Learning how to dynamically decouple by optimizing rotational gates

Arefur Rahman,[1] Daniel J. Egger,[2] and Christian Arenz[1,*]

[1] *School of Electrical, Computer, and Energy Engineering, Arizona State University, Tempe, Arizona 85287, USA*
[2] *IBM Quantum, IBM Research Europe - Zurich, Rüschlikon 8803, Switzerland*

(Received 24 May 2024; revised 13 September 2024; accepted 4 November 2024; published 25 November 2024)

Current quantum computers suffer from noise that stems from interactions between the quantum system that constitutes the quantum device and its environment. These interactions can be suppressed through dynamical decoupling to reduce computational errors. However, the performance of dynamical decoupling depends on the type of system-environment interactions that are present, which often lack an accurate model in quantum devices. We show that the performance of dynamical decoupling can be improved by optimizing its rotational gates to tailor them to the quantum hardware. We find that, compared to canonical decoupling sequences, such as CPMG, XY4, and UR6, the optimized dynamical decoupling sequences yield the best performance in suppressing noise in superconducting qubits. Our work thus enhances existing error suppression methods, which helps increase the circuit depth and result quality on noisy hardware.

DOI: 10.1103/PhysRevApplied.22.054074

I. INTRODUCTION

Noise currently limits quantum computers from harnessing their full potential. In the long term, quantum error correction is expected to overcome this issue [1]. In the short term, noise mitigation and suppression techniques are critical to improve quantum device performance [2]. Error mitigation is designed to reduce noise, typically, in expectation values [2,3]. The computation is executed on the noisy quantum computer multiple times to either extrapolate to a zero-noise limit [4], or to cancel the noise on average [5]. By contrast, error suppression methods, such as dynamical decoupling [6] and pulse-efficient transpilation [7], reduce the presence of noise directly in the quantum circuits.

In this work we focus on the well-established noise and error suppression technique dynamical decoupling (DD). Inspired by nuclear magnetic resonance spectroscopy (NMR) [8], the theory of DD was first developed by Lorenza Viola, Emanuel Knill, and Seth Lloyd in 1998 [6] as an open-loop control technique. DD suppresses errors by decoupling the system from its environment through the application of a sequence of pulses that, in the ideal case, compose to the identity. The utility of DD has been demonstrated in a wide range of quantum systems, such as coupled nuclear and electron spins [9], trapped ions [10], electron spins [11], and superconducting qubits [12]. Since the introduction of the DD framework in the 1990s, DD has also become a viable method to suppress noise and errors in quantum computing [13]. For example, DD can suppress crosstalk [14,15] and improve the performance of superconducting qubit-based quantum devices [12,16] in general. DD sequences can be designed from first principles [17] or with numerical simulations, leveraging tools such as genetic algorithms [18] and machine learning [19].

In recent years, DD has become a major error suppression method for noisy superconducting quantum computers [16]. Indeed, the method is easy to apply as a simple transpilation pass that inserts delays and pulses into a quantum circuit. Furthermore, simple sequences such as X-X already yield excellent results [20]. More elaborate sequences, such as staggered X-X [21] and staggered $XY4$, improve, for instance, the execution of dynamic circuits by canceling crosstalk [22]. Here X and Y are π rotations around the qubit's x and y axes, respectively. Crucially, the performance of a DD sequence depends on the interactions present in the quantum hardware. In superconducting qubits [23], a good model of these interactions is typically not known, a case familiar to optimal control [24–27] that can be overcome with closed-loop optimization [28,29]. Similarly, it is possible to tailor DD sequences to the quantum hardware at hand by learning them through genetic algorithms [30]. However, in hardware, DD pulses are not perfect, potentially introducing additional noise and errors through the controls that implement the DD pulses, thereby diminishing the quality of the designed DD sequence. While control noise-robust DD sequences exist [13], e.g., suppressing errors in rotational angles of the DD gates, such sequences typically require knowledge of the type of control error that is introduced.

[*] Contact author: carenz1@asu.edu

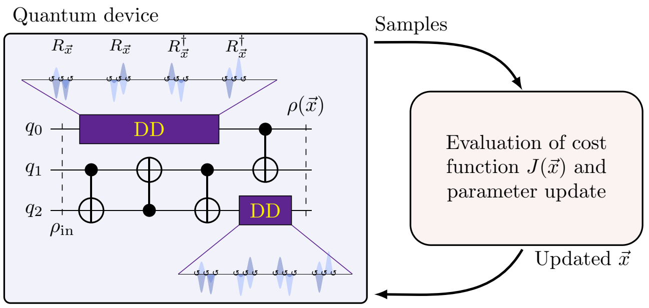

FIG. 1. Illustration of LDD. The left and right blocks indicate a quantum device and a classical computer, respectively, working in tandem. A DD sequence parameterized by $\vec{x}$ is added to idling qubits in a given quantum circuit, here exemplified by a SWAP gate on (q_1, q_2) followed by a CNOT gate on (q_0, q_1). The arbitrary single-qubit rotations $R_{\vec{x}}$ in the DD sequence are realized by three parameterized virtual Z gates (circular arrows) and two $\sqrt{X}$ gates (pulse with in-phase and quadrature shown in light and dark blue, respectively). The noisy output state $\rho(\vec{x})$ is compared to the expected ideal state ρ_{ideal} through the cost function $J(\vec{x})$. A classical optimization algorithm minimizes $J(\vec{x})$ to find the parameters of the gates in the DD sequence that best fit the circuit.

To overcome these limitations, in this work we tailor the DD sequences to the hardware and quantum circuits to execute. The goal is to design DD sequences that suppress potentially unknown pulse imperfections and noise. This is achieved by optimizing the rotational angles of the gates in the DD sequence in a closed loop with the quantum hardware. A classical optimizer is fed the cost function value that is reconstructed from quantum samples and that is sensitive to the quality of the DD pulses; see Fig. 1.

This manuscript is structured as follows. We introduce in Sec. II the three commonly employed DD sequences CPMG, XY4, and UR6. Next, in Sec. III we develop the theoretical framework of how optimal parameters in DD sequences are found on quantum hardware, which we refer to as learning dynamical decoupling (LDD). We demonstrate in Sec. III B the utility of LDD on IBM Quantum hardware by comparing the performance of LDD to CPMG, XY4, and UR6 sequences to suppress noise in two experiments. We show that LDD outperforms CPMG, XY4, and UR6 sequences for suppressing noise present during midcircuit measurements and noise resulting from increasing the depth of a quantum circuit. We conclude in Sec. IV.

II. BACKGROUND: DYNAMICAL DECOUPLING

DD is a well-known strategy that reduces noise by suppressing unwanted interactions with the environment or undesired couplings of the controls with the system. For an overview of DD strategies, we refer the reader to Ref. [13]. DD can in general suppress generic interactions through pulses that rotate around multiple axes [31]. However, DD is most resource efficient when tailored to the specific type of interactions at hand [32,33]. Furthermore, the most effective DD sequence depends on the noise type present in the physical system.

The spin echo [34] in NMR can be seen as a DD experiment where a single Pauli-X gate refocuses coherent errors. The "Carr-Purcell-Meiboom-Gill" (CPMG) DD sequence is an extension of the spin echo with two symmetric insertions of an X gate [35,36]. The resulting pulse sequence is

$$\text{CPMG} \equiv \tau/2 - X - \tau - X - \tau/2. \qquad (1)$$

Here, the total duration of the free evolution is 2τ and $-\tau-$ indicates an idle period of duration τ. Multiple CPMG sequences can be concatenated one after another; see Fig. 2(a). While the CPMG sequence can suppress homogeneous dephasing along one axis, it cannot suppress noise stemming from generic system-environment interactions. By contrast, the XY4 DD sequence [37], defined by

$$\text{XY4} \equiv Y - \tau/2 - X - \tau/2 - Y - \tau/2 - X - \tau/2, \quad (2)$$

is a universal DD sequence that can suppress generic system-environment interactions. Two concatenated XY4 sequences are shown in Fig. 2(b). Both the CPMG and XY4 sequences are sensitive to pulse imperfections [38]. Such errors can be suppressed with a universal rephasing (UR) DD sequence by inserting properly tailored phase gates into, e.g., a CPMG sequence to obtain the symmetric UR6 DD sequence [39]

$$\begin{aligned}\text{UR6} \equiv\ & 2\tau/7 - X - 2\tau/7 - X' - 2\tau/7 - X - 2\tau/7 \\ & - X - 2\tau/7 - X' - 2\tau/7 - X - 2\tau/7. \qquad (3)\end{aligned}$$

The UR6 sequence has four X gates and two phased X gates denoted by $X' = R_z(-2\pi/3)XR_z(2\pi/3)$, where $R_z(\theta)$ is a rotation around the qubit's z axis by an angle θ.

DD sequences with higher-order protection can be built with additional single-qubit pulses, for example, by concatenating existing sequences [40] or considering semiclassical noise models [39]. However, the effectiveness of DD sequences critically depends on the type of noise, i.e., resulting from the detrimental interactions that are present in the system, which is often challenging to infer in superconducting quantum devices [23]. For example, Ezzell *et al.* [12] showed that the performance of different DD sequences is device dependent even though the devices all leverage the cross-resonance interaction [41]. Furthermore, pulse imperfections, such as a detuning in frequency or amplitude, degrade the performance of a DD sequence [39]. As such, the performance of a DD sequence can vary substantially across different devices and even time, if hardware parameters drift.

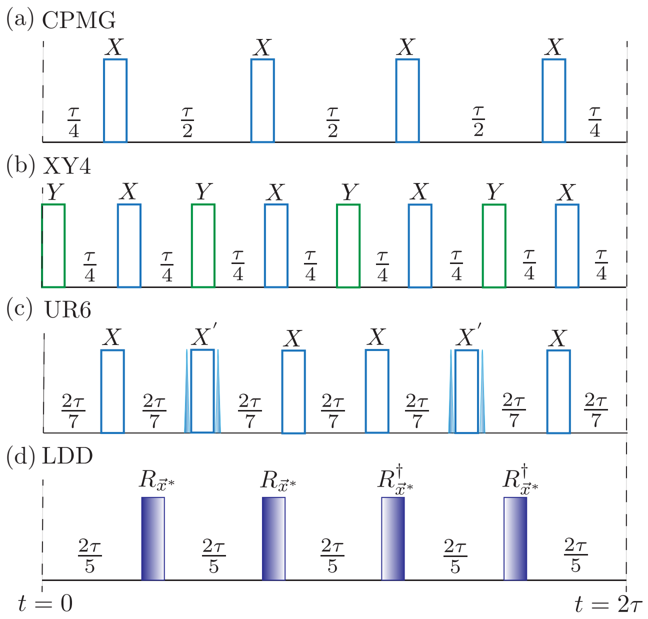

FIG. 2. Schematic diagram of DD sequences. (a) Two concatenated CPMG sequences. (b) Two concatenated XY4 sequences. (c) One UR6 sequence. (d) Two concatenated LDD sequences. All sequences are depicted for an idle period of 2τ, while the gate implementation time is not shown. When running on hardware, the idle times are reduced to account for the duration of the physical pulses of the DD gate. The CPMG sequence uses only X gates (blue rectangle), whereas the XY4 sequence uses both X and Y gates (green rectangles). Here X and Y are π rotations around the qubit's x axis and y axis, respectively. Sequence UR6 uses X' gates that consist of two phase gates with angles $2\pi/3$ (shown as blue pulses) with opposite rotational directions that are inserted before and after an X gate; see Eq. (3). LDD uses parameterized gates, $R_{\vec{x}^*}$ (purple rectangles), where the optimal rotation angles $\vec{x}^*$ are found by classical optimization routines performed in tandem with a quantum device.

III. LEARNING HOW TO SUPPRESS NOISE

We use tools from closed-loop optimal control [24] and optimization [28–30,42] to learn optimal DD sequences without precise knowledge of a noise model. Tong *et al.* [30] demonstrated the usefulness of this approach by optimizing with a genetic algorithm the placement of DD gates on idling qubits in a quantum circuit to achieve, for instance, a higher success probability in the Bernstein-Vazirani algorithm compared to canonical DD sequences. In their approach, they chose DD gates from the fixed set $\{I_\pm, X_\pm, Y_\pm, Z_\pm\}$, where I_+, X_+, Y_+, Z_+ are the Pauli matrices and the minus subscript indicates an added phase of π.

DD gates themselves are prone to errors, such as errors in the rotation angles. Furthermore, the optimal rotation axis of the gates in the DD sequence may depend on the type of noise in the system. We address these issues by adopting a different optimization approach from Tong *et al.* [30]. Instead of fixing a set of DD gates and optimizing over the DD sequence structure to find the best DD sequence [30], we optimize the rotational parameters entering in the chosen DD sequence to improve performance. This approach is similar to finding noise-resilient quantum circuits through machine learning [43]. After introducing the theory behind such a learning approach to dynamical decoupling, we demonstrate on IBM Quantum hardware in two experiments that the optimized LDD sequences are better at suppressing errors than the CPMG, XY4, and UR6 sequences.

A. Theory

We consider a quantum circuit described by the ideal quantum channels $\mathcal{U}_j$ applied in sequence,

$$\rho_{\text{ideal}} = \prod_j \mathcal{U}_j(\rho_0), \tag{4}$$

to an initial state $\rho_0 = |\psi_0\rangle\langle\psi_0|$, which we assume is pure, to create a desired target state $\rho_{\text{ideal}} = |\psi_{\text{ideal}}\rangle\langle\psi_{\text{ideal}}|$. The quantum channels $\mathcal{U}_j$ may correspond to the ideal gates in a circuit. Here, we concern ourselves with noise described by a collection of typically unknown quantum channels $\mathcal{M}_j$, in which we include potential midcircuit measurements, that act between the unitary channels $\mathcal{U}_j$ when qubits are idling. The noisy quantum circuit then takes the form $\prod_j \mathcal{M}_j\mathcal{U}_j$. Each noise channel $\mathcal{M}_j$ can describe unitary and nonunitary errors. To suppress noise induced by $\mathcal{M}_j$, we divide $\mathcal{M}_j$ into $N+1$ channels $\mathcal{M}_{j,k}$ so that $\prod_{k=1}^{N+1}\mathcal{M}_{j,k} = \mathcal{M}_j$, and insert an LDD sequence of the form shown in Fig. 2(c). The LDD sequence is described by parameterized unitary quantum channels $\mathcal{R}_{\vec{x}}(\cdot) = R_{\vec{x}}(\cdot)R^\dagger_{\vec{x}}$ that are applied between the $\mathcal{M}_k$ where the LDD gates $R_{\vec{x}}$ are parameterized by $\vec{x} \in \mathbb{R}^M$. For simplicity, we assume that the LDD parameters $\vec{x}$ are the same for each LDD gate $R_{\vec{x}}$. If we include $N/2$ LDD gates between the noise channels $M_{j,k}$, followed by implementing $N/2$ times the corresponding inverse $R^\dagger_{\vec{x}}$ to insure that the total LDD sequence composes to the identity in the absence of noise, as depicted in Fig. 2, the noise channel $\mathcal{M}_j$ becomes

$$\mathcal{M}_{j,\vec{x}} := \mathcal{M}_{j,N+1}\prod_{k=N/2+1}^{N}\mathcal{R}^\dagger_{\vec{x}}\mathcal{M}_{j,k}\prod_{k=1}^{N/2}\mathcal{R}_{\vec{x}}\mathcal{M}_{j,k}. \tag{5}$$

The state resulting from the noisy circuit with LDD is then given by

$$\rho(\vec{x}) = \prod_j \mathcal{M}_{j,\vec{x}}\mathcal{U}_j(\rho_0). \tag{6}$$

To optimize the parameters $\vec{x}$, we need a cost function $J(\vec{x})$ that is sensitive to the quality of the circuits in Eq. (4). For small circuits, a natural choice for a figure of merit or cost function is the fidelity error

$$J(\vec{x}) = 1 - \langle \psi_{\text{ideal}} | \rho(\vec{x}) | \psi_{\text{ideal}} \rangle. \tag{7}$$

Here, $F = \langle \psi_{\text{ideal}} | \rho(\vec{x}) | \psi_{\text{ideal}} \rangle$ is the fidelity with respect to the ideal state. However, state tomography scales exponentially with system size. For large circuits, a scalable cost function can be built in multiple ways. As done in Refs. [30,44], we can invert the circuit with mirroring in which each $\mathcal{U}_j^\dagger$ is applied in reverse such that $\rho_{\text{ideal}} = \rho_0$. Alternatively, one can reduce the original quantum circuit to a Clifford circuit as done, for instance, in Refs. [45,46]. Here, the single-qubit gates in a circuit [with the controlled-NOT (CNOT) gate as the two-qubit gate] are replaced by Clifford gates such that the whole circuit is a Clifford gate. In hardware where parameters are encoded in virtual-Z gates [47], this will preserve the structure and timing of the underlying pulses, thereby leaving most noise sources such as T_1, T_2 and crosstalk unchanged at the pulse level. We can then compute ρ_{ideal} with efficient Clifford-based simulators.

Solving the optimization problem

$$\min_{\vec{x} \in \mathbb{R}^M} J(\vec{x}) \tag{8}$$

yields the optimal parameter values $\vec{x}^*$ of the LDD sequence that minimize $J(\vec{x})$. Since we do not know the noise processes described by $\mathcal{M}_j$, we minimize J in an iterative, variational quantum algorithm–type fashion by using quantum and classical computing resources in tandem [27,48]. By measuring the output at the end of the quantum circuit we estimate J, while a classical search routine is employed to update the parameters.

B. Case studies on IBM hardware

We study the performance of LDD in two different experiments carried out on IBM Quantum hardware with Bell pairs, a valuable resource. For instance, Bell pairs enable quantum gate teleportation [49] and similar known states generated in a factory of resources enable circuit cutting [22]. In the first experiment we suppress noise during midcircuit measurements. In the second experiment we suppress noise resulting from an increasing circuit depth. In both cases we minimize fidelity loss due to noise by increasing the fidelity of preparing the Bell state $|\psi_{\text{ideal}}\rangle = |\Phi_+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$, on two qubits q_i and q_j within an $n > 2$ qubit system, starting from $\rho_0 = |0\rangle\langle 0|^{\otimes n}$. To infer the value of $J(\vec{x})$ in Eq. (7) for the Bell state, it suffices to measure the expectation values of the X_iX_j, Y_iY_j, and Z_iZ_j Pauli operators with respect to $\rho(\vec{x})$, rather than performing full state tomography [50–56]. The cost function is thus

$$J(\vec{x}) = 1 - \frac{1}{4}\langle \mathbb{1} + X_iX_j - Y_iY_j + Z_iZ_j \rangle_{\rho(\vec{x})}. \tag{9}$$

The three correlators are each estimated with a quantum circuit executed with 400 shots in each evaluation. We employ parameterized decoupling operations given by arbitrary single-qubit rotations

$$R_{\vec{x}} = R(\theta, \phi, \lambda) = e^{-i\theta Z/2} e^{-i\phi Y/2} e^{-i\lambda Z/2}, \tag{10}$$

where in both experiments we repeat $R(\theta, \phi, \lambda)$ and its inverse in total $N = 4$ times, described by Eq. (5). Here, we only consider three angles $\vec{x} = (\theta, \phi, \lambda)$ to parameterize the full LDD sequence. All qubits with an LDD sequence thus share the same parameter values. This limits the size of the search space. On IBM Quantum hardware, these single-qubit rotations are implemented by three parameterized virtual-Z rotations [47] and two $\sqrt{X}$ pulses. Therefore, in LDD we optimize the angles in these virtual-Z rotations. We pick the simultaneous perturbation stochastic approximation (SPSA) gradient descent method [57,58] to solve the optimization problem in Eq. (8), starting with $(\theta, \phi, \lambda) = (0, 0, 0)$. We allow the SPSA a total of 100 iterations. At each iteration the SPSA requires only two estimations of the objective function, regardless of the number of optimization parameters. The hyperparameter "perturbation" in the SPSA is set to the default values and the hyperparameter "learning rate" is calibrated by the optimizer [59].

Below we compare LDD with the CPMG, XY4, and UR6 sequences where DD sequences are inserted when qubits are idling. Throughout the two experiments, we concatenate two LDD, CPMG, and XY4 sequences, as shown in Fig. 2, and adjust the duration τ to match the idling times of the qubits. Therefore, the LDD sequence is built from two $R_{\vec{x}}$ gates and two $R_{\vec{x}}^\dagger$ gates. At the hardware level, each $R_{\vec{x}}$ gate is implemented by three virtual-Z rotations and two $\sqrt{X}$ gates. The X and Y gates are implemented by a single pulse with the same duration as a $\sqrt{X}$ gate. An example of the transpiled DD sequences is shown in Appendix D.

1. *Suppressing noise during midcircuit measurements*

A midcircuit measurement (MCM) involves measuring qubits at intermediate stages within a quantum circuit. MCMs have various applications, including quantum error correction [60], quantum teleportation [61], reducing the depth of a quantum circuit [62], circuit cutting [22], and analyzing complex quantum behaviour [63]. Unfortunately, MCMs may introduce noise on neighboring qubits of the physical device [64].

As depicted in Fig. 3, we consider the task of preparing the Bell state $|\Phi_+\rangle$ between qubit q_0 and qubit q_2, while qubit q_1 is subject to r repeated MCMs. Here,

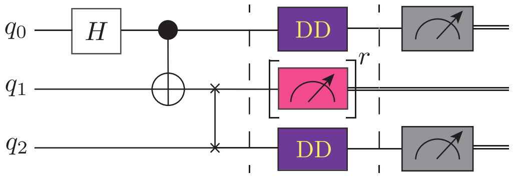

FIG. 3. Suppressing noise during MCMs through DD. A Bell state is prepared on qubits q_0 and q_2. The magenta block on qubit q_1 is an MCM that is repeated r times. The purple blocks on qubits q_0 and q_2 indicate DD sequences inserted during the MCMs to mitigate noise on q_0 and q_2. We vary r and study the performance of different DD sequences on a 127-qubit Eagle device (*ibm_kyiv*) and a 27-qubit Falcon device (*ibm_hanoi*) reported in Appendix B. The corresponding experimental results are shown in Figs. 4 and 7(a).

varying $r \in \{1, \ldots, 15\}$ allows us to amplify the amount of noise introduced in the quantum circuit. To study the effect of DD during noisy measurements, we first identify a measurement that introduces noise on its neighboring qubits (see Appendix A), and thus map $(q_0, q_1, q_2) \rightarrow (q_{120}, q_{121}, q_{122})$.

First, we compute F without any DD, once with r MCMs and once with a delay equivalent to r MCMs. The experimental result are shown in Fig. 4 (red and purple curves) where we report the median and the lower and upper quartiles of ten measurements. In the presence of MCMs we observe a large drop in F at $r = 5$ from $0.813^{+0.016}_{-0.01}$ to $0.396^{+0.061}_{-0.096}$. This implies that an MCM on q_{121} strongly impacts q_{120} and q_{122} aside from adding a 1244-ns delay. Crucially, the fidelity decrease due to the long delays and measurement-induced noise of the r MCMs on q_{121} can be mitigated by DD sequences inserted on neighboring qubits q_{120} and q_{122} during the MCM measurement; see Fig. 3. In Fig. 4 we compare the performance of the LDD (blue), CPMG (green), XY4 (orange), and UR6 (magenta) sequences. The LDD sequence yields the best performance, resulting in, e.g., a fidelity of $0.631^{+0.007}_{-0.007}$ at $r = 15$, while the CPMG, XY4, and UR6 sequences result in $0.625^{+0.019}_{-0.045}$, $0.499^{+0.010}_{-0.021}$, and $0.521\,^{+0.08}_{-0.054}$, respectively. To evaluate the reliability of the LDD sequence, we compute $F(\vec{x}^*)$ 10 times after the learning and report the lower quartile, median, and upper quartile of these ten runs. Because of queuing, these circuits were executed on the hardware two days after the optimal parameters $\vec{x}^*$ were learnt. This indicates that the learned parameters $\vec{x}^*$ are stable in time. We attribute the residual decay in part to T_1, T_2, and state preparation and measurement (SPAM) errors, shown for comparison in Fig. 4 as a solid black, a solid gray, and a dashed gray curve, respectively. The T_1 curve is computed as e^{-t/T_1}, where t is the idling time and $T_1 = 245$ μs is the average T_1 time of qubits q_{120} and q_{122} of *ibm_kyiv*. Similarly, the T_2 curve is computed as e^{-t/T_2}, where $T_2 = 175$ μs is the average T_2 time of qubits q_120 and q_122 of *ibm_kyiv*. The dashed gray line is $0.987e^{-t/T_2}e^{-t/T_1}$, where 0.987 is the product of the readout assignment fidelity of qubits q_{120} and q_{122}, as reported by the backend.

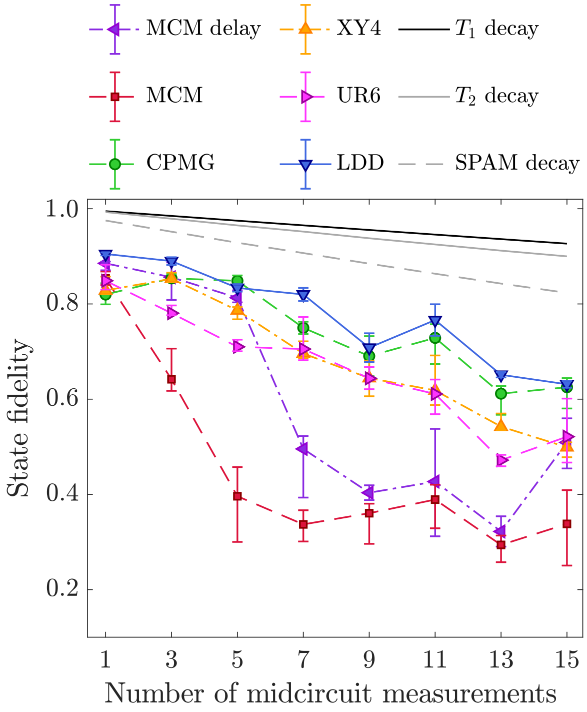

FIG. 4. DD noise suppression during MCMs. Fidelity of a Bell state between two qubits as a function of the number of MCMs performed on another qubit; see Fig. 3. The dashed red curve shows the fidelity with MCMs but without DD and the dash-dot purple curve shows the fidelity with a delay equivalent to MCMs. DD sequences are inserted when the qubits are idling. The dashed green curve corresponds to the CPMG sequence, the dash-dot orange curve corresponds to the XY4 sequence, the dashed magenta curve corresponds to the UR6 sequence, and the solid blue curve corresponds to the optimized LDD sequence. For comparison, the solid black curve, the solid gray curve, and the dashed gray curve show the effect of T_1 decay, T_2 decay, and decay due to SPAM errors, respectively. The error bars show the interquartile range.

2. Suppressing deep circuit noise

Next, we consider suppressing noise through LDD that is introduced due to increasing the depth of a quantum circuit. In particular, we consider the task of preparing the Bell state $|\Phi_+\rangle$ between two qubits located at the edges of a qubit chain with nearest-neighbor interactions.

The corresponding coupling graph of the IBM Quantum device is shown in Fig. 5(a) where the considered qubit chain is highlighted in the dashed gray box. In Fig. 5(b) we show the quantum circuit that prepares a Bell state between qubits q_0 and q_5 by bringing them into proximity with a ladder of SWAP gates. Since the ancilla qubits are in their ground state $|0\rangle$, we can implement each SWAP gate

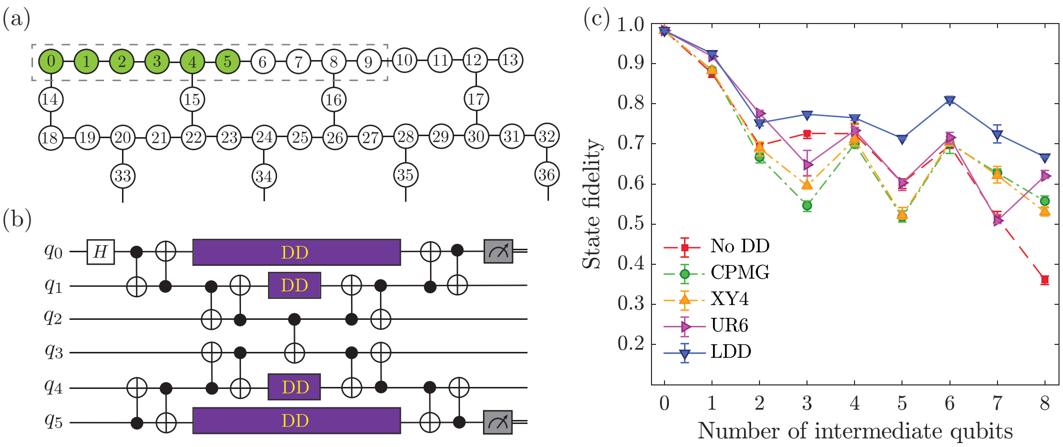

FIG. 5. Suppressing deep circuit noise through DD. (a) A part of the qubit coupling graph of the 127-qubit IBM Eagle device *ibm_strasbourg*. Qubits are shown as circles labeled by numbers. The green circles highlight the qubits in the quantum circuit in (b). The gray dashed box shows the chain of qubits $0, 1, 2, 3, 4, 5, 6, 7, 8$, and 9 used in the experiment. (b) Quantum circuit to prepare a Bell state between qubit q_0 (one edge of the chain) and qubit q_i (here shown for $i = 5$) using a ladder of CNOT gates. The insertion of DD sequences is shown as purple boxes. (c) Fidelity of preparing a Bell state between qubits q_0 and q_i in the chain in the coupling graph (a) as a function of the number of intermediate qubits. The dashed red curve shows the fidelity without DD, the dash-dot orange curve corresponds to the XY4 sequence, the dash-dot green curve corresponds to the CPMG sequence, the dashed magenta curve corresponds to the UR6 sequence, and the solid blue curve corresponds to the optimized LDD sequence. The error bars show the interquartile range.

with two CNOT gates instead of three. Similar gate ladders with a single CNOT gate at each rung often occur in quantum simulation algorithms to create unitaries generated by Pauli strings [65,66]. In Fig. 5(c) we plot the fidelity F for preparing the Bell state between the two edge qubits of the chain in Fig. 5(a) as a function of the intermediate qubits (IQs) shown by the dashed red curve. We see a decrease in fidelity as the number of qubits in the chain increases, ranging from two qubits (zero IQs) to ten qubits (eight IQs).

To mitigate the Bell state fidelity decrease, as depicted in Fig. 5(b), we insert DD sequences on idling qubits and compare their performance. The results shown in Fig. 5(c) for the CPMG sequence (green), the XY4 sequence (orange), the UR6 sequence (magenta), and the LDD sequence (blue) suggest that the best performance is obtained for LDD.

In fact, since the fidelity obtained through inserting XY4 and CPMG sequences can even be below the fidelity obtained without DD, we conclude that DD can increase the noise instead of suppressing it. This situation is avoided by LDD as the DD sequence is tailored to the device.

IV. CONCLUSION

Dynamical decoupling is a powerful noise suppression strategy that averages out detrimental processes by applying properly designed pulses to the system. We introduced the framework of “learning dynamical decoupling.” Instead of considering a DD sequence with fixed rotational gates, LDD optimizes directly on quantum hardware the rotational parameters in the DD gates. We compared the performance of such optimized DD sequences with the known DD sequences CPMG, XY4, and UR6 on IBM Quantum hardware. We found that LDD outperforms all three sequences in suppressing noise that occurs during midcircuit measurements and noise that stems from increasing the depth of a quantum circuit. We believe that this is because the optimal angles in LDD are tailored to the noise of the device. We expect that this performance can be further improved by also optimizing over the DD gate spacings and the number of parameterized gates used. Furthermore, the spacing between the DD gates and the way in which they are compiled into R_Z, X, and $\sqrt{X}$ gates may also impact performance [67].

The LDD sequences that we studied have by design a small number of single-qubit gates and a fixed number of rotational parameters. Appendix C shows that the optimized LDD parameters are robust against perturbations. While we believe that performance can be increased even further by adding more optimization parameters, the results shown in Fig. 5 for different system sizes (i.e., a different number of intermediate qubits) suggest that the number of LDD parameters can remain constant while achieving a similar performance when the system is scaled. As such, the classical optimization overhead does not need to increase when the system size increases. Therefore, the LDD approach considered here is scalable by design. Furthermore, inserting only a small number of single-qubit gates in LDD or DD on idling qubits to suppress noise is important for current quantum devices. Idle times in quantum circuits typically occur when a subset of all qubits undergo two-qubit gates. Therefore, the shorter the two-qubit gate duration is on a device, the more compact the DD sequence needs to be. To illustrate this, consider *ibm_torino* and *ibm_sherbrooke* that have a median two-qubit gate duration of 84 and 533 ns, respectively [68]. The durations of the single-qubit X and $\sqrt{X}$ gates are 32 and 57 ns for *ibm_torino* and *ibm_sherbrooke*, respectively. As such, on *ibm_torino* we can insert up to two X or Y gates or one arbitrary single-qubit rotation during the median two-qubit gate duration. By contrast, on *ibm_sherbrooke* these numbers become eight and four, respectively. Consequently, as the two-qubit gate duration becomes comparable with the single-qubit gate duration, short DD sequences become more important.

In summary, DD is crucial to suppressing errors in noisy hardware. As DD sequences improve—becoming more tailored to the hardware—so do the hardware results. This motivates the strong interest in DD. Future work may include optimizing the spacing of the DD pulses in LDD. Furthermore, one could explore how to protect circuit cutting resources, consumed in teleportation circuits as demonstrated in Ref. [22]. Indeed, these resources are less costly to generate simultaneously. However, this has the drawback that they idle until they are consumed. Finally, we optimized virtual-Z rotations that sandwich $\sqrt{X}$ gates. Future work may thus elect to directly optimize the pulses that implement the DD sequence, e.g., similar to pulse-level variational quantum algorithms [27,69].

ACKNOWLEDGMENTS

C.A. acknowledges support from the National Science Foundation (Grant No. 2231328). A.R. and C.A. acknowledge support from Knowledge Enterprise at Arizona State University. D.J.E. acknowledges funding within the HPQC project by the Austrian Research Promotion Agency (FFG, Project No. 897481) supported by the European Union – NextGenerationEU. We acknowledge the use of IBM Quantum services for this work. The views expressed are those of the authors, and do not reflect the official policy or position of IBM or IBM Quantum.

APPENDIX A: IDENTIFYING QUBITS WITH THE MOST MCM-INDUCED NOISE

The amount of noise an MCM introduces depends on the qubit on which the MCM is performed. To demonstrate that LDD can suppress noise introduced through MCMs, we identify qubits in the device where MCMs have a large effect on the fidelity of a Bell state. This is achieved by identifying qubits on the 127-qubit IBM device (*ibm_kyiv*) for which repeated MCMs significantly introduce more noise than letting the qubit idle for the corresponding amount of time. These qubits were found by preparing a Bell state between two qubits, followed by applying an MCM ($r = 1$) or the delay equivalent to an MCM ($r = 0$) on another qubit, as shown in Fig. 3. In the experiment, we choose the indices of the three qubits such that their corresponding physical qubits are adjacent to each other. This ensures that the noise in the circuit is introduced by the MCM rather than by the circuit transpilation. The Bell state fidelity is shown as a function of the qubit index on which the MCM or delay is applied in Fig. 6. The indices of the qubits with MCM that we consider in *ibm_kyiv* are 1, 4, 7, 10, 19, 22, 25, 28, 31, 40, 43, 46, 49, 58, 61, 64, 67, 76, 79, 82, 85, 88, 97, 100, 103, 106, 115, 118, 121, and 124. These choices ensure that the total three-qubit system forms a chain with nearest-neighbor interactions in the coupling map of the device to avoid any additional SWAP gates that may introduce more noise.

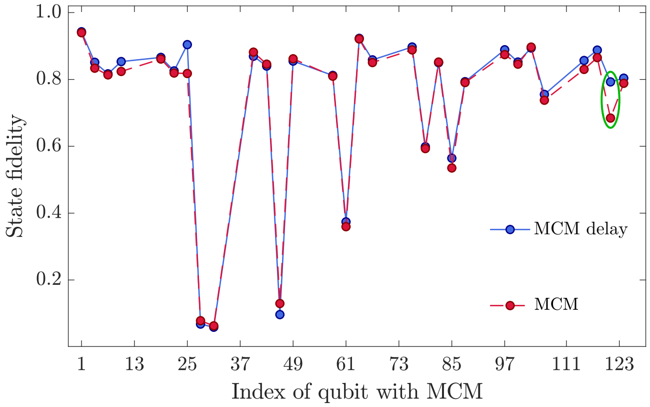

FIG. 6. Noisy MCM identification experiment. Bell state fidelity between qubits q_0 and q_2 as a function of the index of qubit q_1 in Fig. 3 of the main text. The green ellipse shows qubit 121 for which the state fidelity with an MCM (red curve) is significantly lower than when there is a delay equivalent to an MCM (blue curve).

The red curve in Fig. 6 shows the state fidelity in the presence of an MCM, while the blue curve corresponds to the same circuit without an MCM, i.e., the MCM is replaced by a delay with the same duration. The qubit candidate (index 121) chosen for the MCM experiments whose results are shown in Fig. 4 in the main text is highlighted by a green ellipse.

APPENDIX B: ADDITIONAL DATA

1. MCM noise

In addition to the data presented in the main text obtained from the 127-qubit Eagle device *ibm_kyiv*, we also gathered data from the 27-qubit Falcon device *ibm_hanoi* for the LDD, CPMG, and XY4 sequences. Here, we prepared the Bell state between qubits q_0 and q_2, and applied an MCM on qubit q_1. The results for the effect of an MCM on qubits q_0 and q_2 are shown in Fig. 7(a). We report the median and the lower and upper quartiles of ten measurements. In the presence of MCMs

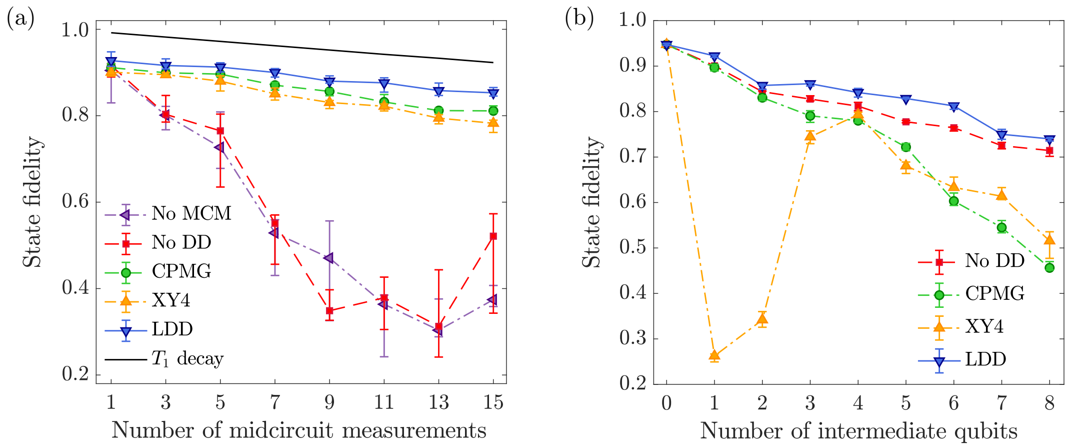

FIG. 7. DD noise suppression during MCMs and deep circuit noise suppression using *ibm_hanoi* and *ibm_cairo*, respectively. (a) Fidelity as a function of the number of MCMs performed. The dashed red curve shows the fidelity with MCMs but without DD and the dash-dot purple curve shows the fidelity with a delay equivalent to MCMs. The dashed green curve corresponds to the CPMG sequence, the dash-dot orange curve corresponds to the XY4 sequence, and the solid blue curve corresponds to the optimized LDD sequence. The optimal LDD parameters are shown in Table I. For comparison, the solid black curve shows the T_1 decay. The error bars show the interquartile range. (b) Fidelity of preparing a Bell state between two qubits in the chain in the coupling graph as a function of the number of intermediate qubits. The dashed red curve shows the fidelity without DD, the dash-dot orange curve corresponds to the XY4 sequence, the dash-dot green curve corresponds to the CPMG sequence, and the solid blue curve corresponds to the optimized LDD sequence. The optimal LDD parameters are shown in Table II. The error bars show the interquartile range.

(red curve), we observe a large drop in F from $0.916^{+0.004}_{-0.026}$ at $r = 1$ to $0.521^{+0.052}_{-0.178}$ at $r = 15$. Without MCMs (purple curve) F drops from $0.905^{+0.015}_{-0.075}$ at $r = 1$ to $0.374^{+0.033}_{-0.016}$ at $r = 15$. This implies that on *ibm_hanoi* an MCM on a single qubit does not have a large impact on the adjacent qubits aside from adding a 820-ns delay. These qubits were not selected using the procedure used to identify noisy MCMs on *ibm_kyiv*.

In Fig. 7(a), we compare the performance of the LDD (blue), CPMG (green), and XY4 (orange) sequences. The LDD sequence, whose corresponding optimal parameters are shown in Table I, yields the best performance, resulting in, e.g., a fidelity of $0.855^{+0.012}_{-0.007}$ at $r = 15$, while the CPMG and XY4 sequences result in $0.811^{+0.012}_{-0.005}$ and $0.782^{+0.007}_{-0.021}$, respectively. We attribute the residual decay to T_1, shown for comparison in Fig. 7(a) as a solid black curve, where $T_1 = 167$ μs is the average T_1 time of qubits q_0 and q_2 of *ibm_hanoi*.

TABLE I. Optimal LDD parameters to suppress noise during MCMs. The angles in the LDD gate given in Eq. (10) are optimized to find the optimal rotational (Euler) angles $\vec{x}^* = (\theta^*, \phi^*, \lambda^*)$. The optimal angles are shown for different numbers of MCMs applied on q_1 that correspond to the idling times of qubits q_0 and q_2. For reference, the Euler angles for the Pauli-X and Pauli-Y gates used in the XY4 sequence are $(\pi, \pi, 0)$ and $(\pi, 0, 0)$, respectively.

		Optimal LDD parameters		
# MCMs	Idle time (μs)	θ^*	ϕ^*	λ^*
1	0.82	0.37π	0.22π	0.05π
3	2.45	-0.32π	-0.35π	0.13π
5	4.09	-0.66π	0.74π	-0.66π
7	5.72	-0.67π	0.40π	-0.54π
9	7.35	-0.28π	-0.53π	-0.29π
11	8.99	-1.13π	1.49π	-1.34π
13	10.62	0.95π	-1.07π	-0.68π
15	12.26	-0.99π	-0.79π	0.50π

2. Deep circuit noise

In addition to the data presented in the main text on the 127-qubit Eagle device *ibm_strasbourg*, we also gathered data on the 27-qubit Falcon device *ibm_cairo* for the LDD, CPMG, and XY4 sequences; see Fig. 7(b). We plot

TABLE II. Optimal LDD parameters for suppressing deep circuit noise. The angles in the LDD gate given in Eq. (10) are optimized to find the optimal rotational (Euler) angles $(\theta^*, \phi^*, \lambda^*)$. The optimal $\vec{x}^* = (\theta^*, \phi^*, \lambda^*)$ are shown for different numbers of IQs.

	Optimal LDD parameters		
# IQs	θ^*	ϕ^*	λ^*
1	-0.41π	0.47π	-0.28π
2	-0.03π	0.72π	0.03π
3	-0.02π	-0.45π	0.02π
4	-0.07π	-0.47π	-0.25π
5	-0.03π	-0.48π	-0.17π
6	-0.02π	0.24π	1.02π
7	1.94π	0.85π	-0.42π
8	-0.01π	1.85π	-0.33π

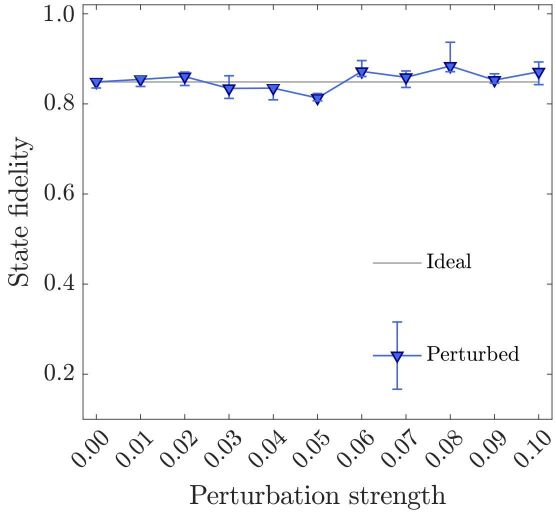

FIG. 8. Robustness of the optimal angles to perturbations. The Bell state fidelity between qubits q_{120} and q_{122} as a function of the strength of perturbation ϵ; see Eqs. (C1) and (C2). The fidelity corresponding to the optimal angles ($\epsilon = 0$) is shown as a gray line.

the fidelity F for preparing the Bell state between the two edge qubits of the qubit chain in *ibm_cairo* as a function of the IQs shown by the dashed red curve. We see a fidelity decrease from $0.948^{+0.002}_{-0.002}$ for a chain consisting of two qubits (i.e., zero IQs) to $0.720^{+0.024}_{-0.013}$ for a chain consisting of ten qubits (i.e., eight IQs).

To mitigate the Bell state fidelity decrease, we insert DD sequences on idling qubits and compare their performance. The results shown in Fig. 7(b) for the CPMG sequence (green), the XY4 sequence (orange), and the LDD sequence (blue) suggest that the best performance is obtained for LDD where the corresponding optimal parameters are given in Table II.

APPENDIX C: ROBUSTNESS OF THE OPTIMAL ANGLES TO PERTURBATIONS

To study the robustness of the optimized angles, we perform additional MCM experiments on *ibm_kyiv*. We first optimize the angles θ, ϕ, λ in an LDD sequence for $r = 1$ as in Fig. 3. Next, we investigate the robustness by perturbing the angles in $R_{\vec{x}} = R(\theta, \phi, \lambda)$ and reporting the effect on the fidelity.

In particular, we uniformly sample three parameters $\Delta\theta$, $\Delta\phi$, and $\Delta\lambda$ from the interval $[-2\pi, 2\pi]$. We normalize the resulting vector $\vec{\delta} = [\Delta\theta, \Delta\phi, \Delta\lambda]$ to have a norm of ϵ that controls the magnitude of the perturbation, i.e.,

$$\vec{\Delta} = \epsilon \frac{\vec{\delta}}{||\vec{\delta}||} = [\Delta\theta', \Delta\phi', \Delta\lambda']. \tag{C1}$$

We perform experiments for the perturbed LDD gates,

$$R_{\vec{x}+\vec{\Delta}} = R(\theta + \Delta\theta', \phi + \Delta\phi', \lambda + \Delta\lambda'), \tag{C2}$$

as a function of the perturbation strength ϵ. The results are shown in Fig. 8 where each data point corresponds to the median of ten samples and the error bars show the upper and lower quartile range. For comparison, we show the fidelity found without perturbing the optimized angles as a gray line.

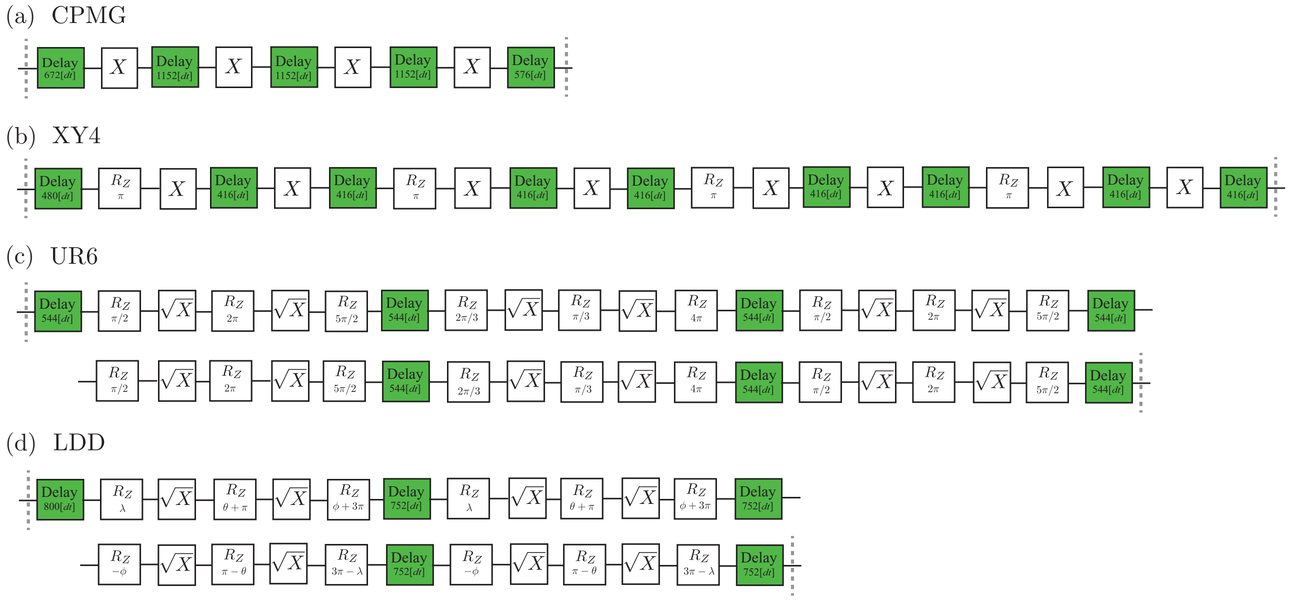
(a) CPMG
Delay 672[dt] X Delay 1152[dt] X Delay 1152[dt] X Delay 1152[dt] X Delay 576[dt]
(b) XY4
Delay 480[dt] R_Z π X Delay 416[dt] X Delay 416[dt] R_Z π X Delay 416[dt] X Delay 416[dt] R_Z π X Delay 416[dt] X Delay 416[dt] R_Z π X Delay 416[dt] X Delay 416[dt]
(c) UR6
Delay 544[dt] R_Z $\pi/2$ $\sqrt{X}$ R_Z 2π $\sqrt{X}$ R_Z $5\pi/2$ Delay 544[dt] R_Z $2\pi/3$ $\sqrt{X}$ R_Z $\pi/3$ $\sqrt{X}$ R_Z 4π Delay 544[dt] R_Z $\pi/2$ $\sqrt{X}$ R_Z 2π $\sqrt{X}$ R_Z $5\pi/2$ Delay 544[dt]
R_Z $\pi/2$ $\sqrt{X}$ R_Z 2π $\sqrt{X}$ R_Z $5\pi/2$ Delay 544[dt] R_Z $2\pi/3$ $\sqrt{X}$ R_Z $\pi/3$ $\sqrt{X}$ R_Z 4π Delay 544[dt] R_Z $\pi/2$ $\sqrt{X}$ R_Z 2π $\sqrt{X}$ R_Z $5\pi/2$ Delay 544[dt]
(d) LDD
Delay 800[dt] R_Z λ $\sqrt{X}$ R_Z $\theta+\pi$ $\sqrt{X}$ R_Z $\phi+3\pi$ Delay 752[dt] R_Z λ $\sqrt{X}$ R_Z $\theta+\pi$ $\sqrt{X}$ R_Z $\phi+3\pi$ Delay 752[dt]
R_Z $-\phi$ $\sqrt{X}$ R_Z $\pi-\theta$ $\sqrt{X}$ R_Z $3\pi-\lambda$ Delay 752[dt] R_Z $-\phi$ $\sqrt{X}$ R_Z $\pi-\theta$ $\sqrt{X}$ R_Z $3\pi-\lambda$ Delay 752[dt]

FIG. 9. Transpiled DD sequences. (a) CPMG, (b) XY4, (c) UR6, and (d) LDD sequences applied in an idle time of the same duration as a midcircuit measurement of duration of 1244 ns, i.e., 5600[dt], where $dt = 0.22$ ns. The hardware native gates are X, $\sqrt{X}$, and $R_Z(\theta)$.

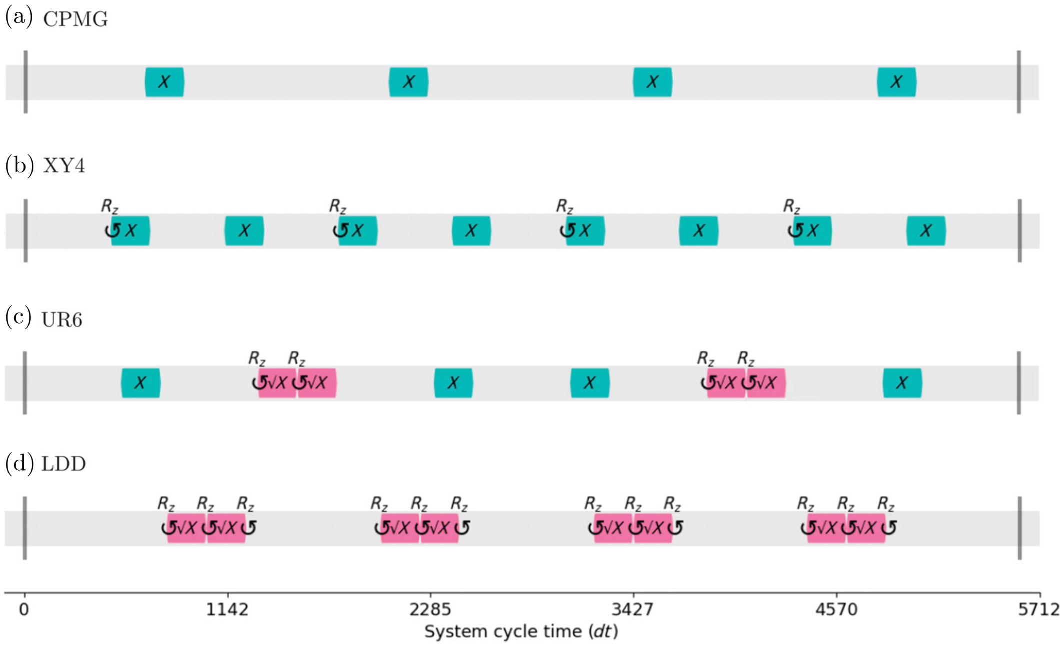

FIG. 10. Timeline drawing of the transpiled DD sequences. The (a) CPMG, (b) XY4, (c) UR6, and (d) LDD sequences are the same as those in Fig. 9 and plotted as a function of time in units of *dt* on the *x* axis. Here, the widths of the X and $\sqrt{X}$ gates, shown in green and red, respectively, match their durations. The $R_Z(\theta)$ gates, shown as curved arrows, are virtual and thus have a duration of zero.

APPENDIX D: DD TRANSPILED CIRCUITS

The circuits in this work are transpiled to the qubit coupling map and the hardware native gates of the quantum device. To insert a DD sequence in a circuit, we first schedule the circuit. That is, we attach to the hardware native circuit instructions for their corresponding durations and schedule them as late as possible. Next, idling times on the qubits that are long enough are replaced by delays and DD gates. The durations of the delays are determined by the total idle time and the DD sequence. In this appendix we exemplify the gate sequences resulting from this transpilation process for a delay of 1244 ns, which corresponds to the duration of a midcircuit measurement on *ibm_kyiv*; see also Sec. III B 1. We implement CPMG, XY4, UR6, and LDD sequences to fit this idling time. Figure 9 shows the transpiled circuits and Fig. 10 shows their timeline diagram with time measured in units of the system cycle time.

[1] S. Bravyi, A. W. Cross, J. M. Gambetta, D. Maslov, P. Rall, and T. J. Yoder, High-threshold and low-overhead fault-tolerant quantum memory, Nature **627**, 778 (2024).

[2] 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).

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

[4] 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).

[5] E. van den Berg, Z. K. Minev, A. Kandala, and K. Temme, Probabilistic error cancellation with sparse Pauli–Lindblad models on noisy quantum processors, Nat. Phys. **19**, 1116 (2023).

[6] L. Viola and S. Lloyd, Dynamical suppression of decoherence in two-state quantum systems, Phys. Rev. A **58**, 2733 (1998).

[7] N. Earnest, C. Tornow, and D. J. Egger, Pulse-efficient circuit transpilation for quantum applications on cross-resonance-based hardware, Phys. Rev. Res. **3**, 043088 (2021).

[8] U. Haeberlen and J. S. Waugh, Coherent averaging effects in magnetic resonance, Phys. Rev. **175**, 453 (1968).

[9] J. J. L. Morton, A. M. Tyryshkin, A. Ardavan, S. C. Benjamin, K. Porfyrakis, S. A. Lyon, and G. A. D. Briggs, Bang–bang control of fullerene qubits using ultrafast phase gates, Nat. Phys. **2**, 40 (2006).

[10] M. J. Biercuk, H. Uys, A. P. VanDevender, N. Shiga, W. M. Itano, and J. J. Bollinger, Optimized dynamical decoupling in a model quantum memory, Nature **458**, 996 (2009).

[11] J. Du, X. Rong, N. Zhao, Y. Wang, J. Yang, and R. Liu, Preserving electron spin coherence in solids by optimal dynamical decoupling, Nature **461**, 1265 (2009).

[12] N. Ezzell, B. Pokharel, L. Tewala, G. Quiroz, and D. A. Lidar, Dynamical decoupling for superconducting qubits: A performance survey, Phys. Rev. Appl. **20**, 064027 (2023).

[13] D. A. Lidar and T. A. Brun, *Quantum Error Correction* (Cambridge university press, Cambridge, 2013).

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

[15] A. Seif, H. Liao, V. Tripathi, K. Krsulich, M. Malekakhlagh, M. Amico, P. Jurcevic, and A. Javadi-Abhari, in *2024 ACM/IEEE 51st Annual International Symposium on Computer Architecture (ISCA)* (2024), p. 310.

[16] B. Pokharel, N. Anand, B. Fortman, and D. A. Lidar, Demonstration of fidelity improvement using dynamical decoupling with superconducting qubits, Phys. Rev. Lett. **121**, 220502 (2018).

[17] G. S. Uhrig, Keeping a quantum bit alive by optimized π-pulse sequences, Phys. Rev. Lett. **98**, 100504 (2007).

[18] G. Quiroz and D. A. Lidar, Optimized dynamical decoupling via genetic algorithms, Phys. Rev. A **88**, 052306 (2013).

[19] M. August and X. Ni, Using recurrent neural networks to optimize dynamical decoupling for quantum memory, Phys. Rev. A **95**, 012335 (2017).

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

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

[22] A. C. Vazquez, C. Tornow, D. Riste, S. Woerner, M. Takita, and D. J. Egger, Combining quantum processors with real-time classical communication, Nature (2024).

[23] P. Krantz, M. Kjaergaard, F. Yan, T. P. Orlando, S. Gustavsson, and W. D. Oliver, A quantum engineer's guide to superconducting qubits, Appl. Phys. Rev. **6**, 021318 (2019).

[24] D. J. Egger and F. K. Wilhelm, Adaptive hybrid optimal quantum control for imprecisely characterized systems, Phys. Rev. Lett. **112**, 240503 (2014).

[25] X.-d. Yang, C. Arenz, I. Pelczer, Q.-M. Chen, R.-B. Wu, X. Peng, and H. Rabitz, Assessing three closed-loop learning algorithms by searching for high-quality quantum control pulses, Phys. Rev. A **102**, 062605 (2020).

[26] Q.-M. Chen, X. Yang, C. Arenz, R.-B. Wu, X. Peng, I. Pelczer, and H. Rabitz, Combining the synergistic control capabilities of modeling and experiments: Illustration of finding a minimum-time quantum objective, Phys. Rev. A **101**, 032313 (2020).

[27] A. B. Magann, C. Arenz, M. D. Grace, T.-S. Ho, R. L. Kosut, J. R. McClean, H. A. Rabitz, and M. Sarovar, From pulses to circuits and back again: A quantum optimal control perspective on variational quantum algorithms, PRX Quantum **2**, 010101 (2021).

[28] J. Kelly *et al.*, Optimal quantum control using randomized benchmarking, Phys. Rev. Lett. **112**, 240504 (2014).

[29] M. Werninghaus, D. J. Egger, F. Roy, S. Machnes, F. K. Wilhelm, and S. Filipp, Leakage reduction in fast superconducting qubit gates via optimal control, Npj Quantum Inf. **7**, 14 (2021).

[30] C. Tong, H. Zhang, and B. Pokharel, Empirical learning of dynamical decoupling on quantum processors, arXiv:2403.02294 [quant-ph].

[31] L. Viola, E. Knill, and S. Lloyd, Dynamical decoupling of open quantum systems, Phys. Rev. Lett. **82**, 2417 (1999).

[32] M. J. Biercuk, H. Uys, A. P. VanDevender, N. Shiga, W. M. Itano, and J. J. Bollinger, Experimental Uhrig dynamical decoupling using trapped ions, Phys. Rev. A **79**, 062324 (2009).

[33] A. J. Berglund, Quantum coherence and control in one- and two-photon optical systems, arXiv:quant-ph/0010001 [quant-ph].

[34] E. L. Hahn, Spin echoes, Phys. Rev. **80**, 580 (1950).

[35] H. Y. Carr and E. M. Purcell, Effects of diffusion on free precession in nuclear magnetic resonance experiments, Phys. Rev. **94**, 630 (1954).

[36] S. Meiboom and D. Gill, Modified spin-echo method for measuring nuclear relaxation times, Rev. Sci. Instrum. **29**, 688 (1958).

[37] A. Maudsley, Modified Carr-Purcell-Meiboom-Gill sequence for NMR fourier imaging applications, J. Magn. Reson. **69**, 488 (1986).

[38] M. A. Ali Ahmed, G. A. Álvarez, and D. Suter, Robustness of dynamical decoupling sequences, Phys. Rev. A **87**, 042309 (2013).

[39] G. T. Genov, D. Schraft, N. V. Vitanov, and T. Halfmann, Arbitrarily accurate pulse sequences for robust dynamical decoupling, Phys. Rev. Lett. **118**, 133202 (2017).

[40] K. Khodjasteh and D. A. Lidar, Fault-tolerant quantum dynamical decoupling, Phys. Rev. Lett. **95**, 180501 (2005).

[41] S. Sheldon, E. Magesan, J. M. Chow, and J. M. Gambetta, Procedure for systematically tuning up cross-talk in the cross-resonance gate, Phys. Rev. A **93**, 060302 (2016).

[42] P. Das, S. Tannu, S. Dangwal, and M. Qureshi, in *MICRO-54: 54th Annual IEEE/ACM International Symposium on Microarchitecture*, MICRO '21 (Association for Computing Machinery, New York, NY, USA, 2021), p. 950.

[43] L. Cincio, K. Rudinger, M. Sarovar, and P. J. Coles, Machine learning of noise-resilient quantum circuits, PRX Quantum **2**, 010324 (2021).

[44] T. Proctor, K. Rudinger, K. Young, E. Nielsen, and R. Blume-Kohout, Measuring the capabilities of quantum computers, Nat. Phys. **18**, 75 (2022).

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

[46] S. H. Sack and D. J. Egger, Large-scale quantum approximate optimization on nonplanar graphs with machine learning noise mitigation, Phys. Rev. Res. **6**, 013223 (2024).

[47] D. C. McKay, C. J. Wood, S. Sheldon, J. M. Chow, and J. M. Gambetta, Efficient Z gates for quantum computing, Phys. Rev. A **96**, 022330 (2017).

[48] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, and L. Cincio *et al.*, Variational quantum algorithms, Nat. Rev. Phys. **3**, 625 (2021).

[49] D. Gottesman and I. L. Chuang, Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations, Nature **402**, 390 (1999).

[50] K. Vogel and H. Risken, Determination of quasiprobability distributions in terms of probability distributions for the rotated quadrature phase, Phys. Rev. A **40**, 2847 (1989).

[51] D. T. Smithey, M. Beck, M. G. Raymer, and A. Faridani, Measurement of the Wigner distribution and the density matrix of a light mode using optical homodyne tomography: Application to squeezed states and the vacuum, Phys. Rev. Lett. **70**, 1244 (1993).

[52] D. Leibfried, E. Knill, S. Seidelin, J. Britton, R. B. Blakestad, J. Chiaverini, D. B. Hume, W. M. Itano, J. D. Jost, and C. Langer *et al.*, Creation of a six-atom 'Schrödinger cat' state, Nature **438**, 639 (2005).

[53] T. J. Dunn, I. A. Walmsley, and S. Mukamel, Experimental determination of the quantum-mechanical state of a molecular vibrational mode using fluorescence tomography, Phys. Rev. Lett. **74**, 884 (1995).

[54] Z. Hradil, Quantum-state estimation, Phys. Rev. A **55**, R1561 (1997).

[55] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, Measurement of qubits, Phys. Rev. A **64**, 052312 (2001).

[56] J. A. Smolin, J. M. Gambetta, and G. Smith, Efficient method for computing the maximum-likelihood quantum state from measurements with additive Gaussian noise, Phys. Rev. Lett. **108**, 070502 (2012).

[57] J. Spall, Multivariate stochastic approximation using a simultaneous perturbation gradient approximation, IEEE Trans. Autom. Control **37**, 332 (1992).

[58] J. Spall, in *Proc. 36th IEEE Conf. Decis. Control* (IEEE, 1997), Vol. 2, p. 1417.

[59] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, Nature **549**, 242 (2017).

[60] B. M. Terhal, Quantum error correction for quantum memories, Rev. Mod. Phys. **87**, 307 (2015).

[61] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels, Phys. Rev. Lett. **70**, 1895 (1993).

[62] E. Bäumer, V. Tripathi, D. S. Wang, P. Rall, E. H. Chen, S. Majumder, A. Seif, and Z. K. Minev, Efficient long-range entanglement using dynamic circuits, PRX Quantum **5**, 030339 (2024).

[63] J. M. Koh, S.-N. Sun, M. Motta, and A. J. Minnich, Measurement-induced entanglement phase transition on a superconducting quantum processor with mid-circuit readout, Nat. Phys. **19**, 1314 (2023).

[64] R. S. Gupta, E. van den Berg, M. Takita, D. Ristè, K. Temme, and A. Kandala, Probabilistic error cancellation for dynamic quantum circuits, Phys. Rev. A **109**, 062617 (2024).

[65] A. Tranter, P. J. Love, F. Mintert, and P. V. Coveney, A comparison of the Bravyi–Kitaev and Jordan–Wigner transformations for the quantum simulation of quantum chemistry, J. Chem. Theory Comput. **14**, 5617 (2018).

[66] G. Li, A. Wu, Y. Shi, A. Javadi-Abhari, Y. Ding, and Y. Xie, in *Proceedings of the 27th ACM International Conference on Architectural Support for Programming Languages and Operating Systems*, ASPLOS '22 (Association for Computing Machinery, New York, NY, USA, 2022), p. 554.

[67] A. Vezvaee, V. Tripathi, D. Kowsari, E. Levenson-Falk, and D. A. Lidar, Virtual Z gates and symmetric gate compilation, arXiv:2407.14782 [quant-ph].

[68] The numbers are as reported from the backends. Furthermore, *ibm _torino* implements two-qubit controlled-*Z* gates with tunable couplers, while *ibm _sherbrooke* uses the cross-resonance interaction.

[69] D. J. Egger, C. Capecci, B. Pokharel, P. K. Barkoutsos, L. E. Fischer, L. Guidoni, and I. Tavernelli, Pulse variational quantum eigensolver on cross-resonance-based hardware, Phys. Rev. Res. **5**, 033159 (2023).