



CHORUS

This is the accepted manuscript made available via CHORUS. The article has been published as:

Effect of noise correlations on randomized benchmarking

Harrison Ball, Thomas M. Stace, Steven T. Flammia, and Michael J. Biercuk

Phys. Rev. A **93**, 022303 — Published 1 February 2016

DOI: [10.1103/PhysRevA.93.022303](https://doi.org/10.1103/PhysRevA.93.022303)

The effect of noise correlations on randomized benchmarking

Harrison Ball,^{1,2} Thomas M. Stace,³ Steven T. Flammia,¹ and Michael J. Biercuk^{1,2,*}

¹*ARC Centre of Excellence for Engineered Quantum Systems,
School of Physics, The University of Sydney, NSW 2006 Australia*

²*Australian National Measurement Institute, West Lindfield, NSW 2070*

³*ARC Centre of Excellence for Engineered Quantum Systems,
School of Physics and Mathematics, The University of Queensland, QLD 4072 Australia*

Among the most popular and well studied quantum characterization, verification and validation techniques is randomized benchmarking (RB), an important statistical tool used to characterize the performance of physical logic operations useful in quantum information processing. In this work we provide a detailed mathematical treatment of the effect of temporal noise correlations on the outcomes of RB protocols. We provide a fully analytic framework capturing the accumulation of error in RB expressed in terms of a three-dimensional random walk in “Pauli space.” Using this framework we derive the probability density function describing RB outcomes (averaged over noise) for both Markovian and correlated errors, which we show is generally described by a gamma distribution with shape and scale parameters depending on the correlation structure. Long temporal correlations impart large nonvanishing variance and skew in the distribution towards high-fidelity outcomes – consistent with existing experimental data – highlighting potential finite-sampling pitfalls and the divergence of the mean RB outcome from worst-case errors in the presence of noise correlations. We use the Filter-transfer function formalism to reveal the underlying reason for these differences in terms of effective coherent averaging of correlated errors in certain random sequences. We conclude by commenting on the impact of these calculations on the utility of single-metric approaches to quantum characterization, verification, and validation.

I. Introduction

Quantum verification and validation protocols are a vital tool for characterizing quantum devices. These take many forms [1–8], but one of the most popular due to its efficiency is randomized benchmarking (RB). In this approach, developed originally by Knill et al. [9], and expanded theoretically by various authors [10–13], the average error probability of a quantum gate (*e.g.* a bit flip) is estimated by implementing a randomly sampled gate sequence from the set of Clifford operations, and measuring the difference between the ideal transformation and the actual result. Averaging over many randomized sequences yields information about the underlying gate fidelity.

RB has become so important in the experimental community [9, 14–23] that despite the experimental complexity it is now common to simply quote a measured gate error, p_{RB} , resulting from a RB measurement for a particular experimental system, and relate this number to tight bounds such as fault-tolerance error thresholds in quantum error correction [24]. Of late, reported values of p_{RB} have compared favorably with these thresholds, and have been used to justify the scalability of particular experimental platforms. Underlying this entire approach is the subtle but central assumption that errors are *uncorrelated* [9]: an assumption which is generally violated by a wide variety of realistic error processes with long temporal correlations [14, 25, 26]. Violation of this assumption has been noted previously as a cause of increased variance of outcomes over randomizations, but until now there has not been a quantitative means to understand the impact of such temporal correlations, a detailed physical mechanism to explain why such distortion of RB results can appear, or a clear under-

standing of the impact of this observation on the applicability of RB.

In this manuscript we examine the impact of relaxing the Markovian-error assumption by studying its effect on the distribution of measurement outcomes over randomizations. We find that while all randomizations (meeting certain criteria) are valid within the RB framework, they exhibit starkly different susceptibility to error when those errors exhibit temporal correlations over multiple individual Clifford operations. We provide a detailed mathematical treatment of error accumulation in RB, using a general model treated in the specific case of dephasing errors. We demonstrate how the reduction of fidelity for a particular sequence may be given a geometric interpretation, taking the form of a random walk in a 3D Cartesian coordinate system. The steps of this walk correspond to the appearance of Pauli \hat{X} , \hat{Y} , \hat{Z} errors derived from interleaved dephasing operators in the sequence of Clifford operations, and the overall statistics are determined by the underlying correlations in the noise process. Our treatment includes both extremal cases of uncorrelated (Markovian) and systematic (DC) errors, as well as generic correlations interpolating between these limiting cases and captured through a noise power spectral density. Our results provide simple analytic forms for the probability density functions of fidelity outcomes over RB randomizations as a gamma distribution, building connections to engineering literature on failure analysis. We describe the impact of these observations on the interpretation of RB measurement outcomes in various noise environments and highlight the disconnect between measured RB error and metrics relevant to fault-tolerance such as worst-case errors.

* Correspondence to: michael.biercuk@sydney.edu.au

II. Randomized Benchmarking Procedure

A common experimental implementation of single-qubit RB involves measuring the fidelity of net operations composed from random sequences of Clifford operations. For a sequence of length J , the net operation is written as the product

$$\mathcal{S}_\eta \equiv \prod_{j=1}^J \hat{\mathcal{C}}_{\eta_j} \quad (1)$$

of Clifford operators $\hat{\mathcal{C}}_{\eta_j}$ (see Table III) indexed by the sequence $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_J)$, where the η_j are random variables uniformly sampled from the set $\{1, 2, \dots, 24\}$ labelling the elements of the Clifford group. Typically one makes the constraint on $\boldsymbol{\eta}$ that, absent any error processes, the ideal RB sequence performs the identity operation $\mathcal{S}_\eta \equiv \mathbb{I}$. Of the 24^J total sequence combinations for a given J , only a small subset of $k \ll 24^J$ random Clifford sequences is implemented in practice. Each sequence is repeated (with appropriate qubit reinitialization) n times to build an ensemble of fidelity measurements sampling and the underlying error process. These measurement outcomes are then averaged together to obtain useful information from projective qubit measurements, with the mean fidelity retained for each random sequence.

We represent the space of measurement outcomes by the $(k \times n)$ matrix

$$\mathbf{F}^{(J)} \equiv \begin{matrix} & \delta_1 & \delta_2 & \dots & \delta_n \\ \begin{matrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_k \end{matrix} & \begin{pmatrix} F_{1,1} & F_{1,2} & \dots & F_{1,n} \\ F_{2,1} & F_{2,2} & \dots & F_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ F_{k,1} & F_{k,2} & \dots & F_{k,n} \end{pmatrix} \end{matrix} \quad (2)$$

where element $F_{i,j} \in [0, 1]$ is the measured fidelity when implementing the i th Clifford sequence $\boldsymbol{\eta}_i = (\eta_{i,1}, \eta_{i,2}, \dots, \eta_{i,J})$ in the presence of the j th realization of the error process, denoted by δ_j . For clarity, we make the distinction between the *measured* fidelity $F_{i,j}$ (e.g. obtained in experiment via projective qubit measurements) and the *calculated* fidelity (see Eq. 5) serving as a proxy for $F_{i,j}$ in our analytic framework. The process of repeating each $\boldsymbol{\eta}_i$ and averaging the fidelity outcomes *over the finite sample of noise realizations* therefore corresponds to a measurement of the “noise-averaged” fidelity for the i th sequence, represented by the quantity

$$\overline{\mathbf{F}}_{i, \langle \cdot \rangle}^{(J)} \equiv \frac{1}{n} \sum_{j=1}^n F_{i,j}, \quad i \in \{1, \dots, k\} \quad (3)$$

where the angle brackets $\langle \cdot \rangle$ in the second subscript denote averaging over the columns of $\mathbf{F}^{(J)}$. Similarly the mean fidelity averaged over both Clifford sequences and errors is denoted

$$\hat{\mu}^{(J)} \equiv \overline{\mathbf{F}}_{\langle \cdot \rangle, \langle \cdot \rangle}^{(J)} = \frac{1}{kn} \sum_{i=1}^k \sum_{j=1}^n F_{i,j} \quad (4)$$

where angle brackets in both subscripts indicate averaging over both rows and columns (that is, all elements) of $\mathbf{F}^{(J)}$. The quantity $\hat{\mu}^{(J)}$ is therefore an *estimator* of the true mean fidelity $\mu^{(J)} \equiv \langle \mathcal{F} \rangle_{\boldsymbol{\eta}, \delta}$, formally obtained as the expectation over all possible fidelity outcomes \mathcal{F} defined on the support of the random variables $\boldsymbol{\eta}$ and δ .

In the standard RB procedure, measurements of $\hat{\mu}^{(J)}$ for increasing J are fitted to an exponential decay from which the mean gate error p_{RB} is extracted as the decay constant. Implicit in this procedure, however, is the assumption that, for any random set of Clifford sequences, the resulting distribution of fidelity outcomes represents the underlying error process *fairly*, and the total mean $\hat{\mu}^{(J)}$ is reasonably representative of any individual sequence. That is, the distribution of values $\overline{\mathbf{F}}_{i, \langle \cdot \rangle}^{(J)}$ is symmetric about $\hat{\mu}^{(J)}$ with small relative variance for any random set $\{\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_k\}$.

In this paper we show $\hat{\mu}^{(J)}$ to be an unbiased and effective estimator only when the error process is truly Markovian. That is, it possesses no temporal correlations. We provide a detailed analytic calculation of the impact of realistic, correlated noise on the distribution of outcomes over different randomizations. To examine this we introduce a physical model compatible with the experimental procedure, and which permits accounting for the presence of noise correlations.

III. Physical model

In this section we develop the mathematical framework used to model and investigate the impact of noise correlations on RB. Our model assumes a Unitary error process with temporal correlations, reflecting typical experimental conditions where qubit rotations, e.g. from a fluctuating classical field, dominate. This process is generically described by a power spectral density (PSD) capturing the various correlation timescales. Our error model is fully general and considers *universal* (multi-axis) errors. For simplicity of technical presentation we treat single-axis dephasing in the main text and provide the full presentation of universal noise in Appendix D.

The main result of this section is the geometric interpretation of error accumulation in RB sequences in terms of a random walk in 3 dimensions (a step taken for each Clifford), with step lengths and directionality inheriting the correlation structure of the error process. This result is independent of any choice of representation of the *Clifford group*, though our numeric simulations later used to verify it do rely on a particular representation (see Table III).

A. Noise-averaged fidelity $\langle \mathcal{F} \rangle$

We begin by introducing our metric to quantify the reduction in fidelity for a given $\boldsymbol{\eta}$ and δ . We employ the trace fidelity

$$\mathcal{F}(\boldsymbol{\eta}, \delta) \equiv \left| \frac{1}{2} \text{Tr} \left(\mathcal{S}_\eta^\dagger \tilde{\mathcal{S}}_{\boldsymbol{\eta}, \delta} \right) \right|^2 = \frac{1}{4} \left| \text{Tr} \left(\tilde{\mathcal{S}}_{\boldsymbol{\eta}, \delta} \right) \right|^2 \quad (5)$$

capturing the overlap between ideal, \mathcal{S}_η , and noise-affected sequences, $\tilde{\mathcal{S}}_{\boldsymbol{\eta}, \delta}$, via the Hilbert-Schmidt inner product. For

inputs η_i and δ_j this provides us with a computational metric corresponding to measurements $F_{i,j}$ obtained in experiment. Consequently our proxy for the measured noise-averaged fidelity $\overline{F}_{i,(\cdot)}^{(J)}$ takes the computational form

$$\langle \mathcal{F} \rangle_{\delta,n} = \frac{1}{4} \left\langle \left| \text{Tr} \left(\tilde{\mathcal{S}}_{\eta,\delta} \right) \right|^2 \right\rangle_{\delta,n} \quad (6)$$

where $\langle \cdot \rangle_{\delta,n}$ denotes an ensemble average over n realizations of δ , and explicit reference to η and δ in the argument of \mathcal{F} has been dropped. Henceforth we also drop the subscripts and denote the calculated noise-averaged fidelity simply by $\langle \mathcal{F} \rangle$. In the following sections we proceed with our main task: deriving the probability density function (PDF) of $\langle \mathcal{F} \rangle$, and its dependence on the correlation structure of δ .

B. Error model

To reiterate, a Clifford sequence η implements the operation $\mathcal{S}_\eta \equiv \prod_{i=1}^J \hat{C}_{\eta_i}$, where \hat{C}_{η_i} is the η_i^{th} Clifford generator. The random variables η_i are uniformly sampled from the set $\{1, 2, \dots, 24\}$, and are assumed to be independent and identically distributed (i.i.d.), subject to the technical constraint $\mathcal{S}_\eta \equiv \mathbb{I}$. Unitary errors are implemented by interleaving \mathcal{S}_η with a sequence of stochastic qubit rotations, yielding the noise-affected operation

$$\tilde{\mathcal{S}}_{\eta,\delta} \equiv U_1 \hat{C}_{\eta_1} U_2 \hat{C}_{\eta_2} \dots U_J \hat{C}_{\eta_J}. \quad (7)$$

This approach holds for arbitrary unitaries, U_j , enacting rotations in any Cartesian direction; a full treatment for universal noise models appears in Appendix D. However, in the following we restrict ourselves to the subset of unitaries enacting a sequence of *dephasing* rotations, $U_j \equiv \exp(-i\delta_j \hat{Z})$, parameterized by the time-series vector $\delta = (\delta_1, \delta_2, \dots, \delta_J)$. We assume the error process is wide-sense stationary (*i.e.* the for errors δ_i and δ_j the two-point correlation function $\langle \delta_i \delta_j \rangle = \langle \delta_i \delta_{i-k} \rangle$ depends only on the *time difference* $k = i - j$), and has zero mean.

Temporal noise correlations are established by introducing correlations between the elements of δ . In this work we treat three distinct cases

- Markovian process*: Elements $\delta_j \sim \mathcal{N}(0, \sigma^2)$ are i.i.d. Gaussian-distributed errors with zero mean and variance σ^2 , and completely uncorrelated between distinct Clifford gates in any sequence \mathcal{S}_η (correlation length 1).
- DC process*: Elements $\delta_j \equiv \delta \sim \mathcal{N}(0, \sigma^2)$ are identical over a given sequence \mathcal{S}_η (maximally correlated, correlation length J), but are i.i.d. (uncorrelated) Gaussian-distributed errors with zero mean and variance σ^2 over different instances.
- Generically-correlated process*: Correlations between elements of δ separated by a time interval of “ k gates” in \mathcal{S}_η are generically specified by an autocorrelation function $C_\delta(k) \equiv \langle \delta_j \delta_{j+k} \rangle$, in terms of which δ may be described by a power spectral density (PSD) $S(\omega)$.

As we will show, the noise interaction “steers” the operator product away from the identity gate performed by the ideal sequence \mathcal{S}_η and reduces the operational fidelity in some way characteristic of the correlation structure of δ .

C. Analytic Expression for Sequence Fidelity

1. Series expansion for Sequence Fidelity

We begin by obtaining an approximation for $\langle \mathcal{F} \rangle$. As RB is intended to estimate errors too small to resolve in a single gate implementation, we assume the noise strength per gate, σ , is small. Over long gate sequences the *accumulation* of errors may be quantified by $J\sigma^2$, which can be large for sufficiently large J . Assuming $J\sigma^2 \lesssim 1$, however, we can make analytical progress by expressing the error unitaries in terms of a power series truncated at $\mathcal{O}(\sigma^4)$. The noise-affected sequence Eq. 7 is therefore approximated by

$$\tilde{\mathcal{S}}_{\eta,\delta} \approx \prod_{j=1}^J \left(\mathbb{I} + i\delta_j \hat{Z} - \frac{\delta_j^2}{2} \hat{Z}^2 - \frac{i\delta_j^3}{6} \hat{Z}^3 + \frac{\delta_j^4}{24} \hat{Z}^4 \right) \hat{C}_{\eta_j} \quad (8)$$

The full expansion of Eq. 8 generates products of order up to $\mathcal{O}(\prod_{j=1}^J \delta_j^4) = \mathcal{O}(\sigma^{4J})$. Let $\xi_{k_1, k_2, \dots, k_m}^{(n)}$ denote the $\mathcal{O}(\prod_{\rho=1}^m \delta_{j_\rho}^{k_\rho}) = \mathcal{O}(\sigma^n)$ product due to cross-multiplying terms like $(\delta_{j_1} \hat{Z})^{k_1} (\delta_{j_2} \hat{Z})^{k_2} \dots (\delta_{j_m} \hat{Z})^{k_m}$, where $\sum_{\rho=1}^m k_\rho = n$. Retaining terms only up to fourth order ($n = 4$), consistent with the order of our original Taylor approximation, we thereby obtain

$$\begin{aligned} \tilde{\mathcal{S}}_{\eta,\delta} \approx & \xi^{(0)} + \xi_1^{(1)} + \xi_{1,1}^{(2)} + \xi_2^{(2)} + \xi_{1,1,1}^{(3)} + \xi_{2,1}^{(3)} + \xi_3^{(3)} \\ & + \xi_{1,1,1,1}^{(4)} + \xi_{1,1,2}^{(4)} + \xi_{2,2}^{(4)} + \xi_4^{(4)} + \mathcal{O}(\sigma^6). \end{aligned} \quad (9)$$

The first term in this expansion is identical to the ideal Clifford sequence, $\xi^{(0)} \equiv \mathcal{S}_\eta$. Higher-order terms capture successive error contributions and are composed of blocks, or subsequences, within \mathcal{S}_η interrupted by one or more \hat{Z} errors, indicated by the subscripts. To evaluate Eq. 5 we must obtain expressions for the quantities $Q_{k_1, k_2, \dots, k_m}^{(n)} \equiv \frac{1}{2} \text{Tr} \left(\xi_{k_1, k_2, \dots, k_m}^{(n)} \right)$. Detailed derivations for these terms are given in Appendix C.

2. Mapping from Clifford Sequence to “Pauli space”

We compute the terms in the power series by relating a given Clifford sequence to a random walk in “Pauli space”. For illustrative purposes, we consider in detail only the quadratic terms:

$$\xi_{1,1}^{(2)} = - \sum_{j < k} \delta_j \delta_k C_{1,j-1} \hat{Z} C_{j,k-1} \hat{Z} C_{k,J} \quad (10)$$

where the ordered summand runs over $1 \leq j < k \leq J$, and we have defined the Clifford *subsequence* operators

$$C_{jk} \equiv \hat{C}_{\eta_j} \dots \hat{C}_{\eta_k}, \quad 1 \leq j \leq k \leq J. \quad (11)$$

We now define the cumulative operators K_m giving the product of the first m Clifford operations

$$K_m \equiv C_{1,m} = \hat{C}_{\eta_1} \dots \hat{C}_{\eta_m}, \quad (12)$$

where $K_0 \equiv K_J \equiv \mathbb{I}$, and $K_m^\dagger = K_m^{-1}$ is also a Clifford operator (since it is a product of Clifford operators). Any subsequence C_{jk} therefore ‘‘factorizes’’ as

$$C_{jk} = \hat{C}_{\eta_{j-1}}^\dagger \dots \hat{C}_{\eta_1}^\dagger \hat{C}_{\eta_1} \dots \hat{C}_{\eta_{j-1}} \hat{C}_{\eta_j} \dots \hat{C}_{\eta_k} = K_{j-1}^\dagger K_k \quad (13)$$

allowing us to rewrite

$$C_{1,j-1} \hat{Z} C_{j,k-1} \hat{Z} C_{k,J} = \mathbf{P}_j \mathbf{P}_k \quad (14)$$

where the operators on the right hand side are defined by

$$\mathbf{P}_m \equiv K_{m-1} \hat{Z} K_{m-1}^\dagger \in \{\pm \hat{X}, \pm \hat{Y}, \pm \hat{Z}\} \quad (15)$$

for $0 \leq m \leq J$, and are always signed Pauli operators (since the Clifford group is the normalizer of the Pauli group). We may therefore express the operators \mathbf{P}_m in the basis of Pauli operators as

$$\mathbf{P}_m = x_m \hat{X} + y_m \hat{Y} + z_m \hat{Z} \quad (16)$$

where $x_m, y_m, z_m \in \{0, \pm 1\}$ subject to the constraint $|x_m|^2 + |y_m|^2 + |z_m|^2 = 1$ (only one nonzero coefficient). The unit vector defined by

$$\hat{\mathbf{r}}_m \equiv (x_m, y_m, z_m), \quad \|\hat{\mathbf{r}}_m\| = 1 \quad (17)$$

therefore points uniformly at random along one of the principle Cartesian axes $\{\pm \hat{x}, \pm \hat{y}, \pm \hat{z}\}$, and maps the ‘‘direction’’ of the operator \mathbf{P}_m in Pauli space. Thus we have constructed a map from a given random Clifford sequence of length J , to a set of J unit vectors each oriented at random along the cartesian axes.

With these insights the error contributions may be recast into more convenient expressions by moving to vector notation. In particular, taking the trace over Eq. 14 and using the cyclic composition properties of the Pauli matrices, we find

$$\frac{1}{2} \text{Tr}(\mathbf{P}_j \mathbf{P}_k) = \hat{\mathbf{r}}_j \cdot \hat{\mathbf{r}}_k. \quad (18)$$

From Eqs. 10, 14 and 18 we therefore obtain

$$Q_{1,1}^{(2)} = - \sum_{j < k} \delta_j \delta_k \hat{\mathbf{r}}_j \cdot \hat{\mathbf{r}}_k. \quad (19)$$

Observing the quantity $\delta_j \delta_k \hat{\mathbf{r}}_j \cdot \hat{\mathbf{r}}_k$ is invariant under exchange of indices, we may recast the restricted sum over $j < k$ into an unrestricted sum, picking up a residual term $Q_2^{(2)} = -\frac{1}{2} \sum_{j=1}^J \delta_j^2$ (see Appendix C), to obtain

$$Q_{1,1}^{(2)} = -\frac{1}{2} \|\vec{\mathbf{R}}\|^2 - Q_2^{(2)}, \quad \vec{\mathbf{R}} \equiv \sum_{j=1}^J \delta_j \hat{\mathbf{r}}_j. \quad (20)$$

Consequently, $\vec{\mathbf{R}}$ is a random walk in ‘‘Pauli space’’.

Taking half the trace of Eq. 9 and substituting in Eq. 20, the term $Q_2^{(2)}$ cancels out in the expression for $\frac{1}{2} \text{Tr}(\tilde{\mathcal{S}}_{\eta, \delta})$. This further simplifies by observing that

$$\frac{1}{2} \text{Tr}(C_{1,j-1} \hat{Z}^k C_{j,J}) = \begin{cases} 0, & k \text{ odd} \\ 1, & k \text{ even} \end{cases} \quad (21)$$

This follows from (a) the cyclic property of the trace, (b) the technical constraint $\mathcal{S}_\eta \equiv C_{1,J} \equiv \mathbb{I}$, and (c) \hat{Z}^k is either Z or \mathbb{I} depending on whether k is odd or even respectively. Using Eq. 21 we find $Q^{(0)} = 1$ and $Q_1^{(1)} = Q_{1,2}^{(3)} = Q_3^{(3)} = 0$, leading to the simplified expression

$$\frac{1}{2} \text{Tr}(\tilde{\mathcal{S}}_{\eta, \delta}) \approx 1 - \frac{1}{2} \|\vec{\mathbf{R}}\|^2 + Q_{1,1,1}^{(3)} + Q_{1,1,1,1}^{(4)} + Q_{1,1,2}^{(4)} + Q_{1,3}^{(4)} + Q_{2,2}^{(4)} + Q_4^{(4)}. \quad (22)$$

Substituting this into Eq. 6 and retaining only terms up to $O(\sigma^4)$ we obtain

$$\langle \mathcal{F} \rangle \approx 1 - \langle \|\vec{\mathbf{R}}\|^2 \rangle + O^{(4)}, \quad (23)$$

where

$$O^{(4)} \equiv \frac{1}{4} \langle \|\vec{\mathbf{R}}\|^4 \rangle + 2 \langle Q_{2,2}^{(4)} \rangle + 2 \langle Q_4^{(4)} \rangle. \quad (24)$$

To a good approximation $O^{(4)}$ may be treated as a small correction in the form of a small constant, with the statistical distribution properties residing in the leading-order term, the random variable $\langle \|\vec{\mathbf{R}}\|^2 \rangle$. In arriving at these expressions we have used the assumption that the error process has zero mean, from which it follows only terms with noise random variables raised to even powers, or those summed over terms raised to even powers, survive the ensemble average with the others reducing to zero. In particular, $\langle Q_{1,1,1}^{(3)} \rangle = \langle Q_{1,1,1,1}^{(4)} \rangle = \langle Q_{1,1,2}^{(4)} \rangle = \langle Q_{1,3}^{(4)} \rangle = 0$ (see Appendix C).

IV. Probability Distribution Functions for RB Outcomes via a Geometric Interpretation of Error Accumulation

In the expressions above we see that the key metric capturing the reduction of fidelity in a randomized benchmarking sequence is

$$\vec{\mathbf{R}} \equiv \sum_{j=1}^J \delta_j \hat{\mathbf{r}}_j \quad (25)$$

which may be interpreted as a random-walk in \mathbb{R}^3 ‘‘Pauli space’’ generated by adding J randomly-oriented steps along the principle Cartesian axes, with step lengths specified by δ . Walks which terminate far from the origin correspond to sequences with large net infidelities while those which ultimately end near the origin have small infidelities. These observations form a key contribution of this work, as we will show.

This geometric picture provides a unique insight into how error accumulates in RB sequences, and facilitates calculation of the distribution of $\langle \mathcal{F} \rangle$ incorporating the effects of correlations in the error process. In particular, the calculation of sequence fidelities maps onto the distribution of random walk, with possible correlations manifesting the random step lengths.

In the following sections we explore how correlations in δ affect the distributions of the terms in Eq. 23, and hence the probability density function (PDF) $f_{\langle \mathcal{F} \rangle}(F)$ of the noise-averaged fidelity $\langle \mathcal{F} \rangle$. Our calculations demonstrate that under all error models treated here $f_{\langle \mathcal{F} \rangle}(F)$ takes the form of a gamma distribution, which is well known in statistics and provides the significant benefit in that explicit analytic forms are available for both the moments of the distribution and the moment-generating functions (see Table I). Interestingly the gamma distribution is known to be useful for failure analysis and life-testing in engineering [27] which bears some similarity to the notion of error accumulation in RB.

A. PDF for Markovian Processes

In the Markovian limit (*i.e.* uncorrelated noise), we assume all noise random variables δ_j are i.i.d. Hence \vec{R} corresponds to a J -length unbiased random walk with step lengths sampled from the normal distribution $\mathcal{N}(0, \sigma^2)$. Since these step lengths are symmetrically distributed about zero, the distributions of the components of the walk vector $\delta_j \hat{r}_j = (\delta_j x_j, \delta_j y_j, \delta_j z_j)$ are invariant with respect to the sign of the coefficients α_j in all Cartesian directions $\alpha \in \{x, y, z\}$. Ignoring the signs we therefore treat the coefficients as binaries $\alpha_j \in \{0, 1\}$, where the zero event simply reduces the number of steps taken in that direction. Let

$$n_\alpha \equiv \sum_{i=1}^J |\alpha_i|, \quad \alpha \in \{x, y, z\}, \quad n_x + n_y + n_z = J \quad (26)$$

count the total number of nonzero components in each Cartesian direction over the sequence of walk vectors $\{\hat{r}_1, \hat{r}_2, \dots, \hat{r}_J\}$. Then

$$\vec{R} = \left(\delta_1^x + \dots + \delta_{n_x}^x, \delta_1^y + \dots + \delta_{n_y}^y, \delta_1^z + \dots + \delta_{n_z}^z \right) \quad (27)$$

where the superscripts in δ_j^α indicate summing only over the subset of δ_j for which the coefficients α_j are nonzero. Thus we have

$$\|\vec{R}\|^2 = \Delta_x^2 + \Delta_y^2 + \Delta_z^2 \quad (28)$$

where we have defined $\Delta_\alpha \equiv (\delta_1^\alpha + \delta_2^\alpha + \dots + \delta_{n_\alpha}^\alpha)$ for $\alpha \in \{x, y, z\}$. Since all $\delta_j \sim \mathcal{N}(0, \sigma^2)$ are i.i.d., the new random variables $\Delta_\alpha \sim \mathcal{N}(0, n_\alpha \sigma^2)$ are also independent and normally distributed, with zero mean and variance $n_\alpha \sigma^2$.

The distribution of the sum of squares of non-identical Gaussians is generally complicated to write down, requiring a generalized chi-square distribution. To avoid this we make the following modest approximation. Since the vectors \hat{r}_j are uniformly-distributed there is a $\frac{1}{3}$ probability of being parallel

to any given Cartesian axis. The probability of finding any particular combination (n_x, n_y, n_z) is therefore given by the multinomial distribution

$$\mathcal{P}(n_x, n_y, n_z) = \frac{J!}{n_x! n_y! n_z!} \left(\frac{1}{3}\right)^{n_x} \left(\frac{1}{3}\right)^{n_y} \left(\frac{1}{3}\right)^{n_z} \quad (29)$$

For $J \gtrsim 5$, however, this is sufficiently peaked around $n_{x,y,z} = J/3$ that we may simply regard these values as fixed without significant error. In this case $\Delta_{x,y,z} \sim \mathcal{N}(0, J\sigma^2/3)$ reduce to i.i.d. random variables. The distribution of $\|\vec{R}\|^2$ consequently reduces to chi-square distribution with 3 degrees of freedom. It is more convenient, however, to express this in more general terms as a member of the two-parameter family of *gamma distributions* (see Eq. A5), of which the chi-square is a special case. Specifically, we obtain

$$\|\vec{R}\|^2 \sim \Gamma(\alpha, \beta), \quad \alpha = \frac{3}{2}, \quad \beta = \frac{2J\sigma^2}{3} \quad (30)$$

with shape parameter α and scale parameter β . An ensemble average over n independent noise realizations is therefore specified by

$$\langle \|\vec{R}\|^2 \rangle_n = \frac{1}{n} \sum_{j=1}^n \|\vec{R}\|_j^2, \quad \|\vec{R}\|_j^2 \sim \Gamma\left(\frac{3}{2}, \frac{2J\sigma^2}{3}\right) \quad (31)$$

where the $\|\vec{R}\|_j^2$ are i.i.d. gamma-distributed random variables. But the sample mean over n gamma-distributed random variables simply yields a rescaled gamma distribution with $\alpha \rightarrow n\alpha$ and $\beta \rightarrow \beta/n$ (see Eq. A7). Consequently

$$\langle \|\vec{R}\|^2 \rangle_n \sim \Gamma\left(\frac{3n}{2}, \frac{2J\sigma^2}{3n}\right) \quad (32)$$

with expectation $\mathbb{E}[\langle \|\vec{R}\|^2 \rangle_n] = J\sigma^2$ and variance $\mathbb{V}[\langle \|\vec{R}\|^2 \rangle_n] = \frac{2}{3}J^2\sigma^4 n^{-1}$.

Higher-order contributions may be included by computing the terms in $O^{(4)}$. For full derivations see Appendix E; here we sketch the result. From the known distribution of $\|\vec{R}\|^2$ in Eq E6, the PDF for $\|\vec{R}\|^4$ is given by a transformation allowing us to compute the expectation and variance. Taking an ensemble average over noise realizations, applying the central limit theorem, and observing the narrowness of the resulting distribution compared to the leading-order term, we make the approximation $\langle \|\vec{R}\|^4 \rangle \approx \mathbb{E}[\|\vec{R}\|^4] = \frac{5}{3}J^2\sigma^4$. The remaining terms yield values $\langle Q_{2,2}^{(4)} \rangle = \frac{1}{8}J(J-1)\sigma^4$ and $\langle Q_4^{(4)} \rangle = \frac{1}{8}J\sigma^4$. Substituting these into Eq. 24 we obtain the 4th order correction $O^{(4)} = \frac{2}{3}J^2\sigma^4$ and the noise-averaged fidelity reduces to

$$\langle \mathcal{F} \rangle \approx 1 - \langle \|\vec{R}\|^2 \rangle_n + \frac{2}{3}J^2\sigma^4 \quad (33)$$

inheriting the gamma distribution of $\langle \|\vec{R}\|^2 \rangle_n$ described in Eq. 32. We linearly transform this expression to produce a

final PDF for noise-averaged *fidelity* in the Markovian regime:

$$f_{\langle \mathcal{F} \rangle}(F) \equiv \nu(F)^{\alpha-1} e^{-\nu(F)/\beta} \beta^{-\alpha} / \Gamma(\alpha), \quad (34)$$

where $\Gamma(x)$ is the gamma function, and the quantities $\nu(F)$, α and β are defined in Table I.

B. PDF for DC (quasi-static) processes

In the DC limit (*i.e.* quasi-static noise) we assume all noise random variables $\delta_j \equiv \delta$ are identical (maximally correlated) over a given sequence \mathcal{S}_η . However over separate instances δ is sampled from the normal distribution $\delta \sim \mathcal{N}(0, \sigma^2)$. Then \vec{R} corresponds to a J -step unbiased random walk with *fixed* step length δ directed along the Cartesian axes. In this case the noise random variables δ and Clifford-dependent random variables \hat{r}_j factorize, allowing us to express

$$\vec{R} = \delta \vec{V}, \quad \vec{V} \equiv \sum_{j=1}^J \hat{r}_j \quad (35)$$

where $\vec{V} \in \mathbb{R}^3$ defines an unbiased random walk on a 3D lattice generated by adding J *unit-length* steps. Thus, in contrast with the Markovian case, here the random walk in Pauli space is unaffected (step-by-step) by the noise interactions. Rather, in a given run, the noise variables δ effectively scale the random walk \vec{V} generated by the Clifford sequence, up to a sign. Since we are interested in the norm square $\|\vec{R}\|^2 = \delta^2 \|\vec{V}\|^2$, however, any sign dependence of δ vanishes. Performing a finite ensemble average over n noise randomizations we therefore obtain $\langle \|\vec{R}\|^2 \rangle_n = \langle \delta^2 \rangle_n \|\vec{V}\|^2$, and Eq. 23 yields

$$\langle \mathcal{F} \rangle \approx 1 - \langle \delta^2 \rangle_n \|\vec{V}\|^2 + O^{(4)} \quad (36)$$

In this case, $O^{(4)}$, includes a term $\frac{1}{4} \langle \delta^4 \rangle_n \|\vec{V}\|^4$ which is now *highly correlated* with the leading-order contribution, so we cannot use the expectation value as a proxy for the whole distribution. However corrections from these terms are $O(\sigma^4)$, so we ignore these terms, and formally study the limit $J\sigma^2 \ll 1$. Numerical evidence indicates that this approximation works well up to $J\sigma^2 \sim 1$.

Since $\delta \sim \mathcal{N}(0, \sigma^2)$ is normally distributed, δ^2 is chi-square-distributed which is a special case of the gamma distribution, $\delta^2 \sim \Gamma(\alpha, \beta)$ with shape parameter $\alpha = 1/2$ and scale parameter $\beta = 2\sigma^2$. Taking the sample mean over n independent gamma-distributed variables again yields a rescaled gamma distribution with $\alpha \rightarrow n\alpha$ and $\beta \rightarrow \beta/n$

$$\langle \delta^2 \rangle_n \sim \Gamma\left(\frac{n}{2}, \frac{2\sigma^2}{n}\right) \quad (37)$$

with expectation $\mathbb{E}[\langle \delta^2 \rangle_n] = \sigma^2$ and variance $\mathbb{V}[\langle \delta^2 \rangle_n] = \frac{2\sigma^4}{n}$.

The random-walk behavior of $\|\vec{V}\|^2$ (Eq. 35) represents a well-studied problem in diffusion statistics. Let the random variable R be the distance from the origin in a symmetric (Bernoulli) 3D random walk after J steps. It is straightforward

	Markovian	DC	Block-correlated
α	$\frac{3}{2}n$	$\frac{3}{2}$	$\frac{3}{2}J/(M-1)$
β	$\frac{2}{3}J\sigma^2/n$	$\frac{2}{3}J\sigma^2$	$\frac{2}{3}(M-1)\sigma^2$
$\nu(F)$	$1 - F + \frac{2}{3}J^2\sigma^4$	$1 - F$	$1 - F$
\mathbb{E}	$1 - J\sigma^2 + \frac{2}{3}J^2\sigma^4$	$1 - J\sigma^2$	$1 - J\sigma^2$
\mathbb{M}	$1 - J\sigma^2 \left(1 - \frac{2}{3n}\right) + \frac{2}{3}J^2\sigma^4$	$1 - \frac{1}{3}J\sigma^2$	$1 - J\sigma^2 \left(1 - \frac{2}{3}\frac{M-1}{J}\right)$
\mathbb{V}	$\frac{2}{3}J^2\sigma^4/n$	$\frac{2}{3}J^2\sigma^4$	$\frac{2}{3}J(M-1)\sigma^4$
\mathbb{S}	$-2\sqrt{2/3n}$	$-2\sqrt{2/3}$	$-2\sqrt{2(M-1)/3J}$

TABLE I. Scale and shape parameters for the respective gamma distributions, and calculated moments for noise-averaged fidelity distributions $f_{\langle \mathcal{F} \rangle}(F)$. Moments obtained from standard gamma distribution after appropriate linear transformation specified by $\nu(F)$: Expectation, $\mathbb{E}[\langle \mathcal{F} \rangle] = \nu + F - \alpha\beta$; Mode, $\mathbb{M}[\langle \mathcal{F} \rangle] = \nu + F - (\alpha - 1)\beta$; Variance, $\mathbb{V}[\langle \mathcal{F} \rangle] = \alpha\beta^2$; Skew, $\mathbb{S}[\langle \mathcal{F} \rangle] = -2/\sqrt{\alpha}$.

ward to show that the PDF for R is

$$f_R(r) = \left(\frac{3}{2\pi J}\right)^{3/2} 4\pi r^2 e^{-\frac{3r^2}{2J}}. \quad (38)$$

This expression describes a random walk of unit step length, and is derived assuming that \hat{r}_j is uniformly and continuously sampled from *all* directions in \mathbb{R}^3 , however Eq. 38 is a good approximation for the PDF of a walk on a 3D lattice [28]. The distribution of the distance square $\|\vec{V}\|^2 \approx R^2$ is then given by the transformation (see Eq. A3)

$$f_{\|\vec{V}\|^2}(x) = \frac{1}{2x^{1/2}} f_R(x^{-1/2}) \quad (39)$$

$$= \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right) \quad (40)$$

where $\alpha = 3/2$ and $\beta = 2J/3$, and $\Gamma(x)$ is the gamma function; again this is a gamma distribution (see Eq. A5) with shape parameter α and scale parameter β . Consequently

$$\|\vec{V}\|^2 \sim \Gamma\left(\frac{3}{2}, \frac{2J}{3}\right). \quad (41)$$

Thus, to leading order, the PDF for $\langle \mathcal{F} \rangle$ with DC noise is specified by the product of two independent gamma-distributed random variables. The closed-form expression can be calculated by direct integration (see Appendix F), however for moderate ensemble sizes $n \gtrsim 50$ it is sufficient to approximate $\langle \delta^2 \rangle_n$ as strongly peaked around its mean, σ^2 , such that $\langle \mathcal{F} \rangle \approx 1 - \sigma^2 \|\vec{V}\|^2$. In this case the PDF $f_{\langle \mathcal{F} \rangle}(F)$ reduces to a linear-transformed gamma distribution associated with $\|\vec{V}\|^2$ and possesses *the same form* as Eq. 34 but with different values of the parameters $\nu(F)$, α and β , as given defined in Table I. This is a remarkable observation given the substantial differences in error correlations between these extreme limits.

C. PDF for Generically-Correlated Processes

For both limiting cases of Markovian and DC noise correlations treated above we showed $\langle \mathcal{F} \rangle$ is described, to first

order, by the gamma distribution with shape and scale parameters dependent on the correlation structure. Assuming continuity of the distribution we also expect $\langle \mathcal{F} \rangle$ to be approximately gamma-distributed for generic, intermediate correlation structures. We can show this formally for a specific class of correlated noise models, in which the noise is *block-correlated*; *i.e.* the error random variables δ_j are constant over blocks, or subsequences, of Clifford operators of fixed length $M \leq J$, and there is no correlation between distinct blocks. The full derivation of the PDF for this case can be found in Appendix G. We find block-correlated noise yields a gamma-distributed fidelity, with parameters $\alpha = 3J/2(M-1)$, $\beta = 2(M-1)\sigma^2/3$, enabling us to interpolate between the Markovian ($M=1$) and DC ($M=J$) limits for arbitrary correlation length M (see Table I).

While block-correlated noise is not stationary (*i.e.* it does not have a stationary power spectrum), it simply and explicitly captures the notion of a correlation length, M . The correlation length thus manifests itself in the distribution of fidelities. With these insights, for brevity we assume that generic noise correlations also give rise to gamma-distributed infidelities. This is supported by the quantitative calculations and qualitative arguments in Appendix G, and by comparison with Monte Carlo numerics. Consequently, the shape and scale factors can be inferred from the mean and variance of the distribution, which we calculate directly.

For simplicity, we assume the error process is sufficiently non-Markovian that the noise ensemble size n and the $\mathcal{O}(\sigma^4)$ contributions may be ignored without introducing large errors, as in the DC case. Thus, we write

$$\langle \mathcal{F} \rangle \approx 1 - \langle \|\vec{\mathbf{R}}\|^2 \rangle, \quad \langle \|\vec{\mathbf{R}}\|^2 \rangle \sim \Gamma[\alpha, \beta]. \quad (42)$$

From the first two moments of the gamma distribution, the expectation and variance of the random-walk variable are $\mathbb{E}[\langle \|\vec{\mathbf{R}}\|^2 \rangle] = \alpha\beta$ and $\mathbb{V}[\langle \|\vec{\mathbf{R}}\|^2 \rangle] = \alpha\beta^2$, so the shape and scale parameters are given by

$$\alpha = \frac{\mathbb{E}^2}{\mathbb{V}}, \quad \beta = \frac{\mathbb{V}}{\mathbb{E}}. \quad (43)$$

The task of defining the gamma distribution therefore reduces to computing the expectation and variance of $\langle \|\vec{\mathbf{R}}\|^2 \rangle$, as a function of the generic correlation structure of the error process, δ .

An arbitrary time-series $x(t)$ may be characterized by an autocorrelation function $C_x(\tau) = \langle x(t)x(t+\tau) \rangle_t$ where the ensemble average is over all t , and τ is the *time difference* between measurements. In this case, invoking the Wiener-Khintchine theorem, the power spectral density (PSD) is given by the Fourier transform $S(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} C_x(\tau) e^{i\omega\tau} d\tau$. We make use of these relations by discretizing the time-series. Let the elements of δ be defined by $\delta_j = x(t_j)$ for $t_j/\tau_g \in \{1, 2, \dots, J\}$ where τ_g is the time taken to perform a Clifford operation. The underlying error process is thereby discretely “sampled” by the Clifford sequence \mathcal{S}_η , and correlations between elements of δ separated by a time interval of “ k gates” are specified by the (discrete)

autocorrelation function

$$C_\delta(k) \equiv \langle \delta_j \delta_{j+k} \rangle_j \quad (44)$$

where the ensemble average is over the index j . Substituting these definitions into Eq. 25 the expectation and variance of $\langle \|\vec{\mathbf{R}}\|^2 \rangle$ are given by (see Appendix H)

$$\mathbb{E} = JC_\delta(0), \quad \mathbb{V} = \frac{4}{3} \sum_{k=1}^{J-1} (J-k) (C_\delta(k))^2. \quad (45)$$

In experiment one typically has access to the PSD $S(\omega)$ rather than the autocorrelation function. However this easily maps to our framework via an inverse Fourier transform

$$C_\delta(k) = \int_{-\infty}^{\infty} S(\omega) e^{-i\omega k\tau_g} d\omega. \quad (46)$$

The PDF therefore has *the same form as* Eq. 34 but with α and β given by Eq. 43, and $\nu(F) = 1 - F$.

With these expressions we now have a complete analytic representation of the distribution of measured noise-averaged fidelities over different RB sequences. In the next section we verify these results with numeric Monte Carlo simulations and discuss the differences in distribution characteristics depending on noise correlations.

D. Comparing PDFs for Various Noise Correlations

The analytic forms for $f_{\langle \mathcal{F} \rangle}(F)$ derived above serve as a tool to analyze the impact of temporal noise correlations on RB experiments. In all cases (Markovian, DC, and Intermediate) the PDF is $\Gamma(\alpha, \beta)$ -distributed, with differences captured in the values of the shape parameter α and scale parameter β . This result is derived from statistics of a 3D random walk.

Plotting the PDF $f_{\langle \mathcal{F} \rangle}(F)$ in Fig. 1 immediately reveals substantial differences in the distribution of outcomes for the two limiting cases. While the distributions in both Markovian and DC cases yield the same value for the mean (the statistic currently used in RB protocols), the higher order moments diverge significantly. For DC noise $f_{\langle \mathcal{F} \rangle}(F)$ is skewed towards high-fidelities and possesses a variance significantly larger than that for Markovian errors of equivalent strength, parameterized by the value of σ^2 . Averaging over a large noise ensemble results in the mode converging to the mean in the Markovian regime, but maintaining a fixed *higher* value of fidelity for the DC regime. By comparison, the variance and skew for Markovian noise diminish with increasing noise averaging, but remain fixed and nonzero in the DC case.

To compare our analytic PDFs with the true distributions we directly simulate the fidelity outcomes associated with the metric in Eq. 5. For a given J , we generate an ensemble $\{\eta_1, \dots, \eta_k\}$ of k random Clifford sequences. The first $J-1$ elements in each sequence η_i are uniformly and independently sampled from the set $\{1, \dots, 24\}$, with the final element chosen such that the total operator product performs the identity $\mathcal{S}_{\eta_i} \equiv \mathbb{I}$. For each sequence η_i we then generate an ensemble $\{\delta_{i,1}, \dots, \delta_{i,n}\}$ of n random and independent realizations of the error process, where each $\delta_{i,j}$ is a sequence

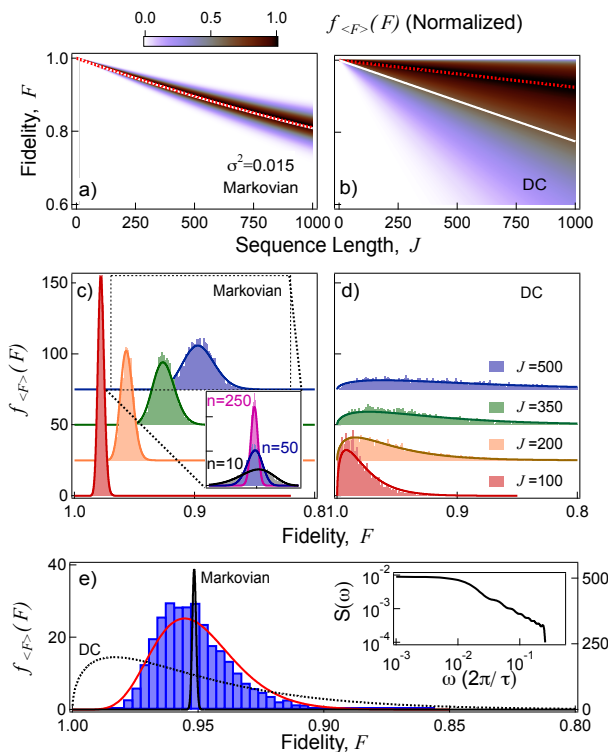


FIG. 1. Analytic PDFs $\langle \mathcal{F} \rangle$ and simulated distributions for different noise regimes using $p_{\text{RB}} \sim 2 \times 10^{-4}$. **a, b**) $f_{\langle \mathcal{F} \rangle}(F)$ calculated as a function of J for Markovian and DC regimes, normalized to unity at the mode for each J for clarity. White line indicates analytic mean, $\mathbb{E}[\langle \mathcal{F} \rangle]$; red dashed line indicates analytic mode, $\mathbb{M}[\langle \mathcal{F} \rangle]$. **c, d**) Compare analytics (solid lines) to numerically simulated histograms of distributions for various J , using $n = 50$ and no free parameters. Curves vertically offset by for clarity by multiples of 25 units. Inset **c**) Markovian distributions varying $n=10, 50, 250$ (black, blue, pink), corresponding to distribution in dotted box. **e**) Analytic PDF $\langle \mathcal{F} \rangle$ (solid red line) and simulated distributions for quasi- $1/\omega$ noise regime with $J = 200$ and $n = 3000$. Comparative PDF for DC (left axis) and Markovian noise (right axis) shown as black dashed and solid lines respectively. Noise parameters chosen such that the mean error is 0.05. PSD shown in inset, constructed using Fourier synthesis as described in [29].

of J noise random variables generated by Monte Carlo sampling from the appropriate correlated-error model. For Markovian and DC processes, this is fully described in Section III. For the intermediate case, random sequences δ with a desired PSD $S(\omega)$ and autocorrelation function $C_\delta(k)$ may be generated by uniformly-phase-randomized Fourier synthesis as described in [29] (see Appendix I).

For each pair η_i and $\delta_{i,j}$ the operator product in Eq. 7 is computed, using Table III to determine the (2×2) unitary matrix representing each Clifford operator. For each pair the trace fidelity $F_{i,j} = \mathcal{F}(\eta_i, \delta_{i,j})$ is then calculated using Eq. 5, generating the array shown in Eq. 2, simulating the measured fidelity outcomes. Averaging over columns, as in Eq. 3, the array reduces to a column vector containing k noise-averaged fidelities $\overline{F}_{i,\langle \cdot \rangle}^{(J)}$, one for each η_i , which we finally plot as a

normalized histogram and compare against our analytic PDFs.

Fig. 1(c) and (d) compares the numerically generated histograms of fidelity against the analytic results, Eq. 34, for Markovian and DC noise respectively. These are in excellent agreement for the different error processes considered. The characteristic long-tailed distribution peaked near high fidelities in the DC limit reproduces key features observed in recent experiments [17]. Agreement with analytics is good (with no free parameters) for $J\sigma^2 \lesssim 1$, beyond which higher order error terms contribute to the distribution.

For correlated noise, similarly good agreement is obtained. We have validated the block-correlated noise model (not shown). For quasi- $1/\omega$ noise, the fidelities are also well described by the gamma distribution, accounting for correlation length, as shown in Fig. 1(e). The dominance of low-frequency components in this PSD (see inset) skews $f_{\langle \mathcal{F} \rangle}(F)$ toward higher fidelities than the mean, but the presence of higher-frequency components (shorter correlation times) reduces the variance and partially restores symmetry. In the limit of power spectra containing substantial high-frequency noise, small shifts due to higher-order terms in Eq. 24 become important.

The underlying physical reason for the differences in the two limiting error models is revealed by examining the filter-transfer functions $G(\omega; \eta)$ for the various Clifford randomizations [30]. The noise-averaged fidelity is given by $\langle \mathcal{F} \rangle_\infty \approx (1 - e^{-\frac{1}{2\pi} \int_0^\infty G(\omega; \eta) S(\omega) \omega^{-2} d\omega})/2$, quantifying the susceptibility to error over a given frequency band, and has demonstrated experimental applications [31]. Using techniques outlined in Refs. [30, 32] we calculate $G(\omega; \eta)$ for 10^3 random sequences η , shown in Fig. 2. In the low-frequency regime we observe variations over several orders-of-magnitude in the vertical offset of $G(\omega; \eta)$ and also variations in slope at higher frequencies, indicative of partial error cancellation and hence substantial variations in susceptibility to correlated errors. The corresponding distribution of fidelities agrees well with $f_{\langle \mathcal{F} \rangle}(F)$ for DC noise (see inset).

These observations arise from the fact that some RB sequences contain coherent, error-suppressing subsequences. In fact RB sequences bear resemblance [33] to randomized dynamical decoupling protocols known to suppress errors in certain limits of correlated noise [34]. This would lead to “artificially” small measured error for correlated noise, thereby increasing the variance and skew in $f_{\langle \mathcal{F} \rangle}(F)$ towards high fidelities for sufficiently low-frequency-dominated noise. Furthermore, this explicit link between the form of $f_{\langle \mathcal{F} \rangle}(F)$ and underlying symmetries in S_η suggests we may downselect RB sequences using, *e.g.* the filter function or appropriate entropy measures on the Clifford sequences to ensure coherent averaging properties are minimized. To the best of our knowledge this is the first quantitative validation of the mechanism underlying the shifts in $f_{\langle \mathcal{F} \rangle}(F)$ for temporally-correlated errors, and the first application of the filter-transfer function formalism for quantum control *at the algorithmic level*.

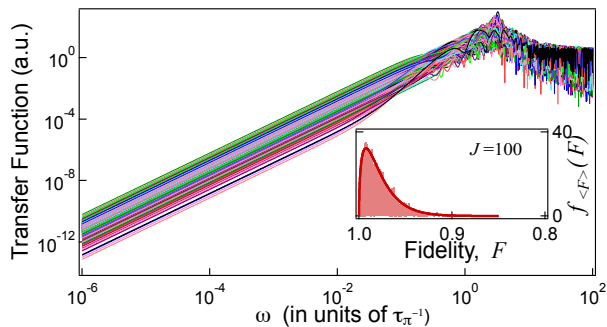


FIG. 2. Filter functions $\{G(\omega; \boldsymbol{\eta})\}$ for ensemble of RB sequences. Dimensionless angular frequency ω normalized to the duration of a bit-flip operation, τ_π ; low-frequency ($\omega < 1$) noise susceptibility is captured by vertical offsets and slopes. Inset: Histogram of calculated fidelities for $\{G(\omega; \boldsymbol{\eta})\}$ using $S(\omega) \propto \delta(\omega - 4\pi \times 10^{-4} \tau_\pi^{-1})$ for quasi-DC noise (scaled to correspond to $\sigma = 0.015$) and overlaid with $f_{\langle \mathcal{F} \rangle}(F)$.

V. Discussion and Conclusion

Our primary observation is that the form of $f_{\langle \mathcal{F} \rangle}(F)$ – and in particular the moments of the distribution – can exhibit strikingly different behaviors in different error regimes. This is true despite the fact that the expectation $\mathbb{E}[\langle \mathcal{F} \rangle] \equiv \mathbb{E}[\mathcal{F}]_J$ approximately converges for Markovian and DC cases for weak noise (see Table I). This has a variety of important impacts in the application and interpretation of RB experiments for quantum information.

For instance, due to the form of $f_{\langle \mathcal{F} \rangle}(F)$ for low frequency noise, the sample estimator $\hat{\mu}^{(J)}$ obtained from an ensemble of $k \ll 24^J$ instances of $\boldsymbol{\eta}$ can differ from the true expectation $\mathbb{E}[\mathcal{F}]_J$, generally leading to *overestimation* of the mean fidelity [12, 13]. This risks systematically underestimating p_{RB} due to insufficient sampling over $\boldsymbol{\eta}$. The difference $|\hat{\mu}^{(J)} - \mathbb{E}[\mathcal{F}]_J|$ may be formally bounded using moment-generating functions for the gamma distribution (see *Supplemental Material*) to provide a more inclusive bound than previous approximations [12, 13]. Ensuring $\hat{\mu}^{(J)}$ falls within an acceptable confidence interval, say within $\pm 10\%$ of $\mathbb{E}[\mathcal{F}]_J$ (and assuming $\sigma = 0.015$, or $p_{\text{RB}}^{(M)} \sim 2 \times 10^{-4}$) requires $k_{\text{min}}^{(M)} > 9$ randomizations be selected in the Markovian case, but $k_{\text{min}}^{(DC)} > 443$ for DC case. These values increase rapidly as confidence bounds tighten. These sample sizes are much larger than typically employed in experimental settings, but may be partially relaxed when calculating p_{RB} by fitting to measurements performed for multiple values of J . In the Markovian regime we also find a tradeoff between finite noise sampling (experimentally achieved by repeating a fixed se-

quence \mathcal{S}_η and averaging over the resulting projective measurements), and the required $k_{\text{min}}^{(M)}$. Increasing the noise averaging, n , reduces the value $k_{\text{min}}^{(M)}$.

Beyond the question of how well measurements performed in strongly correlated environments *estimate* p_{RB} , there is uncertainty surrounding the breadth of applicability of this single proxy metric as a general quantum verification tool for quantum information. One key observation is that while the variance and skew of the $f_{\langle \mathcal{F} \rangle}(F)$ converge to zero for Markovian errors in the limit of infinite noise averaging ($n \rightarrow \infty$), both remain fixed for DC noise. Accordingly our results provide direct evidence of the divergence between p_{RB} and parameters relevant to fault-tolerance [35] such as worst-case errors [36] in noise environments with strong temporal correlations; in the DC limit the worst-case error can be much larger than the average error.

Finally, the fact that some RB sequences are intrinsically “blind” to correlated errors, highlights potential shortcomings in performing experimental gate optimization by maximizing $\hat{\mu}^{(J)}$ at fixed J ; an operator may optimize an experimental parameter (to maximize the RB fidelity) in such a way to *increase* systematic errors in individual gates. Such issues may be partially mitigated by selecting subsets of valid RB sequences using the length of the random walk, $\|\vec{\mathbf{R}}\|$, in the DC limit to ensure fidelities are not overestimated in a RB procedure. Future work will explore the use of entropic measures to associate RB sequence structure with $\|\vec{\mathbf{R}}\|$ without the need to perform full calculations of the random walk.

We conclude that the interpretation and applicability of a measured RB outcome p_{RB} can differ significantly depending on the nature of the underlying errors and measurement parameters experienced in a real experimental situation. This challenge can be partially mitigated through presentation of more complete datasets – specifically $\overline{\mathbf{F}}_{i, \langle \cdot \rangle}$ for each sequence – in order to assist readers making comparisons between reported results. We believe the new insights our calculations have revealed will help bound the utility of p_{RB} in quantum information settings and also help experimentalists ensure that measurements are not subject to hidden biases.

Acknowledgments

Acknowledgements: The authors thank J. M. Chow, J. Gambetta, and A. Yacoby for useful discussions on RB measurements, and L. Viola for discussions on randomized decoupling. Work partially supported by the ARC Centre of Excellence for Engineered Quantum Systems CE110001013, the Intelligence Advanced Research Projects Activity (IARPA) through the ARO, the US Army Research Office under Contracts W911NF-12-R-0012, W911NF-14-1-0098 and W911NF-14-1-0103, and a private grant from H. & A. Harley. STF acknowledges support from an ARC Future Fellowship FT130101744.

[1] James D F V, Kwiat P G, Munro W J and White A G 2001 *Phys. Rev. A* **64**(5) 052312

[2] O’Brien J L, Pryde G J, Gilchrist A, James D F V, Langford N K, Ralph T C and White A G 2004 *Phys. Rev. Lett.* **93**(8)

- 080502
- [3] Emerson J, Alicki R and Życzkowski K 2005 *Journal of Optics B: Quantum and Semiclassical Optics* **7** S347
- [4] Lévi B, López C C, Emerson J and Cory D G 2007 *Phys. Rev. A* **75**(2) 022314
- [5] Bendersky A, Pastawski F and Paz J P 2008 *Phys. Rev. Lett.* **100**(19) 190403
- [6] López C C, Lévi B and Cory D G 2009 *Phys. Rev. A* **79**(4) 042328
- [7] Barz S, Fitzsimons J F, Kashefi E and Walther P 2013 *Nature Physics* **9** 727–731
- [8] Magesan E and Cappellaro P 2013 *Phys. Rev. A* **88**(2) 022127
- [9] Knill E, Leibfried D, Reichle R, Britton J, Blakestad R B, Jost J D, Langer C, Ozeri R, Seidelin S and Wineland D J 2008 *Phys. Rev. A* **77**(1) 012307
- [10] Magesan E, Gambetta J M and Emerson J 2011 *Phys. Rev. Lett.* **106**(18) 180504
- [11] Magesan E, Gambetta J M, Johnson B R, Ryan C A, Chow J M, Merkel S T, da Silva M P, Keefe G A, Rothwell M B, Ohki T A, Ketchen M B and Steffen M 2012 *Phys. Rev. Lett.* **109**(8) 080505
- [12] Wallman J J and Flammia S T 2014 *New Journal of Physics* **16** 103032
- [13] Epstein J M, Cross A W, Magesan E and Gambetta J M 2014 *Phys. Rev. A* **89**(6) 062321
- [14] Biercuk M, Uys H, VanDevender A, Shiga N, Itano W M and Bollinger J J 2009 *Quantum Info & Comp.* **9** 920–949
- [15] Olmschenk S, Chicireanu R, Nelson K D and Porto J V 2010 *New Journal of Physics* **12** 113007
- [16] Gaebler J P, Meier A M, Tan T R, Bowler R, Lin Y, Hanneke D, Jost J D, Home J P, Knill E, Leibfried D and Wineland D J 2012 *Phys. Rev. Lett.* **108**(26) 260503
- [17] Brown K R, Wilson A C, Colombe Y, Ospelkaus C, Meier A M, Knill E, Leibfried D and Wineland D J 2011 *Phys. Rev. A* **84**(3) 030303
- [18] Chow J M, Gambetta J M, Tornberg L, Koch J, Bishop L S, Houck A A, Johnson B R, Frunzio L, Girvin S M and Schoelkopf R J 2009 *Phys. Rev. Lett.* **102**(9) 090502
- [19] Kelly J, Barends R, Campbell B, Chen Y, Chen Z, Chiaro B, Dunsworth A, Fowler A, Hoi I C, Jeffrey E, Megrant A, Mutus J, Neill C, O'Malley P, Quintana C, Roushan P, Sank D, Vainsencher A, Wenner J, White T, Cleland A and Martinis J M 2014 *Phys. Rev. Lett.* **112**(24) 240504
- [20] Harty T P, Allcock D, Ballance C J, Guidoni L, Janacek H A, Linke N M, Stacey D N and Lucas D M 2014 *Phys. Rev. Lett.* **113**(22) 220501
- [21] Xia T, Lichtman M, Maller K, Carr A W, Piotrowicz M J, Isenhowler L and Saffman M 2015 *Phys. Rev. Lett.* **114**(10) 100503
- [22] Lu D, Li H, Trottier D A, Li J, Brodutch A, Krismanich A P, Ghavami A, Dmitrienko G I, Long G, Baugh J and Laflamme R 2015 *Phys. Rev. Lett.* **114**(14) 140505
- [23] Muhonen J T, Laucht A, Simmons S, Dehollain J P, Kalra R, Hudson F E, Freer S, Itoh K M, Jamieson D N, McCallum J C, Dzurak A S and Morello A 2015 *Journal of Physics: Condensed Matter* **27** 154205
- [24] Nielsen M and Chuang I 2010 *Quantum Computation and Quantum Information: 10th Anniversary Edn.* (Cambridge: Cambridge University Press)
- [25] Bylander J, Gustavsson S, Yan F, Yoshihara F, Harrabi K, Fitch G, Cory D G, Nakamura Y, Tsai J S and Oliver W D 2011 *Nat. Phys.* **7** 565
- [26] Anton S M, Birenbaum J S, O'Kelley S R, Bolkhovskiy V, Braje D A, Fitch G, Neeley M, Hilton G C, Cho H M, Irwin K D, Wellstood F C, Oliver W D, Shnirman A and Clarke J 2013 *Phys. Rev. Lett.* **110**(14) 147002
- [27] Papoulis A 1991 *Probability, Random Variables, and Stochastic Processes* (New York: McGraw Hill)
- [28] \vec{V} actually describes a lattice walk and should strictly be described by a probability mass function (PMF). However this introduces unnecessary complications in the form of combinatorial expressions to account for the directional discretization, and in the interest of analytic simplicity it is entirely appropriate to make this continuous approximation. Doing so simply smooths over the discreteness of the PMF while retaining all essential distribution characteristics.
- [29] Soare A, Ball H, Hayes D, Zhen X, Jarratt M C, Sastrawan J, Uys H and Biercuk M J 2014 *Phys. Rev. A* **89** 042329
- [30] Green T J, Sastrawan J, Uys H and Biercuk M J 2013 *New J. Phys.* **15** 095004
- [31] Soare A, Ball H, Hayes D, Sastrawan J, Jarratt M, McLoughlin J, Zhen X, Green T and Biercuk M 2014 *Nat. Phys.* **10** 1
- [32] Green T J, Uys H and Biercuk M J 2012 *Phys. Rev. Lett.* **109** 020501
- [33] Ryan C A, Laforest M and Laflamme R 2009 *New J. Phys.* **11** 013034
- [34] Viola L and Knill E 2005 *Phys. Rev. Lett.* **94**(6) 060502
- [35] Gutiérrez M and Brown K R 2015 *Phys. Rev. A* **91**(2) 022335 (Preprint 1501.00068)
- [36] Sanders Y R, Wallman J J and Sanders B C 2015 *arXiv:1501.04932*

A. Mathematical Preliminaries

1. Linear transformation

Let Y is a continuous, non-negative random variable described by probability density functions $f_Y(y)$. If Z is the random variable defined by $Z = \alpha Y + \beta$, where $\alpha, \beta \in \mathbb{R}$ and $\alpha \neq 0$, then the probability density functions of Z is

$$f_Z(z) = \frac{1}{|\alpha|} f_Y\left(\frac{z - \beta}{\alpha}\right) \quad (\text{A1})$$

2. Product distribution

Let X and Y be to independent, continuous random variables described by probability density functions $f_X(x)$ and $f_Y(y)$. Then the joint distribution of $Z = XY$ is

$$f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{|x|} f_X(x) f_Y(z/x) dx \quad (\text{A2})$$

3. Strictly increasing function of a random variable

Let X be an absolutely continuous non-negative random variable described by probability density function $f_X(x)$. Define the transformed random variable $Y = g(X)$ where $g(x)$ is a strictly increasing function. That is $x_1 > x_2 \iff g(x_1) > g(x_2)$, so that $g(x)$ has a well-defined inverse $g^{-1}(x)$. Then the probability density functions of Y is

$$f_Y(y) = f_X(g^{-1}(y)) \frac{dg^{-1}(y)}{dy} \quad (\text{A3})$$

4. Sum of continuous random variables

Let X and Y be two independent, continuous random variables described by probability density functions $f_X(x)$ and $f_Y(y)$. Then the probability density functions of the sum $Z = X + Y$ is

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z - y) f_Y(y) dy = \int_{-\infty}^{\infty} f_Y(z - x) f_X(x) dx \quad (\text{A4})$$

5. Gamma distribution

The gamma distribution describes a family of continuous probability distributions, related to beta distribution, and arising naturally in processes for which the waiting times between Poisson-distributed events are relevant. The common exponential distribution and chi-squared distribution are special cases. The gamma distribution is a two-parameter distribution, and there are a few different parametrizations in common use. Throughout this paper we parametrize the distribution in terms of its *shape parameter* α and a *scale parameter* β . Let $X \sim \Gamma(\alpha, \beta)$ be a gamma-distributed random variable, then the probability density function is defined by

$$f_X(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right) \quad (\text{A5})$$

where $\Gamma(x)$ is the gamma function. Table A 5 contains some relevant statistics.

mean	$\alpha\beta$
variance	$\alpha\beta^2$
skew	$2/\sqrt{\alpha}$
mode	$(\alpha - 1)\beta$

TABLE II. Statistics of the gamma distribution $\Gamma(\alpha, \beta)$.

6. Distribution of sample mean of gamma-distributed random variables

Let $X_i \sim \Gamma(\alpha, \beta)$, $i \in \{1, \dots, n\}$, be a set of n independent and identically distributed random variables sampled from the gamma distribution with shape parameter α and scale parameter β . Then the sum $Z = \sum_{i=1}^n X_i$ follows the transformed gamma distribution

$$Z \sim \Gamma(n\alpha, \beta) \quad (\text{A6})$$

The ensemble average Z/n has probability density function given by transforming $nf_Z(nx)$. Substituting into Eq. A5, and defining rescaled parameters $\alpha \rightarrow n\alpha$ and $\beta \rightarrow \beta/n$, the distribution of the ensemble average $\langle X_i \rangle_n$ is found to be

$$\langle X_i \rangle_n \sim \Gamma(n\alpha, \beta/n) \quad (\text{A7})$$

7. Lindeberg-Levy central limit theorem

Let $\{X_1, X_2, \dots, X_n\}$ be a set of i.i.d. random variables with expectation $\mathbb{E}[X_i] = \mu$ and variance $\mathbb{V}[X_i] = \rho^2 < \infty$. Then the sample average $S_n = \frac{1}{n} \sum_{i=1}^n X_i$ converges to the normal distribution

$$\lim_{n \rightarrow \infty} \sqrt{n}(S_n - \mu) \sim \mathcal{N}(0, \rho^2) \quad (\text{A8})$$

B. Clifford Group Representation for Single Qubit

A unitary operation \hat{C} is an element of the *Clifford group* if

$$\hat{C}\mathcal{P}\hat{C}^\dagger = \mathcal{P} \quad (\text{B1})$$

where we have defined the *Pauli group*

$$\mathcal{P} = \{\pm\mathbb{I}, \pm\hat{X}, \pm\hat{Y}, \pm\hat{Z}\}. \quad (\text{B2})$$

That is, the Clifford group is the normalizer of the Pauli group, where for every Pauli operation $P \in \mathcal{P}$, there is another $P' \in \mathcal{P}$ such that $\hat{C}P\hat{C}^\dagger = P'$. For a single qubit, the set of all such \hat{C} may be thought of as rotations of the Bloch sphere that permute the orientation of $\pm\hat{X}, \pm\hat{Y}, \pm\hat{Z}$ in the Cartesian basis associated with the Pauli matrices, which we refer to as ‘‘Pauli space.’’ To obtain a clearer physical picture of these operations, consider associating \hat{X} to any of the six Cartesian axes $\{\pm\hat{x}, \pm\hat{y}, \pm\hat{z}\}$. With this axis fixed, we may rotate the axes about \hat{X} into four possible orientations while preserving $\hat{X}\hat{Y}\hat{Z}$ right-handedness. This is the action of the \hat{C} group: the group of rotational symmetries of the cube.

We construct our representation as follows. Let $R_i(\theta)$ represent one of nine elementary unitaries generating a *clockwise* rotation (looking down the axis of rotation toward the origin) through angle $\theta \in \{\pi, \pm\pi/2\}$ about axis $i \in \{x, y, z\}$. The three π rotations correspond to

$$R_{x,y,z}(\pi) \equiv \hat{X}, \hat{Y}, \hat{Z} \quad (\text{B3})$$

and we use the shorthand

$$R_i^\pm \equiv R_i(\pm\pi/2), \quad i \in \{x, y, z\} \quad (\text{B4})$$

for the remaining six $\pi/2$ rotations. For example, the action of the operators R_i^\pm on the Pauli operators/axes is

$$R_x^+ : (\hat{X}, \hat{Y}, \hat{Z}) \rightarrow (\hat{X}, -\hat{Z}, \hat{Y}) \quad (\text{B5})$$

$$R_y^+ : (\hat{X}, \hat{Y}, \hat{Z}) \rightarrow (\hat{Z}, \hat{Y}, -\hat{X}) \quad (\text{B6})$$

$$R_z^+ : (\hat{X}, \hat{Y}, \hat{Z}) \rightarrow (-\hat{Y}, \hat{X}, \hat{Z}) \quad (\text{B7})$$

Products of these nine elementary operations generate a representation of the 24 elements of the single-qubit Clifford group as tabulated in Table III. We use this prescription to generate numerical simulations verifying our analytic calculations above.

#	Gate Name	Action on $(\hat{X}, \hat{Y}, \hat{Z})$	Minimal Sequence(s)	Notes
\hat{C}_1	\mathbb{I}	$(\hat{X}, \hat{Y}, \hat{Z})$	$\hat{X}_i^2, R_i^+ R_i^-, i \in \{1, 2, 3\}$	identity
\hat{C}_2	\hat{X}	$(\hat{X}, -\hat{Y}, -\hat{Z})$	\hat{X}	π rotation
\hat{C}_3	\hat{Y}	$(-\hat{X}, \hat{Y}, -\hat{Z})$	\hat{Y}	
\hat{C}_4	\hat{Z}	$(-\hat{X}, -\hat{Y}, \hat{Z})$	\hat{Z}	
\hat{C}_5	R_x^+	$(\hat{X}, -\hat{Z}, \hat{Y})$	R_x^+	
\hat{C}_6	R_y^+	$(\hat{Z}, \hat{Y}, -\hat{X})$	R_y^+	
\hat{C}_7	R_z^+	$(-\hat{Y}, \hat{X}, \hat{Z})$	R_z^+	
\hat{C}_8	R_x^-	$(\hat{X}, \hat{Z}, -\hat{Y})$	R_x^-	$-\pi/2$ rotations
\hat{C}_9	R_y^-	$(-\hat{Z}, \hat{Y}, \hat{X})$	R_y^-	
\hat{C}_{10}	R_z^-	$(\hat{Y}, -\hat{X}, \hat{Z})$	R_z^-	
\hat{C}_{11}	H	$(-\hat{X}, -\hat{Z}, -\hat{Y})$	$\hat{Z} R_x^+$	Hadamard
\hat{C}_{12}		$(-\hat{X}, \hat{Z}, \hat{Y})$	$\hat{Z} R_x^-$	
\hat{C}_{13}		$(-\hat{Y}, -\hat{X}, -\hat{Z})$	$R_z^+ \hat{X}$	
\hat{C}_{14}		$(\hat{Y}, \hat{X}, -\hat{Z})$	$R_z^- \hat{X}$	
\hat{C}_{15}		$(-\hat{Y}, -\hat{Z}, \hat{X})$	$R_z^+ R_x^+$	
\hat{C}_{16}		$(-\hat{Y}, \hat{Z}, -\hat{X})$	$R_z^+ R_x^-$	
\hat{C}_{17}		$(-\hat{Z}, -\hat{X}, \hat{Y})$	$R_x^+ R_z^-$	
\hat{C}_{18}		$(-\hat{Z}, -\hat{Y}, -\hat{X})$	$\hat{Z} R_y^-$	
\hat{C}_{19}		$(-\hat{Z}, \hat{X}, -\hat{Y})$	$R_z^+ R_y^-$	
\hat{C}_{20}		$(\hat{Z}, -\hat{X}, -\hat{Y})$	$R_z^- R_y^+$	
\hat{C}_{21}		$(\hat{Z}, -\hat{Y}, \hat{X})$	$\hat{Z} R_y^+$	
\hat{C}_{22}		$(\hat{Y}, -\hat{Z}, -\hat{X})$	$R_z^- R_x^+$	
\hat{C}_{23}		$(\hat{Z}, \hat{X}, \hat{Y})$	$R_z^+ R_y^+$	
\hat{C}_{24}		$(\hat{Y}, \hat{Z}, \hat{X})$	$R_z^- R_x^-$	

TABLE III. Representation of the Clifford group for a single qubit from products of elementary rotations. Relevant transformations of the coordinate system $\hat{X}, \hat{Y}, \hat{Z}$ under the action of each Clifford shown in column 3. Minimal sequence of elementary operations needed to generate each Clifford shown in column 4. We also indicate how these geometric rotations map to logical operations of interest for quantum information where relevant.

C. Approximating the Noise-Averaged Fidelity $\langle \mathcal{F} \rangle$

Here we present the full derivation of the approximation to the noise-averaged fidelity $\langle \mathcal{F} \rangle$ given in Eq. 23 of the main text. Let $\xi_{k_1, k_2, \dots, k_m}^{(n)}$ denote the $\mathcal{O}\left(\prod_{\rho=1}^m \delta_{j_\rho}^{k_\rho}\right) = \mathcal{O}(\sigma^n)$ term in the expansion of Eq. 8 due to cross-multiplying terms like $(\delta_{j_1} \hat{Z})^{k_1} (\delta_{j_2} \hat{Z})^{k_2} \dots (\delta_{j_m} \hat{Z})^{k_m}$, where $\sum_{\rho=1}^m k_\rho = n$. Retaining only terms up to fourth order ($n = 4$), consistent with our Talyor approximation of the error unitaries, we thereby obtain

$$\tilde{\mathcal{S}}_{\eta, \delta} \approx \xi^{(0)} + \xi_1^{(1)} + \xi_{1,1}^{(2)} + \xi_2^{(2)} + \xi_{1,1,1}^{(3)} + \xi_{2,1}^{(3)} + \xi_3^{(3)} + \xi_{1,1,1,1}^{(4)} + \xi_{1,1,2}^{(4)} + \xi_{2,2}^{(4)} + \xi_4^{(4)} \quad (\text{C1})$$

where

$$\xi^{(0)} = C_{1,J} \quad (\text{C2})$$

$$\xi_1^{(1)} = \sum_{j=1}^J (i\delta_j) C_{1,j-1} \hat{Z} C_{j,J} \quad (\text{C3})$$

$$\xi_{1,1}^{(2)} = \sum_{j < k} (i\delta_j)(i\delta_k) C_{1,j-1} \hat{Z} C_{j,k-1} \hat{Z} C_{k,J} \quad (\text{C4})$$

$$\xi_2^{(2)} = -\frac{1}{2} \sum_{j=1}^J \delta_j^2 C_{1,j-1} \hat{Z}^2 C_{j,J} \quad (\text{C5})$$

$$\xi_{1,1,1}^{(3)} = \sum_{j < k < l} (i\delta_j)(i\delta_k)(i\delta_l) C_{1,j-1} \hat{Z} C_{j,k-1} \hat{Z} C_{k,l-1} \hat{Z} C_{l,J} \quad (\text{C6})$$

$$\xi_{1,2}^{(3)} = -\frac{1}{2} \sum_{j < k} (i\delta_j) \delta_k^2 C_{1,j-1} \hat{Z} C_{j,k-1} \hat{Z}^2 C_{k,J} \quad (\text{C7})$$

$$\xi_3^{(3)} = -\frac{i}{6} \sum_{j=1}^J \delta_j^3 C_{1,j-1} \hat{Z}^3 C_{j,J} \quad (\text{C8})$$

$$\xi_{1,1,1,1}^{(4)} = \sum_{j < k < l < m} (i\delta_j)(i\delta_k)(i\delta_l)(i\delta_m) C_{1,j-1} \hat{Z} C_{j,k-1} \hat{Z} C_{k,l-1} \hat{Z} C_{l,m-1} \hat{Z} C_{m,J} \quad (\text{C9})$$

$$\xi_{1,1,2}^{(4)} = -\frac{1}{2} \sum_{j < k < l} (i\delta_j)(i\delta_k) \delta_l^2 C_{1,j-1} \hat{Z} C_{j,k-1} \hat{Z} C_{k,l-1} \hat{Z}^2 C_{l,J} \quad (\text{C10})$$

$$-\frac{1}{2} \sum_{j < k < l} (i\delta_j) \delta_k^2 (i\delta_l) C_{1,j-1} \hat{Z} C_{j,k-1} \hat{Z}^2 C_{k,l-1} \hat{Z} C_{l,J} \quad (\text{C11})$$

$$-\frac{1}{2} \sum_{j < k < l} \delta_j^2 (i\delta_k)(i\delta_l) C_{1,j-1} \hat{Z}^2 C_{j,k-1} \hat{Z} C_{k,l-1} \hat{Z} C_{l,J} \quad (\text{C12})$$

$$\xi_{2,2}^{(4)} = \frac{1}{4} \sum_{j < k} \delta_j^2 \delta_k^2 C_{1,j-1} \hat{Z}^2 C_{j,k-1} \hat{Z}^2 C_{k,J} \quad (\text{C13})$$

$$\xi_{1,3}^{(4)} = -\frac{i}{6} \sum_{j < k} (i\delta_j) \delta_k^3 C_{1,j-1} \hat{Z} C_{j,k-1} \hat{Z}^3 C_{k,J} \quad (\text{C14})$$

$$\xi_4^{(4)} = \frac{1}{24} \sum_{j=1}^J \delta_j^4 C_{1,j-1} \hat{Z}^4 C_{j,J} \quad (\text{C15})$$

and we have defined the Clifford *subsequence* operators

$$C_{jk} \equiv \hat{C}_{\eta_j} \dots \hat{C}_{\eta_k}, \quad 1 \leq j \leq k \leq J. \quad (\text{C16})$$

To evaluate Eq. 5 we must obtain an expression for $\frac{1}{2} \text{Tr}(\tilde{\mathcal{S}}_{\eta, \delta})$. By the linearity of the trace, we must therefore obtain expressions for the quantities $Q_{k_1, k_2, \dots, k_m}^{(n)} \equiv \frac{1}{2} \text{Tr} \left(\xi_{k_1, k_2, \dots, k_m}^{(n)} \right)$ for each of the terms above. This requires some manipulation. To begin we highlight the following useful properties

$$\hat{Z}^k = \begin{cases} \hat{Z}, & k \text{ odd} \\ \mathbb{I}, & k \text{ even} \end{cases} \quad (\text{C17})$$

$$C_{1,J} \equiv \mathcal{S}_{\eta} \equiv \mathbb{I}. \quad (\text{C18})$$

In fact since $C_{1,J} = \mathbb{I}$, any cyclic permutation of subsequences C_{jk} also gives the identity. For instance

$$C_{1,J} = C_{1,j-1} C_{j,k-1} C_{k,J} = C_{k,J} C_{1,j-1} C_{j,k-1} = C_{j,k-1} C_{k,J} C_{1,j-1} = \mathbb{I} \quad (\text{C19})$$

and so on for any number of subsequences. Using the cyclic property of the trace, we therefore have

$$\frac{1}{2} \text{Tr} \left(C_{1,j-1} \hat{Z}^k C_{j,J} \right) = \frac{1}{2} \text{Tr} \left(C_{j,J} C_{1,j-1} \hat{Z}^k \right) = \frac{1}{2} \text{Tr} \left(\hat{Z}^k \right) = \begin{cases} 0, & k \text{ odd} \\ 1, & k \text{ even} \end{cases} \quad (\text{C20})$$

The last equality uses Eq. C17, the property of Pauli matrices that $\text{Tr}(\hat{Z}) = 0$, and that the corresponding (2×2) identity has trace $\text{Tr}(\mathbb{I}) = 2$. Now define the cumulative operators K_m giving the product from the first through to the m th Clifford operator in the sequence

$$K_m \equiv C_{1,m} = \hat{C}_{\eta_1} \dots \hat{C}_{\eta_m}, \quad K_0 \equiv \mathbb{I} \equiv K_J \quad (\text{C21})$$

where each K_m is some element of the Clifford group due to the closure property of groups under group composition. In this case any subsequence C_{jk} “factorizes” into the products

$$K_{j-1}^\dagger K_k = \hat{C}_{\eta_{j-1}}^\dagger \dots \hat{C}_{\eta_1}^\dagger \hat{C}_{\eta_1} \dots \hat{C}_{\eta_{j-1}} \hat{C}_{\eta_j} \dots \hat{C}_{\eta_k} = C_{jk} \quad (\text{C22})$$

allowing us to rewrite

$$\begin{aligned} C_{1,j-1} \hat{Z} C_{j,k-1} \hat{Z} C_{k,J} &= \mathbf{P}_j \mathbf{P}_k \\ C_{1,j-1} \hat{Z} C_{j,k-1} \hat{Z} C_{k,l-1} \hat{Z} C_{l,J} &= \mathbf{P}_j \mathbf{P}_k \mathbf{P}_l \\ C_{1,j-1} \hat{Z} C_{j,k-1} \hat{Z} C_{k,l-1} \hat{Z} C_{l,m-1} \hat{Z} C_{m,J} &= \mathbf{P}_j \mathbf{P}_k \mathbf{P}_l \mathbf{P}_m \\ &\dots \end{aligned}$$

where we define the \hat{Z} -conjugating operators

$$\mathbf{P}_m \equiv K_{m-1} \hat{Z} K_{m-1}^\dagger \in \{\pm \hat{X}, \pm \hat{Y}, \pm \hat{Z}\}, \quad 0 \leq m \leq J \quad (\text{C23})$$

and where \mathbf{P}_m is always a signed Pauli matrix due to the property that the Clifford group is the normalizer of the Pauli group. Thus we find

$$Q^{(0)} = 1 \quad (\text{C24})$$

$$Q_1^{(1)} = 0 \quad (\text{C25})$$

$$Q_{1,1}^{(2)} = -\frac{1}{2} \sum_{j < k} \delta_j \delta_k \text{Tr}(\mathbf{P}_j \mathbf{P}_k) \quad (\text{C26})$$

$$Q_2^{(2)} = -\frac{1}{2} \sum_{j=1}^J \delta_j^2 \quad (\text{C27})$$

$$Q_{1,1,1}^{(3)} = -\frac{i}{2} \sum_{j < k < l} \delta_j \delta_k \delta_l \text{Tr}(\mathbf{P}_j \mathbf{P}_k \mathbf{P}_l) \quad (\text{C28})$$

$$Q_{1,2}^{(3)} = 0 \quad (\text{C29})$$

$$Q_3^{(3)} = 0 \quad (\text{C30})$$

$$Q_{1,1,1,1}^{(4)} = \frac{1}{2} \sum_{j < k < l < m} \delta_j \delta_k \delta_l \delta_m \text{Tr}(\mathbf{P}_j \mathbf{P}_k \mathbf{P}_l \mathbf{P}_m) \quad (\text{C31})$$

$$Q_{1,1,2}^{(4)} = \frac{1}{4} \sum_{j < k < l} [\delta_j \delta_k \delta_l^2 \text{Tr}(\mathbf{P}_j \mathbf{P}_k) + \delta_j \delta_k^2 \delta_l \text{Tr}(\mathbf{P}_j \mathbf{P}_l) + \delta_j^2 \delta_k \delta_l \text{Tr}(\mathbf{P}_k \mathbf{P}_l)] \quad (\text{C32})$$

$$Q_{2,2}^{(4)} = \frac{1}{4} \sum_{j < k} \delta_j^2 \delta_k^2 \quad (\text{C33})$$

$$Q_{1,3}^{(4)} = \frac{1}{12} \sum_{j < k} \delta_j \delta_k^3 \text{Tr}(\mathbf{P}_j \mathbf{P}_k) \quad (\text{C34})$$

$$Q_4^{(4)} = \frac{1}{24} \sum_{j=1}^J \delta_j^4. \quad (\text{C35})$$

The nonzero terms may be recast into more convenient expressions by moving to vector notation. We expand the operators \mathbf{P}_m in the basis of Pauli operators,

$$\mathbf{P}_m = x_m \hat{X} + y_m \hat{Y} + z_m \hat{Z}, \quad x_m, y_m, z_m \in \{0, \pm 1\}, \quad |x_m|^2 + |y_m|^2 + |z_m|^2 = 1. \quad (\text{C36})$$

That is, only one nonzero coefficient x_m, y, z_m , equivalent to expressing the fact that they are sampled from the set $\{\pm \hat{X}, \pm \hat{Y}, \pm \hat{Z}\}$. The associated unit vector

$$\hat{\mathbf{r}}_m \equiv (x_m, y_m, z_m), \quad \|\hat{\mathbf{r}}_m\| = 1 \quad (\text{C37})$$

therefore points uniformly at random along one of the principle Cartesian axes $\{\pm \hat{x}, \pm \hat{y}, \pm \hat{z}\}$, capturing the ‘‘direction’’ of the operator \mathbf{P}_m in ‘‘Pauli space’’. With these definitions we may derive

$$\frac{1}{2} \text{Tr}(\mathbf{P}_j \mathbf{P}_k) = \hat{\mathbf{r}}_j \cdot \hat{\mathbf{r}}_k \quad (\text{C38})$$

$$\frac{1}{2} \text{Tr}(\mathbf{P}_j \mathbf{P}_k \mathbf{P}_l) = i (\hat{\mathbf{r}}_j \times \hat{\mathbf{r}}_k) \cdot \hat{\mathbf{r}}_l \quad (\text{C39})$$

$$\frac{1}{2} \text{Tr}(\mathbf{P}_j \mathbf{P}_k \mathbf{P}_l \mathbf{P}_m) = (\hat{\mathbf{r}}_j \cdot \hat{\mathbf{r}}_k) (\hat{\mathbf{r}}_l \cdot \hat{\mathbf{r}}_m) + (\hat{\mathbf{r}}_j \cdot \hat{\mathbf{r}}_m) (\hat{\mathbf{r}}_k \cdot \hat{\mathbf{r}}_l) - (\hat{\mathbf{r}}_j \cdot \hat{\mathbf{r}}_l) (\hat{\mathbf{r}}_k \cdot \hat{\mathbf{r}}_m) \quad (\text{C40})$$

following directly from the trace and the cyclic composition properties of the Pauli matrices $\hat{\sigma}_i \hat{\sigma}_j = i\epsilon_{ijk} \hat{\sigma}_k + \delta_{ij} \mathbb{I}$, where ϵ_{ijk} is the Levi-Civita symbol, δ_{ij} is the Kronecker delta and Einstein summation notation used. Consequently we obtain

$$Q_{1,1}^{(2)} = - \sum_{j < k} \delta_j \delta_k \hat{\mathbf{r}}_j \cdot \hat{\mathbf{r}}_k \quad (\text{C41})$$

$$Q_2^{(2)} = - \frac{1}{2} \sum_{j=1}^J \delta_j^2 \quad (\text{C42})$$

$$Q_{1,1,1}^{(3)} = \sum_{j < k < l} \delta_j \delta_k \delta_l (\hat{\mathbf{r}}_j \times \hat{\mathbf{r}}_k) \cdot \hat{\mathbf{r}}_l \quad (\text{C43})$$

$$Q_{1,1,1,1}^{(4)} = \frac{1}{2} \sum_{j < k < l < m} \delta_j \delta_k \delta_l \delta_m \{ (\hat{\mathbf{r}}_j \cdot \hat{\mathbf{r}}_k) (\hat{\mathbf{r}}_l \cdot \hat{\mathbf{r}}_m) + (\hat{\mathbf{r}}_j \cdot \hat{\mathbf{r}}_m) (\hat{\mathbf{r}}_k \cdot \hat{\mathbf{r}}_l) - (\hat{\mathbf{r}}_j \cdot \hat{\mathbf{r}}_l) (\hat{\mathbf{r}}_k \cdot \hat{\mathbf{r}}_m) \} \quad (\text{C44})$$

$$Q_{1,1,2}^{(4)} = \frac{1}{2} \sum_{j < k < l} [\delta_j \delta_k \delta_l^2 \hat{\mathbf{r}}_j \cdot \hat{\mathbf{r}}_k + \delta_j \delta_k^2 \delta_l \hat{\mathbf{r}}_j \cdot \hat{\mathbf{r}}_l + \delta_j^2 \delta_k \delta_l \hat{\mathbf{r}}_k \cdot \hat{\mathbf{r}}_l] \quad (\text{C45})$$

$$Q_{1,3}^{(4)} = \frac{1}{6} \sum_{j < k} \delta_j \delta_k^3 \hat{\mathbf{r}}_j \cdot \hat{\mathbf{r}}_k \quad (\text{C46})$$

$$Q_{2,2}^{(4)} = \frac{1}{4} \sum_{j < k} \delta_j^2 \delta_k^2 \quad (\text{C47})$$

$$Q_4^{(4)} = \frac{1}{24} \sum_{j=1}^J \delta_j^4. \quad (\text{C48})$$

We expect the major contribution to the distribution of $\langle \mathcal{F} \rangle$ to reside in the term $Q_{1,1}^{(2)}$ since, in the limit $\sigma \ll 1$, higher order terms $Q^{(n>2)}$ will be orders of magnitude smaller. Anticipating this, we recast $Q_{1,1}^{(2)}$ in terms of an unrestricted sum to facilitate more straightforward analysis. Observing the quantity $\delta_j \delta_k \hat{\mathbf{r}}_j \cdot \hat{\mathbf{r}}_k$ is invariant under exchange of indexes we thus obtain

$$Q_{1,1}^{(2)} = - \frac{1}{2} \left(\sum_{j=1}^J \sum_{k=1}^J \delta_j \delta_k \hat{\mathbf{r}}_j \cdot \hat{\mathbf{r}}_k - \sum_{j=1}^J \delta_j^2 \hat{\mathbf{r}}_j \cdot \hat{\mathbf{r}}_j \right) \quad (\text{C49})$$

$$= - \frac{1}{2} \sum_{j,k} \delta_j \delta_k \hat{\mathbf{r}}_j \cdot \hat{\mathbf{r}}_k + \frac{1}{2} \sum_{j=1}^J \delta_j^2 \quad (\text{C50})$$

$$= - \frac{1}{2} \left(\sum_{j=1}^J \delta_j \hat{\mathbf{r}}_j \right) \cdot \left(\sum_{k=1}^J \delta_k \hat{\mathbf{r}}_k \right) - Q_2^{(2)} \quad (\text{C51})$$

$$= - \frac{1}{2} \|\vec{\mathbf{R}}\|^2 - Q_2^{(2)} \quad (\text{C52})$$

where we define the resultant vector

$$\vec{\mathbf{R}} \equiv \sum_{j=1}^J \delta_j \hat{\mathbf{r}}_j. \quad (\text{C53})$$

Hence

$$\frac{1}{2} \text{Tr} \left(\tilde{\mathcal{S}}_{\eta, \delta} \right) \approx 1 - \frac{1}{2} \|\vec{\mathbf{R}}\|^2 + Q_{1,1,1}^{(3)} + Q_{1,1,1,1}^{(4)} + Q_{1,1,2}^{(4)} + Q_{1,3}^{(4)} + Q_{2,2}^{(4)} + Q_4^{(4)}. \quad (\text{C54})$$

Since the above terms are all real, the fidelity is then obtained taking the square of Eq. C54 and truncating any cross terms higher than $\mathcal{O}(\sigma^4)$. Thus we obtain

$$\mathcal{F}(\eta, \delta) \approx 1 - \|\vec{\mathbf{R}}\|^2 + \frac{1}{4} \left(\|\vec{\mathbf{R}}\|^2 \right)^2 + 2Q_{1,1,1}^{(3)} + 2Q_{1,1,1,1}^{(4)} + 2Q_{1,1,2}^{(4)} + 2Q_{1,3}^{(4)} + 2Q_{2,2}^{(4)} + 2Q_4^{(4)}. \quad (\text{C55})$$

Averaging Eq. C55 over an ensemble of noise realizations δ then yields the approximate expression for $\langle \mathcal{F} \rangle$. However we make simplifying observation that only terms raised to even powers, or those summed over terms raised to even powers, survive the ensemble average. Moreover, since the vectors $\hat{\mathbf{r}}_m$ are uniformly distributed over the set $\{\pm\hat{\mathbf{x}}, \pm\hat{\mathbf{y}}, \pm\hat{\mathbf{z}}\}$, summing over compositions

$$(\hat{\mathbf{r}}_j \cdot \hat{\mathbf{r}}_k) \quad j < k < J \quad (\text{C56})$$

$$(\hat{\mathbf{r}}_j \times \hat{\mathbf{r}}_k) \cdot \hat{\mathbf{r}}_l \quad j < k < l < J \quad (\text{C57})$$

$$(\hat{\mathbf{r}}_j \cdot \hat{\mathbf{r}}_k) (\hat{\mathbf{r}}_l \cdot \hat{\mathbf{r}}_m) \quad j < k < l < m < J \quad (\text{C58})$$

will on average resolve to the zero vector. Hence, even in the presence of correlated noise random variables, it is appropriate to set

$$\langle Q_{1,1,1}^{(3)} \rangle \rightarrow 0 \quad (\text{C59})$$

$$\langle Q_{1,1,1,1}^{(4)} \rangle \rightarrow 0 \quad (\text{C60})$$

$$\langle Q_{1,1,2}^{(4)} \rangle \rightarrow 0 \quad (\text{C61})$$

$$\langle Q_{1,3}^{(4)} \rangle \rightarrow 0 \quad (\text{C62})$$

In fact these quantities are random variables sharply (and symmetrically) peaked around 0. However since they are $\mathcal{O}(\sigma^4)$ terms, the spread of their distributions is very small relative to the spread of $\langle \|\vec{\mathbf{R}}\|^2 \rangle$, which scales as $\mathcal{O}(\sigma^2)$. The noise-averaged fidelity to $\mathcal{O}(\sigma^4)$ therefore reduces to

$$\langle \mathcal{F} \rangle \approx 1 - \langle \|\vec{\mathbf{R}}\|^2 \rangle + \mathcal{O}(\sigma^4) \quad (\text{C63})$$

where the final term in Eq. C63 consists only of $\mathcal{O}(\sigma^4)$ terms

$$\mathcal{O}(\sigma^4) \equiv \frac{1}{4} \langle \|\vec{\mathbf{R}}\|^4 \rangle + 2 \langle Q_{2,2}^{(4)} \rangle + 2 \langle Q_4^{(4)} \rangle. \quad (\text{C64})$$

The remaining task is to study how correlations between the noise variables affect the distributions of the terms in Eq. C63, and hence the distribution of $\langle \mathcal{F} \rangle$. In particular how these distributions change depending on whether the noise falls into Markovian or quasi-static regimes. As stated in the main text, our key insight is the interpretation of the vector quantity $\vec{\mathbf{R}}$ in terms of a 3D random walk generated by adding J randomly-oriented steps with step lengths specified by δ . The distribution of $\langle \mathcal{F} \rangle$ then maps onto the distance square of this 3D random walk.

D. Universal Error Model

Here we sketch a generalization of our analytic framework, showing its applicability to *universal errors*, beyond the (perhaps most important) case of dephasing specifically treated above. As in the main text, errors are implemented by interleaving \mathcal{S}_η with a sequence of stochastic unitary rotations, yielding the noise-affected operation

$$\tilde{\mathcal{S}}_{\eta,\delta} \equiv U_1 \hat{C}_{\eta_1} U_2 \hat{C}_{\eta_2} \dots U_J \hat{C}_{\eta_J}. \quad (\text{D1})$$

However we now let the unitaries take the general form

$$U_j \equiv \exp \left[-i \vec{\Delta}_j \cdot \boldsymbol{\sigma} \right] \equiv \exp \left[-i \left(\Delta_j^{(z)} \hat{\sigma}_x + \Delta_j^{(y)} \hat{\sigma}_y + \Delta_j^{(z)} \hat{\sigma}_z \right) \right] \equiv \exp \left[-i \left(\Delta_j^{(z)} \hat{X} + \Delta_j^{(y)} \hat{Y} + \Delta_j^{(z)} \hat{Z} \right) \right] \quad (\text{D2})$$

where $\boldsymbol{\sigma} \equiv (\hat{\sigma}_x, \hat{\sigma}_y, \hat{\sigma}_z) \equiv (\hat{X}, \hat{Y}, \hat{Z})$ denotes a vector of Pauli matrices and the vector $\vec{\Delta}_j = (\Delta_j^{(x)}, \Delta_j^{(y)}, \Delta_j^{(z)})$ contains the error rotations induced about each Cartesian axis. That is, the unitary U_j causes the net rotation through an angle $\|\vec{\Delta}_j\|$ about the axis $\hat{\delta}_j \equiv \vec{\Delta}_j / \|\vec{\Delta}_j\|$ on the Bloch sphere. In this case the error process indicates a list of 3-component *vectors*

$$\boldsymbol{\delta} = (\vec{\Delta}_1, \vec{\Delta}_2, \dots, \vec{\Delta}_J). \quad (\text{D3})$$

Assuming the weak-noise limit, quantified by the perturbative condition $J\mathbb{E}[\|\vec{\Delta}\|^2] < 1$, we make the Taylor approximation

$$U_j \approx \mathbb{I} + i \left(\Delta_j^{(z)} \hat{\sigma}_x + \Delta_j^{(y)} \hat{\sigma}_y + \Delta_j^{(z)} \hat{\sigma}_z \right) + \dots \quad (\text{D4})$$

In the main text we presented a complete treatment up to fourth order to account for both the leading- and higher-order contributions to infidelity, and assuming a dephasing-only environment. Here we demonstrate rigorously how leading-order contributions to the fidelity take the same form when the noise model is universal, as described above. Once again, the leading order contribution derives from the term like Eq. 10 but with a slight amendment:

$$\xi_{1,1}^{(2)} = \sum_{\alpha,\beta} \tilde{\xi}_{1,1}^{(2)}(\mu, \nu) \quad (\text{D5})$$

where

$$\tilde{\xi}_{1,1}^{(2)}(\mu, \nu) = - \sum_{j < k} \Delta_j^{(\mu)} \Delta_k^{(\nu)} C_{1,j-1} \hat{\sigma}_\mu C_{j,k-1} \hat{\sigma}_\nu C_{k,J}. \quad (\text{D6})$$

Writing $C_{jk} = K_{j-1}^\dagger K_k$, and using $K_0 \equiv K_J \equiv \mathbb{I}$, we therefore obtain

$$C_{1,j-1} \hat{\sigma}_\mu C_{j,k-1} \hat{\sigma}_\nu C_{k,J} = K_0^\dagger K_{j-1} \hat{\sigma}_\mu K_{j-1}^\dagger K_{k-1} \hat{\sigma}_\nu K_{k-1}^\dagger K_J \quad (\text{D7})$$

$$= K_{j-1} \hat{\sigma}_\mu K_{j-1}^\dagger K_{k-1} \hat{\sigma}_\nu K_{k-1} \quad (\text{D8})$$

$$= \mathbf{P}_j^{(\mu)} \mathbf{P}_k^{(\nu)} \quad (\text{D9})$$

where the operators in the last equality generalize the definition in Eq.15, namely

$$\mathbf{P}_m^{(\mu)} \equiv K_{m-1} \hat{\sigma}_\mu K_{m-1}^\dagger \in \{\pm \hat{X}, \pm \hat{Y}, \pm \hat{Z}\}. \quad (\text{D10})$$

Once again, $0 \leq m \leq J$, and the $\mathbf{P}_m^{(\mu)}$ are always signed Pauli operators due to the property that the Clifford group is the normalizer of the Pauli group. We may therefore expand the operators in the basis of Pauli operators by writing

$$\mathbf{P}_m^{(\mu)} = x_m^{(\mu)} \hat{X} + y_m^{(\mu)} \hat{Y} + z_m^{(\mu)} \hat{Z} \quad (\text{D11})$$

where $x_m^{(\mu)}, y_m^{(\mu)}, z_m^{(\mu)} \in \{0, \pm 1\}$ with only one nonzero coefficient. Since the Clifford sequences (and hence subsequences) are uniformly random, the operators $\mathbf{P}_m^{(\mu)}$ are uniformly random, independent of the choice of μ . The associated unit vector

$$\hat{\mathbf{r}}_m^{(\mu)} \equiv (x_m^{(\mu)}, y_m^{(\mu)}, z_m^{(\mu)}), \quad \|\hat{\mathbf{r}}_m^{(\mu)}\| = 1 \quad (\text{D12})$$

therefore still points uniformly at random along one of the principle Cartesian axes $\{\pm\hat{x}, \pm\hat{y}, \pm\hat{z}\}$. Taking the trace over Eq. D7, using the cyclic composition properties of the Pauli matrices, and moving to vector notation we therefore obtain

$$\frac{1}{2}\text{Tr}\left(\mathbf{P}_j^{(\mu)}\mathbf{P}_k^{(\nu)}\right) = \hat{\mathbf{r}}_j^{(\mu)} \cdot \hat{\mathbf{r}}_k^{(\nu)}. \quad (\text{D13})$$

Consequently

$$\tilde{Q}_{1,1}^{(2)}(\mu, \nu) \equiv \frac{1}{2}\text{Tr}\left(\tilde{\xi}_{1,1}^{(2)}(\mu, \nu)\right) = -\sum_{j < k} \Delta_j^{(\mu)} \Delta_k^{(\nu)} \hat{\mathbf{r}}_j^{(\mu)} \cdot \hat{\mathbf{r}}_k^{(\nu)} \quad (\text{D14})$$

where leading order contribution derives from the sum of all such terms $\tilde{Q}_{1,1}^{(2)}(\mu, \nu)$

$$Q_{1,1}^{(2)} = \sum_{\mu, \nu=1}^3 \tilde{Q}_{1,1}^{(2)}(\mu, \nu). \quad (\text{D15})$$

Taking an ensemble average over the noise realizations, we obtain

$$\langle Q_{1,1}^{(2)} \rangle = \sum_{\mu, \nu=1}^3 \langle \tilde{Q}_{1,1}^{(2)}(\mu, \nu) \rangle \quad (\text{D16})$$

$$= -\sum_{\mu, \nu=1}^3 \sum_{j < k} \langle \Delta_j^{(\mu)} \Delta_k^{(\nu)} \rangle \hat{\mathbf{r}}_j^{(\mu)} \cdot \hat{\mathbf{r}}_k^{(\nu)}. \quad (\text{D17})$$

We now make a simplification in which we assume the errors in separate quadratures arises from distinct physical mechanisms (*e.g.* dephasing and depolarization noise arise from different mechanisms). In this case the correlation between separate error quadratures is zero, and

$$\langle \Delta_j^{(\mu)} \Delta_k^{(\nu)} \rangle = \tilde{\delta}_{\mu, \nu} \quad (\text{D18})$$

where $\tilde{\delta}_{\mu, \nu}$ is the Kronecker delta. With this assumption, we obtain

$$\langle Q_{1,1}^{(2)} \rangle = -\sum_{\mu=1}^3 \sum_{j < k} \langle \Delta_j^{(\mu)} \Delta_k^{(\mu)} \rangle \hat{\mathbf{r}}_j^{(\mu)} \cdot \hat{\mathbf{r}}_k^{(\mu)}. \quad (\text{D19})$$

Now the quantity $\Delta_j^{(\mu)} \Delta_k^{(\mu)} \hat{\mathbf{r}}_j^{(\mu)} \cdot \hat{\mathbf{r}}_k^{(\mu)}$ is invariant under exchange of lower indices, and we may recast the ordered sum as an unrestricted sum

$$\langle Q_{1,1}^{(2)} \rangle = -\frac{1}{2} \sum_{\mu=1}^3 \left[\sum_{j, k} \langle \Delta_j^{(\mu)} \Delta_k^{(\mu)} \rangle \hat{\mathbf{r}}_j^{(\mu)} \cdot \hat{\mathbf{r}}_k^{(\mu)} - \sum_{k=1}^J \langle (\Delta_j^{(\mu)})^2 \rangle \right] \quad (\text{D20})$$

Defining

$$\vec{\mathbf{R}}^{(\mu)} \equiv \sum_{i=1} \Delta_j^{(\mu)} \hat{\mathbf{r}}_i^{(\mu)} \quad (\text{D21})$$

we obtain

$$\langle Q_{1,1}^{(2)} \rangle = -\frac{1}{2} \sum_{\mu=1}^3 \langle \|\vec{\mathbf{R}}^{(\mu)}\|^2 \rangle - \langle Q_{1,1}^{(2)} \rangle \quad (\text{D22})$$

where the residual term

$$Q_2^{(2)} = -\frac{1}{2} \sum_{\mu=1}^3 \sum_{k=1}^J \left(\Delta_j^{(\mu)} \right)^2 \quad (\text{D23})$$

is independent of the Clifford sequence and cancels out in the full expansion of $\langle \mathcal{F} \rangle$, as in the case of the dephasing error model in the main text. The leading order contribution to the noise-averaged *infidelity* is therefore given by

$$1 - \langle \mathcal{F} \rangle \approx \sum_{\mu=1}^3 \left\langle \|\vec{\mathbf{R}}^{(\mu)}\|^2 \right\rangle \quad (\text{D24})$$

where each of the terms $\langle \|\vec{\mathbf{R}}^{(\mu)}\|^2 \rangle$ are *independent* random variable inheriting the PDF of the random walk, and taking the same form presented in the main text. Consequently each of the $\langle \|\vec{\mathbf{R}}^{(\mu)}\|^2 \rangle$ are mutually independent, gamma-distributed random variables. In general, each of these random walks will be distinct, and the sum in Eq. D24 is over non-identical gamma-distributed random variables. The total PDF may be obtained from successive applications of Eq. A4 by direct integration. In the event that all noise quadratures follow the *same gamma-distribution* $\Gamma(\alpha, \beta)$, the total PDF simply reduces to the rescaled gamma distribution $\Gamma(3\alpha, \beta)$ by an application of Eq. A6.

E. PDF Derivation - Markovian Regime

In the Markovian regime, we assume all noise random variables are i.i.d. Hence $\vec{\mathbf{R}}$ corresponds to a J -length unbiased random walk with step lengths sampled from the normal distribution $\mathcal{N}(0, \sigma^2)$. Since these step lengths are symmetrically distributed about zero, the distributions of the components of the walk vector $\delta_j \hat{\mathbf{r}}_j = (\delta_j x_j, \delta_j y_j, \delta_j z_j)$ are invariant with respect to the sign of the coefficients α_j in all Cartesian directions $\alpha \in \{x, y, z\}$. Ignoring the signs we therefore treat the the coefficients as binaries $\alpha_j \in \{0, 1\}$, where the zero event simply reduces the number of steps taken in that direction. Let

$$n_\alpha \equiv \sum_{i=1}^J |\alpha_i|, \quad \alpha \in \{x, y, z\}, \quad n_x + n_y + n_z = J \quad (\text{E1})$$

count the total number of nonzero components in each Cartesian direction over the sequence of walk vectors $\{\hat{\mathbf{r}}_1, \hat{\mathbf{r}}_2, \dots, \hat{\mathbf{r}}_J\}$. Thus

$$\vec{\mathbf{R}} = \left(\delta_1^x + \dots + \delta_{n_x}^x, \delta_1^y + \dots + \delta_{n_y}^y, \delta_1^z + \dots + \delta_{n_z}^z \right) \quad (\text{E2})$$

where the superscripts in δ_j^α indicate summing only over the subset of δ_j for which the coefficients α_j are nonzero. Thus we have

$$\|\vec{\mathbf{R}}\|^2 = \Delta_x^2 + \Delta_y^2 + \Delta_z^2, \quad \Delta_\alpha \equiv (\delta_1^\alpha + \delta_2^\alpha + \dots + \delta_{n_\alpha}^\alpha), \quad \alpha \in \{x, y, z\}. \quad (\text{E3})$$

Since all $\delta_j \sim \mathcal{N}(0, \sigma^2)$ are i.i.d. in the Markovian regime, so too are all the random variables in set $S_\alpha \equiv \{\delta_1^\alpha, \delta_2^\alpha, \dots, \delta_{n_\alpha}^\alpha\}$, $\alpha \in \{x, y, z\}$. The distribution of their sum is therefore given by

$$\Delta_\alpha \sim \mathcal{N}(0, n_\alpha \sigma^2). \quad (\text{E4})$$

Further, since each $\hat{\mathbf{r}}_j$ projects onto only a single Cartesian direction - and consequently the sets $S_{x,y,z}$ are mutually disjoint - the random variables $\Delta_{x,y,z}$ are mutually independent. The distribution of the sum of squares of non-identical Gaussians involves a generalized generally challenging to write down, requiring a generalized chi-square distribution. To preserve analytic tractability, we make the following simplification. Since the vectors $\hat{\mathbf{r}}_j$ are uniformly-distributed there is a $\frac{1}{3}$ probability of being parallel to any given Cartesian axis. The probability of getting any particular combination (n_x, n_y, n_z) is therefore given by the multinomial distribution

$$\mathcal{P}(n_x, n_y, n_z) = \frac{J!}{n_x! n_y! n_z!} \left(\frac{1}{3}\right)^{n_x} \left(\frac{1}{3}\right)^{n_y} \left(\frac{1}{3}\right)^{n_z}, \quad n_x + n_y + n_z = J \quad (\text{E5})$$

which, for $J \gtrsim 5$, is sufficiently peaked around $n_{x,y,z} = J/3$ that we may regard these values as fixed without significant error. In this case $\Delta_{x,y,z} \sim \mathcal{N}(0, J\sigma^2/3)$ reduce to i.i.d. random variables. The distribution of $\|\vec{\mathbf{R}}\|^2$ consequently reduces to chi-square distribution with 3 degrees of freedom. It is more convenient, however, to express this in more general terms as a member of the two-parameter family of gamma distributions (see Eq. A5), of which the chi-square is a special case. Specifically, we obtain the gamma distribution

$$\|\vec{\mathbf{R}}\|^2 \sim \Gamma(\alpha, \beta), \quad \alpha = \frac{3}{2}, \quad \beta = \frac{2J\sigma^2}{3} \quad (\text{E6})$$

with shape parameter α and scale parameter β . The distribution of a finite noise-ensemble average over $\|\vec{\mathbf{R}}\|^2$ is therefore specified by

$$\langle \|\vec{\mathbf{R}}\|^2 \rangle_n = \frac{1}{n} \sum_{j=1}^n \|\vec{\mathbf{R}}_j\|^2, \quad \|\vec{\mathbf{R}}_j\|^2 \sim \Gamma\left(\frac{3}{2}, \frac{2J\sigma^2}{3}\right) \quad (\text{E7})$$

where now the $\|\vec{\mathbf{R}}_j\|^2$ are i.i.d. random variables. But the sample mean over n gamma-distributed random variables simply yields a rescaled gamma distribution with $\alpha \rightarrow n\alpha$ and $\beta \rightarrow \beta/n$ (see Eq. A7). Consequently

$$\langle \|\vec{\mathbf{R}}\|^2 \rangle_n \sim \Gamma\left(\frac{3n}{2}, \frac{2J\sigma^2}{3n}\right) \quad (\text{E8})$$

with moments

$$\mathbb{E}[\langle \|\vec{\mathbf{R}}\|^2 \rangle_n] = J\sigma^2 \quad (\text{E9})$$

$$\mathbb{V}[\langle \|\vec{\mathbf{R}}\|^2 \rangle_n] = \frac{2}{3}J^2\sigma^4n^{-1}. \quad (\text{E10})$$

From Eq. E6 the PDF for $\|\vec{\mathbf{R}}\|^2$ now has the known form

$$f_{\|\vec{\mathbf{R}}\|^2}(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}\exp\left[-\frac{x}{\beta}\right] \quad (\text{E11})$$

where $\alpha = 3/2$, $\beta = 2J\sigma^2/3$, and $\Gamma(x)$ is the gamma function. The PDF for $\|\vec{\mathbf{R}}\|^4$ therefore given by the transformation (see Eq. A3)

$$f_{\|\vec{\mathbf{R}}\|^4}(x) = \frac{1}{2\sqrt{x}}f_{\|\vec{\mathbf{R}}\|^2}(\sqrt{x}) = \frac{3\sqrt{\frac{3}{2\pi}}e^{-\frac{3\sqrt{x}}{2J\sigma^2}}}{2(J\sigma^2)^{3/2}x^{1/4}} \quad (\text{E12})$$

By direct computation, the first two moments are then given by

$$\mathbb{E}[\|\vec{\mathbf{R}}\|^4] = \frac{5}{3}J^2\sigma^4 \quad (\text{E13})$$

$$\mathbb{V}[\|\vec{\mathbf{R}}\|^4] = \frac{80}{9}J^4\sigma^8 \quad (\text{E14})$$

It is relevant at this point to consider the relative weight of the terms $\langle \|\vec{\mathbf{R}}\|^4 \rangle_n$ and $\langle \|\vec{\mathbf{R}}\|^2 \rangle_n$ in the calculation of $\langle \mathcal{F} \rangle$. An application of the central limit theorem yields the approximation

$$\sqrt{n}\left(\langle \|\vec{\mathbf{R}}\|^4 \rangle_n - \mathbb{E}[\langle \|\vec{\mathbf{R}}\|^4 \rangle_n]\right) \sim \mathcal{N}\left(0, \mathbb{V}[\|\vec{\mathbf{R}}\|^4]\right) \quad (\text{E15})$$

That is, $\langle \|\vec{\mathbf{R}}\|^4 \rangle_n$ is approximately Gaussian-distributed with mean and variance

$$\mathbb{E}[\langle \|\vec{\mathbf{R}}\|^4 \rangle_n] = \mathbb{E}[\|\vec{\mathbf{R}}\|^4] = \frac{5}{3}J^2\sigma^4 \quad (\text{E16})$$

$$\mathbb{V}[\langle \|\vec{\mathbf{R}}\|^4 \rangle_n] = \mathbb{V}[\|\vec{\mathbf{R}}\|^4]/n = \frac{80}{9}J^4\sigma^8n^{-1} \quad (\text{E17})$$

The relative significance of these terms may be captured by the condition $\mathbb{V}[\langle \|\vec{\mathbf{R}}\|^4 \rangle_n] \lesssim \epsilon \mathbb{V}[\langle \|\vec{\mathbf{R}}\|^2 \rangle_n]$ where ϵ is some small fraction. From Eqs. E10 and E17 this condition is met provided

$$J \lesssim \sqrt{\frac{3\epsilon}{40}}\sigma^{-2} \quad (\text{E18})$$

For instance, if $\sigma \sim 0.01$, $\epsilon \lesssim 0.1$ provided we restrict $J \lesssim 1000$. Now, since $\langle \mathcal{F} \rangle$ involves a linear combination of the terms $\langle \|\vec{\mathbf{R}}\|^4 \rangle_n$ and $\langle \|\vec{\mathbf{R}}\|^2 \rangle_n$, the PDF of $\langle \mathcal{F} \rangle$ is approximately given by a convolution over the individual PDFs of these terms. However assuming ϵ is sufficiently small - so that the distribution of $\langle \|\vec{\mathbf{R}}\|^4 \rangle_n$ is sufficiently narrow - the primary contribution of this convolution is to shift the distribution by an amount approximately given by the mean of $\langle \|\vec{\mathbf{R}}\|^4 \rangle_n$. Thus we set

$$\langle \|\vec{\mathbf{R}}\|^4 \rangle_n \rightarrow \mathbb{E}[\langle \|\vec{\mathbf{R}}\|^4 \rangle_n] = \frac{5}{3}J^2\sigma^4 \quad (\text{E19})$$

The other $\mathcal{O}(\sigma^4)$ terms reduce to

$$\langle Q_{2,2}^{(4)} \rangle = \frac{1}{4} \left\langle \sum_{j < k} \delta_j^2 \delta_k^2 \right\rangle = \frac{1}{4} \sum_{j < k} \langle \delta_j^2 \rangle \langle \delta_k^2 \rangle = \frac{1}{4} \frac{J(J-1)}{2} \sigma^4 \quad (\text{E20})$$

$$\langle Q_4^{(4)} \rangle = \frac{1}{24} \left\langle \sum_{j=1}^J \delta_j^4 \right\rangle = \frac{1}{24} \sum_{j=1}^J \langle \delta_j^4 \rangle = \frac{1}{24} J (3\sigma^4) \quad (\text{E21})$$

Substituting in these constants, Eqs. C63 and C64, the noise average fidelity for Markovian errors reduces to

$$\langle \mathcal{F} \rangle \approx 1 - \langle \|\bar{\mathbf{R}}\|^2 \rangle_n + \frac{2}{3} J^2 \sigma^4, \quad \langle \|\bar{\mathbf{R}}\|^2 \rangle_n \sim \Gamma\left(\frac{3n}{2}, \frac{2J\sigma^2}{3n}\right) \quad (\text{E22})$$

Performing the appropriate linear transformations (Eq. A1) to incorporate the constant factors in Eq.E22, and using the definition of the PDF for a gamma-distribution (Eq. A5), the PDF for $\langle \mathcal{F} \rangle$ finally takes the form

$$f_{\langle \mathcal{F} \rangle}(F) = \frac{1}{\Gamma(\alpha)\beta^\alpha} \nu(F)^{\alpha-1} \exp\left[-\frac{\nu(F)}{\beta}\right] \quad (\text{E23})$$

$$\nu(F) = 1 - F + \frac{2}{3} J^2 \sigma^4 \quad (\text{E24})$$

$$\alpha = 3n/2, \quad \beta = 2J\sigma^2/3n \quad (\text{E25})$$

where $\Gamma(x)$ is the gamma function.

F. PDF Derivation - Quasi-Static (DC) Regime

In the DC regime we assume all noise random variables $\delta_j \equiv \delta$ are, in a given instance, identical (maximally correlated). However over separate instances δ is sampled from the normal distribution $\delta \sim \mathcal{N}(0, \sigma^2)$. In a given run, the random walk vector $\vec{\mathbf{R}} = \sum_{j=1}^J \delta \hat{\mathbf{r}}_j = \delta \sum_{j=1}^J \hat{\mathbf{r}}_j$ therefore reduces to a J -length unbiased random walk on a 3D lattice with *fixed* step length δ . In this case the noise random variables δ and Clifford-dependent random variables $\hat{\mathbf{r}}_j$ factorize, allowing us to express

$$\vec{\mathbf{R}} = \delta \vec{\mathbf{V}}, \quad \vec{\mathbf{V}} \equiv \sum_{j=1}^J \hat{\mathbf{r}}_j \quad (\text{F1})$$

where $\vec{\mathbf{V}} \in \mathbb{R}^3$ defines an unbiased random walk on a 3D lattice generated by adding J *unit-length* steps. Since we are interested in the norm square $\|\vec{\mathbf{R}}\|^2 = \delta^2 \|\vec{\mathbf{V}}\|^2$, however, any sign dependence of δ vanishes. Performing a finite ensemble average over noise randomizations we therefore obtain

$$\langle \|\vec{\mathbf{R}}\|^2 \rangle_n = \langle \delta^2 \rangle_n \|\vec{\mathbf{V}}\|^2, \quad \langle \|\vec{\mathbf{R}}\|^4 \rangle_n = \langle \delta^4 \rangle_n \|\vec{\mathbf{V}}\|^4 \quad (\text{F2})$$

Substituting these expressions into Eq. C63 we therefore obtain $\langle \mathcal{F} \rangle \approx 1 - \langle \delta^2 \rangle_n \|\vec{\mathbf{V}}\|^2 + \frac{1}{4} \langle \delta^4 \rangle_n \|\vec{\mathbf{V}}\|^4 + \mathcal{O}(\sigma^4)$ where the term $\mathcal{O}(\sigma^4)$ includes the quantities $\langle Q_{2,2}^{(4)} \rangle$ and $\langle Q_{2,2}^{(4)} \rangle$, which approximately reduce to constants. Since the term in $\|\vec{\mathbf{V}}\|^2$ and $\|\vec{\mathbf{V}}\|^4$ are now *highly correlated*, however, we cannot exploit simplifying properties such as approximate independence between primary and higher-order terms in the expansion. One approach would involve making the approximation that $\langle \delta^2 \rangle_n \rightarrow \sigma^2$ and $\langle \delta^4 \rangle_n \rightarrow 3\sigma^4$ for sufficiently large n , and then completing the square in $\|\vec{\mathbf{V}}\|^2$. In this case the PDF for $\langle \mathcal{F} \rangle$ could be obtained by successively performing linear, square and shifting transformations on the PDF of $\|\vec{\mathbf{V}}\|^2$. However to a good approximation it turns out most of the physics is captured by the first term $\langle \delta^2 \rangle_n \|\vec{\mathbf{V}}\|^2$. Hence we make the approximation

$$\langle \mathcal{F} \rangle \approx 1 - \langle \delta^2 \rangle_n \|\vec{\mathbf{V}}\|^2 \quad (\text{F3})$$

As shown in the main text, proceeding with this truncating the expansion produces good agreement with direct simulation. The PDF for $\langle \mathcal{F} \rangle$ is therefore obtained by incorporating the distributions of $\|\vec{\mathbf{V}}\|^2$ and $\langle \delta^2 \rangle_n$. The PDF of δ is given by the Gaussian $f_\delta(x) := \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{x^2}{2\sigma^2}\right)$. Hence the PDF for δ^2 is given by the transformation (see Eq. A3)

$$f_{\delta^2}(x) = 2 \left[\frac{1}{2x^{1/2}} f_\delta\left(x^{-1/2}\right) \right] \quad (\text{F4})$$

$$= \frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{2\sigma^2}} x^{\frac{1}{2}-1} \exp\left(-\frac{x}{2\sigma^2}\right) \quad (\text{F5})$$

$$= \frac{1}{\Gamma(\frac{1}{2})} \frac{1}{(2\sigma^2)^{1/2}} x^{\frac{1}{2}-1} \exp\left(-\frac{x}{2\sigma^2}\right) \quad (\text{F6})$$

$$= \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right) \quad (\text{F7})$$

where $\alpha = 1/2$ and $\beta = 2\sigma^2$, and $\Gamma(x)$ is the gamma function. Comparing this with Eq. A5, this is the PDF of a Gamma distribution with shape parameter α and scale parameter β . The distribution of a finite noise-ensemble average over δ_2 is therefore specified by

$$\langle \delta^2 \rangle_n = \frac{1}{n} \sum_{j=1}^n \delta_j^2, \quad \delta_j^2 \sim \Gamma\left(\frac{1}{2}, 2\sigma^2\right) \quad (\text{F8})$$

where the δ_j^2 are i.i.d. random variables. Now the sample mean over n gamma-distributed random variables simply yields a rescaled gamma distribution with $\alpha \rightarrow n\alpha$ and $\beta \rightarrow \beta/n$ (see Eq. A7). Consequently

$$\langle \delta^2 \rangle_n \sim \Gamma\left(\frac{n}{2}, \frac{2\sigma^2}{n}\right) \quad (\text{F9})$$

with moments

$$\mathbb{E}[\langle \delta^2 \rangle_n] = \sigma^2 \quad (\text{F10})$$

$$\mathbb{V}[\langle \delta^2 \rangle_n] = \frac{2\sigma^4}{n} \quad (\text{F11})$$

As outlined above, $\|\vec{\mathbf{V}}\|^2$ expresses the distance square of an unbiased random walk on a 3D lattice generated by adding J unit-length steps. Let R be the random variable representing the distance from the origin in a symmetric (Bernoulli) 3D random walk after J steps. Then the PDF for R is known to be

$$f_R(r) = \left(\frac{3}{2\pi J}\right)^{3/2} 4\pi r^2 e^{-\frac{3r^2}{2J}} \quad (\text{F12})$$

The distribution of the distance square $\|\vec{\mathbf{V}}\|^2 = R^2$ is therefore given by the transformation (see Eq. A3)

$$f_{\|\vec{\mathbf{V}}\|^2}(x) = \frac{1}{2x^{1/2}} f_R(x^{-1/2}) \quad (\text{F13})$$

$$= \left(\frac{3}{2\pi J}\right)^{3/2} 2\pi x^{1/2} \exp\left[\frac{-3x}{2J}\right] \quad (\text{F14})$$

$$= \left(\frac{2J}{3}\right)^{-3/2} \left(\frac{\sqrt{\pi}}{2}\right)^{-1} x^{1/2} \exp\left[\frac{-x}{2J/3}\right] \quad (\text{F15})$$

$$= \left(\frac{2J}{3}\right)^{-3/2} \Gamma\left(\frac{3}{2}\right)^{-1} x^{1/2} \exp\left[\frac{-x}{2J/3}\right] \quad (\text{F16})$$

$$= \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right) \quad (\text{F17})$$

where $\alpha = 3/2$ and $\beta = 2J/3$, and $\Gamma(x)$ is the gamma function. But this is the PDF of a Gamma distribution with shape parameter α and scale parameter β . Consequently

$$\|\vec{\mathbf{V}}\|^2 \sim \Gamma\left(\frac{3}{2}, \frac{2J}{3}\right) \quad (\text{F18})$$

with moments

$$\mathbb{E}[\|\vec{\mathbf{V}}\|^2] = J \quad (\text{F19})$$

$$\mathbb{V}[\|\vec{\mathbf{V}}\|^2] = \frac{2J^2}{3} \quad (\text{F20})$$

Thus, to first order the PDF for $\langle \mathcal{F} \rangle$ is specified by the product of independent gamma-distributed random variables. This class of distributions is generally difficult to express. However we may obtain a closed form for the PDF of $\langle \delta^2 \rangle_n \|\vec{\mathbf{V}}\|^2$ by direct integration (see Eq. A2), yielding

$$f_{\langle \mathcal{F} \rangle} = \frac{\kappa^{\frac{n+3}{4}} \left(\frac{\nu}{4}\right)^{\frac{n-1}{4}} K_{\frac{1}{2}(n-3)}(\sqrt{\kappa\nu})}{\Gamma\left(\frac{J}{2}\right)} \quad (\text{F21})$$

where $\nu = 1 - F$, $\kappa = 3n/J\sigma^2$ and $K_n(z)$ gives the modified Bessel function of the second kind. However for fairly reasonable ensemble sizes $n \gtrsim 50$ it is sufficient to approximate $\langle \delta^2 \rangle_n$ as the mean of its distribution, namely σ^2 . In this case the fidelity distribution reduces simply to the (scaled and shifted) gamma distribution associated with which $\|\vec{\mathbf{V}}\|^2$, yielding

$$f_{\langle \mathcal{F} \rangle}(F) = \frac{1}{\Gamma(\alpha)\beta^\alpha} \nu^{\alpha-1} \exp\left[-\frac{\nu}{\beta}\right] \quad (\text{F22})$$

$$\nu = 1 - F \quad (\text{F23})$$

$$\alpha = 3/2, \quad \beta = 2J\sigma^2/3 \quad (\text{F24})$$

G. Probability distribution functions for partially correlated noise

In the main text we derived gamma distributions for $\langle \mathcal{F} \rangle$ in the Markovian and DC limits, and assumed continuity of the distribution to interpolate between these extremal cases. In this appendix we present formal justification of this approach. We treat a specific intermediate-correlation-length model where errors are *block-correlated*, *i.e.* within blocks of length M , the noise value is identical, and there is no correlation between blocks. That is, each noise realization, δ , can be partitioned into blocks of length M , within which the $\delta_i \sim \mathcal{N}(0, \sigma^2)$ are constant. In block k , the corresponding random walk takes steps along the cartesian axes (in either direction), which we count as $m_k^{+x}, m_k^{-x}, m_k^{+y}, m_k^{-y}, m_k^{+z}, m_k^{-z}$. The step counts satisfy the constraint

$$m_k^{+x} + m_k^{-x} + m_k^{+y} + m_k^{-y} + m_k^{+z} + m_k^{-z} = M \quad (\text{G1})$$

and are multinomially distributed,

$$\mathbf{m}_k \equiv (m_k^{+x}, m_k^{-x}, m_k^{+y}, m_k^{-y}, m_k^{+z}, m_k^{-z}) \sim \mathcal{M}(M, \{\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\}). \quad (\text{G2})$$

The displacement vector associated with block k is then

$$\begin{aligned} \vec{\mathbf{R}}_k &= \delta_k (m_k^{+x} - m_k^{-x}, m_k^{+y} - m_k^{-y}, m_k^{+z} - m_k^{-z}) \\ &\equiv \delta_k \vec{\mathbf{V}}_k \end{aligned} \quad (\text{G3})$$

where $\vec{\mathbf{V}}_k$ is the displacement vector of a random walk associated with the k th block, involving M *unit-length* steps along the cartesian axes. The total displacement is then

$$\vec{\mathbf{R}} = \sum_{k=1}^N \vec{\mathbf{R}}_k, \quad N = J/M \quad (\text{G4})$$

where N is the number of M -length blocks in a sequence of total length J . In this picture the Markovian limit corresponds to the case $M = 1, N = J$ (J blocks, each consisting of a single Clifford gate); the DC limit corresponds to the case $M = J, N = 1$ (1 block of length J). We therefore obtain

$$\|\vec{\mathbf{R}}\|^2 = \sum_{i=1}^N \sum_{j=1}^N \vec{\mathbf{R}}_i \cdot \vec{\mathbf{R}}_j \quad (\text{G5})$$

$$= \sum_{k=1}^N \|\vec{\mathbf{R}}_k\|^2 + \sum_{i \neq j}^N \vec{\mathbf{R}}_i \cdot \vec{\mathbf{R}}_j \quad (\text{G6})$$

$$= \sum_{k=1}^N \delta_k^2 \|\vec{\mathbf{V}}_k\|^2 + \sum_{i \neq j}^N \delta_i \delta_j \vec{\mathbf{V}}_i \cdot \vec{\mathbf{V}}_j. \quad (\text{G7})$$

Taking the ensemble average over noise realizations then yields

$$\langle \|\vec{\mathbf{R}}\|^2 \rangle = \sum_{k=1}^N \langle \delta_k^2 \rangle \|\vec{\mathbf{V}}_k\|^2 + \sum_{i \neq j}^N \langle \delta_i \delta_j \rangle \vec{\mathbf{V}}_i \cdot \vec{\mathbf{V}}_j. \quad (\text{G8})$$

Here the angle brackets $\langle \cdot \rangle$ refer to an ensemble average over noise realizations in the limit of an *infinite* ensemble size ($n \rightarrow \infty$). From our assumption that the noise is perfectly block correlated (*i.e.* zero correlation between errors from different blocks), and Gaussian-distributed within each block $\delta_i \sim \mathcal{N}(0, \sigma^2)$, we therefore obtain

$$\langle \delta_i \delta_j \rangle = \sigma^2 \tilde{\delta}_{ij} \quad (\text{G9})$$

where $\tilde{\delta}_{ij}$ is the Kronecker delta. Thus, the second term in Eq. G8 vanishes, yielding

$$\langle \|\vec{\mathbf{R}}\|^2 \rangle = \sum_{k=1}^N \sigma^2 \|\vec{\mathbf{V}}_k\|^2 = \sigma^2 \sum_{k=1}^N [(m_k^{+x} - m_k^{-x})^2 + (m_k^{+y} - m_k^{-y})^2 + (m_k^{+z} - m_k^{-z})^2] \quad (\text{G10})$$

In general, $\langle \|\vec{\mathbf{R}}\|^2 \rangle$, depends on the details of the Clifford sequence $\boldsymbol{\eta}$ through its dependence on the $\vec{\mathbf{V}}_k$, or equivalently, on the step counts \mathbf{m}_k . Considering an ensemble of Clifford sequences, we can compute the statistics of $\langle \|\vec{\mathbf{R}}\|^2 \rangle$. The first two moments are straightforward to calculate exactly. We find

$$\begin{aligned} \mathbb{E}[(m_k^{+x} - m_k^{-x})^2 + (m_k^{+y} - m_k^{-y})^2 + (m_k^{+z} - m_k^{-z})^2] &= \sigma^2 M, \\ \mathbb{V}[(m_k^{+x} - m_k^{-x})^2 + (m_k^{+y} - m_k^{-y})^2 + (m_k^{+z} - m_k^{-z})^2] &= 2\sigma^4 M(M-1)/3, \end{aligned}$$

which, together with Eq. G10, therefore yields

$$\mathbb{E}[\langle \|\vec{\mathbf{R}}\|^2 \rangle] = \sigma^2 J, \quad (\text{G11})$$

$$\mathbb{V}[\langle \|\vec{\mathbf{R}}\|^2 \rangle] = 2\sigma^4 J(M-1)/3. \quad (\text{G12})$$

In the Markovian limit ($M = 1, N = J$), the expectation $\mathbb{E}[\langle \|\vec{\mathbf{R}}\|^2 \rangle] = \sigma^2 J$ is in fact independent of the details of the sequence, since \mathbf{m} consists of a single unit entry, so that the summand in Eq. G10 is identically 1. This is perhaps not surprising: if different Clifford sequences had different noise-averaged fidelities for Markovian noise, then the optimal sequences would be good candidates for dynamical-decoupling sequences to suppress Markovian noise. It is known that dynamical decoupling is not useful for correcting fast noise. The variance is consistent with the $n \rightarrow \infty$ limit, $\mathbb{V}[\langle \|\vec{\mathbf{R}}\|^2 \rangle_\infty] = 0$ in the main text (see Table I). For DC noise ($M = J, N = 1$), we find the same Clifford-averaged expectation as above, however

$$\mathbb{V}[\langle \|\vec{\mathbf{R}}\|^2 \rangle] = 2\sigma^4 J(J-1)/3 \approx 2\sigma^4 J^2/3 \quad (\text{G13})$$

in agreement with earlier results in this limit up to a term $\mathcal{O}(\sigma^4)$ (see Table I).

In a given experimental scenario, it may be that there is a maximum correlation length M_{\max} , beyond which gates are uncorrelated. The above results then suggests that the clifford-averaged variance will vary quadratically with J for $J \ll M_{\max}$, and linearly with J for $J \gg M_{\max}$, with a transition around $J = M_{\max}$. This may be a useful heuristic for identifying correlation lengths in experiments.

For a given sequence length, J , the PDF of $\langle \|\vec{\mathbf{R}}\|^2 \rangle$ is in fact discrete. However, for modest sizes of J and M we can approximate the PDF by a continuous distribution. Given that the PDFs in both the DC and Markovian limits are both well-approximated by gamma distributions, we can use the calculated expectation and variance above to guess the PDF in the regime of intermediate correlation length. Since $\mathbb{E}[\Gamma(\alpha, \beta)] = \alpha\beta$ and $\mathbb{V}[\Gamma(\alpha, \beta)] = \alpha\beta^2$, and using Eqs. G11 and G12, we guess

$$\langle \|\vec{\mathbf{R}}\|^2 \rangle \sim \Gamma\left(\frac{3J}{2(M-1)}, \frac{2(M-1)\sigma^2}{3}\right), \quad (\text{G14})$$

$$\approx \Gamma\left(\frac{3J}{2M}, \frac{2M\sigma^2}{3}\right), \quad \text{for } M \gg 1 \quad (\text{G15})$$

Eq. G15 interpolates between the two limiting cases.

This guess can be formalized by recalling $\langle \|\vec{\mathbf{R}}\|^2 \rangle = \sum_{k=1}^N \sigma^2 \|\vec{\mathbf{V}}_k\|^2$. However, each of the $\vec{\mathbf{V}}_k$ represent the displacement of a *independent* random walk taking M steps of unit length along the cartesian axes, and are formally identical to the random walk vector in the DC limit. From Eq. 41 the displacement square of this random walk, given J steps, is given by the gamma distribution $\|\vec{\mathbf{V}}\|^2 \sim \Gamma\left(\frac{3}{2}, \frac{2J}{3}\right)$. Consequently Eq. G10 describes the sum, scaled by σ^2 , of N *independent and identically distributed* random variables, each following the distribution

$$\|\vec{\mathbf{V}}_k\|^2 \sim \Gamma\left(\frac{3}{2}, \frac{2M}{3}\right). \quad (\text{G16})$$

By an application of Eqs. A6 and A7 we therefore recover

$$\langle \|\vec{\mathbf{R}}\|^2 \rangle \sim \Gamma\left(\frac{3J}{2M}, \frac{2M\sigma^2}{3}\right) \quad (\text{G17})$$

which is asymptotically correct for large M . In showing this we have demonstrated analytically that intermediate correlation structures between the Markovian and DC limits are also described by the gamma distribution.

H. Fidelity Statistics for Generically-Correlated Processes

Here we derive the expectation $\mathbb{E}[\langle \|\vec{\mathbf{R}}\|^2 \rangle]$ and variance $\mathbb{V}[\langle \|\vec{\mathbf{R}}\|^2 \rangle]$ stated in Eq. 45 of the main text, for an error process δ with generic correlation structure specified by an autocorrelation function. Consider an arbitrary time-series $x(t)$ describing some (continuous) time-dependent, wide-sense stationary error process. Then $x(t)$ may be characterized by an autocorrelation function

$$C_x(\tau) = \langle x(t)x(t+\tau) \rangle_t = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t)x(t+\tau) \quad (\text{H1})$$

where $\langle \cdot \rangle_t$ refers to an ensemble average over the time series and τ is the *time difference* between measurements. Invoking the Wiener-Khintchine theorem, $C_x(\tau)$ and the power spectral-density (PSD) $S(\omega)$ form a Fourier-transform pair

$$S(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} C(\tau) e^{i\omega\tau} d\tau \iff C(\tau) = \int_{-\infty}^{\infty} S(\omega) e^{-i\omega\tau} d\omega \quad (\text{H2})$$

We make use of these relations by discretizing the time-series, whereby

$$\vec{\mathbf{R}} = \sum_{j=1}^J \delta_j \hat{\mathbf{r}}_j, \quad \delta_j = x(t_j), \quad t_j/\tau_g \in \{1, 2, \dots, J\} \quad (\text{H3})$$

and τ_g is the time taken to perform a Clifford operation. The underlying error process is thereby discretely ‘‘sampled’’ by the Clifford sequence \mathcal{S}_η , and correlations between elements of δ separated by a time interval of ‘‘ k gates’’ are specified by the (discrete) autocorrelation function

$$C_\delta(k) \equiv \langle \delta_j \delta_{j+k} \rangle_j \quad (\text{H4})$$

Expectation

We begin by expanding

$$\langle \|\vec{\mathbf{R}}\|^2 \rangle = \langle \vec{\mathbf{R}} \cdot \vec{\mathbf{R}} \rangle = \sum_{i=1}^J \sum_{j=1}^J \langle x(t_i)x(t_j) \rangle \hat{\mathbf{r}}_i \cdot \hat{\mathbf{r}}_j. \quad (\text{H5})$$

Here the ensemble average indicated by $\langle \cdot \rangle$ is over the noise random variables, and so does not affect the Clifford-dependent unit vectors $\hat{\mathbf{r}}_i$. To obtain the full expectation we now take a separate expectation over the the random variables $\hat{\mathbf{r}}_i$, which does not affect the noise random variables. Since the Clifford unit vectors $\hat{\mathbf{r}}_i$ are uniformly and randomly distributed over the set $\{\pm \hat{\mathbf{x}}, \pm \hat{\mathbf{y}}, \pm \hat{\mathbf{z}}\}$ the expectation over the inner product $\mathbb{E}[\hat{\mathbf{r}}_i \cdot \hat{\mathbf{r}}_j] = \tilde{\delta}_{ij}$ (Kronecker delta). Hence

$$\mathbb{E}[\langle \|\vec{\mathbf{R}}\|^2 \rangle] = \sum_{i=1}^J \sum_{j=1}^J \langle x(t_i)x(t_j) \rangle \mathbb{E}[\hat{\mathbf{r}}_i \cdot \hat{\mathbf{r}}_j] \quad (\text{H6})$$

$$= \sum_{i=1}^J \sum_{j=1}^J \langle x(t_i)x(t_j) \rangle \tilde{\delta}_{ij} \quad (\text{H7})$$

$$= \sum_{i=1}^J \langle x(t_i)x(t_i) \rangle \quad (\text{H8})$$

$$= \sum_{i=1}^J \langle \delta_i \delta_i \rangle \quad (\text{H9})$$

$$= \sum_{i=1}^J C_\delta(0) \quad (\text{H10})$$

$$= J C_\delta(0) \quad (\text{H11})$$

Variance

For the variance we need to find the expectation of $\langle \|\vec{\mathbf{R}}\|^2 \rangle^2$. Using the above result, we momentarily define

$$\mathcal{E} \equiv \mathbb{E}[\langle \|\vec{\mathbf{R}}\|^2 \rangle] = \sum_{k=1}^J \langle \delta_k \delta_k \rangle = \sum_{k=1}^J \langle \delta_k^2 \rangle \quad (\text{H12})$$

Now we rewrite

$$\langle \|\vec{\mathbf{R}}\|^2 \rangle = \sum_{k=1}^J \langle \delta_k^2 \rangle + \sum_{i \neq j} \langle \delta_i \delta_j \rangle (\hat{\mathbf{r}}_i \cdot \hat{\mathbf{r}}_j) = \mathcal{E} + \sum_{i \neq j} \langle \delta_i \delta_j \rangle (\hat{\mathbf{r}}_i \cdot \hat{\mathbf{r}}_j) \quad (\text{H13})$$

Hence

$$\langle \langle \|\vec{\mathbf{R}}\|^2 \rangle \rangle = \left(\mathcal{E} + \sum_{i \neq j} \langle \delta_i \delta_j \rangle (\hat{\mathbf{r}}_i \cdot \hat{\mathbf{r}}_j) \right) \left(\mathcal{E} + \sum_{i' \neq j'} \langle \delta_{i'} \delta_{j'} \rangle (\hat{\mathbf{r}}_{i'} \cdot \hat{\mathbf{r}}_{j'}) \right) \quad (\text{H14})$$

$$= \mathcal{E}^2 + 2\mathcal{E}^2 \sum_{i \neq j} \langle \delta_i \delta_j \rangle (\hat{\mathbf{r}}_i \cdot \hat{\mathbf{r}}_j) + \sum_{i \neq j} \sum_{i' \neq j'} \langle \delta_i \delta_j \rangle \langle \delta_{i'} \delta_{j'} \rangle (\hat{\mathbf{r}}_i \cdot \hat{\mathbf{r}}_j) (\hat{\mathbf{r}}_{i'} \cdot \hat{\mathbf{r}}_{j'}) \quad (\text{H15})$$

As above, $\mathbb{E}[\hat{\mathbf{r}}_i \cdot \hat{\mathbf{r}}_j] = \tilde{\delta}_{ij}$, so taking the expectation over the the random variables $\hat{\mathbf{r}}_i$ causes the term $2\mathcal{E}^2 \sum_{i \neq j} \langle \delta_i \delta_j \rangle (\hat{\mathbf{r}}_i \cdot \hat{\mathbf{r}}_j)$ vanishes, yielding

$$\mathbb{E}[\langle \langle \|\vec{\mathbf{R}}\|^2 \rangle \rangle] = \mathcal{E}^2 + \sum_{i \neq j} \sum_{i' \neq j'} \langle \delta_i \delta_j \rangle \langle \delta_{i'} \delta_{j'} \rangle \mathbb{E}[(\hat{\mathbf{r}}_i \cdot \hat{\mathbf{r}}_j) (\hat{\mathbf{r}}_{i'} \cdot \hat{\mathbf{r}}_{j'})] \quad (\text{H16})$$

Hence the variance is given by

$$\mathbb{V}[\langle \|\vec{\mathbf{R}}\|^2 \rangle] = \mathbb{E}[\langle \langle \|\vec{\mathbf{R}}\|^2 \rangle \rangle] - \mathbb{E}[\langle \|\vec{\mathbf{R}}\|^2 \rangle]^2 \quad (\text{H17})$$

$$= \mathbb{E}[\langle \langle \|\vec{\mathbf{R}}\|^2 \rangle \rangle] - \mathcal{E}^2 \quad (\text{H18})$$

$$= \sum_{i \neq j} \sum_{i' \neq j'} \langle \delta_i \delta_j \rangle \langle \delta_{i'} \delta_{j'} \rangle \mathbb{E}[(\hat{\mathbf{r}}_i \cdot \hat{\mathbf{r}}_j) (\hat{\mathbf{r}}_{i'} \cdot \hat{\mathbf{r}}_{j'})] \quad (\text{H19})$$

Now for $i \neq j$ and $i' \neq j'$, the expectation over $(\hat{\mathbf{r}}_i \cdot \hat{\mathbf{r}}_j) (\hat{\mathbf{r}}_{i'} \cdot \hat{\mathbf{r}}_{j'})$ is zero unless either

$$i = i' \ \& \ j = j' \quad \implies \quad \mathbb{E}[(\hat{\mathbf{r}}_i \cdot \hat{\mathbf{r}}_j) (\hat{\mathbf{r}}_{i'} \cdot \hat{\mathbf{r}}_{j'})] = 1/3 \quad (\text{H20})$$

$$i = j' \ \& \ j = i' \quad \implies \quad \mathbb{E}[(\hat{\mathbf{r}}_i \cdot \hat{\mathbf{r}}_j) (\hat{\mathbf{r}}_{i'} \cdot \hat{\mathbf{r}}_{j'})] = 1/3 \quad (\text{H21})$$

The variance therefore reduces to

$$\mathbb{V}[\langle \|\vec{\mathbf{R}}\|^2 \rangle] = \sum_{i \neq j} \langle \delta_i \delta_j \rangle \langle \delta_i \delta_j \rangle \frac{1}{3} + \sum_{i \neq j} \langle \delta_i \delta_j \rangle \langle \delta_j \delta_i \rangle \frac{1}{3} = \frac{2}{3} \sum_{i \neq j} \langle \delta_i \delta_j \rangle^2 \quad (\text{H22})$$

Observing the correlations

$$\langle \delta_i \delta_j \rangle \equiv \langle \delta_i \delta_{i+(j-i)} \rangle \equiv C_\delta(j-i) \quad (\text{H23})$$

depend only on the difference index $k = j - i$, which runs between $-(J-1)$ and $J-1$, we may reexpress the sum over all $i \neq j$ as

$$\mathbb{V}[\langle \|\vec{\mathbf{R}}\|^2 \rangle] = \frac{2}{3} \sum_{k=-(J-1)}^{(J-1)} (J-|k|) (C_\delta(k))^2 - J(C_\delta(0))^2 \quad (\text{H24})$$

Since $C_\delta(k) = C_\delta(-k)$, we may reexpress the above simply as a one sided sum

$$\mathbb{V}[\langle \|\vec{\mathbf{R}}\|^2 \rangle] = \frac{4}{3} \sum_{k=1}^{J-1} (J-k) (C_x(k))^2 \quad (\text{H25})$$

where we have removed the contribution $J(C_\delta(0))^2$ by commencing the sum at $k = 1$.

I. Generating Arbitrary PSDs from Fourier Synthesis

Following previous work on engineering noise processes [29], we construct the $x(t)$ as a superposition of phase-randomized cosines

$$x(t) = \frac{\alpha\omega_0}{2} \sum_{q=1}^Q qF(q) \left[e^{i(\omega_q t + \psi_q)} + e^{-i(\omega_q t + \psi_q)} \right], \quad \psi_q \sim \text{UniformDistribution}[0, 2\pi] \quad (11)$$

Here α is a global scaling factor for setting the total power content, ω_0 is the mode separation incrementing the (angular) frequency in the Fourier superposition, and the mode frequencies are given by $\omega_q \equiv \omega_0 q$. The function $F(q)$ specifies the relative weighting of the Fourier components. The autocorrelation function is then given by

$$C_x(\tau) = \langle x(t + \tau)x(t) \rangle_t = \frac{\alpha^2\omega_0^2}{2} \sum_{q=1}^Q (qF(q))^2 \cos(\omega_q \tau) \quad (12)$$

where $\langle \cdot \rangle_t$ denotes averaging over all times t from which the relative lag of duration τ is defined. Invoking the Wiener-Khintchine theorem and moving to the Fourier domain we then obtain the power spectral density

$$S(\omega) = \frac{\pi\alpha^2\omega_0^2}{2} \sum_{q=1}^Q (qF(q))^2 \left[\delta(\omega - \omega_q) + \delta(\omega + \omega_q) \right]. \quad (13)$$

Thus in this formulation the power-spectral density is represented as a Dirac comb of discrete frequency components with the amplitude of the j th tooth determined by the quantity $(qF(q))^2$. It is then straightforward to specify the construction of any power-law PSD by writing the amplitude of the q th frequency component as a power-law, $S(\omega) \propto (q\omega_0)^p$. It therefore follows that the envelope function for the comb teeth in the phase modulation scales as

$$F(q) = q^{\frac{p}{2}-1} \quad (14)$$

Table IV shows the functional form required for $F(q)$ in order to achieve dephasing-noise PSDs of interest.

		Power Laws			
		$1/f^2$	$1/f$	White	Ohmic
p		-2	-1	0	1
$F(q)$		q^{-2}	$q^{-3/2}$	q^{-1}	$q^{-1/2}$

TABLE IV. Functional form of $F(q)$ for well-known power-law PSDs

From the standpoint of the Clifford sequence *sampling* the noise process, the PSD reconstruction $\tilde{S}(\omega)$ is obtained as the band-limited Fourier transform of I2, restricted to the domain $\tau \in [-J, J]$. Hence

$$\tilde{S}(\omega) = \mathcal{F} \left\{ C_x(\tau) [\Theta(\tau + J) - \Theta(\tau - J)] \right\} \quad (15)$$

$$= \frac{J\alpha^2\omega_0^2}{2\sqrt{2\pi}} \sum_{q=1}^Q (qF(q))^2 \left[\text{sinc}(J(\omega - \omega_q)) + \text{sinc}(J(\omega + \omega_q)) \right] \quad (16)$$

where

$$\lim_{J \rightarrow \infty} \tilde{S}(\omega) = S(\omega) \quad (17)$$

J. Confidence Bounds for $\hat{\mu}^{(J)}$

In the main text we have showed the noise averaged fidelity $\langle \mathcal{F} \rangle$ is, to first order, a random variable specified by

$$\langle \mathcal{F} \rangle = \Delta - \nu \quad (\text{J1})$$

where $\nu \sim \Gamma(\alpha, \beta)$ is a gamma-distributed random variable, and the values of the parameters Δ , α and β are dependent on character of the noise. In particular, restricting attention to the Markovian or DC regime, we have

	Markovian	DC
α	$\frac{3}{2}n$	$\frac{3}{2}$
β	$\frac{2}{3}J\sigma^2n^{-1}$	$\frac{2}{3}J\sigma^2$
Δ	$1 + \frac{2}{3}J^2\sigma^4$	1

TABLE V. Input parameters for calculating PDF of $\langle \mathcal{F} \rangle$.

The true mean fidelity formally obtained as an expectation over all possible fidelity outcomes \mathcal{F} defined on the support of the random variables η and δ is given by

$$\langle \mathcal{F} \rangle_{\eta, \delta} \equiv \mathbb{E}[\langle \mathcal{F} \rangle] = \Delta - \alpha\beta \quad (\text{J2})$$

These equivalent expressions for the total expectation value are hereafter denotes simply by $\mathbb{E}[\mathcal{F}]_J$, as in the main text. In the standard RB procedure this expectation value is estimated by the sample mean

$$\hat{\mu}^{(J)} = \frac{1}{k} \sum_{i=1}^k \overline{F}_{i, \langle \cdot \rangle}^{(J)} \quad (\text{J3})$$

The mean gate error p_{RB} is then approximated by the decay constant from an exponential fit for increasing J . It is therefore of interest to quantify the reliability of the estimate $\hat{\mu}^{(J)}$ as a function of the ensemble size. This may be expressed in terms of confidence intervals, treating the measured values $\overline{F}_{i, \langle \cdot \rangle}^{(J)}$ as random variables sampled from the distribution describing $\langle \mathcal{F} \rangle$. The distribution of $\hat{\mu}^{(J)}$ is therefore specified by

$$\hat{\mu}^{(J)} = \Delta - \bar{\nu}, \quad \bar{\nu} \equiv \frac{1}{k} \sum_{j=i}^k \nu_i, \quad \nu_i \sim \Gamma(\alpha, \beta) \quad (\text{J4})$$

where, since the ν_i are i.i.d. random variables we have $\bar{\nu} \sim \Gamma(k\alpha, \beta/k)$. The probability that $\hat{\mu}^{(J)}$ falls outside the confidence interval $[\mathbb{E}[\mathcal{F}]_J - L, \mathbb{E}[\mathcal{F}]_J + U]$ is therefore given by

$$\delta \equiv 1 - \mathcal{P}(\mathbb{E}[\mathcal{F}]_J - L \leq \Delta - \bar{\nu} \leq \mathbb{E}[\mathcal{F}]_J + U), \quad \bar{\nu} \sim \Gamma\left(k\alpha, \frac{\beta}{k}\right) \quad (\text{J5})$$

where $L, U > 0$ specify the lengths of the lower and upper error bars centered on the expectation value $\mathbb{E}[\mathcal{F}]_J$. Substituting $\mathbb{E}[\mathcal{F}]_J = \Delta - \alpha\beta$ and directly integrating the PDF over these confidence bounds we obtain

$$\delta(\alpha, \beta, L, U, k) = 1 - Q\left[k\alpha, k\left(\alpha - \frac{U}{\beta}\right), k\left(\alpha + \frac{L}{\beta}\right)\right] \quad (\text{J6})$$

where $Q(a, z_0, z_1)$ is the generalized regularized incomplete gamma function. This is defined in non-singular cases by

$$Q(a, z_0, z_1) \equiv \frac{\Gamma(a, z_0, z_1)}{\Gamma(a)} \quad (\text{J7})$$

where $\Gamma(a, z_0, z_1) \equiv \Gamma(a, z_0) - \Gamma(a, z_1)$ is the generalized incomplete gamma function, $\Gamma(a, z)$ is the incomplete gamma function, and $\Gamma(a)$ is the Euler gamma function. This expression can be further condensed by scaling the error-bar lengths by

$$L = G_L(\Delta - \mathbb{E}[\mathcal{F}]_J) = G_L\alpha\beta \quad (\text{J8})$$

$$U = G_U(\Delta - \mathbb{E}[\mathcal{F}]_J) = G_U\alpha\beta \quad (\text{J9})$$

where the values $G_L, G_U > 0$ specify the lengths of the error bars as fractions of the expected *infidelity*. In this framework, the reliability of the estimate may be quantified by the requirement that

$$\delta(\alpha, \beta, G_L\alpha\beta, G_U\alpha\beta, k) < \epsilon \quad (\text{J10})$$

where ϵ is a small fraction. Substituting values of α and β into Eq. J6 appropriate for Markovian (M) and DC regimes, we find

$$\delta^{(M)} = 1 - Q \left[\frac{3kn}{2}, \frac{3kn}{2}(1 - G_U), \frac{3kn}{2}(1 + G_L) \right] \quad (\text{J11})$$

$$\delta^{(DC)} = 1 - Q \left[\frac{3k}{2}, \frac{3k}{2}(1 - G_U), \frac{3k}{2}(1 + G_L) \right] \quad (\text{J12})$$

For user-defined G_L, G_U and ϵ , one may then solve the inequality in Eq. J10 for minimum k . Thus one may bound from below the size of the ensemble k necessary to justify the uncertainties quoted for the mean gate errors p_{RB} obtained from RB.
