Dynamically corrected gates suppressing spatiotemporal error correlations as measured by randomized benchmarking

C. L. Edmunds [1,2,*] C. Hempel [1] R. J. Harris,[3] V. Frey,[1,2] T. M. Stace,[3] and M. J. Biercuk[1,2,*]

[1]*ARC Centre for Engineered Quantum Systems, School of Physics, The University of Sydney, New South Wales 2006, Australia*
[2]*Q-CTRL Pty Ltd, Sydney, New South Wales 2000, Australia*
[3]*ARC Centre for Engineered Quantum Systems, School of Physics and Mathematics, The University of Queensland, St Lucia, Queensland 4072, Australia*

Quantum error correction provides a path to large-scale quantum computers, but is built on challenging assumptions about the characteristics of the underlying errors. In particular, the mathematical assumption of statistically independent errors in quantum logic operations is at odds with realistic environments where error sources may exhibit strong temporal and spatial correlations. We present experiments using trapped ions to demonstrate that the use of dynamically corrected gates (DCGs), generally considered for the reduction of error magnitudes, can also suppress error correlations in space and time throughout quantum circuits. We present a first-principles analysis of the manifestation of error correlations in randomized benchmarking and validate this model through experiments performed using engineered errors. We find that standard DCGs can reduce error correlations by $\sim 50\times$ while increasing the magnitude of uncorrelated errors by a factor scaling linearly with the extended DCG duration compared to a primitive gate. We then demonstrate that the correlation characteristics of intrinsic errors in our system are modified by the use of DCGs, consistent with a picture in which DCGs whiten the effective error spectrum induced by external noise.

I. INTRODUCTION

Suppressing and correcting errors in quantum circuits is a critical challenge driving a substantial fraction of research in the quantum information science community. These efforts build on quantum error correction (QEC) and the theory of fault tolerance [1–6] as fundamental developments that support the concept of large-scale quantum computation [7–9]. In combination, these theoretical constructs suggest that so long as the probability of error in each physical quantum information carrier can be reduced below a threshold value, a properly executed QEC protocol can detect and suppress logical errors to arbitrarily low levels, and hence enable arbitrarily large computations. Underlying this proposition is an assumption that errors are statistically independent, i.e., the emergence of a qubit error at a specific time is uncorrelated with errors arising in other qubits or at any other time in the computation. Error correlations that decay with distance between qubits (spatially) can induce simultaneous multiqubit errors [10], and correlations that decay with circuit length (temporally) have

been shown to produce more rapid accumulation of net circuit errors [11,12].

The practicality of the assumption of uncorrelated errors has long been questioned, as laboratory sources of noise commonly exhibit strong temporal correlations, captured through spectral measures exhibiting high weight at low frequencies. As such, coherent errors induced by low-frequency noise and miscalibrations have recently become a larger focus of research, with their detrimental effects on QEC implementations being examined [11,13–15] and first ideas targeting their suppression emerging [16,17]. Attempts to address these errors in the theory of quantum error correction are challenging and results to date suggest that revision of postulated fault-tolerance thresholds may be required [18,19] relative to more optimistic predictions that have recently emerged [20]. Indeed, when implicit assumptions that errors are both spatially and temporally uncorrelated are weakened, the value of a tolerable error threshold can change from some value ε to ε^2, easily leading to order-of-magnitude decreases in the acceptable error rates [7].

The adverse effect of correlated errors on error correction procedures has been observed in the context of a repetition code both experimentally [21] (where they were seen to effectively negate any advantage obtained from iterative error correction) and theoretically [13], where an increase in the logical failure rate was identified. Furthermore, while a recent full-scale numerical simulation has shown that coherent errors at the physical layer can, in fact be overcome by topological error correcting codes [22], large numbers of physical qubits are required with error rates that are *uniformly* subthreshold.

---

*Authors to whom correspondence should be addressed: edmunds.claire@gmail.com; michael.biercuk@sydney.edu.au

The emerging message is that, while correlated errors do not invalidate the use of QEC, their presence can significantly increase the requisite overhead and may reduce the tolerable magnitude of physical qubit errors.

In this paper we demonstrate experimentally that using a low-level abstraction known as a dynamically corrected gate (DCG), we can suppress error correlations in addition to error magnitudes. Replacing "primitive" physical quantum gate operations with logically equivalent DCGs [23–27] forms a "virtual" layer wherein error characteristics can be modified (virtualized) before the application of QEC [28,29]. We present a first-principles analysis of Clifford randomized benchmarking [30,31] in order to quantitatively model the impact of error correlations on simple experimental observables, building on concepts in Ref. [32]. Specifically, we identify that error correlations are manifested in the scaling of the distribution over sequence randomizations, at fixed sequence length, with measurement averaging. We validate this framework using randomized benchmarking experiments performed with a single trapped ytterbium ion. We then demonstrate that the replacement of the individual Clifford operations within each sequence with logically equivalent DCGs modifies the error correlation signatures such that they are experimentally consistent with the presence of uncorrelated errors. Single-qubit experiments performed under engineered noise with tunable correlation characteristics show consistent reduction in the correlated error component when switching from primitive to DCG sequences. We explain this behavior using a framework that describes the action of DCGs at the operator level [27,33,34] as whitening the effective error spectrum experienced by each gate. Finally, we demonstrate that using DCGs in sequence construction reduces spatial error correlations between qubits, through simultaneous randomized benchmarking on five trapped ion qubits. These results provide direct and strong evidence that the use of dynamically protected physical qubit operations in a layered architecture for quantum computing [29] can facilitate the successful application of existing QEC theory with only minimal revision on the path to fault-tolerant quantum computation.

II. IDENTIFYING SIGNATURES OF ERROR CORRELATIONS IN CIRCUITS

We begin by laying out the challenge of establishing clear quantitative metrics allowing the identification of error correlations in quantum circuits. The types of correlated errors treated here concern unitary coherent errors that can be represented by additive environmental coupling terms, e.g., dephasing noise, or fluctuating control field terms in the system Hamiltonian. In general, such error channels exhibit strong temporal correlations when they are induced by colored noise spectra or systematic drifts in control parameters due to insufficiently precise or too infrequent calibrations. Correlated errors between gates can also be caused by incoherent error channels when considering, for example, changes in T_1 times, which lead to slow variations of fundamental gate error rates. We do not consider the latter form of correlated errors here; however, it is worth noting that they can also lead to detrimental and non-Markovian behavior in gate performance.

As a first step we analyze how correlations in a physical noise process translate to correlations in the resultant unitary errors within a circuit of $j = 1, \ldots, J$ gates. In our model, any noisy operation $\tilde{U}_j$ within the circuit can be decomposed into the ideal operator $\hat{U}_j$ and an error operator $\hat{\Lambda}_j$ such that $\tilde{U}_j = \hat{\Lambda}_j \hat{U}_j$. Here $\hat{U}_j \equiv \hat{U}(\boldsymbol{n}_j, \theta_j)$ rotates the state vector by angle θ_j around an arbitrary axis $\boldsymbol{n}_j$ on the Bloch sphere. Considering unitary semiclassical noise processes, the error component in each operation can be written as $\hat{\Lambda}_j = \exp\{i \sum_{\alpha=1}^{\infty} [\boldsymbol{\epsilon}_j]_\alpha \cdot \hat{\boldsymbol{\sigma}}\}$, with $\hat{\boldsymbol{\sigma}}$ the vector of Pauli matrices, α an index denoting the Magnus expansion order [35], and $\boldsymbol{\epsilon}_j$ the error vector characterizing the strength and nature (affected quadrature) of the error [34–37]. A quantum circuit experiences temporally correlated errors if the values of $\boldsymbol{\epsilon}_j$ across the circuit (in space or time) exhibit nonzero correlations.

Our approach to measuring error correlations is built on common quantum verification protocols employed to infer the average behavior of gate operations [11,30,38–47]. Restricting our analysis to the single-qubit case, error correlations between gates may occur in these protocols when physical noise processes exhibit strong correlations in time. We demonstrate this numerically by calculating the error vector $\boldsymbol{\epsilon}_j$ for each operation in a single-qubit randomized benchmarking sequence exposed to detuning ($\hat{\sigma}_z$) noise with a variable block-correlation length $\mathcal{M}_n$; this is defined to be the number of gates over which the noise strength is constant within the sequence. The sequence is assembled from the 24 Clifford operations comprising combinations of π and $\pi/2$ rotations about the x, y, and z axes of the Bloch sphere and an identity gate $\hat{\mathbb{I}}$. Calculating the autocorrelation function of the error vector's magnitude throughout a sequence reveals strong correlations over a length of gates $\mathcal{M}_\varepsilon$ that appear to scale linearly with the correlation length of the input noise process $\mathcal{M}_n$ [Fig. 1(c)]. This behavior suggests a linear mapping from noise correlations to error correlations in conventional settings. As a prelude to future demonstrations in this paper, we note that if the individual Clifford gates are replaced by DCGs, this simple linear mapping from input noise correlations to output error correlations breaks down.

In general, the primary limitation one faces in accessing information about $\mathcal{M}_\varepsilon$ in a physical experiment is that using standard projective measurements at the end of a circuit will limit the ability to probe correlations that arise throughout the circuit's execution. Most experimental quantum verification routines suffer from exactly this limitation, and primarily measure the average difference between a qubit state transformed under an imperfect operation and a predetermined target state at the end of the protocol [Fig. 1(a)]. However, as we will illustrate in the following, there is additional useful information present in the outcomes of randomized benchmarking measurement routines that may be employed to extract insights into error correlations appearing during the sequence.

The key underlying concept is that in a randomized benchmarking sequence built up from many operations, the resultant net state transformation in the presence of noise $\tilde{U}_{eff} |\psi\rangle$ [Fig. 1(b)] is determined by an interplay of both the sensitivity of each individual operation to the noise [35] and the impact of the sequence structure on error accumulation [32,46,48]. Specifically, nominally equivalent randomized benchmarking
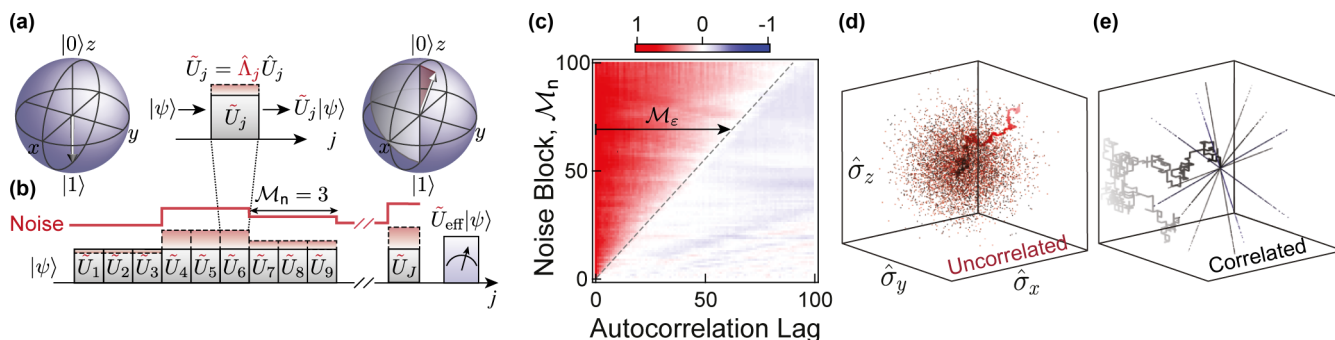
FIG. 1. Translation of noise correlations to error correlations in quantum circuits. (a) A single operation applied to a qubit in the presence of noise $\tilde{U}_j$ can be decomposed into an error operator $\hat{\Lambda}_j$ and the target operation $\hat{U}_j$. Bloch spheres schematically illustrate the effect of an imperfect π rotation about the x axis acting on input state $|1\rangle$, with dark shading indicating an overrotation error. (b) Noise (red line) exhibiting nonzero temporal correlation of length $\mathcal{M}_{\mathrm{n}} = 3$, quantized in units of gate operations, acts on a quantum circuit composed of sequentially applied unitary operations. The resultant errors accumulate and lead to a noisy effective operator $\tilde{U}_{\mathrm{eff}}$, whose effect is determined through a projective measurement at the end of the circuit. (c) Translation of correlations in a noise process to correlations in the magnitude of the circuit error vector $\|\boldsymbol{\epsilon}_j\|$. The error vector for each gate of a randomly composed sequence of 1000 primitive gates under a noise process with noise correlation length $\mathcal{M}_{\mathrm{n}}$ is calculated and the autocorrelation function of the magnitude of the error vector $\mathbb{E}[\|\boldsymbol{\epsilon}_{j_1}\|\|\boldsymbol{\epsilon}_{j_2}\|]$ is shown for the first 100 gates. Random walks are shown for the extreme error correlation cases (d) $\mathcal{M}_{\varepsilon} = 1$ (uncorrelated) and (e) $\mathcal{M}_{\varepsilon} = J$ (fully correlated). Final walk displacements of eight sequences, each with 1000 error realizations, are shown along with the full walk for a single sequence that is common between the two cases.

sequences (constructed to perform the same net operation) exhibit variations in correlated-noise susceptibility that are analytically calculable and verifiable in experiments. We use this variability and the behavior under experimental averaging to extract a signature of error correlations within quantum sequences.

A. Random-walk formalism for error accumulation

We present a first-principles analysis to directly link measurement outcomes for single-qubit randomized benchmarking sequences to the nature of the underlying error correlations quantified by $\mathcal{M}_{\varepsilon}$, expanding the formalism introduced in Ref. [32]. We consider randomized benchmarking sequences composed of J single-qubit Clifford operations $\prod_{j=1}^{J} \hat{C}_{\eta_j} = \hat{\mathbb{I}}$, with the vector $\boldsymbol{\eta}$ containing labels for the 24 Clifford operations, $\eta_j \in \{1, 2, \ldots, 24\}$. A final gate is precalculated to yield a net identity operation for the sequence such that in the absence of error the final qubit state will be the same as the initial state. Due to imperfections in the operations, the physically implemented gates $\tilde{C}_{\eta_j}$ differ from the ideal gates by an error map $\tilde{C}_{\eta_j} = \hat{\Lambda}_j \hat{C}_{\eta_j}$.

The accumulation of errors throughout a sequence can be represented by a sequence-dependent random walk in three-dimensional Pauli-error space; the net walk length can then be related to the final sequence error [32]. For a particular realization of the error i, this walk is captured by the vector

$$\mathbf{R}_{3\mathrm{D}}^{(i)} = \sum_{j=1}^{J} \varepsilon_j^{(i)} \mathbf{r}_{3\mathrm{D}, j}, \qquad (1)$$

with gate error values $\varepsilon_j^{(i)} \sim \mathcal{N}(0, \sigma^2)$ sampled from a zero-mean Gaussian distribution with rms value σ. It will be shown in Sec. II C that this leads to an average randomized benchmarking error per gate proportional to σ^2. Here the values of $\mathbf{r}_{3\mathrm{D}, j}$ are unit-length vectors that define the sequence-specific

random-walk steps; they can be calculated deterministically for any randomized benchmarking sequence, irrespective of the strength or correlation characteristics of the gate errors. In a circumstance where the normalized error takes a consistent value $\varepsilon_j^{(i)} \equiv 1$, the length of the J-step walk created by these steps is an intrinsic property of the sequence and will be shown to act as a proxy for its susceptibility to correlated errors. Examining individual randomized benchmarking sequences reveals the idiosyncratic nature of their walks; certain randomizations exhibit long walks, while others have walks that terminate near the origin, solely determined by the structure of the sequence and the form of the error channel. Accordingly, in the presence of correlated errors we expect a wide variance of outcomes, determined by the underlying structures of the randomly selected sequences. The general framework linking this Pauli walk to accumulated error was experimentally validated in Ref. [46].

B. Signatures of error correlations

We identify that the key measurable signature of error correlations arises in the process of experimental averaging over repetitions of a sequence, and hence over different realizations of the error. In order to understand this, we begin by examining how error correlations impact the random walk introduced above and how the behavior of that walk changes with experimental averaging.

Gate errors induce the mapping $\mathbf{r}_{3\mathrm{D}, j} \rightarrow \mathbf{R}_{3\mathrm{D}}^{(i)}$; the term $\varepsilon_j^{(i)}$ in Eq. (1) can change the direction and scale the magnitude of each step in the random walk. Thus correlations in $\varepsilon_j^{(i)}$ are translated into correlated modifications of the steps in $\mathbf{R}_{3\mathrm{D}}^{(i)}$. To see the effect of correlations in the error process, we calculate the locus of walk termination points for eight different sequences and 1000 error realizations, shown in Figs. 1(d) and 1(e). In the presence of errors whose magnitudes are constant across all gates in a given benchmarking sequence, the error

$\varepsilon_j^{(i)} \equiv \varepsilon^{(i)}$ rescales all steps in the walk uniformly such that all termination points for a given sequence fall on a line [Fig. 1(e)]. The walk terminations for the same sequence are thus dominated by the underlying sequence structure [rays in Fig. 1(e)]. By contrast, in the presence of uncorrelated errors where $\varepsilon_j^{(i)}$ changes randomly for each step, the termination points appear randomly distributed in Pauli space for different realizations of the error [Fig. 1(d)].

These differences will manifest in an experiment that averages the experimental performance of a set of sequences over many different realizations of an error-inducing noise process. In the case of correlated errors, the preservation of sequence-structure dependence in the sequence error leads to a broad distribution of outcomes over different randomized benchmarking sequences. This breadth is maintained even when averaging experiments together over various realizations of the random but temporally correlated errors. In contrast, for uncorrelated errors, the random formless distribution of walk termination points over the same set of sequences implies that averaging over experiments would result in a spread of outcomes that grows narrower as the experiment number increases, consistent with the central limit theorem. It is therefore in the distribution over measured results of noise-averaged randomized benchmarking sequences that the signatures of error correlations between gates within a sequence will appear. In Secs. II C and II D we will describe how this phenomenology can be accessed through a modified analysis of conventional randomized benchmarking experiments.

C. Mapping to measurable quantities

We now link the random-walk framework to measurements commonly performed in the laboratory, a single projective measurement in the qubit basis. Such measurements are unaffected by rotations about the z axis, i.e., they are phase invariant. Consequently, this type of projective measurement is insensitive to the component of the random walk oriented along the $\hat{\sigma}_z$ axis and instead probes a two-dimensional projection of the walk onto the $\hat{\sigma}_x \hat{\sigma}_y$ plane of Pauli-error space [46]. Considering a measurement routine involving averaging a single sequence over n realizations of the error, we may relate the two-dimensional walk length to the projective measurement results as

$$\mathcal{P} = 1 - \langle \|\mathbf{R}_{2D}\|^2 \rangle_n + O(\sigma^4), \tag{2}$$

where $\langle \cdot \rangle_n$ is an average over n instances of the error process, $\mathcal{P} := 1 - \langle P(|1\rangle) \rangle_n$ is the measurable noise-averaged sequence survival probability when the qubit is initialized in the state $|0\rangle$, σ is the rms of the normally distributed errors, and $\mathbf{R}_{2D}$ denotes the random walk in the $\hat{\sigma}_x \hat{\sigma}_y$ plane of Pauli-error space. For simplicity, we will proceed by referring to $\mathbf{R}_{2D}$ and its individual steps $\mathbf{r}_{2D,j}$ simply as $\mathbf{R}$ and $\mathbf{r}_j$, respectively.

We analyze in detail three distinct error correlation regimes for a unitary error channel with values $\varepsilon_j^{(i)} \sim \mathcal{N}(0, \sigma^2)$: (i) $\mathcal{M}_\varepsilon = J$, identically correlated errors with fixed, constant magnitude over a sequence and rms value σ_C; (ii) $\mathcal{M}_\varepsilon = 1$, uncorrelated, normally distributed errors that change randomly between each gate in a sequence with rms value σ_U; and (iii) statistically independent, contemporaneous

correlated and uncorrelated error processes such that the relative strengths σ_C and σ_U determine the effective error correlation length.

The expression for survival probability in Eq. (2) can be used to calculate the distribution of survival probabilities without modification for both regimes (i) and (ii) simply by using the appropriately calculated random walks. In the limit of long sequences and many noise averages (large J and n), the noise-averaged survival probability is Γ distributed over different, nominally equivalent, sequence randomizations [46]; the shape and scale parameters of the distribution, a and b, respectively, can be calculated from first principles using the particulars of the sequence, noise averaging, and error characteristics. For these two limiting cases of identically correlated errors over a sequence and uncorrelated errors changing randomly between gates, the respective survival probabilities are sampled from Γ distributions shaped according to

$$\mathcal{P}_C \sim \Gamma\left(a = 1, b = \frac{2}{3}J\sigma^2\right), \tag{3a}$$

$$\mathcal{P}_U \sim \Gamma\left(a = n, b = \frac{2}{3n}J\sigma^2\right). \tag{3b}$$

From these expressions, the variance and expectation values of the distribution over sequence randomizations can be calculated. To leading order, both distributions exhibit the same mean value $\mathbb{E} = ab$, giving a randomized benchmarking average gate error of $\frac{2}{3}\sigma^2$. However, the distributions diverge in the second moment $\mathbb{V} = ab^2$.

We may now derive the properties of the distribution associated with regime (iii) by considering two independent walks; one is induced by the correlated error component $\mathbf{R}_C^{(i)}$ and the other by the uncorrelated component $\mathbf{R}_U^{(i)}$. To begin, it is convenient to note that in the case of a correlated fixed error process over a sequence, it is possible to factor out the constant error strength from the random walk for a particular realization of the error [32],

$$\mathbf{R}_C^{(i)} = \varepsilon_C^{(i)} \sum_{j=1}^{J} \mathbf{r}_j = \varepsilon_C^{(i)} \mathbf{V}. \tag{4}$$

We thus introduce $\mathbf{V}$ to describe the sequence-specific walk, defined by the steps $\mathbf{r}_j$ that remain invariant under different realizations of the error process [Fig. 1(e)]. This separability is not achievable in the presence of uncorrelated errors due to the randomization of each step in the walk by the error process. The expression for survival probability can then be expanded in terms of these independent walks to second order in σ_C and σ_U as

$$\mathcal{P} = 1 - \langle \|\mathbf{R}_U^{(i)} + \varepsilon_C^{(i)} \mathbf{V}\|^2 \rangle_n$$
$$= 1 - \langle \|\mathbf{R}_U^{(i)}\|^2 \rangle_n - \sigma_C^2 \|\mathbf{V}\|^2, \tag{5}$$

where the cross term is identically zero using $\langle \varepsilon_C^{(i)} \rangle_n = 0$.

For all three correlation regimes, higher-order terms and cross terms contribute to the second moment of the distribution and have been calculated analytically (Table I). These terms reduce to those calculated using the Γ distributions in Eq. (3) in the limit of large J and n, with $J \gg n$. On inspection, we expect that in the presence of uncorrelated errors the variance will narrow with increasing n, while it

TABLE I. Statistical moments for the distribution of noise-averaged sequence survival probabilities with different error correlation lengths, (i) fully correlated across the sequence, (ii) completely uncorrelated values between gates, and (iii) a combination of two independent error processes in the same quadrature, one correlated and one uncorrelated. The variance for case (iii) incorporates contributions from each error source individually, $\mathbb{V}[\mathcal{P}_C]$ and $\mathbb{V}[\mathcal{P}_U]$, as well as a cross term.

Error type	$1 - \mathbb{E}[\mathcal{P}]$	$\mathbb{V}[\mathcal{P}]$
(i) fully correlated $\mathcal{M}_\varepsilon = J$	$\frac{2}{3}J\sigma_C^2$	$\frac{2}{9}\frac{n+2}{n}J(2J-1)\sigma_C^4$
(ii) uncorrelated $\mathcal{M}_\varepsilon = 1$	$\frac{2}{3}J\sigma_U^2$	$\frac{2}{9n}J(4+2J+n)\sigma_U^4$
(iii) correlated +uncorrelated	$\frac{2}{3}J(\sigma_U^2 + \sigma_C^2)$	$\mathbb{V}[\mathcal{P}_U] + \mathbb{V}[\mathcal{P}_C]$ $+\frac{4}{9}J\sigma_C^2\sigma_U^2$

will remain fixed in the presence of correlated errors. Such differences in scaling of a variance measure with averaging are reminiscent of the manifestation of noise correlations in other physical quantities, e.g., the Allan variance used in precision frequency metrology [49,50]. Our analysis therefore highlights that calculating the variance of measurements of randomized benchmarking survival probabilities for different sequences and exploring how this variance changes with experimental averaging can give insights into the underlying error correlations. The functional dependence of the distribution variance with n will be employed throughout the remainder of this work as a key signature of error correlations in standard randomized benchmarking.

In the next section we demonstrate how the model can be updated to connect to realistic laboratory noise models.

D. Modeling realistic laboratory error models

Building on the general framework introduced above, we introduce first-principles calculations connecting the theoretical model for gate *error* with actual error-inducing *noise* in experiments. We determine the sequence walk in the presence of arbitrary unitary error maps, incorporating the possibility of multiaxis and gate-dependent errors. This facilitates the analysis of experimental measurements performed subject to the most common noise sources encountered in the laboratory.

We consider two physically motivated noise processes that can occur throughout a randomized benchmarking sequence. First, frequency detuning noise, on either the qubit's resonant frequency or the frequency of the control field used to drive qubit gate operations, creates an off-resonance error between the qubit and control. Second, amplitude noise, which may arise from coupling-strength variations or drifts and miscalibrations in the control, results in an over- or underrotation error of the qubit state vector. Both of these represent concurrent noise sources, i.e., applied simultaneously with the execution of a gate, which ultimately produce complex gate-dependent errors.

In general, depending on their underlying cause, both frequency detuning and amplitude noise processes may possess temporally correlated and uncorrelated components. Correlated noise sources include miscalibrations, magnetic-field drifts, and temperature drifts in control systems, while uncorrelated noise often stems from electrical noise or local environmental sources, e.g., anomalous heating in ion traps [51] or two-level system fluctuators in superconducting qubits [52,53].

To now examine the impact of these physical noise processes on the behavior of the sequence survival-probability distributions, we proceed by explicitly calculating the translation between the physical noise strength $\delta_j^{(i)} \sim \mathcal{N}(0, \rho^2)$ and the effective sequence errors at the core of our model $\varepsilon = \varepsilon(\delta)$. In our notation, ρ is used to denote the rms magnitude of the noise, distinguishing it from the rms magnitude of the error operator σ. Our calculations incorporate the fact that single-axis noise, e.g., detuning, present during a noncommuting operation generally results in a multiaxis error process. Furthermore, physical implementations of Clifford operations typically employ variable gate durations, resulting in gate-dependent error operators.

In this setting, the error $\varepsilon_j^{(i)}$ employed in Eq. (1) is replaced by the physical noise strength $\delta_j^{(i)}$. As a result, the previously unit-length steps $\mathbf{r}_{3D,j}$ now take more complex, but still analytically calculable, values due to the gate dependence and multiaxis character of the errors induced by concurrent noise processes. For a particular noise process we calculate the associated random walk, which enables a mapping of the rms magnitude of the physical noise ρ to an updated rms value of the error σ. Appendix B describes the formalism to calculate the noise-to-error translation in standard Clifford gates for an arbitrary unitary error process. Table II summarizes the results which, when combined with the expressions from Table I, can be used to predict both the expectation and the variance of the distribution of survival probabilities over sequence randomization.

III. EXPERIMENTAL IMPLEMENTATION

A. Randomized benchmarking on ^{171}Yb$^+$ qubits

We perform experiments using a qubit encoded in the $^2S_{1/2}$ hyperfine ground states of a single laser-cooled ^{171}Yb$^+$ ion confined in a linear Paul trap, with the computational basis states defined as $|0\rangle := |F = 0, m_F = 0\rangle$ and $|1\rangle := |F = 1, m_F = 0\rangle$. Laser cooling, state initialization to $|0\rangle$, and detection are performed using a laser at 369 nm that couples the $^2S_{1/2} |F = 1\rangle$ ground state to the first excited state $^2P_{1/2} |F = 0\rangle$. As the ion selectively fluoresces when it is projected to the upper bright qubit state $|1\rangle$, one can distinguish between the two basis states by counting the number of emitted photons during the detection period. Single-ion qubit state detection is performed in a time-resolved manner [46,54] using an avalanche photodiode; multi-ion data employ an electron multiplying CCD (EMCCD) camera and processing through a random forest classifier from the SCIKIT-LEARN framework [55].

Qubit rotations are driven via a microwave field near 12.6 GHz generated by a vector signal generator (VSG). Using an internal baseband generator, we program arbitrary rotations of the qubit via *IQ* modulation. Rotations about the *z* axis are implemented as instantaneous, precalculated *IQ* frame shifts. Randomized benchmarking sequences composed from Clifford operations are preloaded into the VSG

TABLE II. Translation from the rms value of a physical noise process ρ, with correlation length $\mathcal{M}_n$, to the rms value of the gate error σ used to calculate the first and second moments of noise-averaged sequence survival probabilities. The values ρ_C and ρ_U represent the rms magnitudes of the correlated and uncorrelated noise processes, respectively. Similarly, the terms $\mathbf{r}_{U,j}$ and $\mathbf{r}_{C,j}$ represent the random-walk steps for the different noise processes. Full details of the derivation of the relevant random-walk step expectation values $\mathbb{E}[\|\mathbf{r}_j\|^2]$, $\mathbb{E}[\|\mathbf{r}_j\|^4]$, and $\text{Cov}(\|\mathbf{r}_{U,j}\|^2, \|\mathbf{r}_{C,j}\|^2)$ for the specific noise models employed in our verification experiments are presented in Appendix B 1.

Error type	$\rho \rightarrow \sigma$ translation for $\mathbb{E}$	$\rho \rightarrow \sigma$ translation for $\mathbb{V}$
(i) fully correlated $\mathcal{M}_n = J$	$\sigma_C^2 = \frac{3}{2}\mathbb{E}[\|\mathbf{r}_j\|^2]\rho_C^2$	$\sigma_C^4 = \frac{9}{2}\frac{\mathbb{E}[\|\mathbf{r}_j\|^4]+(J-2)\mathbb{E}[\|\mathbf{r}_j\|^2]^2}{2J-1}\rho_C^4$
(ii) uncorrelated $\mathcal{M}_n = 1$	$\sigma_U^2 = \frac{3}{2}\mathbb{E}[\|\mathbf{r}_j\|^2]\rho_U^2$	$\sigma_U^4 = \frac{9}{2}\frac{(2+n)\mathbb{E}[\|\mathbf{r}_j\|^4]+(J-1-n)\mathbb{E}[\|\mathbf{r}_j\|^2]^2}{4+2J+n}\rho_U^4$
(iii) correlated +uncorrelated	$(\sigma_C^2 + \sigma_U^2) = \frac{3}{2}(\mathbb{E}[\|\mathbf{r}_{U,j}\|^2]\rho_U^2 + \mathbb{E}[\|\mathbf{r}_{C,j}\|^2]\rho_C^2)$	$\sigma_C^2\sigma_U^2 = \frac{9}{2}\text{Cov}(\|\mathbf{r}_{U,j}\|^2, \|\mathbf{r}_{C,j}\|^2)\rho_C^2\rho_U^2$

and mapped to the desired physical operations prior to the recording of each data set. The experiments in this paper are performed using k sequences, each comprising J operations. The first $J-1$ gates are randomly composed Clifford operations $\hat{C}_{\eta_j}$ and the final operation $\hat{C}_{\eta_J} = (\prod_{j=1}^{J-1}\hat{C}_{\eta_j})^\dagger$ is selected such that the sequence implements the identity in the absence of error. A full list of the Clifford operations and their physical implementations can be found in the Supplemental Material of Ref. [32]. Typical single-qubit randomized benchmarking experiments with primitive gates achieve a baseline result of $p_{\text{RB}} \approx 1.9 \times 10^{-5}$ in our system (Appendix A).

B. Verifying error correlation signatures with engineered errors

The key signature of the presence of temporally correlated errors appears in the variance of the distribution over sequence survival probabilities and its scaling with experimental averaging; averaging reduces the variance in the case of uncorrelated errors, but has limited impact when errors exhibit strong temporal correlations.

We begin our experimental study by engineering experimental noise sources to test and verify the predictions of the theoretical model presented in Sec. II. We perform standard randomized benchmarking, but engineer detuning and control-amplitude noise with different user-defined bandwidths. All noise values are generated numerically and are sampled from a zero-mean Gaussian distribution $\mathcal{N}(0, \rho^2)$ with rms strength ρ. Off-resonance errors are induced via fractional detuning noise present during the application of the randomized benchmarking sequence $\delta = (\Delta/\Omega)$ set by the frequency detuning Δ between the qubit transition and the microwave source in units of the Rabi frequency Ω. Overrotation errors are produced by amplitude noise in the microwave control field, effectively changing Ω. Two limiting noise bandwidths are treated: maximally correlated noise $\mathcal{M}_n = J$ and uncorrelated noise $\mathcal{M}_n \leqslant 1$. For the detuning (control-amplitude) noise process, the correlated noise component is engineered using a constant offset in the VSG microwave frequency (amplitude) over the entire sequence and the uncorrelated noise is applied via an external frequency (amplitude) modulation input and changes value every primitive $\pi/2$ time. The relevant random-walk steps calculated for these noise processes and used in modeling our experimental measurements are found in Table III of Appendix B.

Instead of simply calculating the randomized benchmarking decay rate p_{RB} derived from fitting to the mean of the distribution over different values of J, we instead focus on analyzing our data to extract information that is otherwise generally discarded in averaging processes. In each individual measurement, the qubit is initialized in state $|0\rangle$ via optical pumping and one of $k = 50$ randomized benchmarking sequences with $J = 100$ gates is applied in the presence of engineered noise. A final projective measurement in each experiment yields a discretized qubit state measurement, which is used to infer the probability of finding the qubit in state $|1\rangle$ by repeating the experiment $r = 220$ times under application of the same engineered noise realization (reducing quantum projection noise). The survival-probability measurement outcomes for each sequence are then averaged over a variable number up to $n = 200$ different realizations of noise possessing the same engineered correlations. This process is repeated for all $k = 50$ sequences, allowing us to calculate the distribution variance $\mathbb{V}_k^{(n)}$.

Figures 2(a)–2(c) show the distributions over randomized benchmarking sequences of measured noise-averaged survival probabilities in the presence of concurrent detuning noise. The same set of sequences is subject to correlated (gray) or uncorrelated (red) noise sampled from a common distribution. Data are represented as histograms for different fixed values of averaging number n for each sequence. Solid lines are theoretical predictions for the distribution of survival probabilities derived from the updated random-walk framework, as given by the Γ distributions from Eq. (3), and substituting the error rms value σ using the noise-to-error translation for the expectation value shown in Table II. These theoretical predictions, which involve no free parameters, show good agreement with the data in the regimes studied.

These data clearly illustrate the differences in the distributions over the same set of randomized benchmarking sequences when subjected to noise with differing correlation properties. As shown in Ref. [32] and highlighted here in Table I, the distributions possess approximately the same mean value, despite the differing noise correlation properties. The skew to high fidelities in the data taken using correlated noise is a manifestation of the randomized decoupling effects known to exist within some randomized benchmarking sequences [32]. More importantly, the behavior of the variance of the distributions under an increasing number of noise
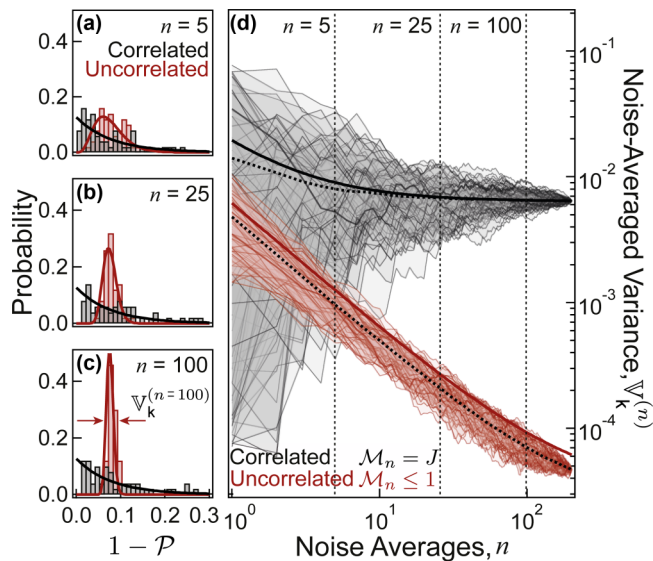
FIG. 2. Signatures of error correlations in randomized benchmarking sequences. (a)–(c) Distribution of measured survival probabilities for $k = 50$ randomly composed sequences averaged over $n = 5$, 25, and 100 noise realizations drawn from $\delta \sim \mathcal{N}(0, \rho^2 = 2 \times 10^{-3})$ for both maximally correlated $\mathcal{M}_n = J$ (gray) and uncorrelated (red) engineered noise processes. Uncorrelated noise possesses a $\pi/2$ bandwidth, i.e., noise values change with a rate set at the inverse of the duration of a primitive $\pi/2$ rotation, and hence can take one or multiple values in a gate ($\mathcal{M}_n \leqslant 1$). Solid lines are normalized Γ distributions plotted with no free parameters. (d) Scaling of cumulatively noise-averaged histogram variances $\mathbb{V}_k^{(n)} \equiv \mathbb{V}_k[\langle P(|1\rangle)\rangle_n]$. Trajectories correspond to different orderings of noise realizations with dotted lines representing the mean of 1000 reorderings and solid lines are theoretical predictions with no free parameters (see the text). Vertical dashed lines indicate the values of n used in (a)–(c).

averages n varies substantially. For small n the distributions are similarly broad despite the differences in their shapes, but with further averaging the distribution measured under uncorrelated noise narrows while the variance of the distribution measured under correlated noise remains approximately constant (as discussed in Sec. II C).

To highlight the effect of noise correlations on the experimental averaging behavior, we plot the variance of the distribution over measured sequence survival probabilities $\mathbb{V}_k^{(n)} \equiv \mathbb{V}_k[\langle P(|1\rangle)\rangle_n]$ as a function of the number of noise averages n [Fig. 2(d)]. Data are represented as a collection of trajectories, each constituting a randomized reshuffling of the data set (indexed by individual noise realizations). Data sets are averaged over n noise realizations sampled from the full set and n is allowed to vary from left to right. This continues until all $n = 200$ noise realizations have been sampled. This procedure is used to mitigate potential unintended systematic bias in the scaling of the noise-averaged variance with n. For instance, if by chance the first several noise samples over which one averages varied substantially, the variance may be artificially inflated from the ensemble average value; accordingly, the resulting deviations between noise-averaging trajectories are largest for small n and vanish at the far right of the graph where all data sets are fully averaged over all available data. For correlated noise $\mathcal{M}_n = J$, the resulting

trajectories are initially broadly distributed and fluctuate before converging with n to a fixed, analytically calculable variance. By contrast, in the case of uncorrelated noise with $\mathcal{M}_n \leqslant 1$, all trajectories show an approximate reduction in $\mathbb{V}_k^{(n)} \propto 1/n$, commensurate with a continued narrowing of the distribution of outcomes over different sequences under averaging [Figs. 2(a)–2(c)].

Solid lines capturing key scaling behaviors observed in both data sets of Fig. 2(d) are derived from the expression for variance in Table I using the noise-to-error translations presented in Tables II and III, calculated for concurrent detuning noise with no free parameters. Overall, agreement with the measured experimental data are good across a wide parameter range and two orders of magnitude in $\mathbb{V}_k^{(n)}$. For correlated noise, small deviations between the theoretical trace and measured mean scaling appear for low values of n. Numerical evidence attributes this to the limited sample size in terms of sequences, which does not always capture the rare highly-error-susceptible sequences that would lead to a larger variance. In the case of uncorrelated noise, there is an overall vertical shift between the theory and the data, which is fully compensated by adjusting the rms noise strength ρ_U by $\sim 6\%$. Numerical simulations and analytic considerations attribute the need for this adjustment to the strong noise employed in these experiments, which violates the theoretical assumption $J\rho_U^2 \ll 1$, such that higher-order terms in the theory cannot be fully ignored.

The uncorrelated noise data begin to deviate from an exact $1/n$ scaling of $\mathbb{V}_k^{(n)}$ at large numbers of noise averages. This behavior is captured by our theoretical model and varies in a predictable way with the applied noise bandwidth and sequence length J (Appendix B 1); we have verified it is not due to fundamental measurement limits in our system or quantum projection noise, as discussed in Appendix D. We are able to attribute this "saturation" in variance scaling for uncorrelated noise to residual sequence dependence, even in the case of purely uncorrelated noise, and the fact that our projective measurement probes only a two-dimensional $\hat{\sigma}_x \hat{\sigma}_y$ plane in Pauli-error space. For example, one can imagine a sequence composed solely of $\hat{\mathbb{I}}$ gates, which, due to an induced off-resonance error, will experience a net phase rotation that cannot be measured by single-axis projective measurements. Hence, no amount of averaging over different noise strength realizations will produce a survival probability that converges to the distribution mean, even in the case of uncorrelated noise.

Overall, we find that our theoretical models predict not only the full distribution of survival probabilities over randomized benchmarking sequences, but also the scaling of this distribution's variance with experimental averaging. The difference between the gray and red data in Fig. 2(d) and the agreement of theory thus constitute key experimental validations of the central theoretical contributions made in this paper.

IV. SUPPRESSING ERROR CORRELATIONS USING DYNAMICALLY CORRECTED GATES

In the next part of our study we explore the ability to modify error correlations within a sequence through deterministic replacement of each Clifford operation in a randomized
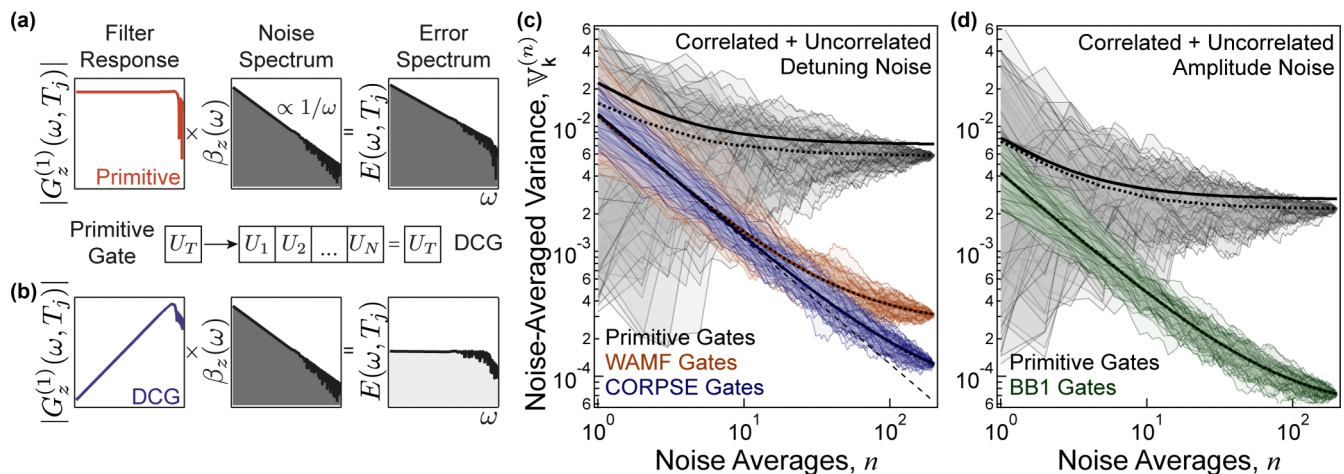
FIG. 3. Suppression of error correlations using dynamically corrected gates. (a) The first-order generalized filter-transfer function for dephasing noise of a primitive operation $G_z^{(1)}(\omega, T_j)$ and the noise spectrum [here $\beta_z(\omega) \propto 1/\omega$] combine to produce an effective error spectrum $E(\omega, T_j)$ for a single gate. (b) The modified filter functions for first-order DCGs scale as ω at low frequencies, which results in a whitening of $E(\omega, T_j)$ relative to the input noise spectrum. (c) and (d) Variance scaling with n for primitive (gray) gates and WAMF (orange), CORPSE (blue), and BB1 (green) DCGs all subjected to noise with both correlated and uncorrelated components. In (c) detuning noise is engineered with strength $\delta_C \sim \mathcal{N}(0, 2 \times 10^{-3})$ and $\delta_U \sim \mathcal{N}(0, 5 \times 10^{-4})$ and in (d) amplitude noise is engineered with strength $\delta_C \sim \mathcal{N}(0, 9 \times 10^{-4})$ and $\delta_U \sim \mathcal{N}(0, 2 \times 10^{-4})$. Dotted lines are means of 1000 trajectories randomized over noise realizations and solid lines for the DCGs are theoretical fits from Table I to the mean with the values of σ_U^2 and σ_C^2 allowed to vary. Black solid lines for primitive gates are derived from the same theory with no free parameters. As with Fig. 2, all data are measured for $k = 50$ sequences of length $J = 100$ with $n = 200$ noise realizations and $r = 220$ repetitions.

benchmarking sequence with an error-suppressing dynamically corrected gate (DCG). Each DCG is implemented by replacing primitive physical rotations with composite pulses comprising multiple physical rotations [33], according to one of several prescriptions [25]. This approach abstracts the target state transformations away from the physical qubit manipulation in a manner that builds in error robustness via coherent averaging. In this way, these composite gates modify the error susceptibility of the target operations and in particular change the relationship between an input correlated-noise process and output gate errors. We therefore refer to their action as virtualizing the Clifford operations, consistent with an abstraction above the physical-layer operations presented in Ref. [29].

The error-virtualization process is described quantitatively by calculating the error vector ϵ_j at the operator level and expressing it in the Fourier domain. In the limit of classical Gaussian dephasing noise, described in the Fourier domain as the spectrum $\beta_z(\omega)$, the leading-order Magnus term ($\alpha = 1$) in the $\hat{\sigma}_z$ quadrature may be written as

$$[\epsilon_{j,z}]_1 = -i \int \frac{d\omega}{2\pi} G_z^{(1)}(\omega, T_j) \beta_z(\omega). \quad (6)$$

Here $G_z^{(1)}(\omega, T_j)$ is an analytically calculable, filter-transfer function that describes the spectral characteristics of a gate active for duration T_j [34]. The *effective* error spectrum experienced by the gate may therefore be represented by the spectral overlap of the filter-transfer function with the noise, written as $G_z^{(1)}(\omega, T_j) \times \beta_z(\omega) \rightarrow E(\omega, T_j)$. Figure 3(a) demonstrates the mapping between input noise and the effective error spectrum schematically for an example $1/\omega$-noise spectrum and a primitive π rotation about the x axis. In this example, correlations in the noise are directly transferred to

the correlations in the effective error spectrum [37] [cf. direct $\mathcal{M}_n$ to $\mathcal{M}_\varepsilon$ translation for primitive gates in Fig. 1(c)].

Replacement of the primitive gate with a logically equivalent DCG virtualizes the effective error spectrum for each operator through the process of noise filtering [27,33,34,37]. Figure 3(b) illustrates this effect, where the DCG's reduced susceptibility to low-frequency noise (captured through its filter-transfer function) results in a "whitening" of the effective error spectrum relative to $\beta_z(\omega)$. We use the term whitening to describe the reduction of low-frequency weight, producing an error spectrum with characteristics closer to a white spectrum than the original input; we do not imply the effective error spectrum after DCG application is formally white (frequency independent). In the current context, this whitening suggests that DCGs should not only reduce overall error magnitudes when the noise is dominated by low-frequency contributions, but they should also suppress the signatures of error correlations between sequentially applied gates.

The particular DCG constructions examined in this work are the compensation for off-resonance with a pulse sequence (CORPSE) [56] and Walsh amplitude modulated filter (WAMF) [57] gates, which suppress detuning errors, and the BB1 pulse family [58], which suppresses overrotation errors. Specific details of DCG construction for the various operations employed here are presented in Appendix C.

A. Modification of variance scaling with engineered errors using DCGs

We begin by performing a detailed quantitative study of the measured signatures of error correlations through the application of engineered noise. We experimentally implement primitive, CORPSE, WAMF, and BB1 gates, where the

first two DCGs are designed to suppress errors arising from frequency detuning noise and the latter is designed to suppress errors arising from amplitude noise. Using the same set of randomly generated randomized benchmarking sequences as in Fig. 2, we now apply a mixed noise spectrum, simultaneously containing uncorrelated rapidly varying noise ($\mathcal{M}_n \leqslant 1$) and quasistatic offsets that are constant over a full sequence giving a strongly correlated component ($\mathcal{M}_n = J$). In addition to performing measurements with primitive gates, we also construct DCG sequences by deterministically replacing each Clifford with its logically equivalent DCG counterpart. The relations for the mixed noise spectrum provided in Tables I and II now permit a direct study of the impact of using DCGs on error correlations appearing within the randomized benchmarking sequences via the averaging behavior of $\mathbb{V}_k^{(n)}$.

Beginning with frequency detuning noise, both DCG implementations shown in Fig. 3(c) exhibit an initial variance scaling with noise averaging $\mathbb{V}_k^{(n)} \propto 1/n$, reminiscent of the application of the purely uncorrelated noise process in Fig. 2(d). The observed saturation in $\mathbb{V}_k^{(n)}$ at large n for the DCG data combines contributions due to both the analytically calculable component occurring in the presence of purely uncorrelated noise introduced above and residual uncompensated error correlations. The general behavior observed for the DCG sequences is to be contrasted with that observed for the same sequences composed of primitive gates where, as in Fig. 2, the strong correlated noise component causes the variance to converge to a large constant value (gray).

Similar behavior is observed when considering the amplitude error quadrature. We demonstrate this through the application of engineered control-amplitude noise in Fig. 3(d), where measurements on sequences composed of DCGs derived from the BB1 family exhibit a similar $\mathbb{V}_k^{(n)} \propto 1/n$ averaging behavior. Again, this is contrasted with the behavior of sequences composed of primitive gates where once more the variance saturates to a high constant value, despite application of the same noise in both settings.

B. Quantitative analysis of error correlation suppression

In order to calculate the change in error correlations realized in randomized benchmarking sequences composed of DCGs, we compare experimental measurements of $\mathbb{V}_k^{(n)}$ with the predictions of the model summarized in Table I. For the primitive gates, we explicitly translate the applied detuning noise strengths to an effective error strength using the noise-to-error relations in Table II; for this, we also use the expected random-walk step expressions calculated and presented in Table III of Appendix B for detuning or amplitude noise with a $\pi/2$ bandwidth in the uncorrelated component. The black solid lines in Figs. 3(c) and 3(d) are then derived using these calculated error strengths, with no free parameters. Agreement between experimental measurements and theoretical predictions for the primitive gate sequences is good, but we observe a small ($\sim$20%) deviation that appears approximately constant over several orders of magnitude in n for both noise processes. Ongoing work is investigating the source of this discrepancy; possible sources include the unaccounted impact of higher-order terms due to the strength of the applied noise

and undersampling of the distribution over noise-averaged sequences.

To extract the relative correlated and uncorrelated error components after DCG application, we fit the data using the theoretical predictions for the scaling of $\mathbb{V}_k^{(n)}$ shown in Table I and use the strengths of the two error components σ_U^2 and σ_C^2 as free parameters. First, for all DCGs we observe a reduction in σ_C^2 coupled with an increase in σ_U^2. Specifically, σ_C^2 is reduced by a factor of 49 times for CORPSE, 6 times for WAMF, and 10 times for BB1 gates, while all experience an increase in σ_U^2 by approximately 6–7 times. The relative performance of the DCGs observed in our experiments is aligned with their documented strengths, as the CORPSE gate is known to more efficiently cancel purely static detuning errors than WAMF gate [33,57], although improved calibration of the pulse-amplitude values used in WAMF gates is expected to improve the efficacy of correlated-error suppression.

The increase in σ_U^2 is approximately consistent with the increase in duration of the DCGs relative to the primitive gate implementations. Considering the high-pass-filtering nature of all DCGs illustrates why uncorrelated noise processes fluctuating rapidly on the scale of the individual DCGs are transmitted by their filters and lead to residual errors that may be amplified by the DCG structure. Overall, these measurements, in particular, the scaling of $\mathbb{V}_k^{(n)}$, are consistent with an interpretation that the action of the noise whitening in the filter-transfer-function framework transforms correlated noise into predominantly uncorrelated residual errors *at the operator level*.

C. Signatures of variable error correlation lengths

To expand on the previous analyses, we experimentally demonstrate that the reduction in effective error correlation indeed resides at the virtual gate layer. Using the same sequences as before and the same engineered ρ_U and ρ_C rms magnitudes for detuning noise, the length of the correlated noise component is now varied in terms of the number of gates at the virtual level, breaking it up into blocks of length $\mathcal{M}_n$. The laboratory-frame durations of the noise blocks therefore now differ by a factor of $\sim$6 between the primitive and the CORPSE gates (the average increase in the duration of the Clifford operations when using the CORPSE gate).

In the case of sequences composed of primitive gates, the signature exhibited by the variance scaling under noise averaging in Fig. 4(a) gradually changes from indicating correlated errors (saturation at high variance) to purely uncorrelated errors ($1/n$-like scaling) as the block length is decreased, consistent with observations in Figs. 2 and 3. By contrast, the sequences composed of CORPSE gates in Fig. 4(b) retain their overall $1/n$-like scaling behavior for all correlated component block lengths, demonstrating that residual uncorrelated errors remain dominant. All traces in Figs. 4(a) and 4(b) have been normalized to the initial mean variance for each engineered noise case to highlight the change in the relative correlated and uncorrelated error components, rather than the net error strength.

As a witness of the suppression of error correlations, Fig. 4(c) shows the ratio of the initial mean variance $\mathbb{V}_k^{(n=1)}$ to the final, fully noise-averaged variance $\mathbb{V}_k^{(n=200)}$. This ratio
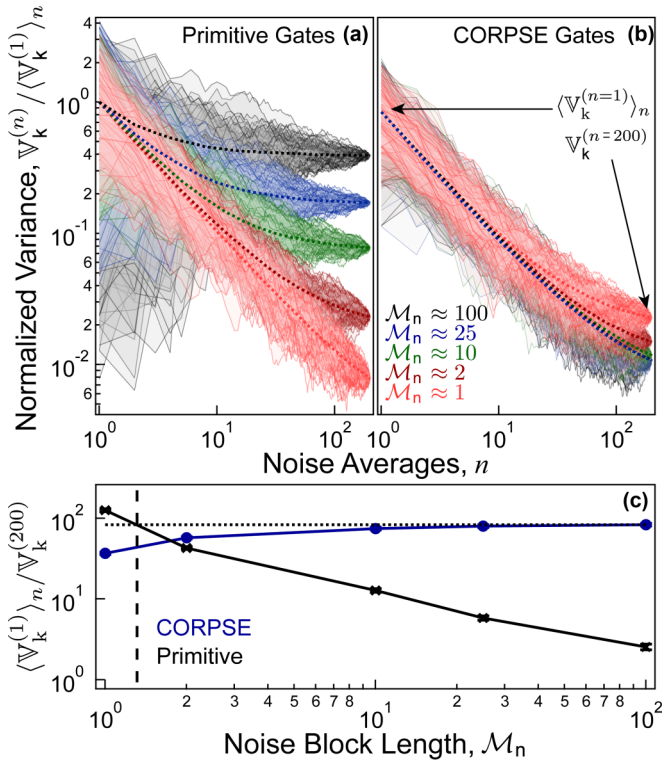
FIG. 4. Suppression of error correlations using DCGs under noise with varying $\mathcal{M}_n$. Variance scaling of $k = 20$ sequences with noise averaging is shown for (a) primitive and (b) CORPSE gates. Traces are normalized to the initial mean variance for each applied noise case. Engineered noise is composed of an uncorrelated component ($\mathcal{M}_n \leqslant 1$) and a block correlated component of length $\mathcal{M}_n$ that is varied from fully correlated ($\mathcal{M}_n = J$) to uncorrelated ($\mathcal{M}_n = 1$) in units of virtual gates. Dotted lines are means of 1000 randomized trajectories. (c) Ratio of initial to final variance in (a) and (b) as $\mathcal{M}_n$ is varied for primitive (black) and CORPSE (blue) gates. The dotted line marks the ratio at which CORPSE gates saturate and the dashed vertical line indicates the value of $\mathcal{M}_n$ where this ratio crosses the scaling trend for primitive gates. Error bars calculated from the standard error of the mean of the 200 initial values of variance and normalized by the fully noise-averaged variance are smaller than point size.

V. THE DCG'S IMPACT ON INTRINSIC ERRORS

After verifying the utility of the theoretical constructs we have introduced in this work, we now turn to characterizing the intrinsic errors limiting the performance of our system. In the trapped $^{171}\text{Yb}^+$ ion experiment described in Sec. III,

we achieve a single-qubit randomized benchmarking average error per gate (EPG) of $(1.89 \pm 0.12) \times 10^{-5}$ (Appendix A). Increasing the number of qubits to 5 and performing simultaneous randomized benchmarking using a global microwave control field reveals a monotonic increase in the EPG across the register, ranging from $(5.7 \pm 0.5) \times 10^{-5}$ to $(1.3 \pm 0.1) \times 10^{-4}$. As such, were we to run multi-ion algorithms that use global state manipulations, e.g., transversal gates in the seven-qubit Steane code [4], we would not see the net error rate scale linearly with respect to the initial single-qubit EPG. This nonlinear scaling with increasing qubit numbers has been observed in many systems and is often due to crosstalk between qubits [59]. It is important to note that this experimental observation of inhomogeneous error rates also violates a common assumption on noise statistics made in studies of error correcting codes, namely, that the noise is independent and identically distributed.

In our case, the underlying cause of the observed error inhomogeneity is a sub-percent-level gradient in the amplitude of the microwave control field across the ion chain, caused by interference from metallic surfaces in the proximity of our in-vacuum antenna. We also observe a small magnetic-field gradient across the qubit chain such that both amplitude and detuning noise are present simultaneously. Spatially correlated errors have recently been studied in Ref. [60], wherein it is noted that previous studies of multiqubit errors tend to assume either spatially independent errors or identically spatially correlated errors, facilitating the use of a decoherence free subspace. Our situation, with a gradient of spatially correlated errors, falls between these two cases, but can still induce simultaneous multiqubit errors that lower the efficacy of QEC.

To characterize the impact of DCGs on spatially correlated errors, we utilize simultaneous randomized benchmarking sequences of length $J = 500$ applied to all five qubits in the register and again explore variance scaling with experimental averaging. We construct DCG sequences using BB1 gates to combat the dominant microwave-control-amplitude errors. Data collection proceeds by interleaving a single sequence implemented using either primitive or BB1 gates to ensure a fair comparison between the sequences in time, in the event that any systematic drifts occur.

We examine the scaling of $\mathbb{V}_k^{(r)}$ with averaging over repetitions r, up to $r = 500$; because noise is native to the system, we make the substitution $n \equiv r$. The signature of the temporally correlated intrinsic errors is observed for all ions when using sequences of primitive gates in Fig. 5(a) (red). We observe a staggered, increasing saturation value for $\mathbb{V}_k^{(r)}$ at $r = 500$, increasing with the spatial distance from qubit 1 [leftmost qubit in Fig. 5(a) inset], which is used to calibrate the gate operations. As expected, the qubit that is furthest from the calibration qubit both suffers the worst randomized benchmarking performance and shows the highest saturation value in variance scaling. By contrast, the overrotation error suppressing BB1 gates (blue) saturate at a value of variance over an order of magnitude lower than achieved by the primitive gates and recover a $1/r$-like scaling for all qubits. We further find that the relationship between the physical positions of the qubits and the ordering of saturation variances has become scrambled. Using the analysis introduced above,

scales approximately inversely with $\mathcal{M}_n$ for primitive gates but remains nearly constant for CORPSE gates. Extrapolation of this ratio for CORPSE gates back towards small $\mathcal{M}_n$ reveals a crossover with the primitive data that lies between $\mathcal{M}_n \approx 1$ and 2. This shows that CORPSE gates can reduce the noise correlation length to an error correlation length commensurate with physical noise $\mathcal{M}_n \approx 1$–2. Because the noise correlation blocks were matched to the duration of the underlying Clifford gates, whether through primitive or composite construction, these data highlight the efficacy of DCGs in virtualizing error characteristics for the logical gates implemented.
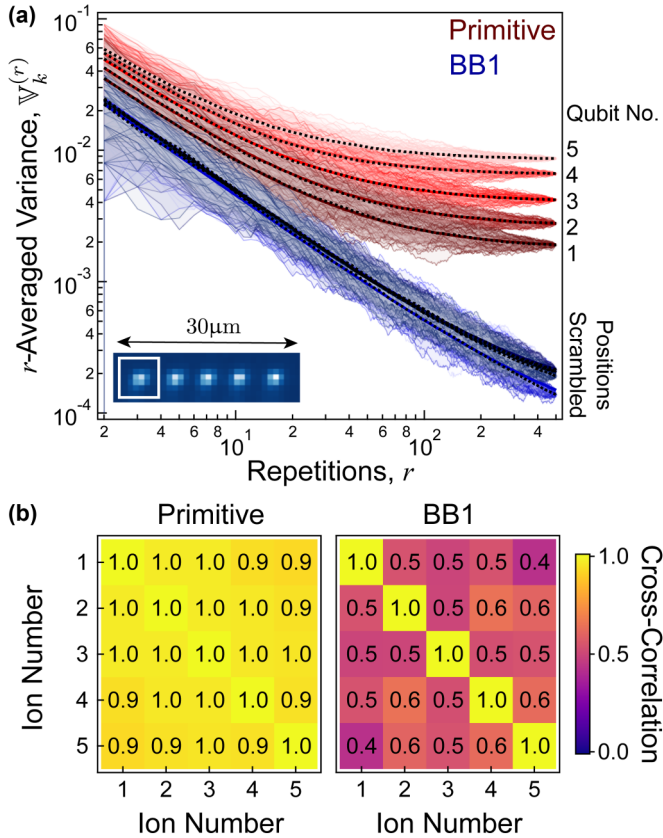
FIG. 5. Intrinsic errors in a five-qubit chain. (a) Variance over noise-averaged sequence survival probabilities for five qubits using $k = 60$ sequences of length $J = 500$, averaged over repetitions r, up to $r = 500$. Each trajectory is produced by shuffling the order of repetitions used in the graph to avoid bias, dotted lines indicate the means of 1000 trajectory randomizations, and solid lines are fits where the correlated and uncorrelated error strengths were free to vary. The correlated error strengths σ_C^2 are $\{1.2, 1.5, 1.9, 2.4, 2.7\} \times 10^{-4}$ from qubit 1 to 5 for the primitive gates and $\{2.3, 2.5, 1.1, 2.2, 2.3\} \times 10^{-5}$ for the BB1 gates. The uncorrelated error strengths σ_U^2 are $\{7.5, 8.1, 8.5, 8.6, 8.7\} \times 10^{-4}$ from qubit 1 to 5 for the primitive gates and $\{6.5, 6.5, 6.3, 6.6, 6.5\} \times 10^{-4}$ for the BB1 gates. The insert shows the EMCCD image of a five-ion chain, spaced over $\sim 30~\mu$m. The control field amplitude and frequency are calibrated with respect to the highlighted leftmost ion. (b) Pairwise cross-correlation coefficients between the five-qubit survival probabilities for primitive gates (left) and BB1 DCGs (right), revealing a $\sim 50\%$ reduction in the correlations between qubit errors when using DCGs.

we fit the mean variance trends with the expression in Table I, allowing the strengths of the error σ_C^2 and σ_U^2 to vary. We extract a reduction in the correlated error strength when using BB1 gates ranging from ~ 5 to 16 times for the five qubits.

To directly probe the action of DCGs in virtualizing the spatially correlated errors, we calculate the pairwise cross-correlation coefficient between the survival probabilities in each experimental realization [Fig. 5(b)]. For primitive gates, all errors are highly correlated between qubits (cross-correlation coefficient equal to or greater than 0.9 for all qubit pairs), whereas for the BB1 gates, a reduction of approximately 50% can be seen between all qubit pairs, further

supporting the evidence that DCGs provide a suppression of error correlations in both time and space.

Separate investigations not presented here using a multi-axis error suppressing DCG called CORPSE in BB1, which combines the detuning-error robustness of the CORPSE gate with the amplitude-error robustness of the BB1 [61], showed no additional benefit. This observation suggests that the off-resonance error created by the magnetic-field gradient was sufficiently small that it was dominated by other larger, but rapidly fluctuating, intrinsic error sources.

VI. OUTLOOK

The results we have presented suggest that the path to the practical implementation of QEC may be facilitated by transforming miscalibrations and common laboratory noise sources exhibiting slow drifts and low-weight noise spectra, into effective error processes with dramatically reduced correlations at the virtual layer using DCGs. We believe this is important as the pursuit of functional quantum computers, even at the mesoscale, will clearly require major advances in the control and suppression of errors, as gate counts quickly exceed 10^{10} for even moderate problems requiring only ~ 200 qubits [62]. Combined with the observation that certain DCGs can mitigate spatial crosstalk in multiqubit systems [63], we believe that our demonstration of the suppression of temporal and spatial error correlations within quantum circuits solidifies the central importance of dynamic error suppression techniques at the virtual level for practical quantum computing.

APPENDIX A: SINGLE-QUBIT RANDOMIZED BENCHMARKING

Using the experiment described in Sec. III, with a single trapped ^{171}Yb$^+$ ion and microwave gates, we achieve a single-qubit error per gate of $p_{\mathrm{RB}} = (1.89 \pm 0.12) \times 10^{-5}$ measured using randomized benchmarking (Fig. 6). The fit to the mean survival probabilities used to extract the error per gate is given by

$$\mathcal{P} = 0.5 + (0.5 - \kappa)e^{-p_{\mathrm{RB}}J}, \qquad \text{(A1)}$$

where $\mathcal{P}$ is the mean survival probability, J is the number of gates in a randomized benchmarking sequence, and κ is the value of our single-qubit state preparation and measurement error, found to be $\kappa = (3.3 \pm 0.1) \times 10^{-3}$.
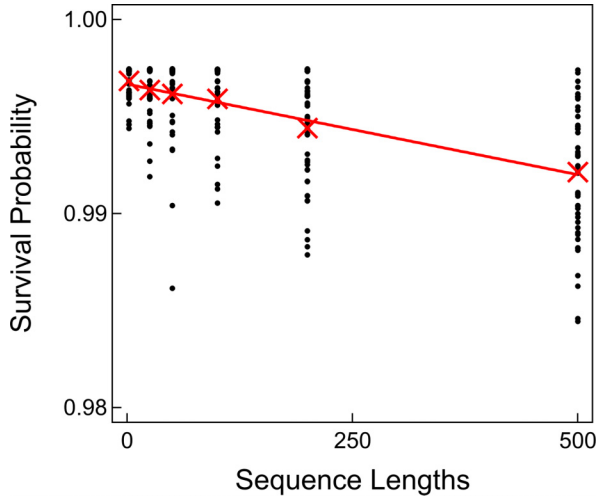
FIG. 6. Single-qubit randomized benchmarking. Randomized benchmarking is performed on a single qubit using a total of 300 sequences composed of primitive gates, with 50 sequences each of length $J = 2, 25, 50, 100, 200,$ and 500. Each sequence was repeated $r = 500$ times to reduce quantum projection noise. Black markers represent individual sequence survival probabilities, red crosses indicate the mean survival probabilities for each sequence length, and the solid red line is a fit to the means to extract the average error per gate.

APPENDIX B: PHYSICAL NOISE TO ERROR STRENGTH TRANSLATION

We verify the model presented in this paper by using primitive Clifford gates under engineered noise, where the strength and effect of the noise are known exactly, allowing for quantitative analysis. For this verification, we need to calculate the translation between the rms magnitude of the physical noise process ρ and that of the resulting error operators σ. The noise is applied concurrently with the gate operations inducing multiaxis and gate-dependent errors for the different Clifford operations, whose lengths differ between π and $\pi/2$ rotations. Due to this introduced gate dependence, an exactly constant noise process will not be directly translated to a constant error process with identical error vectors for every gate, and hence the translation for each gate needs to be considered explicitly.

The method to transform noise strength to error strength for noisy primitive Clifford gates is initially presented here for a general noise process that is static over the duration of a single gate. Each of the single-qubit Clifford gates is made up of rotations on the Bloch sphere with the rotation axis and angle specified by the Clifford gate index $\eta_j \in \{1, \ldots, 24\}$. If the jth gate in a sequence is affected by laboratory noise with value $\delta_j \sim \mathcal{N}(0, \rho^2)$, the resulting noisy gate can be decomposed into an error operator and the ideal gate $\tilde{C}_{\eta_j} = \hat{\Lambda}_j \hat{C}_{\eta_j}$, with

$$\hat{\Lambda}_j = \exp\left(i \sum_{\alpha=1}^{\infty} \delta_j^{\alpha} \big[\boldsymbol{v}_{\eta_j} \big]_{\alpha} \cdot \hat{\boldsymbol{\sigma}} \right)$$
$$\approx \hat{\mathbb{I}} + i\delta_j \big[\boldsymbol{v}_{\eta_j} \big]_1 \cdot \hat{\boldsymbol{\sigma}}, \qquad (B1)$$

where $\hat{\boldsymbol{\sigma}}$ is the vector of Pauli matrices. In the main text, this operator was introduced in terms of the error vector $\boldsymbol{\epsilon}_j$ as $\hat{\Lambda}_j = \exp\{i \sum_{\alpha=1}^{\infty} [\boldsymbol{\epsilon}_j]_{\alpha} \cdot \hat{\boldsymbol{\sigma}}\}$. We have now separated the error vector into two components for the Magnus expansion of order α, $[\boldsymbol{\epsilon}_j]_{\alpha} = \delta_j^{\alpha} [\boldsymbol{v}_{\eta_j}]_{\alpha}$, to explicitly show the dependence on the physical noise strength δ_j, which will change between different realizations of the noise, and the particular gate's susceptibility to the error channel, described by the term $\boldsymbol{v}_{\eta_j}$. There will be 24 gate-specific error vector terms $\boldsymbol{v}_{\eta_j}$ corresponding to the 24 Clifford operations, which can be calculated explicitly for a given noise process. We now consider how these terms affect our ideal randomized benchmarking sequence.

Starting with the standard randomized benchmarking procedure, we compile a sequence of randomly composed single-qubit Clifford operations $\prod_{j=1}^{J} \hat{C}_{\eta_j} = \hat{\mathbb{I}}$, which are mathematically right multiplied to the preceding operator such that they act sequentially on an initial state. Then the complete noisy sequence is given by

$$\tilde{S} = \prod_{j=1}^{J} \hat{\Lambda}_j \hat{C}_{\eta_j}. \qquad (B2)$$

The survival probability for a qubit prepared in $|0\rangle$, averaged over n noise instances, is calculated using

$$\mathcal{P} = 1 - \langle P(|1\rangle) \rangle_n = \langle |\langle 0|\tilde{S}|0\rangle|^2 \rangle_n. \qquad (B3)$$

To approximate the sequence, the method from [32] is employed: The first-order term of each error operator can be translated to a step in Pauli-error space, with the total random walk in three dimensions for a given noise instance i given by

$$\mathbf{R}_{3D}^{(i)} = \sum_{j=1}^{J} \delta_j^{(i)} \mathbf{r}_{3D,j}. \qquad (B4)$$

The jth random-walk step $\mathbf{r}_{3D,j}$ is calculated from the product of the preceding ideal gates modifying the first-order gate-specific error for the jth operation in the sequence $[\boldsymbol{v}_{\eta_j}]_1 \cdot \hat{\boldsymbol{\sigma}}$,

$$\hat{C}_{\eta_1} \cdots \hat{C}_{\eta_{j-1}} \big(\big[\boldsymbol{v}_{\eta_j} \big]_1 \cdot \hat{\boldsymbol{\sigma}} \big) \big(\hat{C}_{\eta_1} \cdots \hat{C}_{\eta_{j-1}} \big)^{\dagger} = \mathbf{r}_{3D,j} \cdot \hat{\boldsymbol{\sigma}}. \qquad (B5)$$

To obtain the sequence survival probability that would be measured via a single-axis projective measurement, the relevant steps are then the projection of $\mathbf{r}_{3D,j}$ in the two-dimensional $\hat{\sigma}_x \hat{\sigma}_y$ plane $\mathbf{r}_{2D,j} \equiv \mathbf{r}_j$ of Pauli-error space. As with the original model, it can be shown that a sequence's survival probability is given by

$$\mathcal{P} = 1 - \langle \|\mathbf{R}\|^2 \rangle_n + O(\rho^4), \qquad (B6)$$

where $\mathbf{R}$ is the two-dimensional random walk. From this expression, the expectation and variance of the distribution over noise-averaged sequence survival probabilities have been calculated for arbitrary step lengths; the results of this calculation are summarized in the noise-to-error translation in Table II of the main text. These expressions are based on the expected random-walk steps induced by the 24 error maps, which are shown in Table III for a range of physical noise processes. We proceed here by showing an example derived for a concurrent detuning error.

TABLE III. Expected step lengths in the Pauli $\hat{\sigma}_x \hat{\sigma}_y$ plane based on the average Clifford gate error for different engineered noise. These quantities are used to predict the statistical moments of the sequence survival-probability distributions in Table II.

Engineered noise process	$\mathbb{E}[\|\mathbf{r}_j\|^2]$	$\mathbb{E}[\|\mathbf{r}_j\|^4]$	$\mathrm{Cov}(\|\mathbf{r}_{U,j}\|^2, \|\mathbf{r}_{C,j}\|^2)$
interleaved dephasing	$\dfrac{2}{3}$	$\dfrac{2}{3}$	$\dfrac{2}{9}$
concurrent detuning with one value per primitive (π or $\pi/2$) gate	$\dfrac{2}{3}\left(\dfrac{1}{2} + \dfrac{\pi^2}{96}\right)$	$\dfrac{2}{3}\left(\dfrac{7}{24} + \dfrac{\pi^4}{384}\right)$	$\dfrac{2}{3}\left(\dfrac{7}{24} + \dfrac{\pi^4}{384}\right) - \dfrac{4}{9}\left(\dfrac{1}{2} + \dfrac{\pi^2}{96}\right)^2$
concurrent detuning with one value every primitive $\pi/2$ gate time	$\dfrac{2}{3}\left(\dfrac{1}{2} + \dfrac{\pi^2}{192}\right)$	$\dfrac{2}{3}\left(\dfrac{1}{4} + \dfrac{\pi^4}{1536}\right)$	$\dfrac{17}{108} + \dfrac{\pi^4}{1152} - \dfrac{4}{9}\left(\dfrac{1}{2} + \dfrac{\pi^2}{192}\right)\left(\dfrac{1}{2} + \dfrac{\pi^2}{96}\right)$
over- and underrotation with one value per primitive (π or $\pi/2$) gate	$\dfrac{\pi^2}{18}$	$\dfrac{5\pi^4}{576}$	$\dfrac{29\pi^4}{5184}$
over- and underrotation with one value every primitive $\pi/2$ gate time	$\dfrac{\pi^2}{36}$	$\dfrac{5\pi^4}{2304}$	$\dfrac{29\pi^4}{10\,368}$

1. Example for concurrent detuning noise

Combined Tables I and II predict the form of the noise-averaged survival-probability distribution for different engineered noise processes, given the expected random-walk steps in the $\hat{\sigma}_x \hat{\sigma}_y$ place of Pauli-error space, $\mathbb{E}[\|\mathbf{r}_j\|^2]$ and $\mathbb{E}[\|\mathbf{r}_j\|^4]$. As an explicit example, these quantities are calculated here for concurrently applied detuning noise, produced by an offset between the qubit frequency and the control field frequency, normalized to the Rabi frequency $\delta = \Delta/\Omega$. An ideal rotation of angle θ about the $\mathbf{n}$ axis of the Bloch sphere is modified by detuning noise as

$$\tilde{U}(\mathbf{n}, \theta, \delta) = e^{-i(\theta \hat{\boldsymbol{\sigma}} \cdot \mathbf{n} + |\theta|\delta\hat{\sigma}_z)/2}. \quad \text{(B7)}$$

From this, the eight physical error maps affecting the Clifford operations are calculated to be

$$\hat{\Lambda}^{(\hat{\mathbb{1}})}(\pi, \delta) = \hat{\mathbb{1}} - i\frac{\pi\delta}{2}\hat{\sigma}_z + O(\delta^2), \quad \text{(B8a)}$$

$$\hat{\Lambda}^{(\hat{\sigma}_x)}(\pi, \delta) = \hat{\mathbb{1}} + i\delta\hat{\sigma}_y + O(\delta^2), \quad \text{(B8b)}$$

$$\hat{\Lambda}^{(\hat{\sigma}_x)}\left(\pm\frac{\pi}{2}, \delta\right) = \hat{\mathbb{1}} \pm \frac{i\delta}{2}\hat{\sigma}_y - \frac{i\delta}{2}\hat{\sigma}_z + O(\delta^2), \quad \text{(B8c)}$$

$$\hat{\Lambda}^{(\hat{\sigma}_y)}(\pi, \delta) = \hat{\mathbb{1}} - i\delta\hat{\sigma}_x + O(\delta^2), \quad \text{(B8d)}$$

$$\hat{\Lambda}^{(\hat{\sigma}_y)}\left(\pm\frac{\pi}{2}, \delta\right) = \hat{\mathbb{1}} \mp \frac{i\delta}{2}\hat{\sigma}_x - \frac{i\delta}{2}\hat{\sigma}_z + O(\delta^2), \quad \text{(B8e)}$$

$$\hat{\Lambda}^{(\hat{\sigma}_z)}(\theta, \delta) = \hat{\mathbb{1}}, \quad \text{(B8f)}$$

more generally expressed for the jth operation in the sequence as

$$\hat{\Lambda}_j = \hat{\mathbb{1}} + \delta_j[\boldsymbol{\nu}_{\eta_j}]_1 \cdot \hat{\boldsymbol{\sigma}} + O(\delta_j^2). \quad \text{(B9)}$$

Only eight error maps are required to treat all 24 Clifford operations due to the error-free nature of $\hat{\sigma}_z$ rotations, which are generally implemented via instantaneous phase changes on the control field. Following the definition of the Clifford operations given in Ref. [32], there is only one non-$\hat{\sigma}_z$ rotation per Clifford, which exactly corresponds to one of the eight error maps described in Eq. (B8). If $\hat{\sigma}_z$ operations were also affected by the noise, the procedure would follow similarly but all error maps would need to be calculated.

To find the expected random-walk steps for this unitary error channel, recall from (B5) that the direction of the Pauli-error steps is determined by the preceding operations in the randomly composed sequence. As such, a given step will remain deterministic in its size, yet be performed along an arbitrary direction in Pauli-error space, determined by the preceding gates. Studying the error maps for concurrent detuning noise, we can write the gate-dependent steps as

$$\hat{\Lambda}^{(\hat{\mathbb{1}})}(\pi) \rightarrow \frac{\pi}{2}\hat{m}_1, \quad \text{(B10)}$$

$$\hat{\Lambda}^{(\hat{\sigma}_x)}(\pi) \rightarrow 1\hat{m}_1, \quad \text{(B11)}$$

$$\hat{\Lambda}^{(\hat{\sigma}_x)}\left(\frac{\pi}{2}\right) \rightarrow \left(\frac{1}{2}\hat{m}_1 + \frac{1}{2}\hat{m}_2\right), \quad \text{(B12)}$$

$$\hat{\Lambda}^{(\hat{\sigma}_y)}(\pi) \rightarrow 1\hat{m}_1, \quad \text{(B13)}$$

$$\hat{\Lambda}^{(\hat{\sigma}_y)}\left(\frac{\pi}{2}\right) \rightarrow \left(\frac{1}{2}\hat{m}_1 + \frac{1}{2}\hat{m}_2\right), \quad \text{(B14)}$$

$$\hat{\Lambda}^{(\hat{\sigma}_z)}(\theta) \rightarrow 0, \quad \text{(B15)}$$

with $\hat{m}_1, \hat{m}_2 \in \pm\{\hat{\sigma}_x, \hat{\sigma}_y, \hat{\sigma}_z\}$. This implies that π rotations about the x and y axes of the Bloch sphere produce a unit-length step in Pauli-error space that will be randomly oriented along one of the six principal axes. Similarly, $\pi/2$ rotations produce a $1/\sqrt{2}$-length step oriented at $45°$ between two principal axes, $\hat{\mathbb{1}}$ gates produce a $\pi/2$-length step along a principal axis, and rotations about the z axis contribute no step due to their error-free nature.

The probability of producing a particular nonzero $\|\mathbf{r}_j\|$ is shown in Table IV, based on the prevalence of different gates in the 24 Clifford gates and the likeliness of their projection into the $\hat{\sigma}_x \hat{\sigma}_y$ plane. Note that these steps are completely independent of the strength of the particular noise realization $\delta_j^{(i)}$. The noise will eventually rescale each step length, but here we only consider the unscaled walk. For this particular noise type and bandwidth, it is not necessary to distinguish between $\mathbb{E}[\|\mathbf{r}_{C,j}\|^2]$ and $\mathbb{E}[\|\mathbf{r}_{U,j}\|^2]$, as both the correlated and uncorrelated error processes are static over the duration of an individual gate and hence will result in the same expected average walk steps; it is only when increasing the bandwidth

TABLE IV. Likelihood of producing particular length random-walk steps in the $\hat{\sigma}_x\hat{\sigma}_y$ plane of Pauli-error space when engineered detuning noise is applied, based on the number of Clifford gates corresponding to the error map and the chance of a randomly oriented step in the $\hat{\sigma}_x\hat{\sigma}_y$ plane.

$\|\mathbf{r}_j\|$	$\mathrm{Pr}_{\hat{\sigma}_x\hat{\sigma}_y}$
1	$\frac{4}{24} \times \frac{2}{3}$
$\frac{1}{\sqrt{2}}$	$\frac{16}{24} \times \frac{4}{12}$
$\frac{1}{2}$	$\frac{16}{24} \times \frac{8}{12}$
$\frac{\pi}{2}$	$\frac{1}{24} \times \frac{2}{3}$

of the uncorrelated noise that they need be distinguished. Using Table IV one finds

$$\mathbb{E}[\|\mathbf{r}_j\|^2] = \frac{2}{3}\left(\frac{1}{2} + \frac{\pi^2}{96}\right), \tag{B16}$$

$$\mathbb{E}[\|\mathbf{r}_j\|^4] = \frac{2}{3}\left(\frac{7}{24} + \frac{\pi^4}{384}\right), \tag{B17}$$

$$\mathbb{E}[\|\mathbf{r}_{U,j}\|^2\|\mathbf{r}_{C,j}\|^2] = \mathbb{E}[\|\mathbf{r}_j\|^4]. \tag{B18}$$

Using Tables I and II, this produces the expectation value

$$\mathbb{E}[\mathcal{P}] \approx J\sigma^2 \frac{2}{3}\left(\frac{1}{2} + \frac{\pi^2}{96}\right) \tag{B19}$$

for both correlated and uncorrelated errors. This again illustrates the equivalence of the distribution mean, which is related to the parameter that standard randomized benchmarking analysis returns, for noise of the same strength despite vastly different correlation lengths. The difference between the correlated and uncorrelated processes becomes evident when looking at the variance over survival probabilities with increased noise averaging.

For uncorrelated errors,

$$\mathbb{V}[\mathcal{P}_U] \approx \frac{J^2\sigma^4}{n}\left\{\frac{4}{9}\left(\frac{1}{2} + \frac{\pi^2}{96}\right)^2 + \frac{1}{J}\left[3\left(\frac{7}{36} + \frac{\pi^4}{576}\right)\right.\right.$$
$$- \frac{8}{9}\left(\frac{1}{2} + \frac{\pi^2}{96}\right)^2\right] + \frac{(n-1)}{J}\left[\frac{7}{36} + \frac{\pi^4}{576}\right.$$
$$\left.\left. - \frac{4}{9}\left(\frac{1}{2} + \frac{\pi^2}{96}\right)^2\right]\right\}, \tag{B20}$$

noting that in the limit $n \to \infty$, the variance scaling saturates at a value proportional to $\frac{1}{J}$ relative to the starting variance. For correlated errors,

$$\mathbb{V}[\mathcal{P}_C] \approx \frac{J^2\sigma^4}{n}\left(\frac{12}{9}\left(\frac{1}{2} + \frac{\pi^2}{96}\right)^2 + \frac{1}{J}\left[3\left(\frac{7}{36} + \frac{\pi^4}{576}\right)\right.\right.$$
$$- \frac{8}{3}\left(\frac{1}{2} + \frac{\pi^2}{96}\right)^2\right] + (n-1)\left\{\frac{4}{9}\left(\frac{1}{2} + \frac{\pi^2}{96}\right)^2\right.$$
$$\left.\left. + \frac{1}{J}\left[\frac{7}{36} + \frac{\pi^4}{576} - \frac{8}{9}\left(\frac{1}{2} + \frac{\pi^2}{96}\right)^2\right]\right\}\right), \tag{B21}$$

again tending towards a constant; however, this occurs at a significantly smaller number of noise averages than for uncorrelated noise and saturates at a much larger variance proportional to $1 + \frac{1}{J}$ relative to the starting variance.

Using the revised model, the noise-averaged survival-probability distributions under correlated noise remain Γ distributed with an updated scale parameter. While this is yet to be shown explicitly for the uncorrelated case, the behavior is approximated in the limit of large n and J, with $n < J$ by modifying the distribution in Eq. (3), to yield

$$\mathcal{P}_C \sim \Gamma\left(a = 1, b = \frac{2}{3}J\sigma^2\left(\frac{1}{2} + \frac{\pi^2}{96}\right)\right), \tag{B22}$$

$$\mathcal{P}_U \sim \Gamma\left(a = n, b = \frac{2}{3n}J\sigma^2\left(\frac{1}{2} + \frac{\pi^2}{96}\right)\right). \tag{B23}$$

The normalized Γ distributions for correlated error processes shown by solid gray lines in Figs. 2(a)–2(c) were calculated from first principles using (B22) with no free parameters. The distributions for the uncorrelated error process in red were calculated from an altered version of (B23), which was modified for higher bandwidth noise that took multiple values of δ in a single error map. This made use of the relation

$$\delta_1 \pm \delta_2 \sim \mathcal{N}(0, 2\rho^2) \equiv \sqrt{2}\mathcal{N}(0, \rho^2) \tag{B24}$$

such that the multiple values of δ could be expressed as

$$\delta_1 \pm \delta_2 \equiv \sqrt{2}\delta, \tag{B25}$$

with $\delta \sim \mathcal{N}(0, \rho^2)$, from which point the previous method can be followed. The equivalence in Eq. (B25) occurs because δ_1 and δ_2 are independent samples from a Gaussian distribution, meaning their combination is also Gaussian distributed.

TABLE V. Gate parameters required to construct a target rotation about the x axis by angle θ_t using different pulse constructions. An additional $\pi/2$ shift in ϕ is required for rotations about the y axis. Here $k = \arcsin\left[\frac{\sin[\theta_t/2]}{2}\right]$, $\phi_k = \arccos\left[\frac{-\theta_t}{4\pi}\right]$, and for WAMF DCGs the target rotations $\theta_t = (\frac{\pi}{4}, \frac{\pi}{2}, \pi)$ have $X_0 = (2\frac{1}{4}, 2\frac{1}{2}, 3)\pi$ and $X_3 = (0.36, 0.64, 1)\pi$ determined explicitly.

Gate construction Gate	$(\theta_1, \Omega_1, \phi_1)$	$(\theta_2, \Omega_2, \phi_2)$	$(\theta_3, \Omega_3, \phi_3)$	$(\theta_4, \Omega_4, \phi_4)$
primitive	$(\theta_t, \Omega, 0)$			
CORPSE	$(2\pi + \theta_t/2 - k, \Omega, 0)$	$(2\pi - 2k, \Omega, \pi)$	$(\theta_t/2 - k, \Omega, 0)$	
WAMF	$(\frac{X_0+X_3}{4}, \Omega, 0)$	$(\frac{X_0-X_3}{2}, \frac{X_0-X_3}{X_0+X_3}\Omega, 0)$	$(\frac{X_0+X_3}{4}, \Omega, 0)$	
BB1	$(\theta_t, \Omega, 0)$	(π, Ω, ϕ_k)	$(2\pi, \Omega, 3\phi_k)$	(π, Ω, ϕ_k)

APPENDIX C: THE DCG CONSTRUCTIONS EMPLOYED IN THIS WORK

Three error suppressing DCGs are utilized in this work: CORPSE and WAMF gates, which suppress detuning errors, and BB1 gates, which suppress overrotation errors. For each of these constructions, the target angle $\theta_t = \pi, \pi/2$ gates are created as multisegment pulses described by the segments' rotation angles θ_i, phase angles ϕ_i, and Rabi frequencies Ω_i normalized to the maximum frequency Ω. The constructions of the different gates are shown in Table V. To ensure that the error suppressing aspects of the DCGs are maintained for all Clifford gates, the identity gate is implemented as a rotary spin echo by concatenating a π rotation about the x axis with its inverse $-\pi$ rotation. While this results in a net zero rotation, effectively identical to the simple wait time used for primitive $\hat{\mathbb{I}}$ gates, it makes the identity operation first-order insensitive to detuning errors during its implementation. The physical motivation here is that if a qubit is remaining idle at any point during a multiqubit circuit, it may be preferable to continuously drive this type of rotary spin echo to ensure that it does not accumulate phase errors during its idle period.

APPENDIX D: INFLUENCE OF QUANTUM PROJECTION NOISE

Quantum projection noise (QPN) describes the intrinsic uncertainty in qubit measurements due to the binomial nature of measurement outcomes [64] and its scaling with the number of samples. The variance of a measurement due to QPN is $\frac{p(1-p)}{r}$, where p is the true state projection onto the z axis of the Bloch sphere and r is the number of identical measurements performed. Our work studies variances over distributions of noise-averaged survival probabilities, and consequently it is necessary to demonstrate that we were not limited by QPN bounds.

We consider the CORPSE data shown in Fig. 3(c); in order to ensure that our results are not measurement artifacts from quantum projection noise, we average each sequence and noise realization combination $r = 220$ times. At this number of repetitions, the largest possible projection noise variance is given by $\frac{0.5(1-0.5)}{220} = 1 \times 10^{-3}$. In addition to the worst-case QPN, we compare the variance scaling results for the
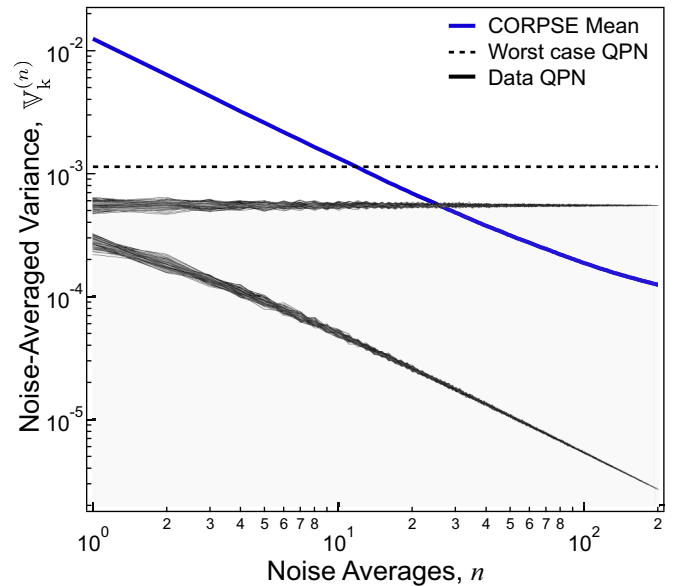

FIG. 7. Quantum projection noise limits for measured survival probabilities with the CORPSE DCG. Comparison of mean CORPSE variance scaling from Fig. 3(c) (blue) to QPN variance limits given by $p(1-p)/r$. The dashed line is the worst case QPN for $r = 220$ when $p = 0.5$. Black lines show additional QPN limits where, for each n, $p(1-p)/r$ is calculated for 100 randomizations of noise realizations. The lower line scaling as $1/n$ is divided by $n \times r$ rather than r.

CORPSE DCG under simultaneously applied correlated and uncorrelated noise to the QPN given by the measured survival probabilities. Figure 7 shows the mean trajectory for the CORPSE variance scaling under the combined noise process presented in Fig. 3(c) in dark blue. The dashed black line gives the worst-case QPN and the two other sets of trajectories are calculated directly from the measured probabilities. For these, the QPN was calculated at each n for 100 randomizations of noise realizations to reduce bias, and the 100 values are plotted. The lower set of trajectories is divided by $n \times r$ rather than just r. Our results are well above this lower limit, suggesting that this is the most valid measurement of setting our QPN limit. Furthermore, we note that the saturation observed at large values of n is not set by any static QPN bound limiting our measurements.

[1] P. W. Shor, Scheme for reducing decoherence in quantum computer memory, Phys. Rev. A **52**, R2493 (1995).

[2] P. W. Shor, *Proceedings of 37th Conference on Foundations of Computer Science* (IEEE, Piscataway, 1996), pp. 56–65.

[3] A. M. Steane, Error Correcting Codes in Quantum Theory, Phys. Rev. Lett. **77**, 793 (1996).

[4] A. Steane, Multiple-particle interference and quantum error correction, Proc. R. Soc. A **452**, 2551 (1996).

[5] D. Aharonov and M. Ben-Or, *Proceedings of the 29th Annual ACM Symposium on Theory of Computing* (ACM, New York, 1997), pp. 176–188.

[6] D. Gottesman, Theory of fault-tolerant quantum computation, Phys. Rev. A **57**, 127 (1998).

[7] J. Preskill, Reliable quantum computers, Proc. R. Soc. A **454**, 385 (1998).

[8] D. A. Lidar and T. A. Brun, *Quantum Error Correction* (Cambridge University Press, New York, 2013).

[9] E. T. Campbell, B. M. Terhal, and C. Vuillot, Roads towards fault-tolerant universal quantum computation, Nature (London) **549**, 172 (2017).

[10] J. Preskill, Sufficient condition on noise correlations for scalable quantum computing, Quantum Inf. Comput. **13**, 0181 (2013).

[11] J. Wallman, C. Granade, R. Harper, and S. T. Flammia, Estimating the coherence of noise, New J. Phys. **17**, 113020 (2015).

[12] T. Proctor, K. Rudinger, K. Young, M. Sarovar, and R. Blume-Kohout, What Randomized Benchmarking Actually Measures, Phys. Rev. Lett. **119**, 130502 (2017).

[13] D. Greenbaum and Z. Dutton, Modeling coherent errors in quantum error correction, Quantum Sci. Technol. **3**, 015007 (2018).

[14] E. Huang, A. C. Doherty, and S. T. Flammia, Performance of quantum error correction with coherent errors, Phys. Rev. A **99**, 022313 (2019).

[15] C. T. Chubb and S. T Flammia, Statistical mechanical models for quantum codes with correlated noise, arXiv:1809.10704.

[16] D. M. Debroy, M. Li, M. Newman, and K. R. Brown, Stabilizer Slicing: Coherent Error Cancellations in Low-Density Parity-Check Stabilizer Codes, Phys. Rev. Lett. **121**, 250502 (2018).

[17] S. Majumder, L. Andreta de Castro, and K. R. Brown, Real-time calibration with spectator qubits, arXiv:1907.03864.

[18] H. K. Ng and J. Preskill, Fault-tolerant quantum computation versus Gaussian noise, Phys. Rev. A **79**, 032318 (2009).

[19] D. Aharonov, A. Kitaev, and J. Preskill, Fault-Tolerant Quantum Computation with Long-Range Correlated Noise, Phys. Rev. Lett. **96**, 050504 (2006).

[20] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, Surface codes: Towards practical large-scale quantum computation, Phys. Rev. A **86**, 032324 (2012).

[21] P. Schindler, J. T. Barreiro, T. Monz, V. Nebendahl, D. Nigg, M. Chwalla, M. Hennrich, and R. Blatt, Experimental repetitive quantum error correction, Science **332**, 1059 (2011).

[22] S. Bravyi, M. Englbrecht, R. König, and N. Peard, Correcting coherent errors with surface codes, npj Quantum Inf. **4**, 55 (2018).

[23] K. R. Brown, A. W. Harrow, and I. L. Chuang, Arbitrarily accurate composite pulse sequences, Phys. Rev. A **70**, 052318 (2004).

[24] K. Khodjasteh and L. Viola, Dynamically Error-Corrected Gates for Universal Quantum Computation, Phys. Rev. Lett. **102**, 080501 (2009).

[25] J. T. Merrill and K. R. Brown, in *Quantum Information and Computation for Chemistry*, edited by S. Kais (Wiley, New Jersey, 2014), pp. 241–294.

[26] X. Wang, L. S. Bishop, J. P. Kestner, E. Barnes, K. Sun, and S. Das Sarma, Composite pulses for robust universal control of singlet-triplet qubits, Nat. Commun. **3**, 997 (2012).

[27] A. Soare, H. Ball, D. Hayes, J. Sastrawan, M. C. Jarratt, J. J. McLoughlin, X. Zhen, T. J. Green, and M. J. Biercuk, Experimental noise filtering by quantum control, Nat. Phys. **10** (2014).

[28] H. K. Ng, D. A. Lidar, and J. Preskill, Combining dynamical decoupling with fault-tolerant quantum computation, Phys. Rev. A **84**, 012305 (2011).

[29] N. C. Jones, R. Van Meter, A. G. Fowler, P. L. McMahon, J. Kim, T. D. Ladd, and Y. Yamamoto, Layered Architecture for Quantum Computing, Phys. Rev. X **2**, 031007 (2012).

[30] J. Emerson, R. Alicki, and K. Życzkowski, Scalable noise estimation with random unitary operators, J. Opt. B **7**, S347 (2005).

[31] C. Dankert, R. Cleve, J. Emerson, and E. Livine, Exact and approximate unitary 2-designs and their application to fidelity estimation, Phys. Rev. A **80**, 012304 (2009).

[32] H. Ball, T. M. Stace, S. T. Flammia, and M. J. Biercuk, Effect of noise correlations on randomized benchmarking, Phys. Rev. A **93**, 022303 (2016).

[33] C. Kabytayev, T. J. Green, K. Khodjasteh, M. J. Biercuk, L. Viola, and K. R. Brown, Robustness of composite pulses to time-dependent control noise, Phys. Rev. A **90**, 012316 (2014).

[34] G. A. Paz-Silva and L. Viola, General Transfer-Function Approach to Noise Filtering in Open-Loop Quantum Control, Phys. Rev. Lett. **113**, 250501 (2014).

[35] T. J. Green, J. Sastrawan, H. Uys, and M. J. Biercuk, Arbitrary quantum control of qubits in the presence of universal noise, New J. Phys. **15**, 095004 (2013).

[36] A. G. Kofman and G. Kurizki, Unified Theory of Dynamically Suppressed Qubit Decoherence in Thermal Baths, Phys. Rev. Lett. **93**, 130406 (2004).

[37] T. Green, H. Uys, and M. J. Biercuk, High-Order Noise Filtering in Nontrivial Quantum Logic Gates, Phys. Rev. Lett. **109**, 020501 (2012).

[38] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, Randomized benchmarking of quantum gates, Phys. Rev. A **77**, 012307 (2008).

[39] E. Magesan, J. M. Gambetta, B. R. Johnson, C. A. Ryan, J. M. Chow, S. T. Merkel, M. P. da Silva, G. A. Keefe, M. B. Rothwell, T. A. Ohki, M. B. Ketchen, and M. Steffen, Efficient Measurement of Quantum Gate Error by Interleaved Randomized Benchmarking, Phys. Rev. Lett. **109**, 080505 (2012).

[40] E. Magesan, J. M. Gambetta, and J. Emerson, Scalable and Robust Randomized Benchmarking of Quantum Processes, Phys. Rev. Lett. **106**, 180504 (2011).

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

[42] S. T. Merkel, J. M. Gambetta, J. A. Smolin, S. Poletto, A. D. Córcoles, B. R. Johnson, C. A. Ryan, and M. Steffen, Self-consistent quantum process tomography, Phys. Rev. A **87**, 062119 (2013).

[43] S. Kimmel, M. P. da Silva, C. A. Ryan, B. R. Johnson, and T. Ohki, Robust Extraction of Tomographic Information via Randomized Benchmarking, Phys. Rev. X **4**, 011050 (2014).

[44] S. Sheldon, L. S. Bishop, E. Magesan, S. Filipp, J. M. Chow, and J. M. Gambetta, Characterizing errors on qubit operations via iterative randomized benchmarking, Phys. Rev. A **93**, 012301 (2016).

[45] 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).

[46] S. Mavadia, C. L. Edmunds, C. Hempel, H. Ball, F. Roy, T. M. Stace, and M. J. Biercuk, Experimental quantum verification in the presence of temporally correlated noise, npj Quantum Inf. **4**, 7 (2018).

[47] E. Onorati, A. H. Werner, and J. Eisert, Randomized Benchmarking for Individual Quantum Gates, Phys. Rev. Lett. **123**, 060501 (2019).

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

[49] D. W. Allan, Statistics of Atomic Frequency Standards, Proc. Inst. Elect. Electron. Eng. **54**, 221 (1966).

[50] J. Rutman, Characterization of phase and frequency instabilities in precision frequency sources: Fifteen years of progress, Proc. IEEE **66**, 1048 (1978).

[51] J. A. Sedlacek, J. Stuart, D. H. Slichter, C. D. Bruzewicz, R. McConnell, J. M. Sage, and J. Chiaverini, Evidence for multiple mechanisms underlying surface electric-field noise in ion traps, Phys. Rev. A **98**, 063430 (2018).

[52] S. Schlör, J. Lisenfeld, C. Müller, A. Bilmes, A. Schneider, D. P. Pappas, A. V. Ustinov, and M. Weides, Correlating Decoherence in Transmon Qubits: Low Frequency Noise by Single Fluctuators, Phys. Rev. Lett. **123**, 190502 (2019).

[53] J. J. Burnett, A. Bengtsson, M. Scigliuzzo, D. Niepce, M. Kudra, P. Delsing, and J. Bylander, Decoherence benchmarking of superconducting qubits, npj Quantum Inf. **5**, 54 (2019).

[54] S. Wölk, C. Piltz, T. Sriarunothai, and C. Wunderlich, State selective detection of hyperfine qubits, J. Phys. B **48**, 075101 (2015).

[55] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, Scikit-learn: Machine learning in Python, J. Mach. Learn. Res. **12**, 2825 (2011).

[56] H. K. Cummins and J. A. Jones, Use of composite rotations to correct systematic errors in NMR quantum computation, New J. Phys. **2**, 6 (2000).

[57] H. Ball and M. J Biercuk, Walsh-synthesized noise filters for quantum logic, EPJ Quantum Technol. **2**, 11 (2015).

[58] S. Wimperis, Broadband, narrowband, and passband composite pulses for use in advanced NMR experiments, J. Magn. Reson. Ser. A **109**, 221 (1994).

[59] T. J. Proctor, A. Carignan-Dugas, K. Rudinger, E. Nielsen, R. Blume-Kohout, and K. Young, Direct Randomized Benchmarking for Multiqubit Devices, Phys. Rev. Lett. **123**, 030503 (2019).

[60] L. Postler, Á. Rivas, P. Schindler, A. Erhard, R. Stricker, D. Nigg, T. Monz, R. Blatt, and M. Müller, Experimental quantification of spatial correlations in quantum dynamics, Quantum **2**, 90 (2018).

[61] M. Bando, T. Ichikawa, Y. Kondo, and M. Nakahara, Concatenated composite pulses compensating simultaneous systematic errors, J. Phys. Soc. Jpn. **82**, 014004 (2013).

[62] M. Reiher, N. Wiebe, K. M. Svore, D. Wecker, and M. Troyer, Elucidating reaction mechanisms on quantum computers, Proc. Natl. Acad. Sci. USA **114**, 7555 (2017).

[63] J. T. Merrill, S. C. Doret, G. Vittorini, J. P. Addison, and K. R. Brown, Transformed composite sequences for improved qubit addressing, Phys. Rev. A **90**, 040301(R) (2014).

[64] W. M. Itano, J. C. Bergquist, J. J. Bollinger, J. M. Gilligan, D. J. Heinzen, F. L. Moore, M. G. Raizen, and D. J. Wineland, Quantum projection noise: Population fluctuations in two-level systems, Phys. Rev. A **47**, 3554 (1993).