates on IBM QCs, we first need a way to model them. The effect of external Markovian noise on the physical qubits can be described by a superoperator called a quantum error channel [5], which is usually applied after the noisy gate and acts on the full density matrix of the system. A general error channel for two qubits can be written as follows:

$$\mathcal{E}_2^{\text{gen}}(\rho) = \sum_{i,j,k,l=0}^{3} p_{ijkl}(\sigma^i \otimes \sigma^j)\rho(\sigma^k \otimes \sigma^l) \qquad (4)$$

with $p_{ijkl}^* = p_{klij}$, and $\{\sigma^0, \sigma^1, \sigma^2, \sigma^3\} \equiv \{\mathbb{I}_2, \sigma^x, \sigma^y, \sigma^z\}$. The trace-preserving condition yields 16 equations:

$$\sum_{k,l} p_{klkl} = 1, \qquad (5)$$

$$\forall i \neq 0, \quad \forall j \neq 0, \quad \sum_k p_{kjk0} = \sum_k p_{ik0k} = p_{ij00} = 0, \quad (6)$$

reducing the naive number of $4^4 = 256$ independent parameters to 240. We do not know *a priori* the values for p_{ijkl}, and determining them would require quantum state tomography protocols with unrealistic overhead. Moreover, this noise originates from many different experimental sources and is not only different for different qubits, but can also change in time, making it impossible to fully characterize.

Randomized compiling is a computational method which allows for the reduction in the complexity of the noise channel [Eq. (4)] to only 15 free parameters, at the expense of running more circuit experiments. The resulting equivalent noise channel reads

$$\mathcal{E}_2^{\text{Pauli}}(\rho) = \sum_{i,j=0}^{3} p_{ij}(\sigma^i \otimes \sigma^j)\rho(\sigma^i \otimes \sigma^j), \qquad (7)$$

where $0 \leqslant p_{ij} \leqslant 1$ and $\sum_{i,j} p_{ij} = 1$. Such an error channel is called a Pauli noise channel.
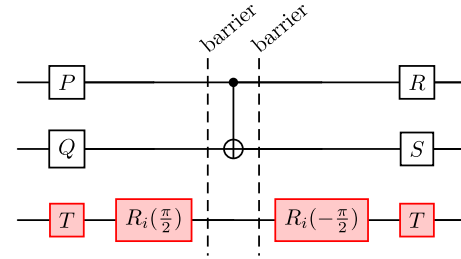

FIG. 5. Representation of the crosstalk RC protocol used for our three-qubit simulations. The white gates are used in the standard RC protocol. In order to twirl crosstalk errors, we extend this scheme by adding the red gates. Gates P, Q, R, S, and T are either the identity matrix or one of the three Pauli matrices. P, Q, and T are drawn randomly from $\mathcal{T}_p$, and R and S are chosen such that the circuit is strictly equivalent to an isolated CNOT. Gates $R_i(\pi/2)$ are $\pi/2$ rotations around the axis $i \in \{x, y, z\}$, drawn at random. Barriers indicate the order of application of gates and prevent annihilation between single-qubit gates.

RC consists of inserting a random layer of single-qubit gates (considered noiseless) before and after each CNOT gate such that the overall action is still equivalent to a CNOT gate:

$$(U_1 \otimes U_2)\text{CNOT}(U_3 \otimes U_4) = \text{CNOT}, \qquad (8)$$

where U_i ($i \in \{1, 2, 3, 4\}$) are random single-qubit gates, with combinations $U_1, \ldots, U_4$ deliberately chosen to satisfy Eq. (8). The set of allowed combinations comprises 16 options that can be constructed by choosing U_1 and U_2 arbitrarily from the set $\mathcal{T}_p = \{\mathbb{I}_2, \sigma^x, \sigma^y, \sigma^z\}$, and then choosing U_3 and U_4 from the same set such that Eq. (8) is satisfied. Table I of Ref. [41] lists the different possibilities for U_i's.

In the presence of noise, one can show [46] that averaging over all twirling configurations $\{U_i\}$ converts the general error channel, Eq. (4), to an effective Pauli noise, Eq. (7). In other words, the average over the expectation values measured from all possible RC circuits coincides with the value that would be obtained from the original circuit with a Pauli noise channel on each CNOT gate.

The number of circuits with different twirl configurations scales exponentially, though: $16^{n_{\text{CNOT}}}$, where n_{CNOT} is the number of CNOT gates in the circuit. Performing an average over this number of circuit experiments is unrealistic in practice. However, in the assumption of noise Markovianity (i.e., the noise on different CNOT gates is not correlated), one can sample very few twirl configurations. For an observable $\in [-1, +1]$, the deviation from the exact average scales $\mathcal{O}(1/\sqrt{N_{\text{RC}}})$ where N_{RC} is the number of RC circuits with different twirl configurations—in the spirit of Monte Carlo sampling. For $N_{\text{RC}} = 300$, error bars in Fig. 6 are below the plot marker size, underscoring the method's accuracy.

We note in passing that the above discussion concerns the cost of "converting" the error channel by means of RC. Further use of error mitigation techniques (such as NEC in Sec. V) on the measured results would require exponential accuracy of the measured results (cf. Fig. 8) and thus impose exponential growth on the required N_{RC} with the number of CNOT gates. Yet, in practice, this is not the limiting factor. For example, in Fig. 12, the number of RC circuits has been
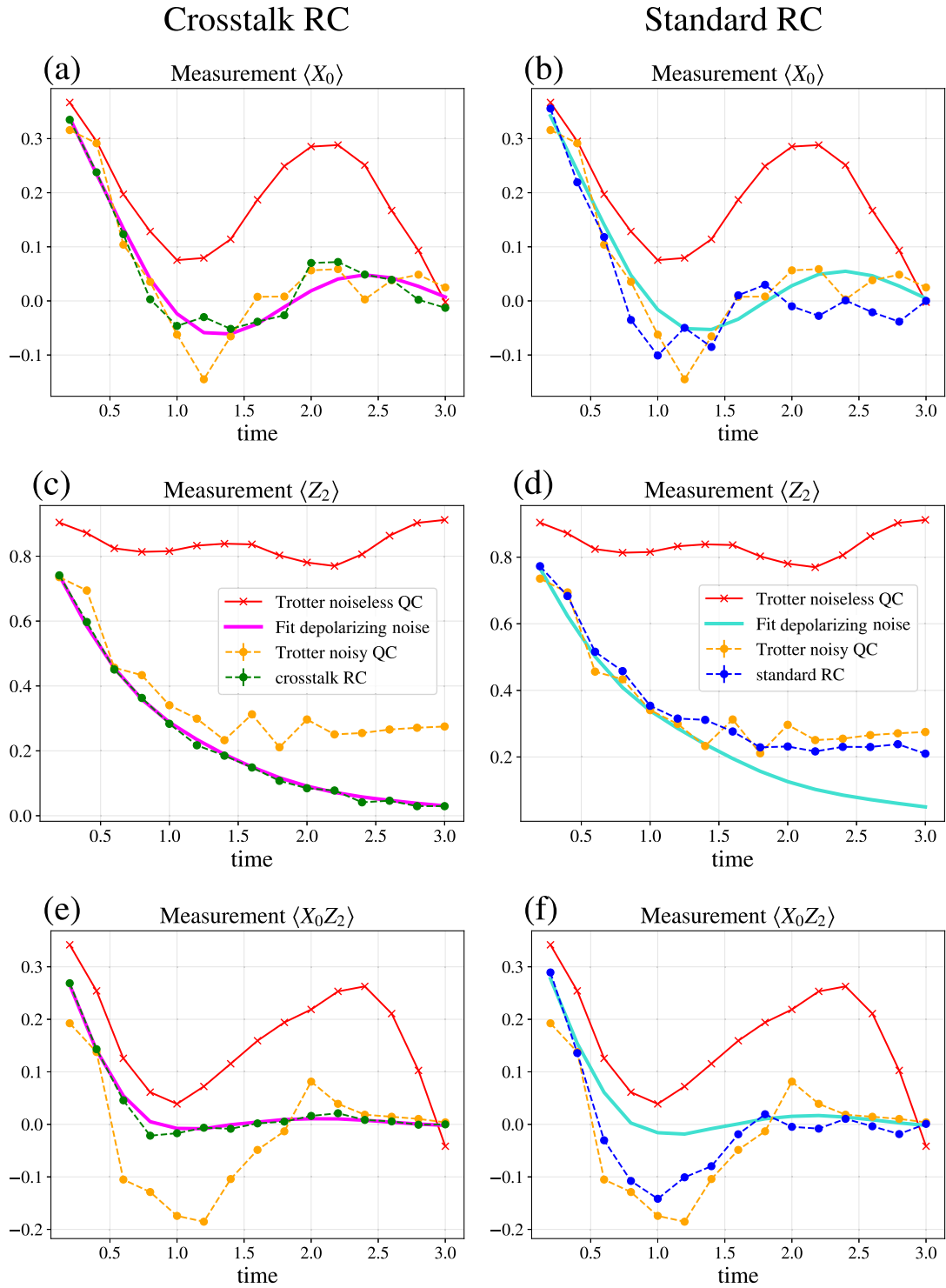
FIG. 6. Time evolution of $\langle X_0 \rangle$ (top panels), $\langle Z_2 \rangle$ (middle panels), and $\langle X_0 Z_2 \rangle$ (bottom panels). The green (blue) dashed curves on the left (right) panels correspond to the crosstalk (standard) RC protocol. The RC protocol has been averaged over 300 different circuits. The magenta (cyan) solid curves on the left (right) panels are the result of a fit of the green (blue) curves made using classical simulations of the circuit with the noise model of Eq. (11) (see main text). Error bars due to shot noise and finite sampling of random twirl configurations are computed in Appendix E. They are smaller than the plot marker size. The red and orange curves correspond respectively to the noiseless and noisy Trotterized dynamics and are plotted here for comparison.
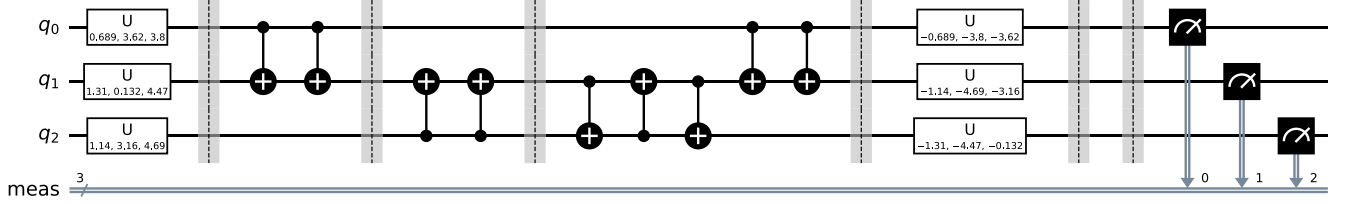
FIG. 7. Example of an estimation circuit for the first Trotter step. To build it, we remove the initial state initialization and single-qubit gates from the example circuit in Fig. 3. A random initial product state is added before the circuit and undone at the end (white boxes). We also perform RC on each CNOT gate (not shown here).

increased to 600 for $t > 1.8$, yet this adjustment does not bring about significant changes in comparison with Fig. 8.

B. Randomized compiling for crosstalk

So far, by restricting ourselves to two-qubit error channel, we have supposed that the noise acting on CNOT gates only affects the qubits on which the gate is applied. In practice, the application of a CNOT is known to also produce errors on neighboring qubits, an effect known as crosstalk [25,26,31,32]. Since the precise noise structure induced by crosstalk is also unknown, we aim to simplify it as well by extending RC to these neighboring qubits.

The crosstalk RC (cRC) procedure for three qubits is represented in Fig. 5. Unlike for the two active qubits on which the CNOT is applied, no logical operation is performed on the neighboring qubits. As a consequence, any single-qubit gates can be part of the twirling set [imagine replacing CNOT with identity in Eq. (8); then one could choose $U_3 = U_1^{-1}$ and $U_4 = U_2^{-1}$ for any U_1 and U_2, not only those belonging to $\mathcal{T}_p$]. We can thus simplify the structure of the noise channel *on the neighboring qubits* even further than for the active qubits by using two distinct twirling sets. The first one is the same twirling set $\mathcal{T}_p$ as in the previous section. The second is composed of the gates $\mathcal{T}_R = \{R_x(\pi/2), R_y(\pi/2), R_z(\pi/2)\}$. We perform two RC procedures in a row. First, we apply the gate drawn randomly from the set $\mathcal{T}_p$, then a second gate drawn from $\mathcal{T}_R$, and undo their action by choosing their corresponding inverse gate after application of the CNOT on the neighboring qubits.

In Appendix D, we derive the effective general n-qubit error channel ($n - 2$ neighbors that can experience crosstalk), as obtained after performing cRC [see Eq. (D23)]. The restriction of this channel to our three-qubit BCS simulations (with only one neighbor) has $16 \times 2^{n-2} - 1 = 31$ free parameters:

$$\mathcal{E}_3^{\mathrm{cRC}}(\rho) = \frac{1}{3} \sum_{i,j=0}^{3} \sum_{s=0,1} p_{i,j,s} \sum_{a=1,2,3} P^{i,j,s,a} \rho P^{i,j,s,a},$$

with $P^{i,j,s,a} = \sigma_1^i \otimes \sigma_2^j \otimes (\sigma_3^a)^s$, $\frac{1}{3} \sum_{i,j=0}^{3} \sum_{s=0,1} p_{i,j,s} = 1$, (9)

where indices 1 and 2 stand for the control and target qubits of the CNOT, s indicates if an error has occurred on the neighboring qubit ($s = 1$) or not ($s = 0$), and a encodes whether it was an X, Y, or Z error ($a = 1, 2$ or 3). Note, in particular,

that taking the partial trace of the density matrix over the neighboring qubit subspace yields the same Pauli channel as in the standard RC protocol presented in the last section as the effective error channel for the active qubits [see Eqs. (7) and (D21)]. Likewise, tracing over the active qubits yields a one-qubit depolarizing noise channel for the neighboring qubit because $p_{i,j,s}$ does not depend on a:

$$\mathcal{E}_1^{\mathrm{dep}}(\rho) = (1 - p)\rho + \frac{p}{3} \sum_{i=1}^{3} \sigma^i \rho \sigma^i = (1 - \lambda)\rho + \lambda \frac{\mathbb{I}}{2},$$

(10)

where $0 \leqslant \lambda = \frac{4p}{3} \leqslant \frac{4}{3}$, $p = \sum_{i,j=0}^{3} p_{i,j,1}$ is the error rate, and λ the depolarizing parameter. The key point about cRC is that it simplifies the error channel of neighboring qubits because it symmetrizes the effect of noise on the Bloch sphere with respect to the different directions. Application of $R_i(\pi/2)$ permutes the eigenstates of σ_j and σ_k: $|0\rangle_j \to |0\rangle_k \to |1\rangle_j \to |1\rangle_k \to |0\rangle_j$. Indices are such that $\epsilon_{ijk} = +1$, where ϵ is the the Levi-Cività tensor. As a result, the depolarizing noise channel is obtained by equally mixing the different coefficients of the Pauli noise channel.

C. Comparing standard and crosstalk RC protocols through BCS simulation

We apply both the standard and crosstalk RC methods from the last sections to the BCS simulation algorithm described in Sec. II. For the crosstalk RC protocol (standard RC protocol), results are shown in Figs. 6(a), 6(c), and 6(e) [Figs. 6(b), 6(d), and 6(f)]. As expected, both RC protocols do not improve the accuracy of the results on their own, as they simplify the CNOT noise structure, but do not eliminate the noise. However, we observe that they tend to flatten the curve and attenuate the unpredictable fluctuations of the expectation values due to the complex structure of the original noise. This could be explained by the fact that the coherent part of the noise, typically responsible for such fluctuations, has indeed been turned incoherent.

To further interpret these results, we try to replicate the QC's output with classical simulations of our Trotter quantum circuits using a simple noise channel. In the following, we assume that the noise acting on the quantum device is close to isotropic (in the sense of the qubit basis), so that the Pauli noise channel restricted to the active qubits can be approximated by a two-qubit depolarizing noise channel. This assumption is rather standard [41] for quantum error mitigation protocols because it allows to relate noisy and
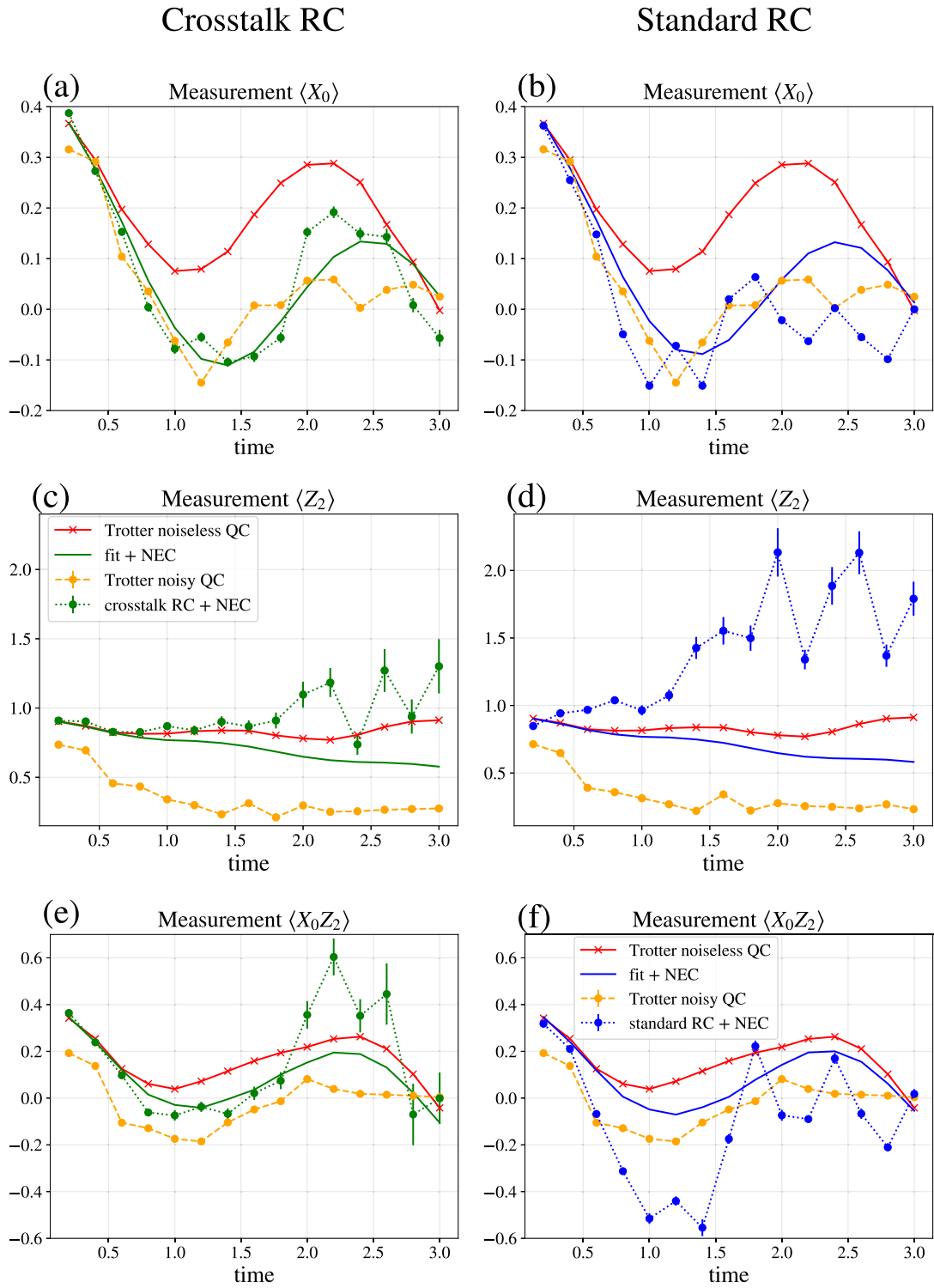
FIG. 8. Evolution of $\langle X_0 \rangle$ (top panels), $\langle Z_2 \rangle$ (middle panels), and $\langle X_0 Z_2 \rangle$ (bottom panels) after quantum error mitigation protocols (NEC + cRC/RC + REC). Green (blue) dotted curves on the left (right) panels are the results of the real IBMQ device using crosstalk (standard) RC protocols. Green (blue) solid curves are the fits obtained in Figs. 6(a), 6(c), and 6(e) (magenta curves) [Figs. 6(b), 6(d), and 6(f) (cyan curves)] mitigated using classical simulation of NEC. Error bars correspond to the uncertainty due to the limited number of shots and the finite sampling of twirl configurations.

noiseless expectation values analytically. Under this assumption, the $(n = 3)$-qubit error channel can be described by only $2^{n-1} - 1 = 3$ free parameters $\{\lambda_{\text{CNOT}}, \lambda_{\text{neigh}}, \lambda_{\text{glob}}\}$:

$$\mathcal{E}_3(\rho) = (1 - \lambda_{\text{CNOT}} - \lambda_{\text{neigh}} - \lambda_{\text{glob}})\rho$$
$$+ \lambda_{\text{CNOT}} \frac{\mathbb{I}_{01}}{4} \otimes \text{Tr}_{01}(\rho)$$
$$+ \lambda_{\text{neigh}} \text{Tr}_2(\rho) \otimes \frac{\mathbb{I}_2}{2} + \lambda_{\text{glob}} \frac{\mathbb{I}_{012}}{8}, \qquad (11)$$

where 0 and 1 denote the indices of the active qubits and 2 the index of the neighboring qubit, and $\text{Tr}_{\{k\}}$ is the partial trace over the set of qubits $\{k\}$. We call this noise channel a quasilocal depolarizing noise.

As displayed in Figs. 6(a), 6(c), and 6(e) (magenta curves), the QC results can be well reproduced by classical simulations (using the density matrix method of the noise simulator QISKIT Aer) performed with a quasilocal depolarizing noise [see Eq. (11)] applying on every CNOT of each junction (between qubits 0 and 1 and between qubits 1 and 2). This leaves us with five free parameters, $\lambda_{\text{CNOT}}^{01}$, $\lambda_{\text{neigh}}^{01}$, $\lambda_{\text{CNOT}}^{12}$, $\lambda_{\text{neigh}}^{12}$, and λ_{glob}, that we assume are the same for each Trotter step. We find a set of parameters that minimizes the square distance between the classical simulation results and all the data points for the seven different observables ($7 \times 15 = 105$ data points). The resulting fit parameters are $\lambda_{\text{CNOT}}^{01} \simeq 0$, $\lambda_{\text{CNOT}}^{12} \simeq 0.014$, $\lambda_{\text{neigh}}^{01} \simeq 0.05$, $\lambda_{\text{neigh}}^{12} \simeq 0.01$, and $\lambda_{\text{glob}} \simeq 0.002$, with the regression standard error $\chi^2 = 2 \times 10^{-4}$. From the fit parameters, it is clear that the noise associated to *neighboring qubits* cannot be neglected. Rather, they seem to confirm that crosstalk plays an important role in IBM's quantum devices and that the scheme we are proposing is essential for the implementation of efficient error mitigation protocols. The noise parameters derived differ noticeably from the CNOT gate error rate (around 0.005–0.01) announced by IBM. We believe that this discrepancy stems from the method used to estimate these parameters. IBM uses the randomized benchmarking protocol [18,19], which does not take crosstalk into account because each junction is considered separately. The resulting error rate does not seem to relate easily to the Pauli noise channel parameters that we obtain through the cRC method.

For comparison, we also plot in Figs. 6(b), 6(d), and 6(f) the fit by classical simulations of the results we obtain using the standard RC protocol (no twirling of neighboring qubits). The resulting fit parameters are $\lambda_{\text{CNOT}}^{01} \simeq 0.007$, $\lambda_{\text{CNOT}}^{12} \simeq 0$, $\lambda_{\text{neigh}}^{01} \simeq 0.07$, $\lambda_{\text{neigh}}^{12} \simeq 0.009$ and $\lambda_{\text{glob}} \simeq 0.0002$ and the

regression standard error is $\chi^2 \simeq 0.02$. The quality of the fit is not as good as in the former case, indicating that the noise model of Eq. (11) is a rough approximation of the real noise occurring on the QC.

The above shows that cRC is *essential* to accurately describe the noise on the system with Eq. (11), ensuring that the effect of crosstalk is reduced to a depolarizing noise channel on the neighboring qubits. In turn, the simpler noise structure facilitates error mitigation, as we illustrate in the next section.

IV. ERROR MITIGATION TECHNIQUES

In order to illustrate the practical implications of our cRC protocol, we implement two auxiliary error mitigation schemes. First, to tackle errors stemming from measurements, we apply a simple readout error correction (REC) procedure detailed in Ref. [66]. Second, to mitigate the noise affecting CNOT gates, we use the NEC method proposed in Ref. [41].

A. Readout error correction

Measurement is expected to introduce an error rate of 2% (data taken from IBM back-end properties). In what follows, we systematically apply the iterative Bayesian unfolding method presented in Ref. [66] to correct readout errors (REC). Note that we invert the whole $2^3 \times 2^3$ measurement matrix, taking into account measurement error correlations. This method scales poorly with the number of measured qubits. However, we have found that performing the inversion qubit by qubit, discarding the correlations, yields similar results.

B. Noise estimation circuits

The NEC technique aims to restore the noiseless value of an observable by dividing the noisy expectation value of an observable by a certain factor which estimates the circuit fidelity. This factor is obtained by running additional circuits derived from the original one, while preserving the same CNOT structure, and in practice allows one to partly get rid of the errors stemming from depolarizing noise channels.

To understand the above statement, first, one has to relate the noisy and the noiseless expectation values of an observable. In the same way as done in Ref. [41] for local and global depolarizing noise (see also Appendix F), we expand the noisy outcome up to the first order in the number of errors for the quasilocal depolarizing noise channel [see Eq. (11)], which yields

$$\langle O \rangle_{\text{noisy}} \simeq (1 - \lambda_{\text{glob}})^{n_{\text{CNOT}}} \langle O \rangle_0 \left[1 + \lambda_{\text{CNOT}} \left(\frac{\langle O \rangle_1}{\langle O \rangle_0} - n_{\text{CNOT}} \right) + \lambda_{\text{neigh}} \left(\frac{\langle O' \rangle_1}{\langle O \rangle_0} - n_{\text{CNOT}} \right) \right] + \mathcal{O}(\lambda_{\text{CNOT}}^2, \lambda_{\text{neigh}}^2), \qquad (12)$$

where n_{CNOT} is the number of CNOT gates in the circuit, and $\langle O \rangle_0$ is the noiseless outcome. $\langle O \rangle_1$ is the sum of expectation values obtained from all circuits where an error occurred on a single CNOT gate, putting the active qubits of the CNOT in the maximally mixed state $\mathbb{I}/4$. $\langle O' \rangle_1$ is the sum of expectation values obtained from all circuits where an error occurred on a single CNOT gate, putting the *neighboring* qubit of the CNOT

in the maximally mixed state $\mathbb{I}/2$. For exposition simplicity, we consider here that noise parameters for different junctions are equal, but the result can be generalized to CNOT gates with different noise parameters for different junctions.

To mitigate these errors, it has been proposed in Ref. [41] to run extra circuits (called noise estimation circuits) whose outcome yields a part of the error terms appearing in Eq. (12).

To construct these NECs, one simply removes all single-qubit gates from the original Trotter circuits (see Fig. 7).

In our Trotter circuits, we can distinguish two types of CNOT gates. The first type is the CNOT gates encoding the interaction terms. After removing all the single-qubit gates, they are now applied in a row with the same control and target qubits. As they always come in pairs, they are now logically equivalent to the identity operator. The second type

is the CNOT gates implementing the SWAP gate. Their action does not change after removing single-qubit gates. The overall unitary operation performed by the estimation circuits, for any number of Trotter steps, is thus the identity operator up to some SWAP operations.

Applying Eq. (12) to the NEC and dividing the noisy Trotter circuits outcome by the fidelity estimated via NEC leads to the mitigated expectation value:

$$\langle O \rangle_{\text{mitigated}} = \frac{\langle O \rangle_{\text{noisy}}}{\langle E \rangle_{\text{noisy}}} \simeq \langle O \rangle_0 \left[1 + \lambda_{\text{CNOT}} \left(\frac{\langle O \rangle_1}{\langle O \rangle_0} - \langle E \rangle_1 \right) + \lambda_{\text{neigh}} \left(\frac{\langle O' \rangle_1}{\langle O \rangle_0} - \langle E' \rangle_1 \right) \right] + \mathcal{O}\left(\lambda_{\text{CNOT}}^2, \lambda_{\text{neigh}}^2 \right), \quad (13)$$

where $\langle E \rangle_1$ and $\langle E' \rangle_1$ are the NEC measured expectation values, which correspond respectively to $\langle O \rangle_1$ and $\langle O' \rangle_1$ in the original circuits and $\langle E \rangle_0 = 1$. Equation (13) shows that the NEC mitigates exactly the noise associated with the global part of our quasilocal depolarizing channel. For the active qubits and neighboring depolarizing channels, it is expected that the coefficients in front of λ_{CNOT} and λ_{neigh} are strongly suppressed.

To show this, we propose in Appendix G an expansion in terms of the Trotter time step Δt (the computation is done in the case of a local depolarizing noise). This computation takes into account any number of errors in the circuit but only relates the noisy result to the noiseless terms at zeroth and first order Δt. Applying the equation found [see Eq. (G11)] to the case of NEC gives an exact result as the only nonvanishing term is of order zero in Δt. Therefore, one finds that the main term of the noisy Trotter circuit is reproduced by the fidelity estimated using NEC. The result holds for any number of errors. In particular, at small time (i.e., for a few Trotter steps), we have

$$\frac{\langle O \rangle_1}{\langle O \rangle_0} \simeq \langle E \rangle_1. \quad (14)$$

As a consequence, NEC is an efficient method to remove noise at small evolution times, or equivalently for small-depth circuits. The results can be generalized also for quasilocal depolarizing noise. At longer evolution times, we expect the noise estimated by NEC to differ from the noise on the Trotter circuits. To remove this remaining noise, it has been proposed in Ref. [41] to use the zero-noise extrapolation method [11,12], which consists in artificially increasing the noise of sets of CNOT gates [39] in order to extrapolate the result to the case of zero noise. In this article, we do not apply this additional mitigation scheme because, as we will see, at longer evolution times the QC result deviates from the classical simulations using depolarizing noise, indicating that the cRC noise channel can no longer be described accurately as an effective quasilocal depolarizing noise channel.

We remind the reader that making the noise describable in terms of Eq. (11) requires the application of RC techniques. Therefore, any circuits we run involve twirling applied in agreement with either RC or cRC protocols described in Sec. III.

Finally, in order to stabilize the fidelity-estimation output over different initial states, we generate random initial product states and undo them at the end after the application of the

noise estimation circuit by inserting the corresponding inverse gates and taking into account the action of the SWAP gates. We produce $N_{\text{RC}} = 300$ such noise estimation circuits, each of which has a different random initial state and a different twirling configuration for the RC (cf. Fig. 7).

V. COMBINING CROSSTALK RC AND ERROR MITIGATION

In order to obtain the error-mitigated results and thus assess the ultimate benefit of replacing RC with cRC, we combine all the techniques described above: REC, cRC/RC, and NEC. Namely, we divide the Trotter result obtained in Sec. III— after averaging over 300 RC circuits and performing REC— by the result of the noise estimation circuits as described in Sec. IV, also averaged over 300 RC circuits and with REC.

The results for the observables $\langle X_0 \rangle$, $\langle Z_2 \rangle$, and $\langle X_0 Z_2 \rangle$ using the crosstalk RC technique are plotted in Figs. 8(a), 8(c), and 8(e). The results for the standard RC protocol are shown in Figs. 8(b), 8(d), and 8(f). While not perfect, the results using cRC are clearly better than the results using the standard RC.

Note the error bars that are sizable in Fig. 8, as opposed to the negligible ones in Fig. 6. This is due to dividing by the fidelity estimate in the NEC protocol, which amplifies the expectation value, but also the errors. For the same reason, the expectation values of the observables may lie outside their range $[-1, 1]$. In practice, such behavior indicates that we have reached the regime where the efficiency of NEC falls off, as unaccounted sources of noise kick in.

Results for all the remaining observables $\langle X_0 \rangle$, $\langle Y_1 \rangle$, $\langle Z_2 \rangle$, $\langle X_0 Y_1 \rangle$, $\langle X_0 Z_2 \rangle$, $\langle Y_1 Z_2 \rangle$, and $\langle X_0 Y_1 Z_2 \rangle$ are gathered in Fig. 12 of Appendix C. The conclusion of cRC producing better results than RC holds for these as well. This is quantitatively demonstrated in Fig. 1(b) by comparing the accuracy averaged over all the observables. Below we provide a detailed analysis of the results.

A. Global accuracy improvement

A comparison between the two columns of Fig. 8 clearly shows that cRC, i.e., the twirling of neighboring qubits using rotation gates, qualitatively improves the accuracy of the noise estimation protocol, while not requiring to run more circuits, compared to the standard RC protocol. These experiments demonstrate that, at least in the context of superconducting

qubit architectures like IBMQ QC, this method should systematically be preferred.

The error mitigation, however, remains imperfect, as simulations still deviate from the noiseless result. These deviations can be separated into two categories. The first comprises the cases where the noiseless expectation values that we are measuring become close to zero. In this case, we observed that the RC result of the Trotter circuit sometimes gets the wrong sign [see, e.g., Fig. 8(a), $1 \leqslant t \leqslant 1.8$, and Fig. 8(e), $0.8 \leqslant t \leqslant 1.4$]. As a consequence, dividing by the fidelity estimate of NEC (which is positive by definition) only enhances this sign problem and makes the result deviate strongly from the noiseless curve, even at intermediate times. In spite of this, the oscillations of the exact evolution are still well reproduced by the mitigated curve. The second category relates to the strong deviations occurring at longer times [see, e.g., Figs. 8(c) and 8(e), $t \geqslant 2$]. We identified the two problems as having different origins. The first one stems from the (expected) effect of a local, two-qubit depolarizing noise [cf. Eq. (11)], while the second issue is caused by the fact that the noise channel is not exactly depolarizing after RC, which negatively affects the output of the estimation circuits. This deviation becomes more important as the time increases since the estimated fidelity becomes very small. The next two sections are devoted to our analysis of these deviations and their origin.

B. Error mitigation enhances the wrong sign

In this section, we discuss why the mitigated result may display a wrong sign as the expectation value approaches zero, inducing unwanted deviations from the perfect result, which are in turn enhanced by the NEC technique. In the case of a pure global depolarizing noise, Eq. (12) [or Eq. (F2)] guarantees that the noisy result stays of the same sign as the noiseless one. However, when the noise is local (but still depolarizing) this may not hold anymore. The equation becomes more involved, and the different terms in the Trotter expansion do not experience the same multiplicative error [see Eq. (G11) of Appendix G]. In Fig. 8, we simulate the Trotter circuit using the local noise channel of Eq. (11) for each CNOT and the depolarizing parameter from the fit of Fig. 6 (magenta and cyan curves on Fig. 6). We then apply the same noise channel to the estimation circuits and divide both results to obtain the mitigated one (green and blue solid curves in Fig. 8). In Fig. 8, we show classical simulations using the noise channel of Eq. (11), which reproduce quite well the observed deviation of the QC results. Therefore, in this case, the deviation from the noiseless dynamics is explained by the local depolarizing noise occurring on qubits. In this regime, the NEC protocol only works approximately, but other error mitigation protocols such as zero-noise extrapolation (ZNE) are more efficient to tackle local depolarizing noise, and can be combined with NEC to improve the results. In Figs. 8(c) and 8(e), at longer time, we observe that mitigated fits (green solid curves) deviate slightly from the noiseless Trotter dynamics, but this deviation does not correspond to the one observed on the QC
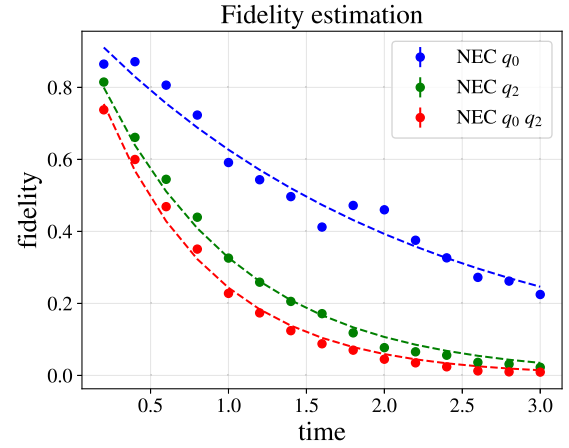

FIG. 9. Effective fidelity estimation for qubits q_0 (blue points), q_2 (green points), and q_0q_2 (red points), obtained by averaging over 300 randomly compiled noise estimation circuits. Solid lines are the results of the fit using the noise model of Eq. (11). Error bars due to shot noise and finite sampling of RC configurations are smaller than the plot marker size.

(green dotted curves), which indicates that it is not caused by the locality of the noise.

C. Analysis of long-time deviations

In Fig. 8, the error bars quantify the statistical fluctuations due to the finite sampling of the RC method and the finite number of shots (see Appendix E for a detailed explanation of the method used to estimate statistical fluctuations). The deviation observed from the classical simulations performed using a quasilocal depolarizing noise are not within these error bars. In Fig. 12, we have increased N_{RC} of NEC from 300 to 600 for $t \geqslant 1.8$ in order to reduce this uncertainty. If one compares Figs. 8(a), 8(c), and 8(e) together with Figs. 12(a), 12(c), and 12(e), data points have moved consistently within the range given by the error bars, and there is no clear improvement of the results. Another source of error must be the cause of this deviation.

Since the deviations are not related to an insufficient number of RC samples, or to shot noise, they must originate from a poor estimation of the effective noise parameters on the Trotter circuits after cRC by the NEC. To understand this, we compare the noise parameters of Eq. (11) estimated using the Trotter circuits with cRC with the ones obtained from the fidelity estimations from the estimation circuits (see Fig. 9). We fit the seven different estimation curves and find that resulting fit parameters obtained are $\lambda_{\mathrm{CNOT}}^{01} \simeq 0.007$, $\lambda_{\mathrm{CNOT}}^{12} \simeq 0.016$, $\lambda_{\mathrm{neigh}}^{01} \simeq 0.05$, $\lambda_{\mathrm{neigh}}^{12} \simeq 0.009$, and $\lambda_{\mathrm{glob}} \simeq 0.002$ and $\chi^2 \simeq 4 \times 10^{-4}$. These results only slightly differ from what we obtained from the randomized compiling of Trotter circuits (see Sec. III C). These small deviations on the noise parameters estimated by the two types of circuits are not relevant at small times but become crucial at longer times. Indeed, the uncertainty on the mitigated result becomes large as the

denominator (estimation circuit) approaches zero. It can be estimated as (see the Supplemental Material of Ref. [41])

$$\sigma_m^2 = \frac{\langle O \rangle_{\text{noisy}}^2 \sigma_e^2 + \langle E \rangle_{\text{noisy}}^2 \sigma_o^2}{\langle E \rangle_{\text{noisy}}^4} \tag{15}$$

$$= \frac{\langle O \rangle_{\text{mitigated}}^2 \sigma_e^2 + \sigma_o^2}{\langle E \rangle_{\text{noisy}}^2}, \tag{16}$$

where $\langle E \rangle_{\text{noisy}}$ is the mean value estimated from the NEC, $\langle O \rangle_{\text{noisy}}$ is the mean value estimated from the Trotter circuits, $\langle O \rangle_{\text{mitigated}} = \frac{\langle O \rangle_{\text{noisy}}}{\langle E \rangle_{\text{noisy}}}$, and σ_e, σ_o, and σ_m their respective standard deviations. Figure 9 shows that the fidelity is exponentially decreasing with the number of CNOT gates. Note, in particular, that the noise estimation method should be less stable for multiqubit observables because one expects their fidelity to be smaller compared to single-qubit observables (as they are more likely to be directly affected by errors). For an example, see Fig. 12(g) in Appendix C, where the error bars (uncertainty due to shot noise and the finite sampling of the RC method) become large for the observable $X_0 Y_1 Z_2$ at long time.

The noise parameter uncertainty stemming from the Trotter and the noise estimation circuits reflects the deviation of the effective noise from a pure quasilocal depolarizing noise even after the application of cRC. This deviation can have two origins. On the one hand, the assumption of a two-qubit depolarizing noise on the CNOT's active qubits is not exact, as the theorem only guarantees a Pauli noise channel. To avoid making this extra assumption, it will be either necessary to find a way to further simplify the Pauli noise channel on active qubits into a depolarizing noise or to use more sophisticated error mitigation protocols (as compared to NEC) which take into account the full effect of the Pauli noise. On the other hand, it may be possible that a single-qubit gate's error cannot be neglected anymore. For the longest circuit using the crosstalk twirling protocol, we have on average (depending on the twirl configuration) $n_1 = 1187$ single-qubit gates and 135 CNOT gates. The single-qubit gate's error rate being estimated at $\epsilon_1 \sim 10^{-4}$, their total effect on the output of a circuit should be around $\sqrt{n_1} \times \epsilon_1 \sim 0.3\%$. However, as we have seen with Eqs. (15) and (16), small deviations can become important when the denominator approaches zero.

VI. CONCLUSION

In this work, we have presented and analyzed simulations of the BCS Hamiltonian for three energy levels on an IBM superconducting quantum computer. To obtain reliable results from the noisy QC, we employ a simple error mitigation technique called NEC, whose efficiency is conditioned by the simplicity of the noise structure. In order to achieve the best possible results, we use the RC technique, which effectively turns the unknown and complex noise channels acting on the QC into a simpler channel.

We use two versions of RC: the standard RC and the crosstalk RC. We find that the latter enables achieving much better accuracy, by further simplifying the noise caused by crosstalk effects, while not requiring extra circuit runs or qubits compared to the former. This demonstrates the significance of crosstalk on IBM quantum computers.

Our work closes a gap in the field of error mitigation on NISQ devices: while the specific structure of the noise on real devices is rather complex and usually not well understood, theoretical proposals for error mitigation techniques usually assume a simple depolarizing error channel. We explicitly demonstrate that the use of crosstalk RC does indeed bring unknown noise closer to depolarizing noise, allowing us to take advantage of existing error mitigation proposals. This is the main reason for the significant overall performance improvements stemming from the use of our crosstalk RC protocol, as shown in Fig. 1.

There is another appeal in our approach: crosstalk errors are notoriously hard to characterize. Being a systematic, coherent noise channel, it is difficult to predict their net effect on the output of a given quantum computation. Using crosstalk RC and our theoretical prediction for the structure of the effective error channel, we show that existing protocols significantly underestimate the real error in superconducting quantum computers. We find in particular that crosstalk error channels map to a depolarizing error channel with error rate per CNOT gate that can be as high as 5%, i.e., almost five times the values obtained by randomized benchmarking for the error rate on the active qubits of the CNOT gates. This result is in good qualitative agreement with the recent papers [67,68], in which crosstalk is also found to be the main source of errors for quantum computation on IBM quantum computers.

The accuracy gain from using crosstalk RC, in comparison to the standard RC, in the explicit simulations of the dynamics of the BCS model on NISQ devices is significant. It allowed us to perform intermediate-time (up to 15 Trotter steps) quantum simulations of an *all-to-all interacting model* with an accuracy beyond what can, to our knowledge, be found in the literature. This brings us one step closer to quantum simulations on NISQ devices matching the capabilities of classical computers.

Data and the code used to generate quantum circuits and analyze the results are available in Ref. [69].

APPENDIX A: BCS HAMILTONIAN IN TERMS OF PAULI MATRICES

In this Appendix, we show how to rewrite the standard Hamiltonian for the BCS model for superconductivity in terms of spin operators. The BCS Hamiltonian couples electron pairs known as Cooper pairs [51] through a phonon-mediated interaction [70]. The effective Hamiltonian,

obtained after eliminating the phonons, can be written in the form [71]

$$\mathcal{H}_{\text{BCS}} = \sum_{j=0}^{L-1} \epsilon_j (c_{j\uparrow}^\dagger c_{j\uparrow} + c_{j\downarrow}^\dagger c_{j\downarrow}) - g \sum_{i,j=0}^{L-1} c_{i\uparrow}^\dagger c_{i\downarrow}^\dagger c_{j\downarrow} c_{j\uparrow},$$

(A1)

where L is the number of energy levels, $c_{j,\sigma}^{(\dagger)}$ is the usual electronic annihilation (creation) operator acting on the energy level j of energy ϵ_j and spin $\sigma \in \{\uparrow, \downarrow\}$. The coupling constant g describes the effective interaction between Cooper pairs. Single electrons are trivially decoupled from the dynamics. We therefore restrict the study to the Cooper pairs' subspace. The latter mimic the statistics of hard-core bosons [72]; their action in terms of creation and annihilation operators is easily related to the algebra of Pauli matrices:

$$\sigma_j^z = 1 - c_{j\downarrow}^\dagger c_{j\downarrow} - c_{j\uparrow}^\dagger c_{j\uparrow},$$

(A2)

$$\sigma_j^+ = \frac{\sigma_j^x - i\sigma_j^y}{2} = c_{j\uparrow}^\dagger c_{j\downarrow}^\dagger,$$

(A3)

$$\sigma_j^- = \frac{\sigma_j^x + i\sigma_j^y}{2} = c_{j\downarrow} c_{j\uparrow},$$

(A4)

where σ_j^a, $a \in \{x, y, z\}$, refers to the operator such that the Pauli matrix σ^a is applied on the energy level j and the identity operator on the other energy levels. The eigenstates $|0\rangle$, $|1\rangle$ of σ_j^z denote respectively the absence or presence of a Cooper pair on the energy level j. Using this transformation, the BCS Hamiltonian can be recast as

$$\mathcal{H}_{\text{BCS}} = -\sum_{j=0}^{L-1} \left(\epsilon_j - \frac{g}{2}\right)\sigma_j^z - \frac{g}{2} \sum_{0 \leqslant i < j \leqslant L-1} (\sigma_i^x \sigma_j^x + \sigma_i^y \sigma_j^y),$$

(A5)

which is Eq. (1). This rewriting in terms of Pauli matrices is closely related to the Anderson 1/2-pseudospin model [73]. The only difference is that the z component of the pseudospin is defined as $S_j^z = -\frac{\sigma_j^z}{2}$, where j is the energy level index. Note that the interaction term rewritten using Pauli matrices does not suffer from long Pauli string terms scaling with the system size as is usually the case for electrons due to the Jordan-Wigner transformation [74,75], which ensures the antisymmetrization of the wave function.

APPENDIX B: CIRCUIT TRANSPILATION

Each term of Eq. (3) needs to be implemented with IBM QC's native gates. They are three single-qubit gates, $R_z(\theta)$, the Pauli matrix σ^x (or simply denoted X), and $\sqrt{X}$, and one two-qubit gate, the CNOT gate [76]. They form a universal set of gate operations which allows to implement any unitary operator. An additional difficulty arises when taking into account the machine's layout: the unitary contains interaction terms that couple qubits which are not directly connected on the QC; we devote Appendix B 3 to this problem.

1. Transpilation of one-body operators

The leftmost product of Eq. (3) corresponds to on-site energies and can therefore be implemented as a product of one-qubit gates. It is indeed a product of $R_z(\theta) = e^{-i\frac{\theta}{2}\sigma^z}$ gates,
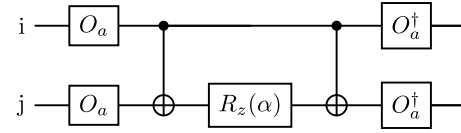

FIG. 10. Quantum circuit decomposition of the unitary operator $e^{-i\frac{\alpha}{2}\sigma_i^a \sigma_j^a}$ with $a \in \{x, y, z\}$. Here i and j denote qubit indices, and the vertical lines represent the two-qubit CNOT gates, for which the control qubit is indicated by the black dot. The other rectangles represent one-qubit gates explicitly mentioned in Eqs. (B3).

which perform a rotation of angle θ around the Z axis on the Bloch sphere. In our case $\theta_j = -\Delta t(2\epsilon_j - g)$, where j is the index of the different qubits. The R_z gate is a native gate for IBM quantum computers.

2. Transpilation of two-body operators

The rightmost product of Eq. (3) describes the interaction between Cooper pairs. To implement them, one needs to use two-qubit gates for entangling both qubits. The CNOT gate is the native two-qubit gate of most IBM QCs.

Using $[\sigma_i^x \sigma_j^x, \sigma_i^y \sigma_j^y] = 0$, we obtain

$$e^{-i\frac{\alpha}{2}(\sigma_i^x \sigma_j^x + \sigma_i^y \sigma_j^y)} = e^{-i\frac{\alpha}{2}\sigma_i^x \sigma_j^x} e^{-i\frac{\alpha}{2}\sigma_i^y \sigma_j^y},$$

(B1)

where $\alpha = -g\Delta t$ in our case. To implement two-body interaction terms of the form

$$e^{-i\frac{\alpha}{2}\sigma_i^a \sigma_j^a},$$

(B2)

where $a \in \{x, y, z\}$, one needs two CNOT gates, a single $R_z(\alpha)$ gate, and a one-qubit gate O_a:

$$O_z = \mathbb{I}, \ O_x = H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

$$\text{and} \quad O_y = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}.$$

(B3)

We represent in Fig. 10 the two-qubit quantum circuit equivalent to exponential (B2). Therefore, implementing Eq. (B1) as a quantum circuit requires four CNOT gates. Note, however, that after implementing such an interaction term using QISKIT [76], we use the `transpile` function with the option `optimization_level=3` (i.e., the most optimized method), which outputs an equivalent quantum circuit using only two CNOT gates (with some single-qubit native gates).

3. Transpilation for the whole dynamics

In this section, we use the one-qubit and two-qubit quantum circuits previously derived to build the circuit implementing the fully Trotterized dynamics. Naively, we could simply concatenate the corresponding elementary circuits on the physical qubits $i, j \in \{0, L - 1\}$ to reproduce the product of Eq. (3). We see then that the number of two-qubit interaction circuits required grows as $\binom{L}{2} = \frac{L(L-1)}{2}$, where L is the number of energy levels, which translates into $L(L-1)$ noisy CNOT gates for each Trotter step. However, an additional difficulty stems from the architecture of the IBM quantum computers (and superconducting quantum computers in general). As shown in Fig. 2, the QCs used (ibm_lagos and
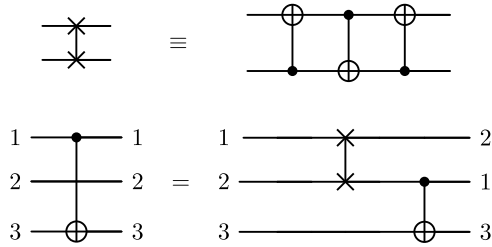
FIG. 11. Top: Logical decomposition of the SWAP gate (left) into three native CNOT gates (right). Bottom: Example of usage of the SWAP gate (represented with crosses at each end). The numbering refers to the position of logical qubits, while the solid horizontal lines are the physical qubits of the QC. The leftmost circuit cannot be realized in a layout where qubits 1, 2, and 3 are on a line, so we have to use the rightmost circuit instead. Because of this, the number of required CNOT gates increases by three, and the positions of the logical qubits are modified.

ibmq_ehningen) have a certain qubit coupling map, and CNOT gates can only be performed on neighboring physical qubits. As a consequence, it is not straightforward to implement an interaction between two distant qubits (in the sense of the coupling map of Fig. 2). To overcome this problem, one has to swap around the logical information contained in physical qubits, so that two interacting logical qubits are brought together on neighboring physical qubits. This swap of information is achieved using the so-called SWAP gate. A simple example of the use of SWAP gates is given in Fig. 11. Note that the SWAP gate is not a native gate of the IBM machines. It requires three CNOT gates as shown in Fig. 11.

To implement an interaction between two qubits initially separated by a distance d (in the sense of the coupling map of Fig. 2), at least $d - 1$ SWAP gates are required, i.e., $3(d - 1)$ CNOT gates. Consequently, the number of CNOT gates—and thus the number of errors—grows very quickly with the size of the system we want to simulate (i.e., the number of qubits).

4. Initial state preparation

In an effort to minimize the number of noisy CNOT gates in the circuit, we made the choice to start with an initial product (nonentangled) state, which can be prepared using only single-qubit gates. A natural product state in the BCS problem is the mean-field ground state which is the ground state of the BCS Hamiltonian in the thermodynamic limit but has a nontrivial dynamic for finite system. To construct it, we use the gap equation derived from the mean-field approximation [77]:

$$1 = g \sum_j \frac{1}{2\sqrt{\epsilon_j^2 + |\Delta|^2}}, \tag{B4}$$

where Δ is the energy gap. The BCS mean-field ground-state wave function can be written as

$$|\text{MF}\rangle = \prod_{i=0}^{L-1} (u_i + v_i c_{i\uparrow}^\dagger c_{i\downarrow}^\dagger) |0\rangle_i$$

$$= \prod_{i=0}^{L-1} (u_i |0\rangle_i + v_i |1\rangle_i), \tag{B5}$$

where

$$\frac{v_i}{u_i} = \frac{\sqrt{\epsilon_i^2 + |\Delta|^2} - \epsilon_i}{\Delta^*} \tag{B6}$$

and we also have the normalization condition $|u_i|^2 + |v_i|^2 = 1$. We rewrite u_i and v_i as

$$u_i = \cos\frac{\theta_i}{2} e^{i\varphi_i/2}, \quad v_i = \sin\frac{\theta_i}{2} e^{-i\varphi_i/2}, \tag{B7}$$

which yields

$$\theta_i = 2\arctan\left(\frac{\sqrt{\epsilon_i^2 + |\Delta|^2} - \epsilon_i}{|\Delta|}\right), \tag{B8}$$

$$\varphi_i = -\arg(\Delta) \equiv \varphi. \tag{B9}$$

Thus, to prepare this initial state we simply apply on each qubit the gate $R_y(\theta_i)$ followed by the gate $R_z(\varphi)$. The first box in Fig. 3 shows an example of this preparation with the set of parameters specified in Sec. II C (these correspond to $\Delta \approx 0.46$).

APPENDIX C: EVOLUTION OF ALL OBSERVABLES MITIGATED

In this Appendix, we present the entirety of the results of our runs on the quantum computer, without RC, with standard RC, and with crosstalk RC. These data were used to obtain the average absolute error plotted in Fig. 1, and only a portion of them are presented in the main text. In Figs. 12 and 13 we show the full time evolution for a first set of seven observables, executed respectively on ibm_lagos and ibmq_ehningen. In Fig. 14, we show the same evolution for a different set of seven observables, also obtained from ibm_lagos. RC averages for both the Trotter circuits and the NEC have been performed over 300 samples, with the exception of the crosstalk RC curves at times $t \geqslant 1.8$ where we averaged over 600 NEC to check that the deviations do not come from a lack of RC circuits. One can see that, over different runs, the noise may change and the efficiency of the mitigated protocol varies.

APPENDIX D: NOISE CHANNEL AFTER RANDOMIZED COMPILING

First, we review the detailed computation of a general n-qubit noise channel twirled over the set $\mathcal{T}_p^{\otimes n} = \{\mathbb{I}_2, \sigma_x, \sigma_y, \sigma_z\}^{\otimes n}$. Then, we argue that this method can be applied to a two-qubit noise channel induced by the application of a noisy CNOT gate. The main difference is that the first twirl can only be applied before the CNOT. Finally, we study the crosstalk RC protocol defined in Sec. III and derived the n-qubit channel. We also show that when traced over the neighboring qubits this channel is reduced to a two-qubit Pauli noise channel and when traced over all the qubits except one neighboring qubit of the CNOT it gives a one-qubit depolarizing noise channel.
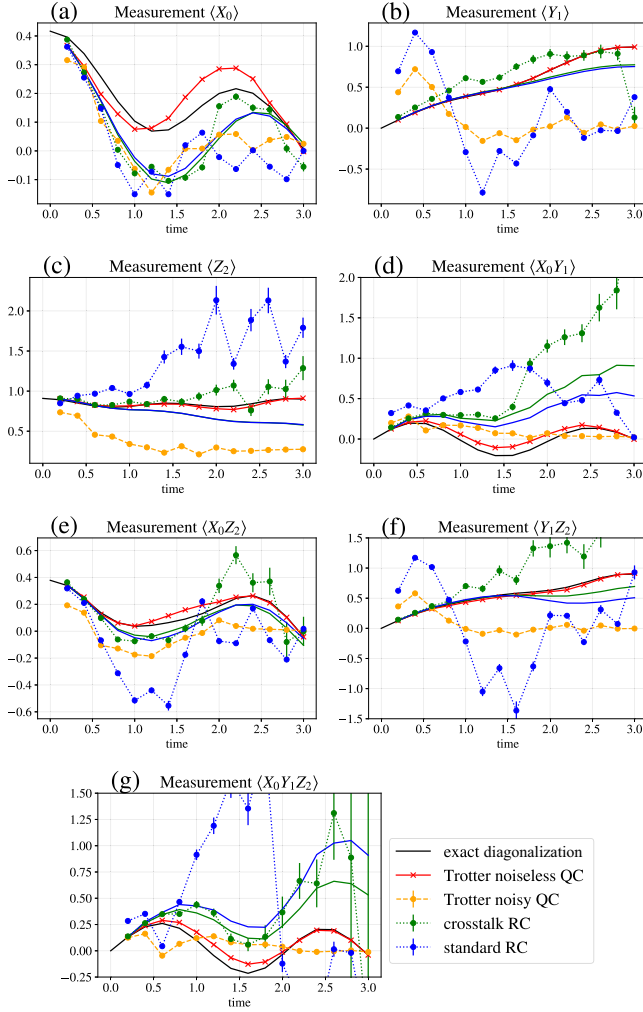
FIG. 12. Evolution of the seven different observables derived from the measurement of `ibm_lagos`'s qubits q_0, q_1, and q_2, respectively, in the directions x, y, and z after application of mitigation protocols (NEC + RC + REC). Blue (green) curves correspond to standard (crosstalk) RC. Dashed curves are the result obtained on a QC and the solid line is the corresponding fit made with the noise model of Eq. (11). Exact evolution (black lines) and noiseless (noisy) Trotter dynamics [red lines (dashed orange curves)] are also plotted for each quantity. Error bars are due to the limited number of shots and the finite sampling over twirl configurations.

FIG. 13. Evolution of the seven different observables derived from the measurement of `ibm_ehningen`'s qubits q_{18}, q_{21}, and q_{23}, respectively, in the directions x, y, and z after application of mitigation protocols (NEC + RC + REC). Blue (green) curves correspond to standard (crosstalk) RC. Dashed curves are the result obtained on a QC. The green solid line is the fit of crosstalk RC made with the noise model of Eq. (11) (fit for the standard RC did not converge). Exact evolution (black lines) and noiseless (noisy) Trotter dynamics [red lines (dashed orange curves)] are also plotted for each quantity. Error bars are due to the limited number of shots and the finite sampling over twirl configurations.

1. Pauli twirling of an *n*-qubit noise channel

Here, we follow the step of computation detailed in Ref. [46]. A general noise channel applied on a density matrix can be written as

$$\mathcal{E}_n^{\text{gen}}(\rho) = \sum_M M\rho M^\dagger \quad \text{such that} \sum_M MM^\dagger = \mathbb{I}. \quad (D1)$$

M is the so-called Kraus operator and we should stress that for a given noise channel this decomposition in terms of Kraus operators is not unique. Each Kraus operator can be decomposed over the Pauli
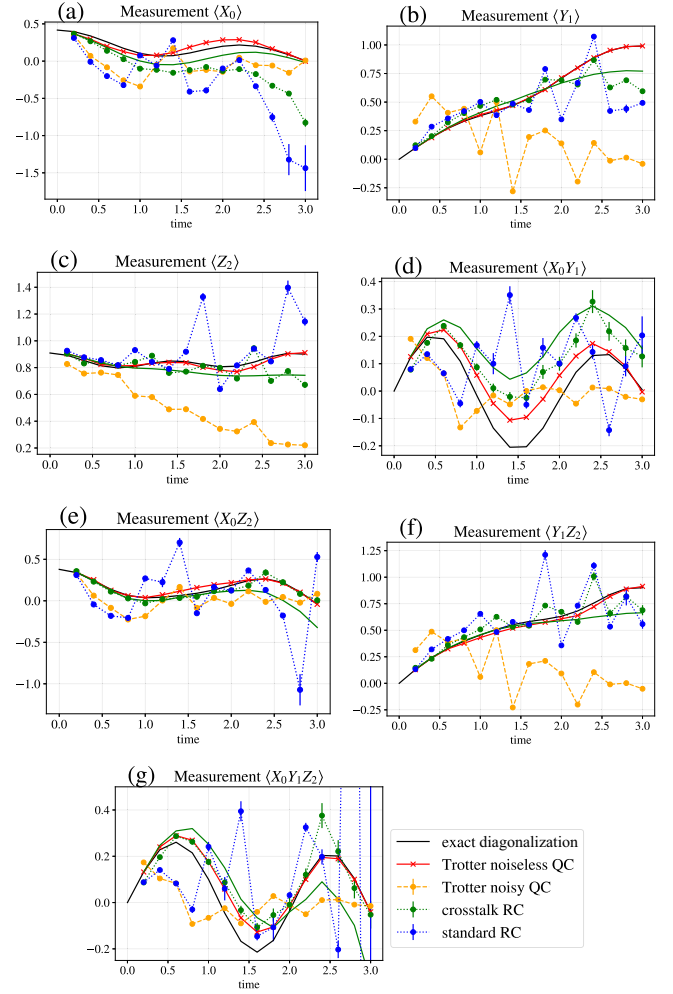
basis,

$$M = \frac{1}{2^n} \sum_{p \in P} \text{Tr}(Mp)p, \quad (D2)$$

where $P = \{\mathbb{I}_2, \sigma_x, \sigma_y, \sigma_z\}^{\otimes n}$. The general noise channel reads

$$\mathcal{E}_n^{\text{gen}}(\rho) = \frac{1}{4^n} \sum_M \sum_{p_1, p_2 \in P} \text{Tr}(Mp_1)\text{Tr}(Mp_2)^* p_1 \rho p_2. \quad (D3)$$

The twirling of the noise channel over $\mathcal{T}_p^{\otimes n} = P$ consists of applying Pauli matrices [*w* in Eq. (D4)] before and after the noise channel and averaging the result over the twirling
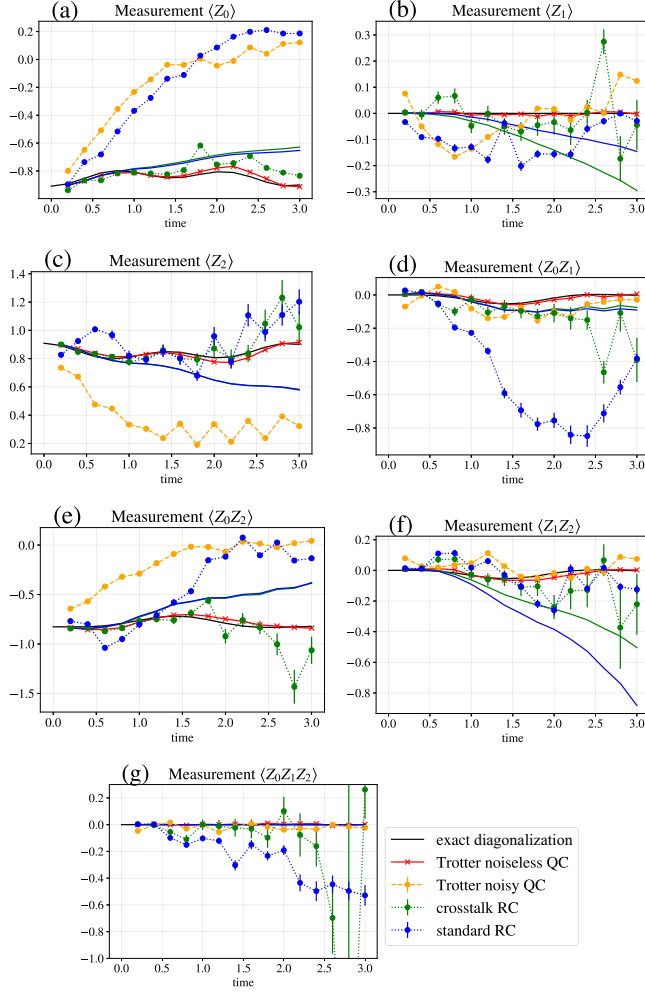
FIG. 14. Evolution of the seven different observables derived from the measurement of `ibm_lagos`'s qubits q_0, q_1, and q_2 in the direction z after application of mitigation protocols (NEC + RC + REC). Blue (green) curves correspond to standard (crosstalk) RC. Dashed curves are the result obtained on a QC and the solid line is the corresponding fit made with the noise model of Eq. (11). Exact evolution (black lines) and noiseless (noisy) Trotter dynamics [red lines (dashed orange curves)] are also plotted for each quantity. Error bars correspond to the uncertainty due to shot noise and the finite sampling over twirl configurations.

set:

$$\mathcal{T}\big[\mathcal{E}_n^{\mathrm{gen}}\big](\rho) = \frac{1}{4^n} \sum_M \sum_{p_1, p_2 \in P} \mathrm{Tr}(Mp_1)\mathrm{Tr}(Mp_2)^*$$

$$\times \frac{1}{4^n} \sum_{w \in \mathcal{T}_p^{\otimes n}} w p_1 w \rho w p_2 w \qquad \text{(D4)}$$

$$= \frac{1}{4^{2n}} \sum_M \sum_{p_1, p_2, w \in P} \mathrm{Tr}(Mp_1)\mathrm{Tr}(Mp_2)^*$$

$$\times p_1 \rho p_2 \zeta(w, p_1)\zeta(w, p_2), \qquad \text{(D5)}$$

where $\zeta(w, p)$ is the sign of the commutation between w and p: $+1$ if both commute, -1 if they anticommute. In particular,

one has the relation

$$\zeta(w, p_1 p_2) = \zeta(w, p_1)\zeta(w, p_2). \qquad \text{(D6)}$$

One inserts Eq. (D6) in Eq. (D5), and splits the sums over p_1 and p_2 in two parts: the first one when $p_1 = p_2$ and the second one when $p_1 \neq p_2$. The first sum evaluates $\zeta(w, p_1^2) = \zeta(w, \mathbb{I}_n) = +1$. For the second sum when $p_1 \neq p_2$, we decompose the Pauli matrices over the different qubits $w = \bigotimes_{a=0}^{n-1} w^a$ and $p = p_1 p_2 = \bigotimes_{a=0}^{n-1} p^a$. The sum over w is rewritten as

$$\sum_{w \in P} \zeta(w, p_1 p_2) = \prod_{a=0}^{n-1} \sum_{w^a \in \mathcal{T}_p} \zeta(w^a, p^a). \qquad \text{(D7)}$$

When $p_1 \neq p_2$, there exists at least one qubit a such that $p^a \neq \mathbb{I}_2$. For that qubit, the terms $w^a = \mathbb{I}_2$ and $w^a = p^a$ evaluate to $\zeta(w^a, p^a) = +1$ while the two other terms give -1. As a consequence, Eq. (D7) vanishes. Only the diagonal term $(p_1 = p_2)$ is nonzero in Eq. (D5) and gives

$$\mathcal{T}\big[\mathcal{E}_n^{\mathrm{gen}}\big](\rho) = \frac{1}{4^n} \sum_{p \in P} \left(\sum_M |\mathrm{Tr}(Mp)|^2 \right) p \rho p. \qquad \text{(D8)}$$

Any noise channel is mapped, after twirling, to a Pauli noise channel with coefficients $\frac{1}{4^n} \sum_M |\mathrm{Tr}(Mp)|^2$ for each Pauli matrix p where the M's are the Kraus operators of the original noise channel.

2. Randomized compiling for CNOT gates

In practice, a noise channel emerges from the application of a noisy gate. Consequently, RC cannot be directly applied to the noise channel; rather, it is performed on the entire noisy gate, modeled by its noiseless counterpart U followed by the corresponding error channel $\mathcal{E}_U$. Equation (D4) is rewritten as

$$\mathcal{T}[\mathcal{E}_U](\rho) = \frac{1}{4^n} \sum_M \sum_{p_1, p_2 \in P} \mathrm{Tr}(Mp_1)\mathrm{Tr}(Mp_2)^*$$

$$\times \frac{1}{4^n} \sum_{w \in P} w p_1 U w' \rho w' U^\dagger p_2 w, \qquad \text{(D9)}$$

where $w' \in P$ and such that $wU = U w'$. In general, for any U it is not guaranteed to find such w'. Following the main text, here the noisy gate U_{noisy} will be the two-qubit CNOT gate and for all w there exists such w'. Let us denote CNOT_{ij} the CNOT gate acting on qubits i and j and $w = w^i \otimes w^j \bigotimes_{a \neq i, j} w^a$. For any $(w^i, w^j) \in \mathcal{T}_p^{\otimes 2}$, Table I of Ref. [41] gives the corresponding (w'^i, w'^j). For $a \neq i, j$, the action of the perfect CNOT is trivial and then $w'^a = w^a$. The rest of the computation follows the above derivation.

This twirling of noisy gates is called randomized compiling because, as it is explained in the main text, the space of twirling gates grows exponentially with the number of CNOT gates in the circuit and one usually samples this space with a finite number of different twirl configurations by choosing randomly gates in the twirling set.

3. Rotation twirling

In this section, we derive the noise channel obtained after Pauli twirling and apply the rotation twirling on neighboring qubits as described in Sec. III B (see Fig. 5). First, we start with a simpler case: we consider a one-qubit noise channel and apply this double RC procedure. The noise channel of Eq. (D4) is modified as follows:

$$
\mathcal{T}^*[\mathcal{E}_1^{\text{gen}}](\rho) = \frac{1}{16} \sum_M \sum_{p_1, p_2, w \in \mathcal{T}_p} \text{Tr}(Mp_1)\text{Tr}(Mp_2)^*
$$
$$
\times \frac{1}{3} \sum_{R \in \mathcal{T}_R} wR^\dagger p_1 Rw \rho wR^\dagger p_2 Rw, \quad \text{(D10)}
$$

where $\mathcal{T}_R = \{R_x(\pi/2), R_y(\pi/2), R_z(\pi/2)\}$ as defined in Sec. III B. For each rotation R, the Pauli matrices p_1 and p_2 are mapped to the Pauli matrices $p_1' = R^\dagger p_1 R$ and $p_2' = R^\dagger p_2 R$ and we rewrite Eq. (D10):

$$
\mathcal{T}^*[\mathcal{E}_1^{\text{gen}}](\rho) = \frac{1}{48} \sum_{R \in \mathcal{T}_R} \sum_M \sum_{p_1', p_2', w \in \mathcal{T}_p} \text{Tr}(MRp_1'R^\dagger)
$$
$$
\times \text{Tr}(MRp_2'R^\dagger)^* w p_1' w \rho w p_2' w. \quad \text{(D11)}
$$

The same derivation as in the above section can be done for p_1' and p_2' and therefore only the diagonal term $p_1' = p_2'$ does not vanish. We get

$$
\mathcal{T}^*[\mathcal{E}_1^{\text{gen}}](\rho) = \frac{1}{12} \sum_{p \in P} \sum_M \sum_R |\text{Tr}(MRpR^\dagger)|^2 p\rho'p. \quad \text{(D12)}
$$

The effect of the rotation matrices when the identity matrix is applied ($p = \mathbb{I}_2$) does not change the matrix:

$$
RpR^\dagger = p = \mathbb{I}_2 \quad \forall R \in \mathcal{T}_R. \quad \text{(D13)}
$$

When $p = \sigma^x$, σ^y, or σ^z, rotation matrices $R \in \{R_x(\pi/2), R_y(\pi/2), R_z(\pi/2)\}$ equally mix the different directions:

$$
R_k(\pi/2)\sigma^l R_k(-\pi/2) = \sum_{m=x,y,z} \epsilon_{klm}\sigma^m + \delta_{kl}\sigma^l, \quad \text{(D14)}
$$

where $k, l = x, y, z$ and ϵ_{klm} is the Levi-Cività tensor. Equation (D12) reduces to

$$
\mathcal{T}^*[\mathcal{E}_1^{\text{gen}}](\rho) = \frac{1}{4} \sum_M |\text{Tr}(M)|^2 \rho
$$
$$
+ \frac{1}{3}\left(\sum_{i=x,y,z} \frac{1}{4} \sum_M |\text{Tr}(M\sigma^i)|^2\right) \sum_{j=x,y,z} \sigma^j \rho \sigma^j. \quad \text{(D15)}
$$

Equation (D15) is a one-qubit depolarizing channel. Adding a rotation twirling set equally mixes the different coefficients of the Pauli noise channel to obtain a depolarizing channel.

Let us now consider the real case where a CNOT is applied and the rotation twirling is performed over the neighboring

qubits. The twirled channel reads

$$
\mathcal{T}^*[\mathcal{E}_{\text{CNOT}}](\rho) = \frac{1}{4^{2n}} \sum_M \sum_{p_1, p_2, w \in P} \text{Tr}(Mp_1)\text{Tr}(Mp_2)^*
$$
$$
\times \frac{1}{3^{n-2}} \sum_{R \in \mathcal{T}_R^{\otimes n-2}} wR^\dagger p_1 Rw \rho' wR^\dagger p_2 Rw, \quad \text{(D16)}
$$

where $\rho' = \text{CNOT} \cdot \rho \cdot \text{CNOT}$. Here n denotes the number of qubits on which the noise channel applies. When considering crosstalk, n is equal to 2 (the qubits on which the CNOT is applied) plus the number of direct neighbors. The same trick can be applied as in Eq. (D11), and after a similar derivation we end up with

$$
\mathcal{T}^*[\mathcal{E}_{\text{CNOT}}](\rho) = \frac{1}{4^n} \sum_{p \in P} \sum_M \frac{1}{3^{n-2}} \sum_R |\text{Tr}(MRpR^\dagger)|^2 p\rho'p. \quad \text{(D17)}
$$

The probability to apply to the density matrix the Pauli gate p is $\frac{1}{4^n} \sum_M \frac{1}{3^{n-2}} \sum_R |\text{Tr}(MRpR^\dagger)|^2$.

Let us consider for the rest of the derivation that 1 and 2 denote the index of the control and target qubits of the CNOT gate; hence, there is no rotation matrix applied on these qubits. We decompose the whole Pauli matrix and rotation matrix on each individual qubit: $P = \bigotimes_{a=0}^{n-1} p^a$ and $R = \bigotimes_{a \neq 1,2} r^a$. Using Eqs. (D13) and (D14), the coefficients in Eq. (D17) only differ depending when $p^b = \mathbb{I}_2$ or $p^b \neq \mathbb{I}_2$, where b is the index of a neighboring qubit. Therefore, the number of free parameters decreases thanks to this additional twirling. It goes from $4^n - 1$ for a Pauli noise channel on n qubits to $16 \times 2^{n-2} - 1 = 2^{n+2} - 1$.

One can also trace over all the qubits except one neighboring qubit b. The probability to apply the gate p^b is

$$
\frac{1}{4^n \times 3} \prod_{a \neq b} \sum_{p^a \in \mathcal{T}_p} \sum_M \sum_{r^b \in \mathcal{T}_R} \left|\text{Tr}\left(M \bigotimes_{c \neq b} p^c \otimes r^b p^b (r^b)^\dagger\right)\right|^2. \quad \text{(D18)}
$$

It reduces, when $p^b = \mathbb{I}_2$, to

$$
1 - q_b = \frac{1}{4^n} \prod_{a \neq b} \sum_{p^a \in \mathcal{T}_p} \sum_M \left|\text{Tr}\left(M \bigotimes_{c \neq b} p^c \otimes \mathbb{I}_2\right)\right|^2. \quad \text{(D19)}
$$

When $p^b = \sigma^x$, σ^y, or σ^z, the probabilities are the same:

$$
\frac{q_b}{3} = \frac{1}{4^n \times 3} \prod_{a \neq b} \sum_{p^a \in \mathcal{T}_p} \sum_M \sum_{p^b \in \{\sigma^x, \sigma^y, \sigma^z\}} \left|\text{Tr}\left(M \bigotimes_{c=0}^{n-1} p^c\right)\right|^2, \quad \text{(D20)}
$$

where q_b relates to the depolarizing parameter $\lambda_b = \frac{4q_b}{3}$.

Therefore, it mimics a depolarizing channel when looking at only one neighboring qubit. When tracing over all the neighboring qubits and looking at the noise channel on qubits 1 and 2, one has still a Pauli noise channel where the probability to apply the gate $p^1 \otimes p^2$ is

$$
\frac{1}{4^n} \prod_{a \neq 1,2} \sum_{p^a \in \mathcal{T}_p} \sum_M \left|\text{Tr}\left(M \bigotimes_{c=0}^{n-1} p^c\right)\right|^2. \quad \text{(D21)}
$$

The general form of the noise channel obtained after the double RC protocol is

$$\mathcal{E}_n^{\text{crosstalk RC}}(\rho) = \sum_{i,j=0}^{3} \sum_{\vec{s}} \frac{p_{i,j,\vec{s}}}{3} \sum_{\vec{a}} P^{i,j,\vec{s},\vec{a}} \rho P^{i,j,\vec{s},\vec{a}}, \quad \text{(D22)}$$

$$P^{i,j,\vec{s},\vec{a}} = \sigma_1^i \otimes \sigma_2^j \bigotimes_{k=3}^{n} \left(\sigma_k^{a_k}\right)^{s_k}, \quad \text{(D23)}$$

$$\text{with } \vec{s} \in \{0,1\}^{n-2}, \quad \vec{a} \in \{1,2,3\}^{n-2}, \quad \text{(D24)}$$

where $\vec{s} = (s_3, \ldots, s_n)$ is a binary vector indicating if an error occurred ($s_k = 1$) on the neighboring qubit of index k and $\vec{a}$ indicates if it was an X, Y, or Z error ($a_k = 1$, 2, or 3).

APPENDIX E: STATISTICAL FLUCTUATIONS DUE TO FINITE SAMPLING

In this Appendix, we estimate the uncertainty due to the shot noise and the finite sampling of the twirl configurations. The quantity we are interested to estimate is

$$O^{\text{expt}} = \frac{1}{N_t} \sum_{r=1}^{N_t} O_r^{\text{expt}}, \quad \text{(E1)}$$

where N_t is the number of twirl configurations (in the main text $N_t = 300$; in Appendix C for the NEC of the crosstalk RC and $t \geqslant 1.8$, $N_t = 600$). O_r^{expt} is the experimental estimation of the expectation value of an observable O for the twirled configuration circuit C_r. It is estimated by running N_s times the circuit C_r:

$$O_r^{\text{expt}} = \frac{1}{N_s} \sum_{m=1}^{N_s} O_{rm}^{\text{expt}}, \quad \text{(E2)}$$

where N_s is the number of shots ($N_s = 32\,000$ in the main text). O_{rm}^{expt} is a bimodal random variable which is equal to $+1$ with probability $p(r)$ and -1 with probability $1 - p(r)$. The total expression is

$$O^{\text{expt}} = \frac{1}{N_t N_s} \sum_{r=1}^{N_t} \sum_{m=1}^{N_s} O_{rm}^{\text{expt}}. \quad \text{(E3)}$$

There are two different averages. The first one is used to estimate the expectation value of an observable O_r corresponding to the circuit C_r, denoted as $\langle \cdots \rangle$:

$$\langle O_r^{\text{expt}} \rangle = \frac{1}{N_s} \sum_{m=1}^{N_s} \langle O_{rm}^{\text{expt}} \rangle = 2p(r) - 1 \quad \text{(E4)}$$

$$= \lim_{N_s \to \infty} O_r^{\text{expt}}. \quad \text{(E5)}$$

The second average is over the different twirl configurations and is denoted $\overline{\cdots}$. The probability that the expectation value of an observable corresponding to a certain circuit C_r outputs $O_r^{\text{expt}} = o \in [-1, 1]$ is given by the unknown probability distribution $f(o)$:

$$\overline{O^{\text{expt}}} = \frac{1}{N_t} \sum_{r=1}^{N_t} \overline{O_r^{\text{expt}}} = \int_{-1}^{1} o \times f(o)\,\mathrm{d}o \quad \text{(E6)}$$

$$= \lim_{N_t \to \infty} O^{\text{expt}}. \quad \text{(E7)}$$

To estimate the deviation from both averages due to the finite sampling it is useful to compute

$$\langle (O_r^{\text{expt}})^2 \rangle - \langle O_r^{\text{expt}} \rangle^2 = \frac{1}{N_s} \text{Var}_{\langle \cdots \rangle}(O_r^{\text{expt}}) \quad \text{(E8)}$$

$$= \frac{1 - \langle O_r^{\text{expt}} \rangle^2}{N_s}, \quad \text{(E9)}$$

$$\overline{(O^{\text{expt}})^2} - \overline{O^{\text{expt}}}^2 = \frac{1}{N_t} \text{Var}_{\overline{\cdots}}(\langle O^{\text{expt}} \rangle), \quad \text{(E10)}$$

where $\text{Var}_{\langle \cdots \rangle}$ denotes the variance corresponding to the average over the different shots and $\text{Var}_{\overline{\cdots}}$ denotes the variance corresponding to the average over the different twirled configurations. The total deviation is the sum of both quantities where we average the deviation due to shot noise over the different twirl configurations:

$$\overline{\langle (O^{\text{expt}})^2 \rangle} - \overline{\langle O^{\text{expt}} \rangle}^2 = \overline{\langle (O^{\text{expt}})^2 \rangle - \langle O^{\text{expt}} \rangle^2} + \overline{\langle O^{\text{expt}} \rangle^2} - \overline{\langle O^{\text{expt}} \rangle}^2 \quad \text{(E11)}$$

$$= \frac{1}{N_s N_t} \overline{\text{Var}_{\langle \cdots \rangle}(O^{\text{expt}})} + \frac{1}{N_t} \text{Var}_{\overline{\cdots}}(\langle O^{\text{expt}} \rangle) \quad \text{(E12)}$$

$$\simeq \frac{1}{N_s N_t^2} \sum_{r=1}^{N_t} 1 - \langle O_r^{\text{expt}} \rangle^2 + \frac{1}{N_t} \text{Var}_{\overline{\cdots}}(\langle O^{\text{expt}} \rangle). \quad \text{(E13)}$$

To obtain the last equation, we estimate the average over the twirl configurations empirically ($\overline{\cdots} \simeq \frac{1}{N_t} \sum_{r=1}^{N_t} \cdots$). $\langle O_r^{\text{expt}} \rangle \simeq O_r^{\text{expt}}$ and $\text{Var}_{\overline{\cdots}}$ are also estimated empirically from the data. Error bars in the plots correspond to the square root of Eq. (E13). For the mitigated results, we combine this result with Eq. (15) to derive its uncertainty.

APPENDIX F: MITIGATION FOR GLOBAL AND LOCAL DEPOLARIZING CHANNEL

In this Appendix, we recall the result obtained for the noise estimation circuit in Ref. [41] in the case of a global and local depolarizing noise channel. First, to understand how the noise estimation circuit protocol works, it is interesting to express the noisy expectation value in terms of the noiseless one.

1. Relation between noisy and noiseless outcome

In the simple case of a global depolarizing channel, this can be done exactly. For a local depolarizing channel, the relation can be expanded in terms of number of errors occurring in the circuit and one keeps only the first-order term. After considering these two noisy channels, the result generalizes to the case of Eq. (11) and yields Eq. (12), which includes these two cases plus a noise on the neighboring qubit.

A global depolarizing channel on each CNOT reads

$$\mathcal{E}_n^{\text{dep}}(\rho) = (1 - \lambda_{\text{glob}})\rho + \lambda_{\text{glob}} \frac{\mathbb{I}_{2^n}}{2^n}, \quad \text{(F1)}$$

where n is the number of qubits in the circuit and $0 \leqslant \lambda \leqslant \frac{4^n}{4^n - 1}$. If an error occurs in the circuit, it outputs the maximally mixed state. The noisy result of a traceless observable is therefore expressed as [41]

$$\langle O \rangle_{\text{noisy}}^{\text{glob}} = (1 - \lambda_{\text{glob}})^{n_{\text{CNOT}}} \langle O \rangle_{\text{noiseless}}, \quad \text{(F2)}$$

where n_{CNOT} is the number of CNOT gates in the circuit.

In the case of a depolarizing channel acting only on the qubits i, j where the CNOT is applied (called a local depolarizing channel), the error channel is written as

$$\mathcal{E}_2^{\text{dep}}(\rho) = (1 - \lambda_{\text{CNOT}})\rho + \lambda_{\text{CNOT}}\frac{\mathbb{I}_4}{4} \otimes \text{Tr}_{ij}(\rho), \quad \text{(F3)}$$

where Tr_{ij} is the partial trace over the active qubits i, j. An exact expression between the noisy result and the noiseless cannot be derived but one can expand it in terms of the number of errors:

$$\langle O \rangle_{\text{noisy}}^{\text{local}} = \sum_{i=0}^{\infty}(1 - \lambda_{\text{CNOT}})^{n_{\text{CNOT}}-i}\lambda_{\text{CNOT}}^i \langle O \rangle_i$$

$$= (1 - \lambda_{\text{CNOT}})^{n_{\text{CNOT}}} \sum_{i=0}^{\infty}\left(\frac{\lambda_{\text{CNOT}}}{1 - \lambda_{\text{CNOT}}}\right)^i \langle O \rangle_i$$

$$\simeq \langle O \rangle_0 \left[1 + \lambda_{\text{CNOT}}\left(\frac{\langle O \rangle_1}{\langle O \rangle_0} - n_{\text{CNOT}}\right)\right]$$

$$+ \mathcal{O}(\lambda_{\text{CNOT}}^2), \quad \text{(F4)}$$

where $\langle O \rangle_i$ is the outcome such that i errors occurred in the circuit; it is a sum of expectation values of all circuit where i CNOT gates output the maximally mixed state on the active qubits. In the last line we expand the expression up to one error occurring in the circuit.

2. Noisy expectation value of the noise estimation circuit

In the case of a global depolarizing noise, applying Eq. (F2) to the NEC leads to

$$\langle E \rangle_{\text{noisy}}^{\text{glob}} = (1 - \lambda_{\text{glob}})^{n_{\text{CNOT}}}, \quad \text{(F5)}$$

where $\langle E \rangle_{\text{noisy}}^{\text{glob}}$ is the outcome of the NEC (the noiseless outcome being 1). In this simple case, the NEC exactly reproduces the noise of the Trotter circuit. Dividing the Trotter circuits by the NEC yields the noiseless result:

$$\frac{\langle O \rangle_{\text{noisy}}^{\text{glob}}}{\langle E \rangle_{\text{noisy}}^{\text{glob}}} = \langle O \rangle_{\text{noiseless}}^{\text{glob}}. \quad \text{(F6)}$$

For the local noise, we apply Eq. (F4) in the case of NEC and divide both results, which yields

$$\frac{\langle O \rangle_{\text{noisy}}^{\text{local}}}{\langle E \rangle_{\text{noisy}}^{\text{local}}} \simeq \langle O \rangle_0 \left[1 + \lambda_{\text{CNOT}}\left(\frac{\langle O \rangle_1}{\langle O \rangle_0} - \langle E \rangle_1\right)\right]$$

$$+ \mathcal{O}(\lambda_{\text{CNOT}}^2). \quad \text{(F7)}$$

We show in the next Appendix that the term in front of λ_{CNOT} is small.

APPENDIX G: FIRST-ORDER TROTTER EXPANSION OF AN OBSERVABLE UNDER A LOCAL DEPOLARIZING NOISE

In this Appendix, we show that, assuming a local depolarizing noise on each CNOT gate, the noise of the NEC reproduces part of the noise of the Trotter circuit up to the first Trotter order.

Every linear quantum circuit can be written as a sequence of a two-body unitary gate [see Fig. 15(a) for an example].

We start by considering only one qubit measured (qubit m) along the direction $d = x, y, z$ [in Fig. 15(a), $m = 2$]. The density matrix just before the measurement is denoted as ρ_0. Thus, one computes the quantity $\text{Tr}(\rho_0 \sigma_m^d)$. Each U_{ij}^k unitary gate contains an arbitrary number of CNOT gates r_{ij}^k. Here i and j are the qubit indices on which the unitary gate is applied and k corresponds to the $2k - 1$ layer of gates if i is odd and the $2k$ layer if i is even. We suppose a local depolarizing noise after each CNOT gate with parameter ε:

$$\Lambda_{ij}(\rho) = (1 - \varepsilon)\rho + \epsilon\frac{\mathbb{I}_{ij}}{4} \otimes \text{Tr}_{ij}(\rho),$$

where Tr_{ij} denotes the partial trace over qubits i and j. No matter on which CNOT gate the noise occurs inside a U_{ij}^k unitary gate, the noisy outcome will be $\frac{\mathbb{I}_{ij}}{4} \otimes \text{Tr}_{ij}(\rho)$. Then the noisy channel after a U_{ij}^k unitary gate is

$$\text{E}_{ij}^k(\rho) = (1 - \varepsilon)^{r_{ij}^k}\rho + (1 - (1 - \epsilon)^{r_{ij}^k})\frac{\mathbb{I}_{ij}}{4} \otimes \text{Tr}_{ij}(\rho).$$

Suppose we have noise on the last layer of unitary gates (in the above example on U_{34}^3 and U_{12}^3). If the noise is on a unitary gate which is not on the measured qubit, this gate is not in the light cone of the measured qubit and therefore its noise cannot affect the result. If the noise is on a unitary gate which directly acts on the measured qubit, the resulting output vanishes because the Pauli matrix σ_m^d is traceless, $\text{Tr}((\frac{\mathbb{I}_{ij}}{4} \otimes \text{Tr}_{ij}(\rho))\sigma_m^d) = 0$, with $i = m$ or $j = m$. One can then write the noisy result of the circuit with respect to the noiseless one. For the example shown in Fig. 15(a), it gives

$$\text{Tr}(\rho_0 \sigma_2^d) = (1 - \varepsilon)^{r_{12}^3}\text{Tr}(\rho_0^p \sigma_2^d), \quad \text{(G1)}$$

where ρ_0^p denotes the "perfect" result without noise; i.e., the unitary gates of the last layer have indeed been applied perfectly, $\rho_0^p = U_{12}^3 U_{34}^3 \rho_1 (U_{34}^3)^{\dagger}(U_{12}^3)^{\dagger}$.

Let us denote ρ_1 as the density matrix just before the application of the last layer of unitary gates. All unitary gates which do not act on the measured qubit $m = 2$ commute with the Pauli matrix σ_2^d; then

$$\text{Tr}(\rho_0 \sigma_2^d) = (1 - \varepsilon)^{r_{12}^3}\text{Tr}(\rho_1 (U_{12}^3)^{\dagger}\sigma_2^d U_{12}^3). \quad \text{(G2)}$$

In the following, we will make the assumption that all unitary gates are parametrized with a small time step Δt as in the Trotter algorithm such that

$$U_{ij}^k = e^{-i\Delta t \sum_{a,b=0}^3 \alpha_{ab}^{ijk}\sigma_i^a \otimes \sigma_j^b}, \quad \text{(G3)}$$

where σ_i^0 denotes the identity matrix on qubit i and σ_i^a are the Pauli matrices where $a = 1, 2, 3$ are identified respectively to the directions x, y, and z.

For computation at first order in Δt the quantity $(U_{12}^3)^{\dagger}\sigma_2^d U_{12}^3$ yields three terms:

$$(U_{12}^3)^{\dagger}\sigma_2^d U_{12}^3 \simeq \sigma_2^d - 2\Delta t(\mathbb{I}_1 \otimes A_2^{213} + B_2^{213}), \quad \text{(G4)}$$

where $A_2^{213} = \sum_{b,c=1}^3 \alpha_{0b}^{123}\epsilon_{bdc}\sigma_2^c$ and $B_2^{213} = \sum_{a,b,c=1}^3 \alpha_{ab}^{123}\epsilon_{bdc}\sigma_1^a \otimes \sigma_2^b$ and ϵ_{bdc} is the Levi-Città tensor.
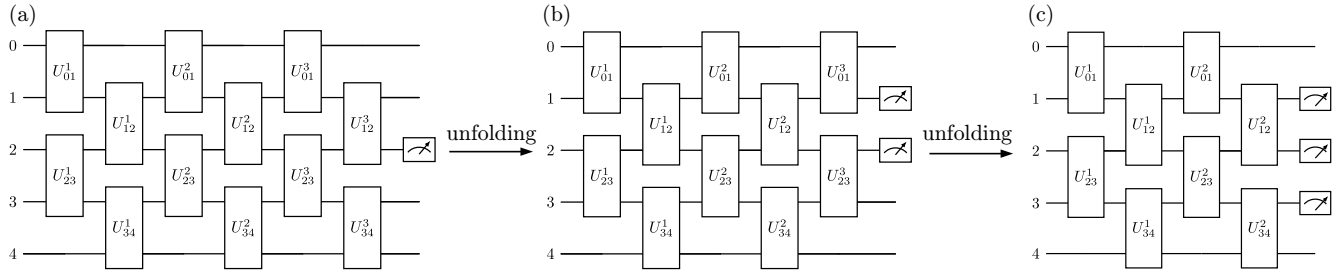
FIG. 15. Schematic representation of the expansion performed. We start from a general circuit (a) and expand the different terms with respect to the last layer of gates and keep terms only up to first order in the Trotter time step Δt. (a) Example of a general circuit with a linear connection between qubits. We have five qubits and a six-layer depth. (b) Circuit representation of the terms $\mathrm{Tr}(\rho_1 \sigma_2^d)$, $\mathrm{Tr}(\rho_1 A_2^{213})$, and $\mathrm{Tr}(\rho_1 B_2^{213})$ [see Eq. (G4)] as a circuit. For the two first terms, measurement is performed solely on qubit 2 while the last term measures both qubits 1 and 2. (c) Circuit representation of the terms in Eq. (G10). Measurement is performed only on qubit 2 for the three first terms of this equation. The fourth term measures qubits 1 and 2 and the last term qubits 2 and 3.

Substituting Eq. (G4) in Eq. (G2), one obtains

$$\mathrm{Tr}\big(\rho_0 \sigma_2^d\big) \simeq (1-\varepsilon)^{r_{12}^3}\big[\mathrm{Tr}\big(\rho_1 \sigma_2^d\big) - 2\Delta t\big(\mathrm{Tr}\big(\rho_1 A_2^{123}\big) + \mathrm{Tr}\big(\rho_1^p B_2^{123}\big)\big)\big]. \tag{G5}$$

The two first terms on the right-hand side of Eq. (G5), i.e., $\mathrm{Tr}(\rho_1 \sigma_2^d)$ and $\mathrm{Tr}(\rho_1 A_2^{213})$, correspond to the circuit when only qubit 2 is measured. The last term of Eq. (G5) [$\mathrm{Tr}(\rho_1^p B_2^{213})$] corresponds to the case where qubits 1 and 2 are measured. A graphical representation of these terms as a quantum circuit is given in Fig. 15(b).

Then, such as in Eq. (G1), when considering the last layer of gates only the noise of the gates which acts directly on the qubits measured will matter. This leads us to

$$\mathrm{Tr}\big(\rho_0 \sigma_2^d\big) \simeq (1-\varepsilon)^{r_{12}^3 + r_{23}^3}\big[\mathrm{Tr}\big(\rho_1^p \sigma_2^d\big) - 2\Delta t\big(\mathrm{Tr}\big(\rho_1^p A_2^{213}\big) + (1-\varepsilon)^{r_{01}^3}\mathrm{Tr}\big(\rho_1^p B_2^{213}\big)\big)\big]. \tag{G6}$$

Therefore, one can factorize the noise coming from the gate U_{23}^3 because it acts on qubit 2 which is a common measured

qubit for all the terms of Eq. (G5). For the last term there is one extra noise term coming from the CNOT acting on qubit 1 (gate U_{01}^3).

One continues the expansion of the expression until we reach the first layer of gates. For the next step we get $\rho_1^p = U_{01}^3 U_{23}^3 \rho_2 (U_{23}^3)^\dagger (U_{01}^3)^\dagger$. One keeps only the first order in Δt. Therefore, the evolution of the terms A_2^{123} and B_2^{123} remains trivial for the rest of the computation (i.e., identity operator):

$$\big(U_{23}^3\big)^\dagger \big(U_{01}^3\big)^\dagger A_2^{123} U_{01}^3 U_{23}^3 \simeq A_2^{213}, \tag{G7}$$

$$\big(U_{23}^3\big)^\dagger \big(U_{01}^3\big)^\dagger B_2^{123} U_{01}^3 U_{23}^3 \simeq B_2^{213}. \tag{G8}$$

Only the main term proportional to Δt^0 can produce a nontrivial term. For the next layers of gate it couples to qubit 3 thanks to the gate U_{23}^3 [see Fig. 15(c)]. We obtain a similar equation as in Eq. (G4):

$$\big(U_{23}^3\big)^\dagger \big(U_{01}^3\big)^\dagger \sigma_2^d U_{01}^3 U_{23}^3 = \sigma_2^d - 2\Delta t\big(A_2^{233} + B_2^{233}\big). \tag{G9}$$

Figure 15(c) represents the different term of this expansion.

After application of the noise of the unitary U_{12}^2 and U_{34}^2 we obtain

$$\mathrm{Tr}\big(\rho_0 \sigma_2^d\big) \simeq (1-\varepsilon)^{r_{12}^3 + r_{23}^3 + r_{12}^2}\big[\mathrm{Tr}\big(\rho_2^p \sigma_2^d\big) - 2\Delta t\big(\mathrm{Tr}\big(\rho_2^p A_2^{213}\big) + \mathrm{Tr}\big(\rho_2^p A_2^{233}\big) + (1-\varepsilon)^{r_{01}^3}\mathrm{Tr}\big(\rho_2^p B_2^{213}\big) + (1-\varepsilon)^{r_{34}^2}\mathrm{Tr}\big(\rho_2^p B_2^{233}\big)\big)\big], \tag{G10}$$

where the general definitions of $A_k^{mm\pm1j}$ and $B_k^{mm\pm1j}$ are given Eqs. (G12) and (G13).

One can then generalize the above formula for a circuit of $2L$ layers and measured qubit m:

$$\mathrm{Tr}(\rho_0 \sigma_m^d) = \prod_{j=1}^{L}(1-\epsilon)^{r_{m-1m}^j}(1-\epsilon)^{r_{mm+1}^j}\bigg[\mathrm{Tr}\big(\rho_{\mathrm{ini}}\sigma_m^d\big) - 2\Delta t\sum_{j=1}^{L}\mathrm{Tr}\big(\rho_{\mathrm{ini}}A_m^{mm-1j}\big) + \mathrm{Tr}\big(\rho_{\mathrm{ini}}A_m^{mm+1j}\big)$$

$$- 2\Delta t\sum_{j=1}^{L}(1-\varepsilon)^{\sum_{k=1}^{j}r_{m-2m-1}^k}\mathrm{Tr}\big(\rho_{\mathrm{ini}}B_m^{mm-1j}\big) + (1-\varepsilon)^{\sum_{k=1}^{j}r_{m+1m+2}^k}\mathrm{Tr}\big(\rho_{\mathrm{ini}}B_m^{mm+1j}\big)\bigg] + \mathcal{O}(\Delta t^2), \tag{G11}$$

where ρ_{ini} is the initial density matrix before the first layer of unitary gates and

$$A_k^{mm\pm1j} = \sum_{b,c=1}^{3}\alpha_{0b}^{mm\pm1j}\epsilon_{bdc}\sigma_k^c, \tag{G12}$$

$$B_k^{mm\pm1j} = \sum_{a,b,c=1}^{3}\alpha_{ab}^{mm\pm1j}\epsilon_{bdc}\sigma_{m\pm1}^a \otimes \sigma_k^b. \tag{G13}$$

To implement interaction between two non-neighboring qubits, one needs to use SWAP gates which are a succession of three CNOT gates. These SWAP gates need to be taken into account at order zero because they do not depend on any small parameter. This modifies the above formula: with SWAP gates the measured qubit does not remain in the same physical qubit; we need to follow it during the course of the circuit. In the formula of Eq. (G11), the measured qubit m becomes a function of the layers, $m(j)$, which indicate the location of the measured qubit at layer j.

Equation (G11) can be applied to estimate the noise measured by the noise estimation circuit. In this case, the general U_{bc}^a gates reduce to the identity operator up to some SWAP operations and $\Delta t = 0$. Therefore, these circuits estimate only the noise term in factor of the whole expression $\prod_{j=1}^{L}(1 - \epsilon)^{r_{m-1m}^j}(1 - \epsilon)^{r_{mm+1}^j}$, i.e., the noise coming from the CNOT gate acting on the measured qubit. Then, it shows that if one divides the noise of the Trotter circuit by the noise of the estimation circuit, one can remove all the noise of the main term and a part of the first-order Trotter term.

The number of higher-order terms in Δt scale exponentially with the total time evolution and therefore nothing guarantees that these terms will remain small. Their noise will deviate strongly as time goes on from the noise estimation circuit [see Fig. 8 for a simulation of the noise estimation method using the error channel of Eq. (11) (quasilocal depolarizing noise)]. Therefore, the noise estimation circuit is a method which delays the effect of the noise and allows simulation to longer evolution time.

Equation (G11) can be generalized to multiple measured qubits. For each term containing the index m for the measured qubit, one should add a product running over the different measured qubits. The formula for the noise estimation circuit also gets this modification and is still valid in that case.

This result generalizes to an error channel of the form of Eq. (11). As long as a depolarizing channel is applied on qubits whenever an error occurs, the main term of the Trotter expansion will be reproduced by the estimation circuit.

[1] J. Preskill, Quantum computing and the entanglement frontier, arXiv:1203.5813.

[2] J. Preskill, Quantum computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

[3] E. Knill, R. Laflamme, and W. H. Zurek, Resilient quantum computation, Science **279**, 342 (1998).

[4] D. A. Lidar and T. A. Brun, *Quantum Error Correction* (Cambridge University Press, Cambridge, U.K., 2013).

[5] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, U.K., 2012).

[6] IBM Quantum, https://quantum-computing.ibm.com/.

[7] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O'Brien, Quantum error mitigation, Rev. Mod. Phys. **95**, 045005 (2023).

[8] Suguru Endo, Z. Cai, S. C. Benjamin, and X. Yuan, Hybrid quantum-classical algorithms and quantum error mitigation J. Phys. Soc. Jpn. **90**, 032001 (2021).

[9] P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio, Error mitigation with Clifford quantum-circuit data, Quantum **5**, 592 (2021).

[10] S. McArdle, X. Yuan, and S. C. Benjamin, Error-mitigated digital quantum simulation Phys. Rev. Lett. **122**, 180501 (2019).

[11] K. Temme, S. Bravyi, and J. M. Gambetta, Error mitigation for short-depth quantum circuits, Phys. Rev. Lett. **119**, 180509 (2017).

[12] Y. Li and S. C. Benjamin, Efficient variational quantum simulator incorporating active error minimization, Phys. Rev. X **7**, 021050 (2017).

[13] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, and A. Aspuru-Guzik, Noisy intermediate-scale quantum algorithms, Rev. Mod. Phys. **94**, 015004 (2022).

[14] F. Leymann and J. Barzen, The bitter truth about gate-based quantum algorithms in the NISQ era, Quantum Sci. Technol. **5**, 044007 (2020).

[15] Y. Zhou, E. M. Stoudenmire, and X. Waintal, What limits the simulation of quantum computers? Phys. Rev. X **10**, 041038 (2020).

[16] T. Ayral, T. Louvet, Y. Zhou, C. Lambert, E. M. Stoudenmire, and X. Waintal, Density-matrix renormalization group algorithm for simulating quantum circuits with a finite fidelity, PRX Quantum **4**, 020304 (2023).

[17] A. Kandala, K. X. Wei, S. Srinivasan, E. Magesan, S. Carnevale, G. A. Keefe, D. Klaus, O. Dial, and D. C. McKay, Demonstration of a high-fidelity CNOT gate for fixed-frequency transmons with engineered ZZ suppression, Phys. Rev. Lett. **127**, 130501 (2021).

[18] E. Magesan, J. M. Gambetta, and J. Emerson, Scalable and robust randomized benchmarking of quantum processes, Phys. Rev. Lett. **106**, 180504 (2011).

[19] E. Magesan, J. M. Gambetta, and J. Emerson, Characterizing quantum gates via randomized benchmarking, Phys. Rev. A **85**, 042311 (2012).

[20] K. Georgopoulos, C. Emary, and P. Zuliani, Modeling and simulating the noisy behavior of near-term quantum computers, Phys. Rev. A **104**, 062432 (2021).

[21] S. Johnstun and J.-F. Van Huele, Understanding and compensating for noise on IBM quantum computers Am. J. Phys. **89**, 935 (2021).

[22] K. Georgopoulos, C. Emary, and P. Zuliani, Quantum computer benchmarking via quantum algorithms, arXiv:2112.09457.

[23] J. K. Iverson and J. Preskill, Coherence in logical quantum channels New J. Phys. **22**, 073066 (2020).

[24] Strictly speaking, the term "crosstalk" encapsulates a wide variety of effects where qubits are affected by their neighbors, as described extensively in Ref. [25]. In this work, we designate by "crosstalk" the result of the application of an imperfect quantum

gate on the neighboring qubits, usually referred to as "spillover crosstalk.".

[25] M. Sarovar, T. Proctor, K. Rudinger, K. Young, E. Nielsen, and R. Blume-Kohout, Detecting crosstalk errors in quantum information processors, Quantum **4**, 321 (2020).

[26] P. Zhao, K. Linghu, Z. Li, P. Xu, R. Wang, G. Xue, Y. Jin, and H. Yu, Quantum crosstalk analysis for simultaneous gate operations on superconducting qubits, PRX Quantum **3**, 020301 (2022).

[27] A. Ketterer, T. Wellens, Characterizing crosstalk of superconducting transmon processors, Phys. Rev. Appl. **20**, 034065 (2023).

[28] K. Rudinger, C. W. Hogle, R. K. Naik, A. Hashim, D. Lobser, D. I. Santiago, M. D. Grace, E. Nielsen, T. Proctor, S. Seritan, S. M. Clark, R. Blume-Kohout, I. Siddiqi, and K. C. Young, Experimental characterization of crosstalk errors with simultaneous gate set tomography, PRX Quantum **2**, 040338 (2021).

[29] V. Tripathi, H. Chen, M. Khezri, K.-W. Yip, E. M. Levenson-Falk, and D. A. Lidar, Suppression of crosstalk in superconducting qubits using dynamical decoupling, Phys. Rev. Appl. **18**, 024068 (2022).

[30] Y. Ding, P. Gokhale, S. F. Lin, R. Rines, T. Propson, and F. T. Chong, Systematic crosstalk mitigation for superconducting qubits via frequency-aware compilation, in *53rd Annual IEEE/ACM International Symposium on Microarchitecture (MICRO), Athens, Greece* (IEEE, New York, 2020), pp. 201–214.

[31] P. Murali, D. C. Mckay, M. Martonosi, and A. Javadi-Abhari, Software mitigation of crosstalk on noisy intermediate-scale quantum computers, in *Proceedings of the 25th International Conference on Architectural Support for Programming Languages and Operating Systems* (Association for Computing Machinery, New York, 2020), pp. 1101–1016.

[32] L. Xie, J. Zhai, and W. Zheng, Mitigating crosstalk in quantum computers through commutativity-based instruction reordering, in *58th ACM/IEEE Design Automation Conference (DAC), San Francisco, CA* (IEEE, New York, 2021), pp. 445–450.

[33] A. J. C. Woitzik, L. Hoffmann, A. Buchleitner, and E. G. Carnio, An energy estimation benchmark for quantum computers, arXiv:2302.04144.

[34] R. Blume-Kohout, J. K. Gamble, E. Nielsen, J. Mizrahi, J. D. Sterk, and P. Maunz, Robust, self-consistent, closed-form tomography of quantum logic gates on a trapped ion qubit, arXiv:1310.4492.

[35] R. Blume-Kohout, J. K. Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier, and P. Maunz, Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography, Nat. Commun. **8**, 14485 (2017).

[36] S. Endo, S. C. Benjamin, and Y. Li, Practical quantum error mitigation for near-future applications, Phys. Rev. X **8**, 031027 (2018).

[37] E. Nielsen, J. K. Gamble, K. Rudinger, T. Scholten, K. Young, and R. Blume-Kohout, Gate set tomography, Quantum **5**, 557 (2021).

[38] E.F. Dumitrescu, A.J. McCaskey, G. Hagen, G.R. Jansen, T.D. Morris, T. Papenbrock, R.C. Pooser, D.J. Dean, and P. Lougovski, Cloud quantum computing of an atomic nucleus, Phys. Rev. Lett. **120**, 210501 (2018).

[39] A. He, B. Nachman, W. A. de Jong, and C. W. Bauer, Zero-noise extrapolation for quantum-gate error mitigation with identity insertions Phys. Rev. A **102**, 012426 (2020).

[40] T. Giurgica-Tiron, Y. Hindy, R. LaRose, A. Mari, and W. J. Zeng, Digital zero noise extrapolation for quantum error mitigation, in *International Conference on Quantum Computing and Engineering (QCE), Denver, CO* (IEEE, New York, 2020), pp. 306–316.

[41] M. Urbanek, B. Nachman, V. R. Pascuzzi, A. He, C. W. Bauer, and W. A. de Jong, Mitigating depolarizing noise on quantum computers with noise-estimation circuits, Phys. Rev. Lett. **127** 270502, 270502 (2021).

[42] Y. Quek, D. S. França, S. Khatri, J. J. Meyer, and J. Eisert, Exponentially tighter bounds on limitations of quantum error mitigation, arXiv:2210.11505.

[43] O. Kern, G. Alber, and D. L. Shepelyansky, Quantum error correction of coherent errors by randomization, Eur. Phys. J. D **32**, 153 (2005).

[44] J. J. Wallman and J. Emerson, Noise tailoring for scalable quantum computation via randomized compiling, Phys. Rev. A **94**, 052325 (2016).

[45] A. Hashim, R. K. Naik, A. Morvan, J.-L. Ville, B. Mitchell, J. M. Kreikebaum, M. Davis, E. Smith, C. Iancu, K. P. O'Brien, I. Hincks, J. J. Wallman, J. Emerson, and I. Siddiqi, Randomized compiling for scalable quantum computing on a noisy superconducting quantum processor, Phys. Rev. X **11**, 041039 (2021).

[46] Z. Cai and S. C. Benjamin, Constructing smaller Pauli twirling sets for arbitrary error channels, Sci. Rep. **9**, 11281 (2019).

[47] E. Rosenberg, P. Ginsparg, and P. L. McMahon, Experimental error mitigation using linear rescaling for variational quantum eigensolving with up to 20 qubits, Quantum Sci. Technol. **7**, 015024 (2022).

[48] T. Kurita, H. Qassim, M. Ishii, H. Oshima, S. Sato, and J. Emerson, Synergetic quantum error mitigation by randomized compiling and zero-noise extrapolation for the variational quantum eigensolver, Quantum **7**, 1184 (2023).

[49] A. C. Vazquez, D. J. Egger, D. Ochsner, and S. Woerner, Well-conditioned multi-product formulas for hardware-friendly Hamiltonian simulation, Quantum **7**, 1067 (2023).

[50] Y. Kim, C. J. Wood, T. J. Yoder, S. T. Merkel, J. M. Gambetta, K. Temme, and A. Kandala, Scalable error mitigation for noisy quantum circuits produces competitive expectation values, Nat. Phys. **19**, 752 (2023).

[51] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Theory of superconductivity, Phys. Rev. **108**, 1175 (1957).

[52] R. W. Richardson, A restricted class of exact eigenstates of the pairing-force Hamiltonian, Phys. Lett. **3**, 277 (1963).

[53] R. W. Richardson and N. Sherman, Exact eigenstates of the pairing-force Hamiltonian, Nucl. Phys. **52**, 221 (1964).

[54] J. Dukelsky, S. Pittel, and G. Sierra, Colloquium: Exactly solvable Richardson-Gaudin models for many-body quantum systems, Rev. Mod. Phys. **76**, 643 (2004).

[55] S. Rombouts, D. Van Neck, and J. Dukelsky, Solving the Richardson equations for fermions, Phys. Rev. C **69**, 061303(R) (2004).

[56] A. Faribault, P. Calabrese, and J.-S. Caux, Bethe ansatz approach to quench dynamics in the Richardson model, J. Math. Phys. **50**, 095212 (2009).

[57] D. Lacroix, Symmetry-assisted preparation of entangled many-body states on a quantum computer, Phys. Rev. Lett. **125**, 230502 (2020).

[58] A. Khamoshi, F. A. Evangelista, and G. E. Scuseria, Correlating AGP on a quantum computer, Quantum Sci. Technol. **6**, 014004 (2021).

[59] E. A. Ruiz Guzman and D. Lacroix, Accessing ground-state and excited-state energies in a many-body system after symmetry restoration using quantum computers, Phys. Rev. C **105**, 024324 (2022).

[60] J. Ruh, R. Finsterhoelzl, and G. Burkard, Digital quantum simulation of the BCS model with a central-spin-like quantum processor, Phys. Rev. A **107**, 062604 (2023).

[61] M. Suzuki, Generalized trotter's formula and systematic approximants of exponential operators and inner derivations with applications to many-body problems, Commun. Math. Phys. **51**, 183 (1976).

[62] A. M. Childs, D. Maslov, Y. Nam, N. J. Ross, and Y. Su, Toward the first quantum simulation with quantum speedup, Proc. Natl. Acad. Sci. USA **115**, 9456 (2018).

[63] K. Seetharam, D. Biswas, C. Noel, A. Risinger, D. Zhu, O. Katz, S. Chattopadhyay, M. Cetina, C. Monroe, E. Demler, and D. Sels, Digital quantum simulation of NMR experiments, Sci. Adv. **9**, 46 (2023).

[64] After each Trotter step, logical qubits have been swapped around and are not at their initial location. To simulate strictly the operator of Eq. (3), one should put them back to their original configuration at the end of each Trotter step again using additional SWAP gates. However, the order in which the interaction is applied does not matter, as we neglected the noncommuting part of the evolution operator within the Trotter approximation. To limit the number of CNOT gates, we keep track of the positions of the logical qubits and apply the following Trotter step accordingly.

[65] P. Jurcevic *et al.*, Demonstration of quantum volume 64 on a superconducting quantum computing system, Quantum Sci. Technol. **6**, 025020 (2021).

[66] B. Nachman, M. Urbanek, W. A. de Jong, and C. W. Bauer, Unfolding quantum computer readout noise, npj Quantum Inf. **6**, 84 (2020).

[67] J. Hines, M. Lu, R. K. Naik, A. Hashim, J.-L. Ville, B. Mitchell, J. M. Kriekebaum, D. I. Santiago, S. Seritan, E. Nielsen, R. Blume-Kohout, K. Young, I. Siddiqi, B. Whaley, and T. Proctor, Demonstrating scalable randomized benchmarking of universal gate sets, Phys. Rev. X **13**, 041030 (2023).

[68] H. Landa L. Shirizly, and G. Misguich, Dissipative dynamics of graph-state stabilizers with superconducting qubits, Phys. Rev. Lett. **132**, 010601 (2024).

[69] https://github.com/Perrin35/BCS-error-mitigation.git.

[70] P. W. Anderson, Theory of dirty superconductors, J. Phys. Chem. Solids **11**, 26 (1959).

[71] E. A. Yuzbashyan, B. L. Altshuler, V. B. Kuznetsov, and V. Z. Enolskii, Solution for the dynamics of the BCS and central spin problems, J. Phys. A: Math. Gen. **38**, 7831 (2005).

[72] M. De Llano and J. F. Annett, Generalized Cooper pairing in superconductors Int. J. Mod. Phys. B **21**, 3657 (2007).

[73] P. W. Anderson, Random-phase approximation in the theory of superconductivity, Phys. Rev. **112**, 1900 (1958).

[74] P. Jordan and E. Wigner, Über das paulische Equivalenzverbot, Z. Phys. **47**, 631 (1928).

[75] E. Ovrum and M. Hjorth-Jensen, Quantum computation algorithm for many-body studies, arXiv:0705.1928.

[76] G. Aleksandrowicz *et al.*, QISKIT: An open-source framework for quantum computing (2019).

[77] M. Tinkham, *Introduction to Superconductivity, 2nd ed.* (McGraw-Hill, New York, 1996).