Randomized Benchmarking for Non-Markovian Noise

Pedro Figueroa-Romero[1,*] Kavan Modi[2,†] Robert J. Harris[3] Thomas M. Stace[3] and
Min-Hsiu Hsieh[1]

[1]*Hon Hai Quantum Computing Research Center, Taipei, Taiwan*

[2]*ARC Centre for Engineered Quantum Systems & School of Physics and Astronomy, Monash University, Victoria, Australia*

[3]*ARC Centre for Engineered Quantum Systems & School of Mathematics and Physics, The University of Queensland, Brisbane, Australia*

Estimating the features of noise is the first step in a chain of protocols that will someday lead to fault-tolerant quantum computers. The randomized benchmarking (RB) protocol is designed with this exact mindset, estimating the average strength of noise in a quantum processor with relative ease in practice. However, RB, along with most other benchmarking and characterization methods, is limited in scope because it assumes that the noise is temporally uncorrelated (Markovian), which is increasingly evident not to be the case. Here, we combine the RB protocol with a recent framework describing non-Markovian quantum phenomena to derive a general analytical expression of the average sequence fidelity (ASF) for non-Markovian RB with the Clifford group. We show that one can identify non-Markovian features of the noise directly from the ASF through its deviations from the Markovian case, proposing a set of methods to collectively estimate these deviations, non-Markovian memory time scales, and diagnose (in)coherence of non-Markovian noise in a RB experiment. Finally, we demonstrate the efficacy of our proposal by means of several proof-of-principle examples. Our methods are directly implementable and pave the way for a better understanding of correlated noise in quantum processors.

I. INTRODUCTION

The biggest challenge faced in any quantum computation can almost unequivocally be said to be the presence of errors. Among these, noise arising from interactions with the surroundings of a system represent an important class that is still far from being well understood. Given the current widespread interest in designing complex fault-tolerant quantum systems, together with the fundamental restriction that no system can ever be fully isolated from its surroundings, the need to advance our understanding of this type of noise cannot be understated.

Over the last decade, the approach known as randomized benchmarking (RB) [1–5] has become the gold standard to certify the performance of gate sets and characterize the noise in computations involving these sets. RB generally refers to an experimental protocol allowing estimation of the error rates of a gate set by quantifying their control fidelity as a function of the number of gates [6]. Moreover, it does so in an efficient way that is robust to state preparation and measurement (SPAM) errors, as opposed to approaches such as quantum process tomography (QPT) [7]. It is important to point out, however, that the two approaches are rather complementary [8], as RB extracts less information about the noise, namely average error rates of average gates, but requires little resources for high confidence [9], while QPT allows noise to be fully reconstructed but with a higher resource cost [10]. Aside from QPT and RB, there is a plethora of other methods lying in between, such as gate-set tomography [11], compressed sensing [12,13], or direct fidelity estimation [14–16], to name a few, to characterize quantum devices. The main reason why RB has become an essential tool for quantum technologies is thus its practicality and applicability to realistic experimental settings.

The most common versions of RB protocols are executed for sequences of Clifford gates [3], and consider noise that is both time and gate independent, in particular, Markovian and context independent. In this case, it is observed that the so-called average sequence fidelity (ASF), i.e., a figure of merit relating to the gate fidelity of the noise [17], behaves as a decaying exponential in

---

*pedro.romero@foxconn.com
†kavan.modi@monash.edu

the number of gates applied in the sequence. Nevertheless, progress for time-dependent [9] and gate-dependent noise [18–20], as well as different gate sets [20–22] or other figures of merit has also been made [23,24]. Despite this, RB has generally remained elusive to a characterization in the presence of temporally correlated, so-called non-Markovian noise, and has rather been identified when the ASF does not behave as a decaying exponential in numerical and experimental studies [25–29]. Hence it is not an overstatement that overcoming the Markovianity assumption in RB remains one of the most important hurdles to clear towards fault tolerance in quantum computers.

Correlated noise has been thoroughly examined in particular scenarios, such as that of dephasing noise. For classical correlations, e.g., in Refs. [29,30] (and similarly in Ref. [31]), the noise is modeled as rotations of a qubit around the *z* axis as determined by a classical random variable, and deviations from the uncorrelated case are found. For the quantum counterpart, in Ref. [32] this is generalized to correlations being mediated by a bath, modeled as a multimode bosonic field interacting with the qubit. Similarly, correlations arising as interaction between neighboring qubits, so-called crosstalk [33] have been addressed in multiqubit RB protocols, generally noticing that averaging over a single qubit generally leads to a nonexponential decay of the ASF.

The study of temporal correlations in quantum systems necessarily require in its foundations a theory of quantum stochastic processes. The development of such a theory has much older origins than RB but has often been contentious and faced conceptual problems still widely discussed in the community [34]. Nevertheless, approaches in terms of higher-order maps [35,36] have proved successful in providing a general theory of quantum stochastic processes [34], in particular, unambiguously establishing a Markov condition [37,38] and providing an operational framework to characterize non-Markovian processes [39].

In this paper, we derive an analytical expression for the ASF of a RB experiment with the Clifford group under non-Markovian gate-independent noise. This allows the study of the behavior of ASF decays due to non-Markovianity, and particularly of deviations from exponential decays, given a model for the noise. We also discuss ways in which the relevant time scales, i.e., sequence lengths, for finite non-Markovian noise can be determined, and deviations from a Markovian decay can be quantified, both with or without an *a priori* model of the noise. The main limitation to these methods is precisely the RB protocol itself, as the non-Markovian ASF—as we show—is not a simple function of sequence length anymore. Nevertheless, just as in the standard Markovian case, the relevance of RB lies in its simplicity, as it allows non-Markovian features to be analyzed and quantified from experiment with relative ease. Overcoming these restrictions thus can be a focus of future research towards a practical and

more complete characterization of temporally correlated noise.

The paper is structured as follows. In Sec. II we introduce the RB protocol and discuss the theoretical setting employed in the remainder of the paper. In Sec. III we introduce the process-tensor framework and elaborate on how it is a natural framework for non-Markovian RB. In Sec. IV we present our main result within Eq. (7) and discuss some of its properties and consequences, including containment of the Markovian case, the issue of initial correlations and the impact of SPAM errors. In Sec. V we introduce a theoretical measure for non-Markovian RB by means of Eq. (15), discussing the case of classical correlations and the possibility of blindness to non-Markovian noise by RB. In Sec. VI we discuss the more realistic scenario of finite non-Markovian noise, with which we can operationally approach the problem of determining sequence lengths, i.e., time scales, at which temporal correlations in the noise are relevant, as well as quantifying deviations from an exponential decay whenever a model for the noise is unknown. In Sec. VII we show a proof-of-principle numerical example finding agreement with our analytical result, and discuss the effect of SPAM errors and non-Markovianity blindness. Finally, in Sec. VIII we demonstrate numerically how the memory length of a finite non-Markovian noise process can be estimated in practice, non-Markovian deviations quantified, and how to diagnose (in)coherence of non-Markovian noise. We conclude in Sec. IX with an overview of our results and a perspective for future work.

II. RANDOMIZED BENCHMARKING

While there are many variants of RB, and a general framework encompassing these can be established [6], for concreteness here we consider a RB protocol employing the Clifford group. This has been the most common approach in RB mainly because the elements on the Clifford group can be realized efficiently on a quantum processor [40–42]. The RB protocol is then as follows:

1. Prepare an initial state ρ.
2. Sample m distinct elements, $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_m$, uniformly at random from the Clifford group. Let $\mathcal{G}_{m+1} := \circ_{i=m}^{1} \mathcal{G}_i^\dagger = \mathcal{G}_1^\dagger \circ \cdots \circ \mathcal{G}_m^\dagger$, where $\circ$ denotes composition of maps and $\mathcal{G}^\dagger(\cdot) = G^\dagger(\cdot)G$ for any Kraus representation with unitaries G of the map $\mathcal{G}$. We refer to $\mathcal{G}_{m+1}$ as an undo gate.
3. Apply the composition $\circ_{i=1}^{m+1} \mathcal{G}_i$ on ρ. In practice, this amounts to applying a noisy sequence $\mathcal{S}_m := \circ_{i=1}^{m+1} \hat{\mathcal{G}}_i$ of length m on ρ, where $\hat{\mathcal{G}}_i$ are the physical noisy gates associated to $\mathcal{G}$.
4. Estimate the probability $f_m = \operatorname{tr}[\mathcal{M}\mathcal{S}_m(\rho)]$ via a positive operator-valued measure (POVM) element $\mathcal{M}$.

5. Repeat steps 1 to 4 n times for the same initial state ρ, same POVM element $\mathcal{M}$, and different sets of gates chosen uniformly at random $\{\mathcal{G}_i^{(1)}\}_{i=1}^m, \{\mathcal{G}_i^{(2)}\}_{i=1}^m, \ldots, \{\mathcal{G}_i^{(n)}\}_{i=1}^m$ from the Clifford group to obtain the probabilities $f_m^{(1)}, f_m^{(2)}, \ldots f_m^{(n)}$. Compute the average $\mathcal{F}_m = 1/n \sum_{i=1}^n f_m^{(i)}$. We refer to $\mathcal{F}_m$ as an average sequence fidelity.

6. Examine the behavior of the ASF $\mathcal{F}_m$ over different sequence lengths m.

The important insight in the RB protocol is that the ASF contains the average noise rate of the applied sequences, which can be extracted efficiently by analyzing it over varying sequence lengths. Specifically, when the noise is approximated as both independent of the gates applied and the time step at which these are applied, the ASF is given by

$$\mathcal{F}_m = Ap^m + B, \qquad (1)$$

where the error rate of the noise, or so-called noise strength, is given by $p \in [0, 1]$ and A, B are constants determined by state preparation and measurement errors [5]. This implies that having performed a RB experiment, the data of the experimental ASFs can be fitted to an exponential, from which the noise strength p and the SPAM factors can be extracted. The noise strength is directly related [17] to the gate fidelity of the noise with respect to the identity [28], and hence the labeling of $\mathcal{F}_m$ as a fidelity, but similarly other figures of merit can be used to learn average error rates through RB [23,24].

It is important to mention that SPAM errors are implicit to steps 1 and 4, that is, in an execution of the protocol, neither the initial state preparation nor the measurement of the output state might be perfect. In the time- and gate-independent scenario for the Clifford group, however, as seen in Eq. (1), SPAM errors are constants both independent of the error rate and the sequence length.

The exponential decay in Eq. (1) can be obtained by modeling each noisy gate as $\hat{\mathcal{G}}_i := \Lambda \circ \mathcal{G}_i$ for some completely positive trace-preserving (CPTP) map Λ; then the analytical average of the survival probabilities is given by the average over gates $\mathcal{G}_i$. For our purposes, we just care that the gates belong to a unitary 2-design, i.e., any distribution of gates replicating up to the second moment of the unitary group with the uniform Haar measure [43], such as the Clifford group. This implies that averaging over gates can be replaced with that over the Haar measure to obtain $\mathcal{F}_m$, and similarly the use of higher unitary designs could serve to characterize higher-order statistical properties of noise in RB [44]. Detail about how such averaging is carried out can be seen in Appendix B.

Importantly, one sees deviations from an exponential decay for more complex noise profiles, including

non-Markovian noise. Nevertheless, within a Markovian assumption, RB generally renders a linear combination of exponential decays for the ASF [6], with the particular case of gate dependence rendering a single perturbation term that decays exponentially as well with the sequence length [18,19]. Here, together with the assumption that the gates $\mathcal{G}$ belong either to the multiqubit Clifford group or to a 2-design, we make the assumption that the noise modeled by the maps Λ is gate independent. Other than these two assumptions, we are interested in temporal correlations in the noise described as being mediated by an external environment.

III. QUANTUM PROCESSES AND NON-MARKOVIANITY

The setting we consider is that of a bipartite quantum system, labeled SE, composed of a d_S-dimensional system S and a d_E-dimensional environment E. An experimenter, in principle, would apply the sequence $\mathcal{S}_m = \circ_{i=1}^{m+1} \mathcal{G}_i$ of Clifford gates $\mathcal{G}_i$ solely on S, and not have access to E. We consider different scenarios for the initial state ρ, which is solely prepared on S, but can nevertheless get correlated with E afterwards, accounting for a new type of SPAM error. We now can model the noisy gates as $\hat{\mathcal{G}}_i = \Lambda_i \circ (\mathcal{I}_E \otimes \mathcal{G}_i)$, where Λ_i acts on the full SE system and $\mathcal{I}_E$ is an identity map on E. In particular, we require the (gate-independent) noise maps Λ_i to be at least completely positive (CP) trace nonincreasing and allow a further time dependence, $\Lambda_i \neq \Lambda_j$ for $i \neq j$; this can further be constrained to requiring trace preservation (e.g., if the device never fails), unitarity (e.g., if the device is perfectly isolated), or time independence (the noise does not change between time steps).

The sequence $\mathcal{S}_m$ can be understood as a particular example of a quantum stochastic process where the underlying dynamics are given by the noise inherent to the computation on the whole SE. Motivated by what is done operationally in a laboratory, the process-tensor framework [34,38,39,45] provides the means by which we can treat the underlying noise source separately from what the experimenter has control over, which are the gates they apply. This effectively means that we can treat the whole noise in the sequence, together with the initial state, as a tensor Υ_m. We may contract this tensor with the set of Clifford gates, which too can be incorporated in a tensor $\mathfrak{C}_m$. This can be depicted as in the circuit of Fig. 1.

These tensors, Υ_m and $\mathfrak{C}_m$, just as any quantum map, can have different representations [46]. Here we employ the Choi-state representation, which simply is a generalization of the Choi-Jamiołkowski isomorphism for quantum channels [47]. Specifically, these can be written as

$$\Upsilon_m = \mathrm{tr}_E \left\{ \left[\underset{i=1}{\overset{m+1}{\bigcirc}} (\Lambda_i \otimes \mathcal{I}_{\mathrm{aux}}) \circ \mathscr{S}_i \right] \rho \otimes \psi^{\otimes m+1} \right\}, \qquad (2)$$
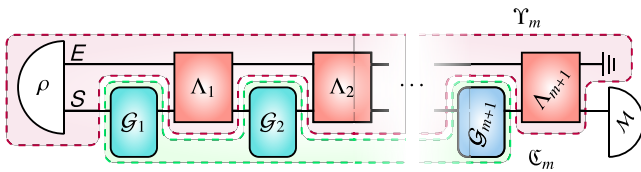
FIG. 1. **A non-Markovian noisy** RB **sequence as a process-tensor contraction**. An initial system-environment (*SE*) state ρ is acted on with a random Clifford element $\mathcal{G}_1$ on *S* alone, inducing a noise map Λ_1 on *SE*, followed by a random Clifford $\mathcal{G}_2$ inducing noise Λ_2, and so on until the undo map $\mathcal{G}_{m+1}$ is applied with some noise Λ_{m+1}. Finally, a POVM element $\mathcal{M}$ is applied on *S* and the environment is traced out (grounding symbol). This can be described by the contraction of a tensor Υ_m (upper box) comprising the initial state together with the noise process, with a tensor $\mathfrak{C}_m$ describing the sequence of applied Clifford gates, all followed by applying $\mathcal{M}$.

where $\mathcal{I}_{\text{aux}}$ is an identity map on an auxiliary space aux $= A_1 B_1 \cdots A_{m+1} B_{m+1} \simeq S^{\otimes 2(m+1)}$ composed of $m+1$ pairs of *S* systems, $\mathscr{S}_i$ is a swap gate between *S* and one of these pairs in the *i*th auxiliary space, say A_i, and $\psi := \sum |ii\rangle\langle jj|$ is an unnormalized maximally entangled state; on the other hand, for the gate sequence [48],

$$\mathfrak{C}_m = \mathbb{1}_S \otimes \left[\bigotimes_{i=1}^{m+1} \left(\mathcal{I}_{A_i} \otimes \mathcal{G}_i \right) \right] \psi^{\otimes m+1}. \qquad (3)$$

Detail about the definitions in Eqs. (2) and (3) can be consulted in Appendix A; we highlight, however, that we do not require the physical construction of these tensors nor access, e.g., to the space aux or the states ψ, but rather they are part of the theoretical framework that will let us study the RB protocol when the Λ_i maps are temporally correlated by means of the environment *E*. More broadly, the process-tensor framework generalizes the notion of a stochastic process to the quantum domain in a fully consistent way [49,50], resolving problems such as the initial correlation and not-CP problems, and fully accounting for memory effects [34].

The notion of Markovianity is formalized in the process-tensor framework through a proper operational Markov condition [38] as an independence of past observations, in turn containing the classical definition of Markovianity and unifying all quantum Markov conditions that had been proposed thus far [34,38,51]. Markovianity, and hence the absolute absence of temporal correlations in a process tensor, implies that no information is passed through *E* between time steps. This is mathematically manifest in the Choi state, which takes the form of a product of individual Choi states of quantum channels joining each step, as for $\mathfrak{C}_m$ in Eq. (A8). That is, temporal correlations in the process tensor correspond to spatial correlations in the

Choi-state representation, and more precisely then, a process tensor $\Upsilon_m^{(M)}$ will contain only Markovian noise if and only if there are noise maps $\Lambda_i^{(M)}$ acting solely on *S* such that

$$\Upsilon_m^{(M)} = \rho_S \otimes \left[\bigotimes_{i=1}^{m+1} \left(\Lambda_i^{(M)} \otimes \mathcal{I}_{B_i} \right) \right] \psi^{\otimes m+1}. \qquad (4)$$

Non-Markovianity can then naturally be quantified by means of any operationally meaningful distinguishability measure *D* with

$$\mathcal{N} := \min_{\Upsilon_m^{(M)}} D\left(\Upsilon_m, \Upsilon_m^{(M)} \right), \qquad (5)$$

where the choice of such a distance measure is rather a matter of practicality, as the minimization over all Markovian processes will often make the computation of $\mathcal{N}$ unfeasible. This can be alleviated either by choosing a measure *D* such as relative entropy, where the min argument is just a product of marginals, $\Upsilon_m^{(M)} = \rho_S \bigotimes_{i=1}^{m+1} \text{tr}_{\overline{i:i-1}} [\Upsilon_m]$, where $\text{tr}_{\overline{j:i}}$ means trace over all except between steps *i* to *j*, or otherwise placing relevant bounds on $\mathcal{N}$ for Schatten-norm measures, as done in Ref. [52,53] to study some statistical properties of non-Markovian processes. Here, we care about quantifying how non-Markovian a RB experiment is, which will boil down to quantifying how distinguishable a non-Markovian ASF is from a sensible Markovian counterpart.

We can now write the probability with *m* noisy gates as per step 4 of the RB protocol, $f_m = \text{tr}\left[\mathcal{M} \mathcal{S}_m(\rho) \right]$, in terms of the process tensor with

$$\mathcal{S}_m(\rho) = \text{tr}_E \left\{ \left[\mathop{\bigcirc}\limits_{i=1}^{m+1} \Lambda_i \circ (\mathcal{I}_E \otimes \mathcal{G}_i) \right] \rho \right\}$$
$$= \text{tr}_{\overline{S}} \left(\Upsilon_m \mathfrak{C}_m^T \right), \qquad (6)$$

where $\text{tr}_{\overline{S}}$ here means a partial trace over all intermediate input and output systems except the final *S* and *T* denotes a transpose. Computing the ASF, $\mathcal{F}_m$, then amounts to computing the average of $\mathfrak{C}^T$ over the applied gates, $\mathcal{G}_i$. This is a simplification allowing us to deal with the average over gates separately from the underlying noise. Furthermore, given that here we deal with the Clifford group, as explained in Sec. II, we can replace averaging over Clifford gates with averaging over the unitary group with the uniform Haar measure. To finally obtain $\mathcal{F}_m$, we have to contract the average gate sequence tensor with the noise tensor Υ_m, which will contain the noise inherent to the RB sequence, and, in particular, can be labeled as non-Markovian if the individual noise is correlated between time steps or Markovian otherwise.

We now present a general expression for the ASF $\mathcal{F}_m$ for RB of the Clifford group under non-Markovian noise and explore some of its consequences.

IV. AVERAGE SEQUENCE FIDELITY FOR NON-MARKOVIAN NOISE

Given a RB sequence with m Clifford gates affected by non-Markovian noise, we can construct the noise and gate-sequence process tensors, compute the average gate tensor, and contract with the noise tensor to get the average sequence fidelity. This yields the following:

Theorem: *Let ρ be the initial state of a system-environment, SE, composite with $d_S d_E = \dim(SE)$. Let $\mathcal{S}_m(\rho)$ describe a randomized benchmarking sequence of length m over Clifford gates with the* CP *map Λ_n acting on SE being the associated noise at the nth time step. Then, the average sequence fidelity $\mathcal{F}_m$ with a* POVM *element $\mathcal{M}$ is given by*

$$\begin{aligned} \mathcal{F}_m &= \mathrm{tr}\{\mathcal{M}\,\mathbb{E}[\mathcal{S}_m(\rho)]\} \\ &= \mathrm{tr}\left[\mathcal{M}\,\mathrm{tr}_E \circ \Lambda_{m+1} \circ (\mathscr{A}_m + \mathscr{B}_m)\,\rho\right], \end{aligned} \quad (7)$$

where $\mathbb{E}$ denotes average over Clifford gates, $\circ$ denotes composition of maps, and

$$\mathscr{A}_m(\rho) := \frac{\overset{m}{\underset{n=1}{\bigcirc}} \left(\$_{\Lambda_n} - \Theta_{\Lambda_n}\right) \otimes \mathcal{I}_S}{\left(d_S{}^2 - 1\right)^m} \left(\rho - \rho_E \otimes \frac{\mathbb{1}}{d_S}\right), \quad (8)$$

$$\mathscr{B}_m(\rho) := \left(\overset{m}{\underset{n=1}{\bigcirc}} \Theta_{\Lambda_n}\right) \rho_E \otimes \frac{\mathbb{1}}{d_S}, \quad (9)$$

with $\rho_E := \mathrm{tr}_S(\rho)$ being the reduced initial state in E; here $\$_{\Lambda_n}, \Theta_{\Lambda_n}$ are maps acting solely on E as defined by

$$\$_{\Lambda_n}(\varepsilon) := \sum_{s,s'=1}^{d_S} \langle s|\Lambda_n(\varepsilon \otimes |s\rangle\langle s'|)|s'\rangle \quad (10)$$

$$\Theta_{\Lambda_n}(\varepsilon) := \mathrm{tr}_S\left[\Lambda_n\left(\varepsilon \otimes \frac{\mathbb{1}}{d_S}\right)\right], \quad (11)$$

for any operator ε acting on E.

The proof can be found in full in Appendix C. As stated before, this amounts to writing the average sequence fidelity as the contraction of tensors $\mathcal{F}_m = \mathrm{tr}\left\{\mathcal{M}\,\mathrm{tr}_{\overline{S}}\left[\Upsilon_m\,\mathbb{E}\left(\mathfrak{C}_m^T\right)\right]\right\}$, where the average $\mathbb{E}\left(\mathfrak{C}_m^T\right)$ can be evaluated via the second moment of the unitary group with the Haar measure, given that the Clifford group constitutes a unitary 2-design.

We first notice that in the strict noiseless limit, $\Lambda_1 = \Lambda_2 = \ldots = \Lambda_n = \mathcal{I}$, we recover $\mathcal{F}_m \to \mathrm{tr}[\mathcal{M}\,\rho_S]$, where here $\rho_S = \mathrm{tr}_E(\rho)$, so that indeed Eq. (7) is bounded by one. For the ideal case of SE being a closed system, each Λ_n is a unitary. If there is no external time dependence on the noise and all temporal correlations are described by E, then $\Lambda_n = \Lambda$ for all n.

The two relevant terms to gain some insight about Eq. (7) are $\mathscr{A}_m$ and $\mathscr{B}_m$ in Eqs. (8) and (9), respectively, where the depolarizing effect of the noise on S is manifest, with $\mathscr{A}_m$ being partially depolarizing in S and $\mathscr{B}_m$ completely depolarizing in S. The action of $\mathscr{B}_m$, in particular, is independent of the initial state on S and picks up noise solely over E. Furthermore, if the initial state is uncorrelated, the effect of averaging a sequence of m gates in S is to totally decouple S from E, so that both $\mathscr{A}_m$ and $\mathscr{B}_m$ give a product state, with E carrying all the noise factors. Finally upon applying $\mathrm{tr}_E \circ \Lambda_{m+1}$ on $\mathscr{A}_m$, this would render a factor analogous to a product of noise strengths $p_1 p_2 \cdots p_m$.

The notation we use for $\mathscr{A}$ and $\mathscr{B}$, which here are quantum maps, is suggestive in that these reduce to the corresponding Ap^m and B, respectively, in the time-independent Markovian limit. In a Markovian scenario the environment is superfluous and we would have $\Lambda_n \to \mathcal{I}_E \otimes \Lambda_n^{(M)}$ together with $\rho \to \rho_E \otimes \rho_S$, i.e., the noise at each step is a CP map acting on S alone and the initial state on SE is completely uncorrelated. Then, if the noise is trace preserving as well, Eq. (7) reduces to the Markovian time-dependent ASF derived in Ref. [9],

$$\mathcal{F}_m^{(M)} = p_1 \cdots p_m A + B, \quad (12)$$

where,

$$p_n = \frac{\mathrm{tr}\left[\Lambda_n^{(M)}\right] - 1}{d_S{}^2 - 1}, \quad (13)$$

$$A = \mathrm{tr}\left[\mathcal{M}\Lambda_{m+1}^{(M)}\left(\rho_S - \frac{\mathbb{1}}{d_S}\right)\right], \quad B = \mathrm{tr}\left[\mathcal{M}\Lambda_{m+1}^{(M)}\left(\frac{\mathbb{1}}{d_S}\right)\right]. \quad (14)$$

That is, we get $\mathscr{A}_m(\rho) \to p_1 \cdots p_m (\rho - \mathbb{1}/d_S)$ and $\mathscr{B}_m(\rho) \to \mathbb{1}/d_S$ in this limit, which makes it clear that $\mathscr{B}$ renders only SPAM and non-Markovian noise contributions. Here $\mathrm{tr}\left[\Lambda_n^{(M)}\right] = \sum_\mu |\mathrm{tr}\,\lambda_\mu^{(M)}|^2$ where $\lambda_\mu^{(M)}$ are the Kraus operators of $\Lambda_n^{(M)}$. Furthermore, despite being complicated in the general case [54], the map $\$_{\Lambda_n^{(M)}}$ simply picks up a noise multiplicative factor, $\$_{\Lambda_n^{(M)}}(\varepsilon) = \mathrm{tr}\left[\Lambda_n^{(M)}\right]\varepsilon$ and $\Theta_{\Lambda_n^{(M)}}$ becomes an identity map, $\Theta_{\Lambda_n^{(M)}}(\varepsilon) = \varepsilon$, in this limit. Finally, Eq. (12) implies that we recover the decaying exponential in Eq. (1) for time-independent Markovian noise. The recovery of the standard ASF in this limit is shown in detail in Appendix D.

On the other hand, a unique feature when considering non-Markovian noise is initial correlations [55,56]; these could be particularly relevant in a non-Markovian RB experiment because the averaging over S gates depolarizes only the noise in S after the first gate is applied, but does nothing to correlations in the initial state. Furthermore, as pointed out before, if the initial state is uncorrelated, the ASF reduces to a quantity of the form $\mathcal{F}_m \to$

$\text{tr}\left[\mathcal{M}\,\text{tr}_E \circ \Lambda_{m+1}\left(\sigma_E^{(m)} \otimes \sigma_S^{(m)}\right)\right]$, and tracing the environment part would give a term analogous to a product of noise strengths $p_1 p_2 \cdots p_m$. This implies that, in general, when benchmarking non-Markovian errors with RB, the impact of SPAM errors could potentially be relevant, in general, in the error rates if such errors are large and generate initial correlations. In principle, the presence of such errors could also be diagnosed by an offset in the average sequence fidelity $\mathcal{F}_m$, as we exemplify numerically in Appendix G.

Finally, nonexponential decays in RB have often been attributed to non-Markovianity [27–29,32]: by mere inspection, setting $\Lambda_n = \Lambda$ on all steps n, we get $\bigcirc_{n=1}^m (\$_{\Lambda_n} - \Theta_{\Lambda_n}) = (\$_\Lambda - \Theta_\Lambda)^{\bigcirc m}$, which will generally not render an exponential decay in the ASF. It is important to point out that while non-Markovianity generally leads to nonexponential decays, there can also be other contextual factors [57], such as gate dependence or other rather arbitrary external time dependence leading to such behavior.

V. QUANTIFYING NON-MARKOVIANITY IN RANDOMIZED BENCHMARKING

Non-Markovianity in a quantum process can encompass both classical and quantum correlations; the latter is manifest in the Choi state of a process tensor whenever its components are entangled [58,59]. As examples of classical correlations, in Appendix F we reproduce the ASF of the model in Ref. [30], where classical temporal correlations are modeled via dephasing noise determined by a classical stochastic process; this effectively renders an ASF analogous to one that is Markovian time dependent with the noise parameter being a random variable. We also illustrate this via a shallow pocket model [34,60–62], where the time dependence in the ASF is explicit but the treatment as a Markov ASF decay remains the same. These examples suggest that while the general measure of non-Markovianity $\mathcal{N}$ for a process tensor in Eq. (5) is sensitive to any sort of temporal correlation, this might not necessarily be the case for the ASF.

A RB experiment could be blind to non-Markovianity in the sense of producing equivalent data of some Markovian noise model. It is, of course, a possibility for there to be a subclass of time-independent non-Markovian processes leading to exponential or almost exponential behavior, although as mentioned above, in general a time-independent noise does not lead to an exponential behavior unless the environment is superfluous. In Appendix G we exemplify this numerically with a spin interaction as the source of non-Markovian noise. Whether, in general, there exists a whole class of non-Markovian processes that can be classified as RB blind, together with criteria to decide
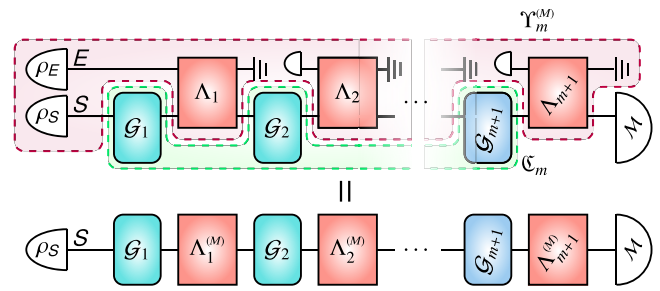

FIG. 2. **A Markovianized counterpart of a non-Markovian RB sequence.** Given a non-Markovian RB sequence with a noise process Υ_m, a Markovianized counterpart $\Upsilon^{(M)}$ can be given by one where the information carried in E is dissipated or lost between each step. The corresponding RB sequence has Markovian noise given by the CP maps $\Lambda_n^{(M)}$ acting as $\Lambda_n^{(M)}(\sigma) = \text{tr}_E \circ \Lambda_n(\varepsilon_n \otimes \sigma)$ for any pure state ε_n on E.

RB blindness, is an open question that could potentially be addressed in the near future.

There could be instances where having a non-Markovian noise process and being able to quantify its general non-Markovianity $\mathcal{N}$ with Eq. (5), we really only care about how much its associated ASF for the Clifford group deviates from a Markovian one. As, in principle, the set of possible Markovian processes to compare with would be restricted to a class specific to the given device to be benchmarked, ideally, we would look for a direct Markovian counterpart of the original non-Markovian noise process that we have. Thus, we propose to look at deviations from the ASF generated by the Markovianized process $\Upsilon^{(M)}$ where each noise map in the original non-Markovian noise Λ_n at time step n dissipates its E part: this amounts to taking a Markovian process with the initial state being uncorrelated $\rho \rightarrow \rho_E \otimes \rho_S$, and with dynamics at each step being given by the CP map $\Lambda_n^{(M)}$ on system S acting as $\Lambda_n^{(M)}(\sigma) = \text{tr}_E \circ \Lambda_n(\varepsilon_n \otimes \sigma)$ for an arbitrary pure state ε_n. This is depicted in Fig. 2.

Definition: *Let $\mathcal{F}_m$ be the average sequence fidelity of a randomized benchmarking experiment over the Clifford group with gate-independent non-Markovian noise. We define the* RB *non-Markovianity as*

$$\mathcal{N}_q^{\mathcal{F}_m} := \left\| \mathcal{F}_m - \mathcal{F}_m^{(M)} \right\|_q$$

$$= \left(\sum_{n=1}^m \left| \text{tr}\left\{ \mathcal{M}\,\text{tr}_{\bar{S}}\left[\left(\Upsilon_n - \Upsilon_n^{(M)}\right) \mathbb{E}\left(\mathfrak{C}_n^T\right) \right] \right\} \right|^q \right)^{1/q},$$

$$(15)$$

where $\mathcal{F}_m^{(M)}$ is the ASF *of the Markovian noise process associated to Υ_m, given by $\Upsilon_m^{(M)} := \rho_S \otimes$*

$$\left[\bigotimes_{i=1}^{m+1} \left(\Lambda_i^{(M)} \otimes \mathcal{I}_{B_i} \right) \right] \psi^{\otimes m+1}, \text{ where}$$

$$\Lambda_n^{(M)}(\sigma) := \mathrm{tr}_E \circ \Lambda_n(\varepsilon_n \otimes \sigma), \qquad (16)$$

for any σ acting on S and an arbitrary pure state ε_n on E.

The measure $\mathcal{N}_q^{\mathcal{F}_m}$ boils down to how well the POVM element $\mathcal{M}$ can distinguish $\mathrm{tr}_E \circ \Lambda \circ \mathscr{A}_m(\rho)$ from $p^m \Lambda^{(M)}(\rho_S - \mathbb{1}/d_S)$, as well as $\mathrm{tr}_E \circ \Lambda \circ \mathscr{B}(\rho)$ from $\Lambda^{(M)}(\mathbb{1}/d_S)$ for CPTP noise. Generic bounds can also potentially become possible with this non-Markovianity quantifier. Of course, the RB non-Markovianity measure in Eq. (15) also already makes it manifest that if an underlying noise process in a RB sequence is Markovian, then $\mathcal{N}_q^{\mathcal{F}_m} = 0$. The converse, however, might not necessarily be true or deviations could be negligible in practice [63], as we exemplify below numerically.

There could be several scenarios where Eq. (15) could be computed or estimated. One might be where either the full Markov process $\Upsilon_m^{(M)}$, or just an error rate is known, but once the RB experiment is run, deviations from $\mathcal{F}_m^{(M)}$ are observed, which most plausibly could be explained by non-Markovianity. This means we could actually compute $\mathcal{N}_q^{\mathcal{F}_m}$ directly from the experimental data and, e.g., analyze the observed ASF as a time-dependent RB decay. On the other hand, another scenario could be that we have a plausible model for the non-Markovian noise process Υ_m, and thus know the expression for the non-Markovian ASF $\mathcal{F}_m$ in Eq. (7). Then we may construct the Markovian counterpart $\mathcal{F}_m^{(M)} = p_1 \cdots p_m A + B$ of the ASF, compute $\mathcal{N}_q^{\mathcal{F}_m}$ in Eq. (15) and compare with the actual RB data.

Perhaps the most common case, however, will be that a RB experiment is run without *a priori* knowledge of a model for the noise and a nonexponential curve for the ASF is observed. At the same time, the observed statistics for a given physical process often depend only on a portion of their history rather than on their full past, implying that the relevant temporal correlations in the noise would likely be manifest in RB only over a finite sequence length. This notion of a finite memory within the noise will allow us to estimate, in practice, the amount of non-Markovian effects that are being observed in a RB experiment, as well as to operationally construct an analog of a Markovianized ASF, $\mathcal{F}_m^{(M)}$, to estimate deviations from Markovianity in RB.

VI. MODELS OF FINITE NON-MARKOVIAN NOISE

A possible scenario is to have an underlying noise process that is non-negligibly non-Markovian up to a given finite sequence length, with the remaining noise being effectively almost Markovian. This is related to the notion of finite quantum Markov order [45,64,65], which similar to the classical concept of finite Markov order, describes a quantum process where future statistics depend only on a

finite number of the previous operations on the system and its outcomes. We have then the following.

Corollary 1: (Initial non-Markovian noise). *Let ρ be an initial state on a system-environment, SE, composite and let $\mathcal{S}_m^{\ell:1}(\rho)$ describe a RB sequence of length m with noise described by CP maps Λ_n on SE for all n up to a sequence length $\ell < m$, with the rest of the sequence having noise CPTP maps $\Lambda_{\tilde{n}}^{(M)}$ on S and associated noise strengths $p_{\tilde{n}}$. Then the average sequence fidelity upon acting with a POVM element $\mathcal{M}$ is given by*

$$\begin{aligned} \mathcal{F}_m &= \mathrm{tr}\{ \mathcal{M} \, \mathbb{E}[\mathcal{S}_m^{\ell:1}(\rho)] \} \\ &= p_{\ell+1} \cdots p_m \, \mathrm{tr}\left[\mathcal{M} \Lambda_{m+1}^{(M)} \circ \mathrm{tr}_E \circ \mathscr{A}_\ell(\rho) \right] \\ &\quad + B \, \mathrm{tr}\left[\mathscr{B}_\ell(\rho) \right], \end{aligned} \qquad (17)$$

where $B = \mathrm{tr}\left[\mathcal{M} \Lambda_{m+1}^{(M)} (\mathbb{1}/d_S) \right]$ and $\mathscr{A}_n, \mathscr{B}_n$ are defined in Eqs. (8) and (9), respectively.

This implies that after a sequence length ℓ, non-Markovian noise will be manifest in an RB experiment as SPAM errors and not affect the subsequent decay, which for time-independent noise, would remain exponential. The assumption that the noise suddenly stops acting jointly on *SE* is at best an approximation, but one that can effectively be used whenever the non-Markovian noise effects are relevant only over some finite sequence length ℓ.

The main reason why this is important is twofold: first, detecting non-Markovian effects with a RB experiment will most likely be efficient for short sequence lengths, in the sense of requiring a small amount of fidelity samples, since there is no compounding error, so for small ℓ any significant non-Markovian noise effects can be resolved through RB; and second, the time scale of the memory effects displayed by the noise, i.e., the length ℓ inherent in the noise process, can then potentially be determined through a RB experiment. This would also be related to determining the order of a finite quantum Markov order process [66].

In Sec. VIII we show one such example where the sequence length ℓ of non-Markovian noise can be estimated from a RB experiment's data alone, and where a sensible time-independent Markovianized ASF, $\tilde{\mathcal{F}}_m^{(M)}$, can be constructed so as to operationally estimate non-Markovian deviations in such an experiment. This follows by noticing the following. Whenever we have finite non-Markovian noise, say over an initial sequence length ℓ, described by CPTP maps Λ_n, and an initial uncorrelated state, by choosing to fix $\ell - 1$ Cliffords after the first one to be identities, by Eq. (17), we get a Markovian decay with

$$\mathcal{F}_m = p_{\ell+1} \cdots p_m (p_{\ell:1} A) + B, \qquad (18)$$
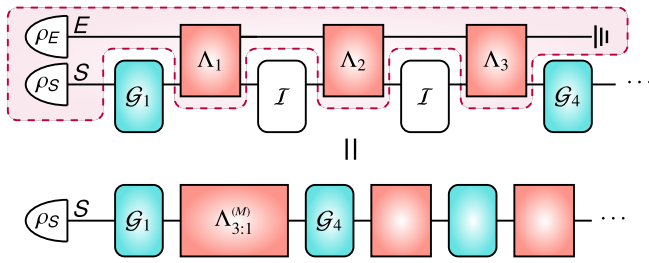
FIG. 3. **Determining time scales of finite non-Markovian noise.** A noise process with initial finite non-Markovian noise over a sequence length $\ell = 5$ will decay as described by a Markovian ASF after such a step, with the non-Markovian part contributing as SPAM error factors. By fixing to identity the gates of at least time steps, steps 2 and 3, the decay of the ASF corresponding to such a sequence becomes entirely Markovian; this allows us to operationally determine the time scales of finite non-Markovian noise as well as to construct sensible Markovian ASFs to quantify RB non-Markovianity, as exemplified in Sec. VIII.

where,

$$p_{\ell:1} := \frac{\text{tr}\left[\Lambda_{\ell:1}^{(M)}\right] - 1}{d_S^2 - 1},$$

$$\Lambda_{\ell:1}^{(M)}(\cdot) := \text{tr}_E\left[\underset{n=1}{\overset{\ell}{\bigcirc}} \Lambda_n\left(\rho_E \otimes \cdot\right)\right], \qquad (19)$$

that is, the initial block of finite non-Markovian noise looks like a single noise map $\Lambda_{\ell:1}^{(M)}$ if we randomize over a single Clifford within this block, with the remaining ones set to identities. This is more clearly seen in Fig. 3. This is, again, at best an approximation, but one that serves effectively to estimate the time scales for non-Markovian noise in a RB experiment. Of course, presumably, in realistic cases this would be more complicated and possibly all of the noise process be time dependent, albeit with small non-Markovianity effects.

Another scenario could be to have an almost Markovian noise initially, up to a sequence length ℓ, after which non-Markovianity turns significant. Then we have the following.

Corollary 2: (Late non-Markovian noise). *Let ρ be an initial state on a system-environment, SE, composite and let $\mathcal{S}_m^{m:\ell+1}(\rho)$ describe a RB sequence of length m with noise described by CPTP maps $\Lambda_n^{(M)}$ on S and noise strengths p_n for all n up to a sequence length $\ell < m$, with the rest of the sequence having noise CP maps $\Lambda_{\tilde{n}}$ on SE. Then the average sequence fidelity upon acting with a POVM element $\mathcal{M}$ is given by*

$$\mathcal{F}_m = \text{tr}\{\mathcal{M} \, \mathbb{E}[\mathcal{S}_m^{m:\ell+1}(\rho)]\}$$
$$= p_1 \cdots p_\ell \, \text{tr}[\mathcal{M} \, \text{tr}_E \circ \Lambda_{m+1} \circ \mathscr{A}_{m:\ell+1}(\rho)]$$
$$+ \text{tr}[\mathcal{M} \, \text{tr}_E \circ \Lambda_{m+1} \circ \mathscr{B}_{m:\ell+1}(\rho)], \qquad (20)$$

where

$$\mathscr{A}_{m:k}(\rho) := \frac{\underset{n=k}{\overset{m}{\bigcirc}} \left(\$_{\Lambda_n} - \Theta_{\Lambda_n}\right) \otimes \mathcal{I}_S}{\left(d_S^2 - 1\right)^{m-k+1}} \left(\rho - \rho_E \otimes \frac{\mathbb{1}}{d_S}\right),$$
$$(21)$$

$$\mathscr{B}_{m:k}(\rho) := \underset{n=k}{\overset{m}{\bigcirc}} \Theta_{\Lambda_n}(\rho_E) \otimes \frac{\mathbb{1}}{d_S}, \qquad (22)$$

with $\$_{\Lambda_n}$ and Θ_{Λ_n} defined in Eqs. (10) and (11), respectively.

This case might be relevant in practice whenever the sequence length ℓ is relatively small, both because non-Markovian noise would affect relevant computations and because the onset of such non-Markovian deviations could be resolved by a RB experiment.

Furthermore, in the middle of these two cases, we have the possibility of noise being intermittently non-Markovian, i.e., being displayed significantly over blocks of some finite sequence length. We have then the following.

Corollary 3: (Blocks of finite non-Markovian noise). *Let ρ be an initial state on a system-environment, SE, composite and let $\mathcal{S}_m^{\{m:\ell+1,\ell:1\}}(\rho)$ describe a RB sequence of length m with noise given by CP maps Λ_n on SE for all n up to a sequence length $\ell < m$, then at the ℓth step by $\Lambda_\ell(\cdot) \to \varepsilon \otimes \text{tr}_E \circ \Lambda_\ell(\cdot)$ for some E state ε, and with the rest of the sequence having noise CP maps $\Lambda_{\tilde{n}}$ on SE. Then the average sequence fidelity upon acting with a POVM element $\mathcal{M}$ is given by*

$$\mathcal{F}_m = \text{tr}\left\{\mathcal{M} \, \mathbb{E}\left[\mathcal{S}_m^{\{m:\ell+1,\ell:1\}}(\rho)\right]\right\}$$
$$= \text{tr}\left\{\mathcal{M} \, \text{tr}_E \circ \Lambda_{m+1} \circ \mathscr{A}_{m:\ell+1}\left[\varepsilon \otimes \text{tr}_E \circ \mathscr{A}_\ell(\rho)\right]\right\}$$
$$+ \text{tr}[\mathscr{B}_\ell(\rho)] \, \text{tr}\left\{\mathcal{M} \, \text{tr}_E \circ \Lambda_{m+1} \circ \mathscr{B}_{m:\ell+1}\left(\varepsilon \otimes \frac{\mathbb{1}}{d_S}\right)\right\},$$
$$(23)$$

with $\mathscr{A}_{\ell:k}$ and $\mathscr{B}_{\ell:k}$ defined as in Corollary 2.

This turns into a much more complicated ASF, but in essence any other combination considering finite non-Markovian noise can be considered. Of course, experimentally, there would be other challenges involved to study these more complicated finite non-Markovian noise processes, such as being restricted to short sequence lengths and/or requiring a larger amount of observations.

All cases in Corollaries 1, 2, and 3, are derived in detail in Appendix E. We now turn to study two numerical examples of non-Markovian RB.

VII. NUMERICAL MODEL: TWO-QUBIT FULLY NON-MARKOVIAN SPIN NOISE

As a proof of principle, we now test Eq. (7) with a qubit in S subject to time-independent unitary noise $\Lambda(\cdot) = \lambda(\cdot)\lambda^\dagger$, where $\lambda = \exp(-i\delta H)$, due to interaction with another qubit, identified as E, where H given by the two-spin interaction

$$H = J\, X_1 X_2 + h_x(X_1 + X_2) + h_y(Y_1 + Y_2), \qquad (24)$$

with X_i, Y_i being Pauli matrices acting on the ith site. Even though we use this as a simple theoretical construction and illustration, similar noise dynamics, albeit with many more considerations, come upon in real spin qubit quantum computers, e.g., as undesired crosstalk [67].

We take $J = 1.7$, $h_x = 1.47$, and $h_y = -1.05$ arbitrarily, for which we compute the ASF $\mathcal{F}_m$ as a function of m, both by numerical averaging and employing Eq. (7) with $\delta = 0.029\,475$. We take $\rho = |00\rangle\langle 00|$ and $\mathcal{M} = |0\rangle\langle 0|$ and ignore SPAM errors. We display the results in Fig. 4 together with its Markovianized ASF, $\mathcal{F}_m^{(M)}$, whereby the time-independent noise is modeled as a CP map given by $\Lambda^{(M)}(\cdot) = \mathrm{tr}_E \circ \Lambda(\varepsilon \otimes \cdot)$; specifically there we perform the numerical average over 50 samples of numerical sequence fidelities computed by sampling Haar random one-qubit unitaries, with the bars denoting the standard deviation of the mean.

We can verify that Eq. (7) effectively predicts the correct ASF, which is a rather complicated decaying function of m, clearly nonexponential. The numerical data remains reasonably well around the analytical prediction, with deviations becoming apparent for larger sequence lengths, which can be understood as compounded error. Despite these deviations being relatively small, they are significant enough that they can be probed numerically with a reasonable sample size for small sequence lengths, say for at least $m \lesssim 50$. This also makes manifest that for larger sequence lengths, many more sample runs would be needed to reveal non-Markovianity deviations. The RB non-Markovianity, $\mathcal{N}_q^{\mathcal{F}_m}$ with respect to the Markovian counterpart can also be swiftly computed through the sum of absolute values of the differences between $\mathcal{F}_m$ and $\mathcal{F}_m^{(M)}$; in particular, in Fig. 4 the RB non-Markovianity is not particularly high (between $\mathcal{N}_1^{\mathcal{F}_{100}} \approx 2.1$ and $\mathcal{N}_\infty^{\mathcal{F}_{100}} \approx 0.04$) but it is enough to be distinguished numerically for small sequence lengths.

Let us now consider the effect of SPAM errors. Suppose the initial state ρ is affected by the same Λ error for some small $\delta = \Delta_1$, and that $\mathcal{M}$ is slightly rotated via $\exp(-i\Delta_2 Y)$ for a small Δ_2. In Fig. 5 we show examples for both mild, $\Delta_1 = 0.04232$ and $\Delta_2 = 0.09321$, and much stronger noise with, $\tilde{\Delta}_1 = 0.2932$ and $\tilde{\Delta}_2 = 0.10321$.
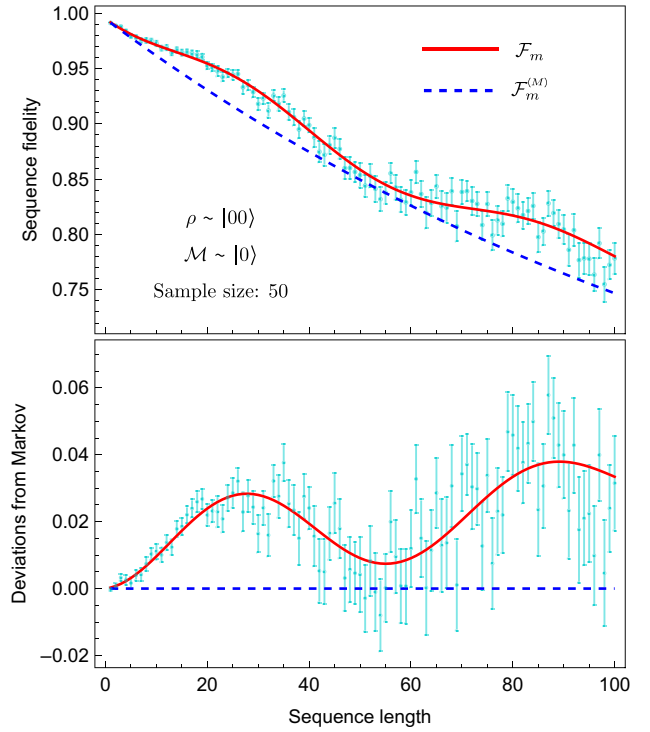

FIG. 4. **Average sequence fidelity for time-independent unitary non-Markovian noise and deviations from its Markovianized counterpart.** We consider the noise model described by the two-spin interaction of Eq. (24) with $\delta \approx 0.03$, $J = 1.7$, $h_x = 1.47$, and $h_y = -1.05$, for a single qubit as system S. We take $\rho = |00\rangle\langle 00|$ and $\mathcal{M} = |0\rangle\langle 0|$. Top: the continuous (red) line denotes the analytical ASF given by Eq. (7), with each point joined for clarity, the dots denote the numerical average of the ASF over 50 samples, with bars being the standard deviation of the mean (uncertainty of the numerical mean from the true mean), and the dashed (blue) line denotes the analytical ASF of the Markovianized process with time-independent noise $\Lambda^{(M)}(\cdot) = \mathrm{tr}_E \circ \Lambda(\varepsilon \otimes \cdot)$, here with $\varepsilon = |0\rangle\langle 0|$. Bottom: deviations of the ASF by both the analytical data (continuous red line) and numerical data produced by Eq. (7), from the Markovianized ASF, $\mathcal{F}_m^{(M)}$ (dashed blue line).

In Appendix G, we also show the case where the preparation affects only S by some rotation $\exp(-i\gamma X)$ with a small γ, but somehow does not generate correlations with E. Similar to Markovian noise models, add an offset to the average fidelities. In the non-Markovian case, however, the error rates do seem to be affected, presumably mainly because of the initial correlations induced by the preparation errors, as argued before in Sec. IV. This is still an aspect that would need to be examined closely, as when SPAM errors are significant, the offset also appears larger in the non-Markovian case, making it more difficult to distinguish non-Markovian errors from Markovian ones numerically.

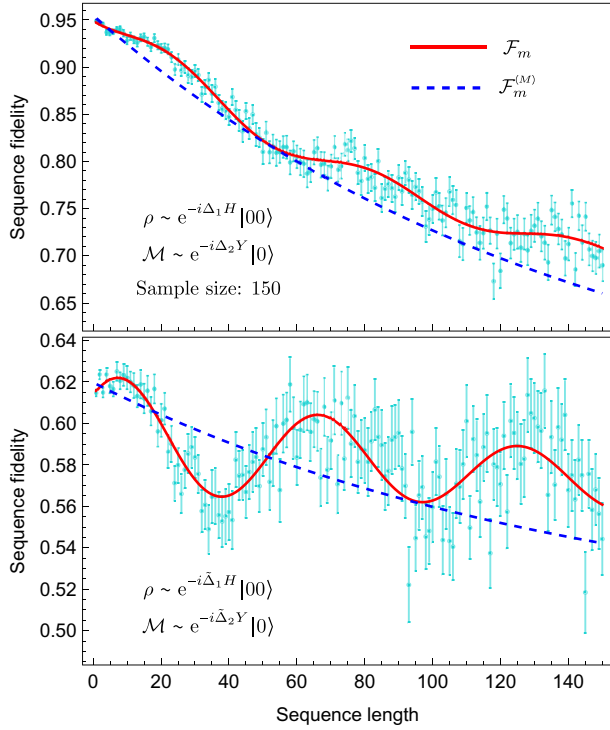We also notice in Appendix G that the non-Markovian effect of deviating from an exponential seems to fade in

FIG. 5. **Effect of** SPAM **errors in the two-qubit spin noise in Eq. (24).** We consider the noise model described by the two-spin interaction of Eq. (24) with $\delta \approx 0.03$, $J = 1.7$, $h_x = 1.47$, and $h_y = -1.05$, for a single qubit as system S. On both plots, the continuous (red) line denotes analytical ASF in Eq. (7), with each point joined for clarity, dots denote numerical average of the ASF over 150 samples, with bars being the standard deviation of the mean (uncertainty of the numerical mean from the true mean), and the dashed (blue) line denotes the analytical ASF of the Markovianized process. Top: the initial state $\rho = |00\rangle\langle00|$ is affected by the sequence noise $\Lambda \sim \exp(-i\Delta_1 H)$ for a small $\Delta_1 \approx 0.04$ and $\mathcal{M} = |0\rangle\langle0|$ is slightly rotated via $\Lambda \sim \exp(-i\Delta_2 Y)$ with a small $\Delta_2 \approx 0.09$. Bottom: $\tilde{\Delta}_1 \approx 0.29$ and $\Delta_2 \approx 0.10$ are increased considerably, amounting to large SPAM errors. In all cases the sample size is 100.

increasing E qubits; this is expected but this too would need to be thoroughly studied in realistic scenarios where the dimension of the environment is effectively finite [68,69]. On the other hand, we notice as well that an XX-spin chain displays practically no deviations from an exponential ASF decay presumably because of the absence of the external field, i.e., while the noise is non-Markovian, $\mathcal{N} \neq 0$, RB displays only minimal deviations, $\mathcal{N}_q^{\mathcal{F}_m} \approx 0$, and the behavior is almost exponential for all sequence lengths.

While this is mainly a numerical test of our main result, we now show an example and propose how to analyze a plausible realistic scenario for a RB experiment displaying finite non-Markovian noise, and having no prior knowledge of a model for such noise.

VIII. NUMERICAL EXAMPLE: NOISE MEMORY TIME SCALES, MARKOVIANIZED AVERAGE SEQUENCE FIDELITY, AND COHERENT NOISE

Consider now again a pair of qubits that up to some sequence length ℓ display an ASF that is mostly non-Markovian and subsequently turns almost Markovian. Here we model the underlying noise with

$$\Lambda_n^{(\ell)} = q_{n-\ell}\Lambda + (1 - q_{n-\ell})\Lambda^{(M)}, \qquad (25)$$

where $q_{n-\ell} = [1 + \exp(n - \ell)]^{-1}$ and both Λ, $\Lambda^{(M)}$ are determined as in the previous example with Eq. (24) with the same constants, $J = 1.7$, $h_x = 1.47$, and $h_y = -1.05$, but we now fix $\delta \approx 0.03$ for Λ and $\delta^{(M)} = 2.5\delta$ for $\Lambda^{(M)}$. In particular, we notice that q_k converges rapidly to 1 for $k < 0$, i.e., for a sequence lengths $m < \ell$, similarly converges rapidly to 0 for the remaining $k > 0$, meaning sequence lengths $m > \ell$, and finally $q_k = 0.5$ at $k = 0$, i.e., for a sequence length $m = \ell$.

Henceforth we assume that an experimenter would not know both what the noise maps $\Lambda_n^{(\ell)}$ are, nor what the non-Markovian finite sequence length ℓ is. Given Corollary 1, however, we know that whenever we have finite time-independent non-Markovian noise, within the Markovian part the decay will be practically exponential with the non-Markovian part acting as SPAM errors. Specifically, here we would get an ASF of the form of Eq. (18) for almost time-independent noise (i.e., with almost equal noise strengths $p_{\ell+1} \approx \ldots \approx p_m$) after such sequence length ℓ. Our expression assumes that the transition to Markovian noise occurs from step to step, however, even if dissipation occurs smoothly and non-Markovianity never entirely fades, we can still estimate at which sequence length the memory of the noise stops being relevant by identifying exponential decays. This also allows identification of a Markovianized time-independent ASF with which the experimenter can estimate the impact of non-Markovian errors.

A way to achieve this in practice is by fixing Clifford gates to identity wherever the decay appears nonexponential; this will give an exponential decay of the ASF whenever there is at most one random Clifford within the non-Markovian sequence. In Fig. 6 we display the ASF, $\mathcal{F}_m$, both analytical and numerical, for a finite noise memory process with noise modeled by Eq. (25), again taking $\rho = |00\rangle\langle00|$ and $\mathcal{M} = |0\rangle\langle0|$. We also display numerical ASFs, denoted $\mathcal{F}_{m/\{i,\ldots,j\}}$, with fixed identities at sequence lengths $i, \ldots, j$. The corresponding ASFs $\mathcal{F}_{m/\{i,\ldots,j\}}$ will normally be decreasing as $\mathcal{F}_m > \mathcal{F}_{m/\{1\}} > \mathcal{F}_{m/\{1,2\}} > \cdots > \mathcal{F}_{m/\{1,2,\ldots,10\}}$ given that fixing identities at subsequent steps is equivalent to set compounding error over such steps, which can be thought of simply as leaving the noise as a dynamical process to accumulate in time.

The non-Markovian sequence length can be identified by matching approximately the decay rate $p_{m/\{i,\ldots,j\}}$ of one
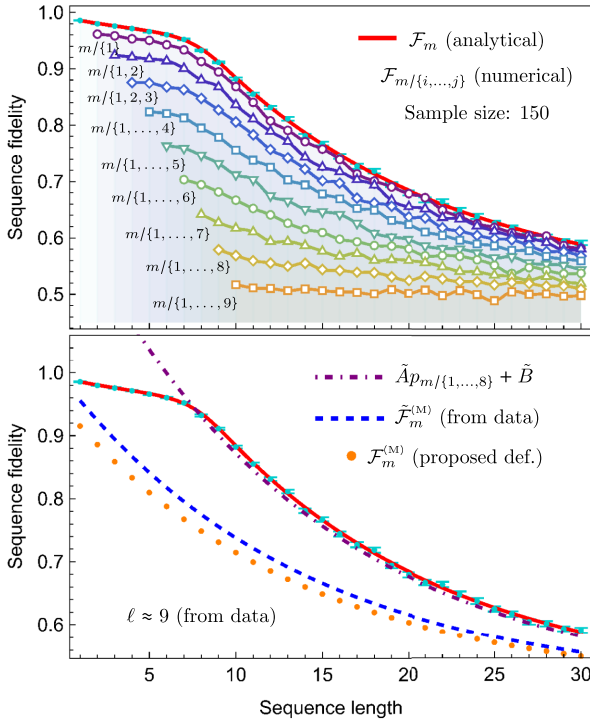
FIG. 6. **Determining the sequence length of finite non-Markovian noise.** Top: a RB experiment might display deviations from an exponential over a finite sequence length, as shown by the first ASF from the top, with the continuous red line denoting the underlying analytical ASF, $\mathcal{F}_m$. Such non-Markovian noise sequence length can be determined experimentally by fixing Cliffords to identities, running corresponding RB protocols to obtain ASFs $\mathcal{F}_{m/\{i,\dots,j\}}$, where $\{i,\dots,j\}$ denotes steps taken to identity (shown joined in the plot for clarity), fitting exponentials to these, and approximately matching their decay rates $p_{m/\{i,\dots,j\}}$ with the one in the manifestly exponential part in the original data. Bottom: the non-Markovian noise sequence length is determined to be $\ell \approx 9$; the dot-dash purple line denotes the curve with the decay rate $p_{m/\{1,\dots,8\}}$ and constants A, B of the fitted exponential of the original data starting at $m = 9$. Once ℓ is determined, a sensible Markovianized ASF, shown as a dashed blue line, can be taken with $p = p_{m/\{1,\dots,8\}}$ and reasonable criteria for fixing A, B; here we choose $A \approx B$ assuming low SPAM errors. The analytical Markovianized time-independent ASF of the form we propose in Sec. V is shown with orange dots just as a comparison.

of these sequences $\mathcal{F}_{m/\{i,\dots,j\}}$ with the corresponding one of the manifestly Markovian part in the full sequence. Once the decay rate is determined, a sensible time-independent Markovianized ASF, $\tilde{\mathcal{F}}_m^{(M)}$, can be constructed by making reasonable assumptions for the SPAM factors A and B. Detail of this process is shown in Appendix G. For the case of the RB experiments in Fig. 6, the non-Markovian noise sequence length is determined to be $\ell \approx 9$ by approximately matching $p_{m/1,\dots,8}$ with the corresponding one for the exponential fit between sequence lengths $15 \le m \le 30$ of the original data. Notice that in our model in Eq. (25),

at sequence length $m = 9$ the noise still has half probability of acting jointly on SE; the found $\ell \approx 9$ just says that after such sequence length the decay turns mostly exponential. We then finally constructed a Markovianized ASF with $\tilde{\mathcal{F}}_m^{(M)} = A p_{m/\{1,\dots,8\}} + B$ with $A \approx B$ supposing SPAM errors to be small; we compare this with a Markovianized construction as proposed in Sec. V, with time-independent noise given throughout by $\Lambda^{(M)}$.

This practical approach can work reasonably well, as we show in this example, and allow us both to determine the amount of memory within the noise, i.e., for how long the noise is being meaningfully non-Markovian, as well as to operationally construct a time-independent Markovianized ASF with which the impact of non-Markovianity in the noise can be quantified. The approach is consistent as well, in the sense that applying it to an exponential decay yields $\ell = 1$ and at most a numerical error due to fixing an identity on the first step.

There are, however, two apparent downsides to this approach, one is having to run another set of experiments requiring a higher amount of samples, given that the noise accumulates and makes it harder to get reliable data, and the second is that the ASFs with fixed identities $\mathcal{F}_{m/\{i,\dots,j\}}$ can eventually get too low if the noise memory is too high and not provide useful information. These are issues that could be resolved easily or otherwise depending on the particular case at hand.
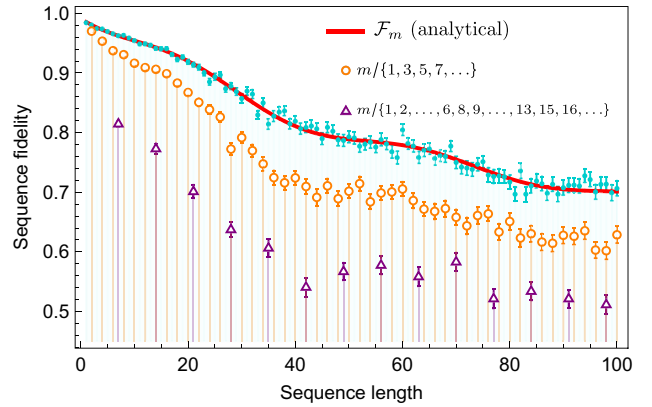

FIG. 7. **Diagnosing coherent non-Markovian noise.** Interleaving identities can allow determination of whether non-Markovian noise is coherent: in a continuous red line we show the analytical ASF with the same model of Eq. (24), with $\rho \sim |00\rangle$, $\mathcal{M} \sim |0\rangle$, and same parameters as displayed in Fig. 4, with numerical data displayed by teal points with bars denoting uncertainty around the mean; in circles and triangles are shown numerical ASFs $\mathcal{F}_{m/\{i,\dots,j\}}$ with Cliffords at time steps $\{1, 3, 5, \dots\}$ and $\{1, 2, \dots, 6, 8, 9, \dots, 13, 15, 16, \dots\}$ set to identity, respectively. Here we show only two examples of $\mathcal{F}_{m/\{i,\dots,j\}}$, not describing an exponential decay, for clarity, but similar behavior occurs interleaving identities at any set of steps. Numerical averages are done with 200 sequence fidelity samples.

Finally, while this approach cannot be used generally on fully non-Markovian noise, i.e., one over all sequence lengths, to determine operationally a sensible Markovianized ASF $\tilde{\mathcal{F}}_m^{(M)}$, it can nevertheless tell us whether the non-Markovian noise we are dealing with is coherent. This is important because whenever coherent noise can be diagnosed and characterized, e.g., with via unitary measures [8,24,70,71] or otherwise, in principle, it could be addressed and calibrated if we have access to the *E* qubits. Precisely then, we may tell if the noise is unitary over the whole *SE* if we get a general nonexponential behavior described by Eq. (7) no matter how many identities we fix, or if some dissipation is occurring and we rather have a scenario closer to that of Corollary 3 of finite non-Markovian blocks of noise. We use the model of the previous section in Eq. (24) to exemplify this, as shown in Fig. 7. The way we can proceed is to run RB experiments with a given number of identities interleaved; if the deviations from an exponential disappear, or fade considerably, this might point out to some dissipation, otherwise we would be able to identify the noise as highly coherent. Here once again the challenge is rather with numerical precision and compounded error, as interleaving identities highly degrades the ASF.

IX. CONCLUSIONS AND DISCUSSION

We have, (i) derived a general analytical expression for the average sequence fidelity of a randomized benchmarking experiment with the Clifford group subject to gate-independent non-Markovian noise, (ii) proposed a theoretical measure to quantify non-Markovian deviations in an ASF, (iii) derived the ASF for the case of finite non-Markovian noise, allowing us to operationally estimate both non-Markovian noise time scales and the measure of deviations from Markovianity, and (iv) exemplified all these with two proof-of-principle numerical examples. Along the paper we also discuss the effect of state preparation and measurement errors in RB with non-Markovian noise, as well as the case of classical correlations, which we argue can be treated as a Markovian time-dependent problem, and more generally the idea of RB blindness to a subclass of non-Markovian noise processes.

The ASF in our main result makes the depolarizing effect of averaging over Clifford gates on the system of interest manifest, while taking all of the noise in the sequence to the environment. The reduction of our main result to the standard Markovian gate-independent scenario is then straightforward with the trace over the environment giving rise to the noise strength and the SPAM error constants, and similarly one may consider cases where non-Markovian noise is finite over a subset of sequence lengths. Our main result also makes it clear that, in general, non-Markovian noise will display nonexponential behavior, although we point out that there could

be a subclass of non-Markovian models that do display an almost exponential decay that in practice would be almost impossible to resolve. We exemplify numerically how for small sequence lengths, deviations from Markovianity can be observed efficiently, as well as how the relevant time frames for finite non-Markovianity can be operationally determined and non-Markovian deviations in the ASF quantified.

We highlight as well that the methods to quantify non-Markovian effects, determine memory time scales, and diagnose coherence of non-Markovian noise, could be implemented beyond the randomized benchmarking framework on other noise benchmarking, characterization or mitigation approaches whenever temporal correlations should be taken into account, e.g., for crosstalk or leakage errors [72–74], similar to how it is done within the framework of resource theories in Ref. [75] with the notion of temporal coarse graining.

Needless to say, there are countless ways to move forward in the study of time-correlated errors in quantum computing. Arguably, the clearest ones arising from our paper within the RB procedure, would be to have a model-independent ASF, similar to the Markovian case, to benchmark other experimentally relevant groups or more generally arbitrary gate sets, to study context-dependent errors, with one possible way being the one we propose in Fig. 8, or to incorporate non-Markovianity in the general RB framework carefully constructed in Ref. [6]. All of these extensions have already been studied in quite some depth for Markovian errors and doing the same for the non-Markovian case would be a natural step forward. Other than this, there are questions that still would need to be understood such as the impact of non-Markovianity in decay rates as a function of sequence length, or explicitly how a higher or lower amount of non-Markovianity affects the average gate fidelity. More generally, there is
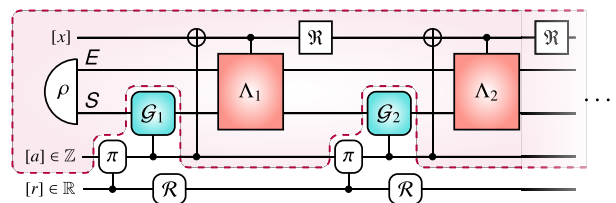

FIG. 8. **A circuit describing both gate-dependence and non-Markovianity.** The full system consists of a *SE* system in state ρ and classical registers [x], [a] and [r]. The π operations are permutations, $\mathcal{R}$ are randomizing operations and $\mathfrak{R}$ are reset operations. Vertical lines joining with • denote control with the corresponding classical register: essentially the CNOT gates (control with ⊕ in the [x] extreme) will carry the dependence from the applied gates at every step. Finally in RB the inverse operation (with an associated error) would be applied and a measurement in *S* would be performed.

still a need to deeply understand errors arising from temporal correlations, and our result represents a step in this direction.

APPENDIX A: THE PROCESS TENSOR OF THE NOISE AND GATE SEQUENCES

The process tensor is a multilinear map taking CP maps as the input and giving a single quantum state as the output. The operational scenario is the following: an initial quantum state ρ on the joint SE composite is acted on with an operation $\mathcal{G}_1$ solely on system S, which, in general, is given by a CP map; subsequently the whole composite evolves unitarily through a unitary map $\mathcal{U}_1$, after which an operation $\mathcal{G}_2$ is performed on S, then the whole evolves unitarily under a unitary map $\mathcal{U}_2$, and so on, until an intervention $\mathcal{G}_k$, followed finally by a unitary map $\mathcal{U}_k$. This means the final state in system S will be given by

$$\rho_S^{(k)} = \mathrm{tr}_E\left[\left(\overset{k}{\underset{i=1}{\bigcirc}} \mathcal{U}_i \circ \mathcal{G}_i\right)\rho\right], \qquad (A1)$$

where here we implicitly write $\mathcal{G}_\ell$ for $\mathcal{I}_E \otimes \mathcal{G}_\ell$. The process tensor is thus a map $\mathcal{T}_{k:1} : \mathcal{B}(\mathscr{H}_S)^{\otimes 2k} \to \mathcal{B}(\mathscr{H}_S)$, where $\mathcal{B}(\mathscr{H})$ means space of bounded linear operators over the Hilbert space $\mathscr{H}$, taking k CP maps as arguments and giving a quantum state as the output at time step k, i.e.,

$$\mathcal{T}_{k:1}\left[\vec{\mathcal{G}}_{k:1}\right] = \rho_S^{(k)}, \qquad (A2)$$

where $\vec{\mathcal{G}}_{k:1} = (\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_k)$. Such operations $\mathcal{G}_i$ are said to form an intervention and belong to an instrument, which can be understood as a generalization of a POVM, and the particular outcomes of each intervention yield a joint probability distribution describing a stochastic process.

The generalization of a Choi state, as given by the Choi-Jamiołkowski isomorphism [47,76], for a k-step process tensor follows by introducing k maximally entangled states $\psi_{A_iB_i} \in \mathcal{B}(\mathscr{H}_{A_i} \otimes \mathscr{H}_{B_i})$, where $\mathscr{H}_{A_i} \cong \mathscr{H}_S$ and similarly for B, and letting half of each (that is, the part of either subspace A_i or B_i) act as an input at every step by swapping the input spaces with the corresponding auxiliary space. This is more clearly illustrated in Fig. 9.
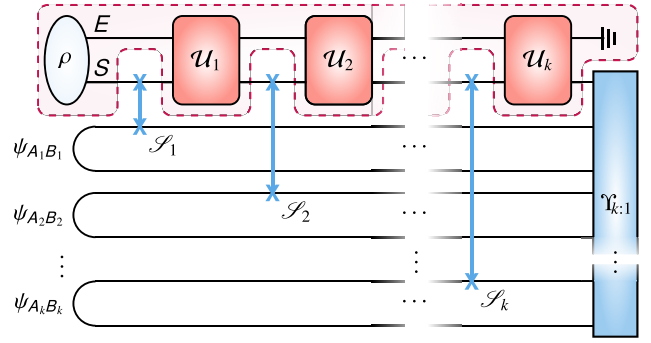

FIG. 9. **The Choi-state representation of a k-step process tensor**, denoted $\Upsilon_{k:1}$, can be obtained by swapping out the system, $\mathscr{S}_i$ with half a maximally entangled state, $\psi_{A_iB_i}$, at each step i. The final state is an unnormalized many-body state acting on a d_S^{2k+1}-dimensional system.

Specifically, the Choi-state of the process tensor takes the form

$$\Upsilon_{k:1} = \mathrm{tr}_E\left[\left(\overset{k}{\underset{i=1}{\bigcirc}} \mathcal{U}_i \circ \mathscr{S}_i\right)\rho \otimes \psi^{\otimes k}\right], \qquad (A3)$$

where here we are implicitly writing $\mathcal{U}_i$ for $\mathcal{U}_i \otimes \mathcal{I}_{A_1B_1\cdots A_kB_k}$ and $\psi^{\otimes k} = \psi_{A_1B_1} \otimes \cdots \otimes \psi_{A_kB_k}$. The generalized swap $\mathscr{S}_i$ between system S and an auxiliary space A_i at time step i is defined by $\mathscr{S}_i(\cdot) := \varsigma_i(\cdot)\varsigma_i$, where here

$$\varsigma_i := \sum_{\ell,j=1}^{d_S} \mathcal{I}_E \otimes |\ell\rangle\langle j| \otimes \mathcal{I}_{A_1B_1\cdots A_{i-1}B_{i-1}} \otimes |j\rangle\langle\ell|$$

$$\otimes \mathcal{I}_{B_iA_{i+1}B_{i+1}\cdots A_kB_k}. \qquad (A4)$$

Similar to the case of quantum channels, the isomorphism between the action of the process tensor and its Choi representation is manifest [77] through the relationship

$$\mathcal{T}_{k:1}\left[\vec{\mathcal{G}}_{k:1}\right] = \mathrm{tr}_{\bar{S}}\left[\Upsilon_{k:1}\left(\mathbb{1}_S \otimes \mathfrak{Y}_{k:1}^T\right)\right], \qquad (A5)$$

where here

$$\mathfrak{Y}_{k:1} = \left(\bigotimes_{i=1}^{k} \mathbb{1}_{A_i} \otimes \mathcal{G}_i\right)\psi^{\otimes k}, \qquad (A6)$$

is the Choi state for the operations $\vec{\mathcal{G}}_{k:1}$, the notation $\mathrm{tr}_{\bar{S}}$ stands for partial trace over all except output in S, and T denotes a transpose.

The tensor $\mathfrak{Y}_{k:1}$ is an example of a Markovian process tensor in the sense that it does not have any temporal correlations and thus assumes a product form. For a Markovian dynamical k-step process, $\Upsilon_{k:1}^{(M)} = \bigotimes_{i=1}^{k} \Phi_{i:i-1} \otimes \rho_S$, the Choi states $\Phi_{j:i}$ can either correspond to a closed system dynamics between steps i and j, or in general to
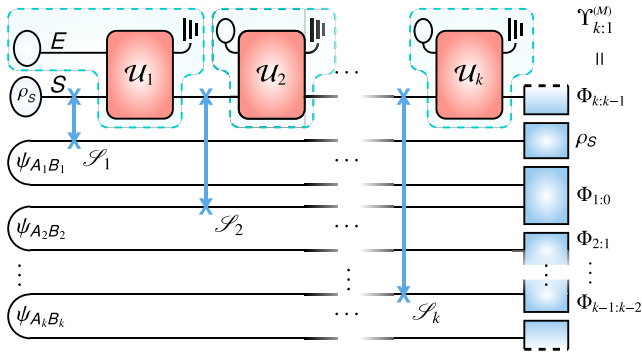
FIG. 10. **Process tensor for a Markov process** $\Upsilon_{k:1}^{(M)}$, having no temporal correlations, can be described by the Choi state $\Upsilon_{k:1}^{(M)}$ taking a product of Choi states $\Phi_{j:i}$ connecting adjacent time steps i to j. Each $\Phi_{j:i}$ corresponds to a CP map describing the (open) system evolution between time step i and j.

a CP dynamics, e.g., that of an open system, where the environment is discarded between each step and no information passes on to the next step, as shown in Fig. 10. The order of the spaces will be relevant whenever two tensors are contracted and can be written generally through swaps with auxiliary spaces and half maximally entangled states.

In this paper we describe a noise RB sequence as a process tensor with the dynamics being described by the noise at each step, $\mathcal{U}_i \to \Lambda_i$. Explicitly, in the main text we denote the tensor Choi state for a $m+1$-step process for the noise as

$$\Upsilon_m = \mathrm{tr}_E\left\{\left[\overset{m+1}{\underset{i=1}{\bigcirc}}(\Lambda_i \otimes \mathcal{I}_{\mathrm{aux}}) \circ \mathscr{S}_i\right]\rho \otimes \psi^{\otimes m+1}\right\}, \quad (A7)$$

where $\mathcal{I}_{\mathrm{aux}}$ is an identity map on an auxiliary space aux $= A_1B_1 \cdots A_{m+1}B_{m+1} \simeq S^{\otimes 2(m+1)}$ composed of $m+1$ pairs of S systems, $\mathscr{S}_i$ is a swap gate between S and one of these pairs in the ith auxiliary space, say A_i, and $\psi := \sum |ii\rangle\langle jj|$ is an unnormalized maximally entangled state. The tensor state for the gate sequence, on the other hand, can be defined simply

$$\mathfrak{C}_m = \mathbb{1}_S \otimes \left[\overset{m+1}{\underset{i=1}{\bigotimes}}(\mathcal{I}_{A_i} \otimes \mathcal{G}_i)\right]\psi^{\otimes m+1}, \quad (A8)$$

where strictly speaking, here we are defining $\mathfrak{C}_m = \mathbb{1} \otimes \mathfrak{A}_m$, where $\mathfrak{A}_m$ is the Choi state of the sequence of gates $\mathcal{G}_i$, and $\mathcal{G}_i$ can be defined to act on either auxiliary space A_i or B_i, the choice only depends on what auxiliary space the swap $\mathscr{S}_i$ on the definition of Υ_m swaps with, so that the contraction $\Upsilon_m\mathfrak{C}_m^T$ contracts the correct spaces.

APPENDIX B: AVERAGE GATE SEQUENCE

Representing the noise and gate sequences as process tensors implies that computing the ASF just requires computing the average gate sequence, $\mathbb{E}\left(\mathfrak{C}_m^T\right)$.

Whenever the gates entering this sequence belong to at least a unitary 2-design, we can simply replace the average over gates by that over unitaries distributed uniformly, i.e., according to the Haar measure, say μ, over the d_S-dimensional unitary group, which we denote $\mathbb{U}(d_S)$. The Haar measure is the unique measure on $\mathbb{U}(d)$ satisfying invariance under left and right multiplication, i.e., it is invariant under arbitrary rotations. Specifically, given a subset $\mathbb{V} \subseteq \mathbb{U}(d)$, we have $\mu(\mathbb{W}) = \int_{\mathbb{W}} d\mu(U)$ for the Haar measure, μ, with the (left-right invariance) property

$$\mu(\mathbb{W}) = \mu(V\mathbb{W}) = \int_{\mathbb{W}} d\mu(VU) = \int_{\mathbb{W}} d\mu(UV) = \mu(\mathbb{W}V), \tag{B1}$$

for any fixed $V \in \mathbb{U}(d)$. For any quantity f depending on a unitary $U \in \mathbb{U}(d)$, we denote integration over such unitary by $\mathbb{E}[f(U)]$ and refer to it as the Haar or uniform average of f.

Let the action of the unitary maps $\mathcal{G}_i$ be given by $\mathcal{G}_i(\cdot) = G_i(\cdot)G_i^\dagger$, then let us compute

$$\mathbb{E}\left[\mathfrak{C}_m^\dagger\right] = \mathbb{1}_S \bigotimes_{i=1}^{m+1} \sum_{\ell_i,k_i=1}^{d_S} \int_{\mathbb{U}(d_S)}\left(|\ell_i\rangle\langle k_i| \otimes G_i^\dagger|\ell_i\rangle\langle k_i|G_i\right)$$
$$\times d\mu(G_1)\cdots d\mu(G_m), \tag{B2}$$

where crucially, $G_{m+1} = G_1^\dagger G_2^\dagger \cdots G_m^\dagger$. This means that we need to be able to compute integrals with two pairs of G_i and $G_i^\dagger$ terms. One way to do this is by employing the 2-moment of $\mathbb{U}(d)$, given by [78,79]

$$\int_{\mathbb{U}(d)} U_{v_1u_1}U_{v_1'u_1'}^* U_{v_2u_2}U_{v_2'u_2'}^* \, d\mu(U)$$
$$= \sum_{\sigma,\tau\in\$_2} \delta_{u_1u_{\sigma(1)}'}\delta_{u_2u_{\sigma(2)}'}\delta_{v_1v_{\tau(1)}'}\delta_{v_2v_{\tau(2)}'}\mathrm{Wg}(\tau\sigma^{-1},d), \tag{B3}$$

where here the subindices $v_1u_1, v_1'u_1', v_2u_2, v_2'u_2'$ refer to components of the same matrix U (with U^* denoting complex conjugate), with the extra subindices $1, 2$ being mere labels, and where $\$_2$ is the symmetric group on two elements. The symbol Wg is the Weingarten function, which here takes the values

$$\mathrm{Wg}[(1)(2), d] = \frac{1}{d^2-1}, \quad \mathrm{Wg}[(1,2), d] = -\frac{1}{d(d^2-1)}, \tag{B4}$$

on the two possible permutations $\tau\sigma^{-1} \in \$_2$.

Then we can let $G_\ell = \sum G_{v^{(\ell)}u^{(\ell)}}|v^{(\ell)}\rangle\langle u^{(\ell)}|$ for each G_ℓ and employ the 2-moment above; let us take the integral over G_1 first,

$$\mathbb{E}\left[\mathfrak{C}_m^\dagger\right] = \mathbb{1}_S \otimes \sum_{u,v=1}^{d_S} \sum_{\sigma,\tau\in\mathbb{S}_2} \mathrm{Wg}(\tau\sigma^{-1},d_S)|v_1'^{(1)}u_1'^{(1)}\rangle\langle v_{\tau(1)}'^{(1)}u_{\sigma(1)}'^{(1)}|$$

$$\bigotimes_{\ell=2}^m \sum_{i_\ell,j_\ell} \int_{\mathbb{U}(d)} \left(|i_\ell\rangle\langle j_\ell| \otimes G_\ell^\dagger|i_\ell\rangle\langle j_\ell|G_\ell\right) \otimes |v_{\tau(2)}'^{(1)}\rangle\langle v_2'^{(1)}| \otimes G_m\cdots G_2|u_{\sigma(2)}'^{(1)}\rangle\langle u_2'^{(1)}|G_2^\dagger\cdots G_m^\dagger \, d\mu(G_2)\cdots d\mu(G_m), \quad \text{(B5)}$$

then we can do similarly with all remaining unitaries by also labeling each permutation mapping σ_ℓ and τ_ℓ for the corresponding integral over each G_ℓ, i.e.,

$$\mathbb{E}\left[\mathfrak{C}_m^\dagger\right] = \mathbb{1}_S \otimes \sum \mathrm{Wg}_1\cdots\mathrm{Wg}_m|v_1^{(1)}u_1^{(1)}\cdots v_1^{(m)}u_1^{(m)}u_{\sigma_m(2)}^{(m)}v_{\tau_1(2)}^{(1)}\rangle\langle v_{\tau_1(1)}^{(1)}u_{\sigma_1(1)}^{(1)}\cdots v_{\tau_m(1)}^{(m)}u_{\sigma_m(1)}^{(m)}u_2^{(m)}v_2^{(1)}|$$

$$\delta(u_2^{(1)},v_2^{(2)})\cdots\delta(u_2^{(m-1)},v_2^{(m)})\,\delta(u_{\sigma_1(2)}^{(1)},v_{\tau_2(2)}^{(2)})\cdots\delta(u_{\sigma_{m-1}(2)}^{(m-1)},v_{\tau_m(2)}^{(m)}), \quad \text{(B6)}$$

where for easiness of notation we dropped the primes and denoted $\mathrm{Wg}_\ell := \mathrm{Wg}(\tau_\ell\sigma_\ell^{-1},d_S)$, where a sum is implicit over basis vectors and permutations on $\mathbb{S}_2$, and where $\delta(a,b)$ stands for the usual Kronecker δ_{ab}. We finally notice that $\mathbb{E}\left[\mathfrak{C}_m^\dagger\right] = \mathbb{E}\left[\mathfrak{C}_m^T\right]$.

APPENDIX C: AVERAGE SEQUENCE FIDELITY

The Markovian case. As a first case let us verify that the average gate sequence given by Eq. (B6) reproduces an ASF described by a decaying exponential in the number of gates when the noise is Markovian.

Consider first a single gate, $m = 1$. We have

$$\mathbb{E}\left[\mathfrak{C}_m^T\right] = \mathbb{1}_S \otimes \sum \mathrm{Wg}|v_1u_1u_{\sigma(2)}v_{\tau(2)}\rangle\langle v_{\tau(1)}u_{\sigma(1)}u_2v_2|, \quad \text{(C1)}$$

where Wg implicitly depends on $\tau\sigma^{-1}$, with each τ and σ being summed over the symmetric group $\mathbb{S}_2$, and with an implicit sum over each u and v.

For the Markovian process tensor, we consider noise described by some d_S-dimensional CP map $\Lambda_n^{(M)}$ at time step n with Kraus representation $\Lambda_n^{(M)}(\cdot) = \sum_i \lambda_{i_n}^{(M)}(\cdot)\lambda_{i_n}^{(M)}$, so that

$$\Upsilon_1^{(M)} = (\Lambda_2^{(M)} \otimes \mathcal{I}) \circ \mathscr{S}_2 \circ (\Lambda_1^{(M)} \otimes \mathcal{I}) \circ \mathscr{S}_1\left(\rho \otimes \psi^{\otimes 2}\right)$$

$$= \sum (\lambda_{i_2}^{(M)} \otimes \mathbb{1})\varsigma_2(\lambda_{i_1}^{(M)} \otimes \mathbb{1})\varsigma_1(\rho \otimes \psi^{\otimes 2})\varsigma_1^\dagger(\lambda_{i_1}^{(M)\,\dagger} \otimes \mathbb{1})\varsigma_2^\dagger(\lambda_{i_2}^{(M)\,\dagger} \otimes \mathbb{1})$$

$$= \sum \lambda_{i_2}^{(M)}|\alpha_2\rangle\langle\beta_2|\lambda_{i_1}^{(M)}|\alpha_1\rangle\langle\beta_1|\rho|\delta_1\rangle\langle\gamma_1|\lambda_{i_1}^{(M)\,\dagger}|\delta_2\rangle\langle\gamma_2|\lambda_{i_2}^{(M)\,\dagger} \otimes |\beta_1\alpha_1\beta_2\alpha_2\rangle\langle\delta_1\gamma_1\delta_2\gamma_2|. \quad \text{(C2)}$$

Let us simply denote d_S as d, as there is no environment to care about. Then we obtain the average sequence

$$\mathbb{E}[\mathcal{S}_1(\rho)] = \mathrm{tr}_{\overline{S}}\left\{\Upsilon_1^{(M)}\,\mathbb{E}\left[\mathfrak{C}_m^T\right]\right\}$$

$$= \Lambda_2^{(M)}\left[\sum \mathrm{Wg}|v_2\rangle\langle u_2|\lambda_{i_1}^{(M)}|u_{\sigma(1)}\rangle\langle v_{\tau(1)}|\rho|v_1\rangle\langle u_1|\lambda_{i_1}^{(M)\,\dagger}|u_{\sigma(2)}\rangle\langle v_{\tau(2)}|\right]$$

$$= \Lambda_2\left[\frac{1}{d^2-1}\left(\underbrace{\mathrm{tr}\left(\sum_i \lambda_{i_1}^{\dagger(M)}\lambda_{i_1}^{(M)}\right)\mathbb{1}}_{\sigma=\tau=\mathbb{1}} + \sum_i \underbrace{\mathrm{tr}(\lambda_{i_1}^{(M)})\,\mathrm{tr}(\lambda_{i_1}^{(M)\,\dagger})\,\rho}_{\sigma=\tau=(1,2)}\right)\right.$$

$$\left. - \frac{1}{d(d^2-1)}\left(\underbrace{\mathrm{tr}\left(\sum_i \lambda_{i_1}^{\dagger(M)}\lambda_{i_1}^{(M)}\right)\rho}_{\sigma=\mathbb{1},\,\tau=(1,2)} + \sum_i \underbrace{\mathrm{tr}(\lambda_{i_1}^{(M)})\,\mathrm{tr}(\lambda_{i_1}^{(M)\,\dagger})\,\mathbb{1}}_{\sigma(1,2),\,\tau=(1)(2)}\right)\right]$$

$$= \Lambda_2 \left[\frac{d \operatorname{tr}\left(\sum_i \lambda_{i_1}^{\dagger(M)} \lambda_{i_1}^{(M)}\right) - \sum_i |\operatorname{tr}(\lambda_{i_1}^{(M)})|^2}{d^2-1} \left(\frac{\mathbb{1}}{d}\right) + \frac{d \sum_i |\operatorname{tr}(\lambda_{i_1}^{(M)})|^2 - \operatorname{tr}\left(\sum_i \lambda_{i_1}^{\dagger(M)} \lambda_{i_1}^{(M)}\right)}{d(d^2-1)} \rho \right]. \tag{C3}$$

Now, if the noise is trace-preserving as well, we have $\operatorname{tr}\left(\sum_i \lambda_{i_1}^{\dagger(M)} \lambda_{i_1}^{(M)}\right) = \operatorname{tr}(\mathbb{1}) = d$. Then we get

$$\mathbb{E}[\mathcal{S}_1(\rho)] = \Lambda_2^{(M)} \circ \mathscr{D}_p(\rho), \tag{C4}$$

where we define $\mathscr{D}_p(X) := pX + (1-p)\mathbb{1}/d$ as a depolarizing map with the so-called noise strength,

$$p := \frac{\sum_i |\operatorname{tr}(\lambda_{i_1}^{(M)})|^2 - 1}{d^2-1} \in [0,1], \tag{C5}$$

which has to be constrained to $[0,1]$. If we denote the noise map $\Lambda_n^{(M)} = \sum_i \lambda_{i_n}^{(M)} \otimes \lambda_{i_n}^{(M)\dagger}$, with the Kraus operators acting on the respective system S Hilbert space and conjugate space, respectively, we can simply write

$$p = \frac{\operatorname{tr}\left[\Lambda_1^{(M)}\right] - 1}{d^2-1}. \tag{C6}$$

The noise strength can be shown to be related to the gate fidelity of Λ_1 with respect to the identity [28], i.e., $f_{\Lambda_1,\mathcal{I}} = \int d\psi \langle\psi|\Lambda_1^{(M)}(|\psi\rangle\langle\psi|)|\psi\rangle$, as $p = df_{\Lambda_1,\mathcal{I}} - 1/d - 1$. This is the relevant parameter, which in practice can be recovered by running several sequences for different lengths and averaging the resulting probabilities.

To generalize to an arbitrary number of time steps, we now use the fact that the action of the depolarizing channel can be written as

$$\mathscr{D}_p(X) = \sum \operatorname{Wg}|v_2\rangle\langle u_2|\lambda_i^{(M)}|u_{\sigma(1)}\rangle\langle v_{\tau(1)}|X|v_1\rangle$$
$$\times \langle u_1|\lambda_i^{(M)\dagger}|u_{\sigma(2)}\rangle\langle v_{\tau(2)}|, \tag{C7}$$

for any X, which follows from Eq. (C3). This then implies that for an arbitrary sequence length,

$$\mathbb{E}[\mathcal{S}_m(\rho)] = \operatorname{tr}_{\bar{S}}\left\{\Upsilon_m^{(M)}\mathbb{E}\left[\mathfrak{C}_m^T\right]\right\}$$
$$= \Lambda_{m+1} \circ \mathscr{D}_{p_m} \circ \cdots \circ \mathscr{D}_{p_1}(\rho)$$
$$= p_1 p_2 \cdots p_m \Lambda_{m+1}^{(M)}\left(\rho - \frac{\mathbb{1}}{d}\right) + \Lambda_{m+1}^{(M)}\left(\frac{\mathbb{1}}{d}\right), \tag{C8}$$

where here now $p_n := \operatorname{tr}[\Lambda_n^{(M)}] - 1/d_S^2 - 1$, as expected. The case $p_1 \neq p_2 \neq \cdots \neq p_m$ corresponds to the Markovian time-dependent noise case as in Ref. [9]. When the noise strengths are the same this gives the usual fitting model for the average probabilities

$$\mathcal{F}_m = p^m \operatorname{tr}\left[\mathcal{M}\Lambda_{m+1}^{(M)}\left(\rho - \frac{\mathbb{1}}{d}\right)\right]$$
$$+ \operatorname{tr}\left[\mathcal{M}\Lambda_{m+1}^{(M)}\left(\frac{\mathbb{1}}{d}\right)\right] := Ap^m + B, \tag{C9}$$

with $A := \operatorname{tr}\left[\mathcal{M}\Lambda^{(M)}(\rho - \mathbb{1}/d)\right]$ and $B := \operatorname{tr}\left[\mathcal{M}\Lambda^{(M)}(\mathbb{1}/d)\right]$, which relate to state preparation and measurement errors.

General non-Markovian gate-independent noise. We now consider the general situation where the noise is correlated across each step through an external environment as depicted in Fig. 1. Let us take first the simplest case $m = 1$; the process tensor for the noise sequence is

$$\Upsilon_1 = \operatorname{tr}_E\left[(\Lambda_2 \otimes \mathcal{I}) \circ \mathscr{S}_2 \circ (\Lambda_1 \otimes \mathcal{I}) \circ \mathscr{S}_1\left(\rho \otimes \psi^{\otimes 2}\right)\right]$$
$$= \operatorname{tr}_E\left[(\lambda_2 \otimes \mathbb{1})\varsigma_2(\lambda_1 \otimes \mathbb{1})\varsigma_1(\rho \otimes \psi^{\otimes 2})\varsigma_1^\dagger(\lambda_1^\dagger \otimes \mathbb{1})\varsigma_2^\dagger(\lambda_2^\dagger \otimes \mathbb{1})\right]$$
$$= \operatorname{tr}_E\left[\lambda_2 \zeta_{\alpha_2\beta_2}\lambda_1 \zeta_{\alpha_1\beta_1}\rho \zeta_{\delta_1\gamma_1}\lambda_1^\dagger \zeta_{\delta_2\gamma_2}\lambda_2^\dagger\right] \otimes |\beta_1\alpha_1\beta_2\alpha_2\rangle\langle\delta_1\gamma_1\delta_2\gamma_2|, \tag{C10}$$

where here $\zeta_{ab} := \mathbb{1}_E \otimes |a\rangle\langle b|$, hence we get

$$\mathbb{E}[\mathcal{S}_1(\rho)] = \operatorname{tr}_{\bar{S}}\left\{\Upsilon_1 \mathbb{E}\left(\mathfrak{C}_1^T\right)\right\}$$
$$= \operatorname{tr}_E\left\{\Lambda_2\left[\sum \operatorname{Wg} \zeta_{v_2 u_2}\lambda_{i_1}\zeta_{u_{\sigma(1)}v_{\tau(1)}}\rho \zeta_{v_1 u_1}\lambda_{i_1}^\dagger \zeta_{u_{\sigma(2)}v_{\tau(2)}}\right]\right\}. \tag{C11}$$

Let us write the Kraus operators of the nth noise map, Λ_n, as

$$\lambda_{i_n} := \sum_{e_n^{(\prime)}=1}^{d_E} \sum_{s_n^{(\prime)}=1}^{d_S} L_{i_n}^{s_n s_n' \; e_n e_n'} |e_n s_n\rangle\langle e_n' s_n'|, \tag{C12}$$

where the e and s indices refer to systems E and S, respectively; the subindex n is simply a label for the nth Kraus operator. Then

$$\mathbb{E}\left[\mathcal{S}_1(\rho)\right] = \sum L_{i_1}^{s_1 s_1' \; e_1 e_1'} \chi^{ee'}_{ss'} \, L_{i_1}^{*\; \zeta_1 \zeta_1' \; \epsilon_1 \epsilon_1'} \mathrm{Wg} \, \mathrm{tr}_E \left\{ \Lambda_2 \left[\left(|e_1\rangle\langle e_1'| \otimes |v_2\rangle\langle u_2| s_1\rangle\langle s_1'|u_{\sigma(1)}\rangle\langle v_{\tau(1)}| \right) \rho \left(|\epsilon_1'\rangle\langle\epsilon_1| \otimes |v_1\rangle\langle u_1|\zeta_1'\rangle\langle \zeta_1|u_{\sigma(2)}\rangle\langle v_{\tau(2)}| \right) \right] \right\}, \tag{C13}$$

and let us now similarly write the initial state as

$$\rho = \sum_{e,e'=1}^{d_E} \sum_{s,s'=1}^{d_S} \chi^{ee'}_{ss'} |es\rangle\langle e's'|, \qquad \text{where} \qquad \sum \chi^{ee}_{ss} = 1, \tag{C14}$$

then also

$$\mathbb{E}\left[\mathcal{S}_1(\rho)\right]$$

$$= \sum L_{i_1}^{s_1 s_1' \; e_1 e_1'} \chi^{ee'}_{ss'} \, L_{i_1}^{*\; \zeta_1 \zeta_1' \; \epsilon_1 \epsilon_1'} \, \mathrm{tr}_E \left\{ \Lambda_2 \left[|e_1\rangle\langle e_1'| e\rangle\langle e'|\epsilon_1'\rangle\langle\epsilon_1| \otimes \sum \mathrm{Wg}|v_2\rangle\langle u_2| s_1\rangle\langle s_1'|u_{\sigma(1)}\rangle\langle v_{\tau(1)}|s\rangle\langle s'|v_1\rangle\langle u_1|\zeta_1'\rangle\langle \zeta_1|u_{\sigma(2)}\rangle\langle v_{\tau(2)}| \right] \right\}$$

$$:= \sum L_{i_1}^{s_1 s_1' \; e_1 e} \chi^{ee'}_{ss'} \, L_{i_1}^{*\; \zeta_1 \zeta_1' \; \epsilon_1 e'} \, \mathrm{tr}_E \left\{ \Lambda_2 \left[|e_1\rangle\langle\epsilon_1| \otimes \Phi_{s_1\zeta_1}(|s\rangle\langle s'|) \right] \right\}$$

$$= \sum L_{i_2}^{s_2 s_2' \; e_2 e_2'} L_{i_1}^{s_1 s_1' \; e_1 e} \chi^{ee'}_{ss'} \, L_{i_1}^{*\; \zeta_1 \zeta_1' \; \epsilon_1 e'} L_{i_2}^{*\; \zeta_2 \zeta_2' \; \epsilon_2 \epsilon_2'} \, \mathrm{tr}\left[|e_2\rangle\langle e_2'|e_1\rangle\langle\epsilon_1|\epsilon_2'\rangle\langle\epsilon_2| \right] \, |s_2\rangle\langle s_2'|\Phi_{s_1\zeta_1}(|s\rangle\langle s'|)|\zeta_2'\rangle\langle\zeta_2|$$

$$= \sum L_{i_2}^{s_2 s_2' \; e_2 e_1} L_{i_1}^{s_1 s_1' \; e_1 e} \chi^{ee'}_{ss'} \, L_{i_1}^{*\; \zeta_1 \zeta_1' \; \epsilon_1 e'} L_{i_2}^{*\; \zeta_2 \zeta_2' \; e_2 \epsilon_1} \, \langle s_2'|\Phi_{s_1\zeta_1}(|s\rangle\langle s'|)|\zeta_2'\rangle \, |s_2\rangle\langle\zeta_2|, \tag{C15}$$

where the second line follows by Eq. (C7) and by defining

$$\Phi_{s_n\zeta_n}(X) := \frac{d_S\, \delta_{s_n s_n'}\delta_{\zeta_n\zeta_n'} - \delta_{s_n\zeta_n}\delta_{s_n'\zeta_n'}}{d_S(d_S^2-1)} X + \frac{d_S\, \delta_{s_n\zeta_n}\delta_{s_n'\zeta_n'} - \delta_{s_n s_n'}\delta_{\zeta_n\zeta_n'}}{d_S^2-1} \left(\frac{\mathbb{1}_S}{d_S}\right). \tag{C16}$$

Now let

$$\alpha_{s_n^{(\prime)}\zeta_n^{(\prime)}} := \frac{d_S\delta_{s_n s_n'}\delta_{\zeta_n\zeta_n'} - \delta_{s_n\zeta_n}\delta_{s_n'\zeta_n'}}{d_S(d_S^2-1)}, \qquad \beta_{s_n^{(\prime)}\zeta_n^{(\prime)}} := \frac{\delta_{s_n\zeta_n}\delta_{s_n'\zeta_n'}}{d_S} - \alpha_{s_n\zeta_n} \tag{C17}$$

so that

$$\Phi_{s_n\zeta_n}(X) = \alpha_{s_n^{(\prime)}\zeta_n^{(\prime)}} X + \beta_{s_n^{(\prime)}\zeta_n^{(\prime)}}\left(\frac{\mathbb{1}_S}{d_S}\right) = \alpha_{s_n^{(\prime)}\zeta_n^{(\prime)}}\left(X - \frac{\mathbb{1}_S}{d_S}\right) + \frac{\delta_{s_n\zeta_n}\delta_{s_n'\zeta_n'}}{d_S}\left(\frac{\mathbb{1}_S}{d_S}\right). \tag{C18}$$

Now, we can also define

$$\mathcal{E}^{(2)}_{\mathbf{s}^{(\prime)}\zeta^{(\prime)}} := \sum_{i=1}^{d_\lambda} \sum_{e=1}^{d_E} L_{i_2}^{s_2 s_2' \; e_4 e_2} L_{i_1}^{s_1 s_1' \; e_2 e_0} \chi^{e_0 e_1}_{ss'} \, L_{i_1}^{*\; \zeta_1 \zeta_1' \; e_3 e_1} L_{i_2}^{*\; \zeta_2 \zeta_2' \; e_4 e_3}$$

$$= \sum \langle e_4 s_2|\lambda_{i_2}|e_2 s_2'\rangle\langle e_2 s_1|\lambda_{i_1}|e_0 s_1'\rangle\langle e_0 s|\rho|e_1 s'\rangle\langle e_1 \zeta_1'|\lambda_{i_1}^\dagger|e_3 \zeta_1\rangle\langle e_3 \zeta_2'|\lambda_{i_2}^\dagger|e_4 \zeta_2'\rangle$$

$$= \mathrm{tr}\left[(\mathbb{1}_E \otimes \langle s_2|)\lambda_{i_2}(\mathbb{1}_E \otimes |s_2'\rangle\langle s_1|)\lambda_{i_1}(\mathbb{1}_E \otimes |s_1'\rangle\langle s|)\rho(\mathbb{1}_E \otimes |s'\rangle\langle\zeta_1'|)\lambda_{i_1}^\dagger(\mathbb{1}_E \otimes |\zeta_1\rangle\langle\zeta_2'|)\lambda_{i_2}^\dagger(\mathbb{1}_E \otimes |\zeta_2'\rangle) \right], \tag{C19}$$

where summation is over all $i_1, i_2, \dots$ and $e_0, e_1, \dots$, and which contains all information about the noise within the whole SE and the correlations in between the two. We can simply write this as $\mathcal{E}^{(2)}_{s^{(\prime)}\zeta^{(\prime)}} =$ $\mathrm{tr}\left[\langle s_2|\lambda_{i_2}|s_2'\rangle\langle s_1|\lambda_{i_1}|s_1'\rangle\langle s|\rho|s'\rangle\langle\zeta_1'|\lambda_{i_1}^\dagger|\zeta_1\rangle\langle\zeta_2'|\lambda_{i_2}^\dagger|\zeta_2\rangle\right]$ as in the main text, where an identity on E is implicit. With this we can write Eq. (C15) as

$$\mathbb{E}\left[\mathcal{S}_1(\rho)\right] = \sum \mathcal{E}^{(2)}_{s^{(\prime)}\zeta^{(\prime)}}\left[\alpha_{s_1\zeta_1}\left(\delta_{ss_2'}\delta_{s'\zeta_2'} - \frac{\delta_{ss'}\delta_{s_2'\zeta_2'}}{d_S}\right) + \frac{\delta_{s_1\zeta_1}\delta_{s_1'\zeta_1'}}{d_S}\left(\frac{\delta_{ss'}\delta_{s_2'\zeta_2'}}{d_S}\right)\right]|s_2\rangle\langle\zeta_2|. \tag{C20}$$

In general, for an arbitrary sequence length, we have

$$\mathcal{E}^{(m)}_{s^{(\prime)}\zeta^{(\prime)}} := \sum_{\boldsymbol{e},\boldsymbol{\epsilon}} \chi^{ee'}_{ss'}\, \mathrm{tr}\left[\left(\prod_{\eta=m}^{1} \mathrm{L}^{e_\eta e_\eta'}_{i_\eta\, s_\eta s_\eta'}|e_\eta\rangle\langle e_\eta'|\right)|e\rangle\langle e'|\left(\prod_{n=1}^{m} \mathrm{L}^{*\,\epsilon_n\epsilon_n'}_{i_n\, \zeta_n\zeta_n'}|\epsilon_n'\rangle\langle\epsilon_n|\right)\right]$$

$$= \sum_{i=1}^{d_\lambda} \mathrm{tr}\left[\left(\prod_{\eta=m}^{1}\langle s_\eta|\lambda_{i_\eta}|s_\eta'\rangle\right)\langle s|\,\rho\,|s'\rangle\left(\prod_{n=1}^{m}\langle\zeta_n'|\lambda_{i_n}^\dagger|\zeta_n\rangle\right)\right], \tag{C21}$$

so that,

$$\mathbb{E}\left[\mathcal{S}_m(\rho)\right] = \sum \mathcal{E}^{(m+1)}_{s^{(\prime)}\zeta^{(\prime)}}\langle s_{m+1}'|\Phi_{s_m\zeta_m}\circ\cdots\circ\Phi_{s_1\zeta_1}(|s\rangle\langle s'|)|\zeta_{m+1}'\rangle\,|s_{m+1}\rangle\langle\zeta_{m+1}|. \tag{C22}$$

The sequential application of Φ maps is given by

$$\Phi_{s_m\zeta_m}\circ\cdots\circ\Phi_{s_1\zeta_1}(X) := \boldsymbol{\alpha}^{(m)}_{\boldsymbol{s},\boldsymbol{\zeta}}\left(X - \frac{\mathbb{1}}{d_S}\right) + \boldsymbol{\Delta}^{(m)}_{\boldsymbol{s},\boldsymbol{\zeta}}\,\frac{\mathbb{1}}{d_S}, \tag{C23}$$

where

$$\boldsymbol{\alpha}^{(m)}_{\boldsymbol{s},\boldsymbol{\zeta}} := \prod_{n=1}^{m} \alpha_{s_n^{(\prime)}\zeta_n^{(\prime)}}, \tag{C24}$$

and $\boldsymbol{\Delta}^{(m)}_{\boldsymbol{s},\boldsymbol{\zeta}}$ is a sum of all m-term product combinations of $\alpha_{s_1^{(\prime)}\zeta_1^{(\prime)}},\dots,\alpha_{s_m^{(\prime)}\zeta_m^{(\prime)}}$ and $\beta_{s_1^{(\prime)}\zeta_1^{(\prime)}},\dots,\beta_{s_m^{(\prime)}\zeta_m^{(\prime)}}$, that is,

$$\boldsymbol{\Delta}^{(1)}_{\boldsymbol{s},\boldsymbol{\zeta}} \sim \alpha_1 + \beta_1, \tag{C25}$$

$$\boldsymbol{\Delta}^{(2)}_{\boldsymbol{s},\boldsymbol{\zeta}} \sim \alpha_1\alpha_2 + \alpha_1\beta_2 + \beta_1\alpha_2 + \beta_1\beta_2, \tag{C26}$$

$$\boldsymbol{\Delta}^{(3)}_{\boldsymbol{s},\boldsymbol{\zeta}} \sim \alpha_1\alpha_2\alpha_3 + \alpha_1\alpha_2\beta_3 + \alpha_1\beta_2\alpha_3 + \alpha_1\beta_2\beta_3 + \beta_1\alpha_2\alpha_3 + \beta_1\alpha_2\beta_3 + \beta_1\beta_2\alpha_3 + \beta_1\beta_2\beta_3,$$

$$\vdots \tag{C27}$$

where $\alpha_i = \alpha_{s_i^{(\prime)}\zeta_i^{(\prime)}}$ and similarly for β_i; in general, there are 2^m of these summands on $\boldsymbol{\Delta}^{(m)}_{\boldsymbol{s},\boldsymbol{\zeta}}$. However, notice that as $\beta_i \sim 1/d_S\delta_{s_i\zeta_i}\delta_{s_i'\zeta_i'} - \alpha_i$, every term simplifies to products of deltas, i.e.,

$$\boldsymbol{\Delta}^{(1)}_{\boldsymbol{s},\boldsymbol{\zeta}} \sim \frac{\delta_{s_1\zeta_1}\delta_{s_1'\zeta_1'}}{d_S}, \quad \boldsymbol{\Delta}^{(2)}_{\boldsymbol{s},\boldsymbol{\zeta}} \sim \frac{\delta_{s_1\zeta_1}\delta_{s_1'\zeta_1'}\delta_{s_2\zeta_2}\delta_{s_2'\zeta_2'}}{d_S^2}, \quad \cdots \quad, \boldsymbol{\Delta}^{(m)}_{\boldsymbol{s},\boldsymbol{\zeta}} \sim \frac{\prod_{i=1}^{m}\delta_{s_i\zeta_i}\delta_{s_i'\zeta_i'}}{d_S^m}. \tag{C28}$$

Thus we can rewrite Eq. (C22) as

$$\mathbb{E}\left[\mathcal{S}_m(\rho)\right] = \sum \mathcal{E}^{(m+1)}_{s^{(\prime)}\zeta^{(\prime)}}\left[\boldsymbol{\alpha}^{(m)}_{\boldsymbol{s},\boldsymbol{\zeta}}\left(\delta_{ss_{m+1}'}\delta_{s'\zeta_{m+1}'} - \frac{\delta_{ss'}\delta_{s_{m+1}'\zeta_{m+1}'}}{d_S}\right) + \boldsymbol{\Delta}^{(m)}_{\boldsymbol{s},\boldsymbol{\zeta}}\left(\frac{\delta_{ss'}\delta_{s_{m+1}'\zeta_{m+1}'}}{d_S}\right)\right]|s_{m+1}\rangle\langle\zeta_{m+1}|, \tag{C29}$$

and so for a measurement $\mathcal{M}$, on average,

$$\mathcal{F}_m = \sum_{s,\zeta,s',\zeta'=1}^{d_S} \mathcal{E}_{\boldsymbol{s}^{()}\boldsymbol{\zeta}^{()}}^{(m+1)} \left(\mathcal{A}_{\boldsymbol{s}^{()}\boldsymbol{\zeta}^{()}}^{(m+1)} + \mathcal{B}_{\boldsymbol{s}^{()}\boldsymbol{\zeta}^{()}}^{(m+1)} \right), \tag{C30}$$

where

$$\mathcal{A}_{\boldsymbol{s}^{()}\boldsymbol{\zeta}^{()}}^{(m+1)} := \boldsymbol{\alpha}_{s,\zeta}^{(m)} \left(\delta_{s'_{m+1}s}\delta_{s'\zeta'_{m+1}} - \frac{\delta_{ss'}\delta_{s'_{m+1}\zeta'_{m+1}}}{d_S} \right) \langle \zeta_{m+1}|\mathcal{M}|s_{m+1}\rangle, \quad \mathcal{B}_{\boldsymbol{s}^{()}\boldsymbol{\zeta}^{()}}^{(m+1)} := \boldsymbol{\Delta}_{s,\zeta}^{(m)} \left(\frac{\delta_{ss'}\delta_{s'_{m+1}\zeta'_{m+1}}}{d_S} \right) \langle \zeta_{m+1}|\mathcal{M}|s_{m+1}\rangle. \tag{C31}$$

This expression contains $4d_S(m+1)$ terms, and could potentially be useful whenever the underlying noise model is not known, as all this information will be contained solely on the factors $\mathcal{E}_{\boldsymbol{s}^{()}\boldsymbol{\zeta}^{()}}^{(m+1)}$.

We can, however, write this expression in a more succinct way. We have

$$\sum_{s,\zeta,s',\zeta'=1}^{d_S} \mathcal{E}_{\boldsymbol{s}^{()}\boldsymbol{\zeta}^{()}}^{(m+1)} \mathcal{A}_{\boldsymbol{s}^{()}\boldsymbol{\zeta}^{()}}^{(m+1)} = \sum \mathrm{tr}\left[\left(\prod_{\eta=m+1}^{1} \langle s_\eta|\lambda_{i_\eta}|s'_\eta\rangle \right) \langle s| \rho |s'\rangle \left(\prod_{n=1}^{m+1} \langle \zeta'_n|\lambda_{i_n}^\dagger|\zeta_n\rangle \right) \right] \left(\frac{\prod_{N=1}^{m} \left(d_S\delta_{s_N s'_N}\delta_{\zeta_N \zeta'_N} - \delta_{s_N \zeta_N}\delta_{s'_N \zeta'_N} \right)}{d_S^{\,m}(d_S^{\,2}-1)^m} \right)$$

$$\left(\delta_{s'_{m+1}s}\delta_{s'\zeta'_{m+1}} - \frac{\delta_{ss'}\delta_{s'_{m+1}\zeta'_{m+1}}}{d_S} \right)$$

$$\langle \zeta_{m+1}|\mathcal{M}|s_{m+1}\rangle, \tag{C32}$$

so now let us define the following. Let

$$\$_{\Lambda_n}(\varepsilon) := \sum_{i=1}^{d_\lambda} \mathrm{tr}_S(\lambda_{i_n})\, \varepsilon\, \mathrm{tr}_S(\lambda_{i_n}^\dagger), \tag{C33}$$

$$\Theta_{\Lambda_n}(\varepsilon) := \mathrm{tr}_S\left[\Lambda_n \left(\varepsilon \otimes \frac{\mathbb{1}}{d_S} \right) \right], \tag{C34}$$

for any operator ε acting on E. Then we notice that

$$\frac{1}{d_S^{\,m}} \sum \left(\prod_{\eta=m}^{1} \langle s_\eta|\lambda_{i_\eta}|s'_\eta\rangle \right) \langle s|\rho|s'\rangle \left(\prod_{n=1}^{m} \langle \zeta'_n|\lambda_{i_n}^\dagger|\zeta_n\rangle \right) \prod_{N=1}^{m} \left(d_S\delta_{s_N s'_N}\delta_{\zeta_N \zeta'_N} - \delta_{s_N \zeta_N}\delta_{s'_N \zeta'_N} \right)$$

$$= \frac{1}{d_S^{\,m}} \sum \left(\prod_{\eta=m}^{2} \langle s_\eta|\lambda_{i_\eta}|s'_\eta\rangle \right) \left(d_S\, \mathrm{tr}_S(\lambda_{i_1})\langle s|\rho|s'\rangle\, \mathrm{tr}_S(\lambda_{i_1}^\dagger) - \mathrm{tr}_S[\Lambda_1(\langle s|\rho|s'\rangle \otimes \mathbb{1})] \right) \left(\prod_{n=2}^{m} \langle \zeta'_n|\lambda_{i_n}^\dagger|\zeta_n\rangle \right)$$

$$\prod_{N=2}^{m} \left(d_S\delta_{s_N s'_N}\delta_{\zeta_N \zeta'_N} - \delta_{s_N \zeta_N}\delta_{s'_N \zeta'_N} \right)$$

$$= \frac{1}{d_S^{\,m-1}} \sum \left(\prod_{\eta=m}^{2} \langle s_\eta|\lambda_{i_\eta}|s'_\eta\rangle \right) (\$_{\Lambda_1} - \Theta_{\Lambda_1}) \langle s|\rho|s'\rangle \left(\prod_{n=2}^{m} \langle \zeta'_n|\lambda_{i_n}^\dagger|\zeta_n\rangle \right) \prod_{N=2}^{m} \left(d_S\delta_{s_N s'_N}\delta_{\zeta_N \zeta'_N} - \delta_{s_N \zeta_N}\delta_{s'_N \zeta'_N} \right)$$

$$\vdots$$

$$= \left[\bigcirc_{n=1}^{m} (\$_{\Lambda_n} - \Theta_{\Lambda_n}) \right] \langle s|\rho|s'\rangle, \tag{C35}$$

where as before there are implicit identities, which should be clear by context, i.e., for example, $\mathrm{tr}_S(\lambda_i)$ means $\mathrm{tr}_S(\lambda_i) \otimes \mathbb{1}_S$. Then this means that

$$\sum_{s,\zeta,s',\zeta'=1}^{d_S} \mathcal{E}_{s^{()}\zeta^{()}}^{(m+1)} \mathcal{A}_{s^{()}\zeta^{()}}^{(m+1)} = \frac{\mathrm{tr}\left\{ \mathcal{M} \; \mathrm{tr}_E \circ \Lambda_{m+1}\left[\overset{m}{\underset{n=1}{\bigcirc}} \left(\$_{\Lambda_n} - \Theta_{\Lambda_n}\right) \otimes \mathcal{I}_S \right] \left(\rho - \rho_E \otimes \frac{\mathbb{1}}{d_S}\right)\right\}}{\left(d_S^2 - 1\right)^m}, \tag{C36}$$

where $\rho_E := \mathrm{tr}_S(\rho)$. Now for the second term, similarly (again we omit implicit identity operators),

$$\sum_{s,\zeta,s',\zeta'=1}^{d_S} \mathcal{E}_{s^{()}\zeta^{()}}^{(m+1)} \mathcal{B}_{s^{()}\zeta^{()}}^{(m+1)}$$

$$= \frac{1}{d_S^{m+1}} \sum_{s,s'=1}^{d_S} \sum_{i=1}^{d_\lambda} \mathrm{tr}\left[\langle s_{m+1}|\lambda_{i_{m+1}}|s'_{m+1}\rangle \left(\prod_{\eta=m}^{1} \langle s_\eta|\lambda_{i_\eta}|s'_\eta\rangle\right) \langle s| \rho |s\rangle \left(\prod_{n=1}^{m} \langle s'_n|\lambda_{i_n}^\dagger|s_n\rangle\right) \langle s'_{m+1}|\lambda_{i_{m+1}}^\dagger|\zeta_{m+1}\rangle \right]$$

$$\langle \zeta_{m+1}|\mathcal{M}|s_{m+1}\rangle = \mathrm{tr}\left[(\mathbb{1}_E \otimes \mathcal{M}) \circ \Lambda_{m+1} \circ \left(\overset{m}{\underset{n=1}{\bigcirc}} \Theta_{\Lambda_n} \otimes \mathcal{I}_S\right) \left(\rho_E \otimes \frac{\mathbb{1}}{d_S}\right)\right]. \tag{C37}$$

Thus we can finally write

$$\mathcal{F}_m = \mathrm{tr}\left[\mathcal{M} \; \mathrm{tr}_E \circ \Lambda_{m+1} \circ (\mathscr{A}_m + \mathscr{B}_m)\,\rho\right], \tag{C38}$$

where

$$\mathscr{A}_m(\rho) := \frac{\overset{m}{\underset{n=1}{\bigcirc}} \left(\$_{\Lambda_n} - \Theta_{\Lambda_n}\right) \otimes \mathcal{I}_S}{\left(d_S^2 - 1\right)^m} \left(\rho - \rho_E \otimes \frac{\mathbb{1}}{d_S}\right), \tag{C39}$$

$$\mathscr{B}_m(\rho) := \overset{m}{\underset{n=1}{\bigcirc}} \Theta_{\Lambda_n}(\rho_E) \otimes \frac{\mathbb{1}}{d_S}, \tag{C40}$$

with $\$_{\Lambda_n}$ and Θ_{Λ_n} defined in Eqs. (C33) and (C34), respectively.

APPENDIX D: MARKOVIAN LIMIT

For the Markovian limit we take $\Lambda_n \to \mathcal{I}_E \otimes \Lambda_n^{(M)}$ and $\rho = \rho_E \otimes \rho_S$. First, let us notice that, assuming $\Lambda_n^{(M)}$ are CPTP,

$$\$_{\Lambda_n^{(M)}}(\varepsilon) = \mathrm{tr}\left[\Lambda_n^{(M)}\right] \varepsilon, \qquad \Theta_{\Lambda_n^{(M)}}(\varepsilon) = \mathrm{tr}\left[\Lambda_n^{(M)}\left(\frac{\mathbb{1}}{d_S}\right)\right] \varepsilon = \varepsilon, \tag{D1}$$

for any operators ε acting on E and σ on S. Then this implies that

$$\mathrm{tr}_E \circ \mathscr{A}_m(\rho_E \otimes \rho_S) \to \mathrm{tr}_E \circ \frac{\overset{m}{\underset{n=1}{\bigcirc}} \left(\$_{\Lambda_n^{(M)}} - \Theta_{\Lambda_n^{(M)}}\right)}{\left(d_S^2 - 1\right)^m} (\rho_E) \otimes \left(\rho_S - \frac{\mathbb{1}}{d_S}\right)$$

$$= \frac{\mathrm{tr}\left[\Lambda_1^{(M)}\right] - 1}{\left(d_S^2 - 1\right)^m} \mathrm{tr}_E \circ \left[\overset{m}{\underset{n=2}{\bigcirc}} \left(\$_{\Lambda_n^{(M)}} - \Theta_{\Lambda_n^{(M)}}\right)\right](\rho_E) \otimes \left(\rho_S - \frac{\mathbb{1}}{d_S}\right)$$

$$\vdots$$

$$= \frac{\prod_{n=1}^{m}\left(\mathrm{tr}\left[\Lambda_n^{(M)}\right] - 1\right)}{\left(d_S^2 - 1\right)^m} \left(\rho_S - \frac{\mathbb{1}}{d_S}\right)$$

$$= p_1 \cdots p_m \left(\rho_S - \frac{\mathbb{1}}{d_S}\right), \tag{D2}$$

where here as well $p_n := \frac{\text{tr}\left[\Lambda_n^{(M)}\right]-1}{d_S{}^2-1}$ is the noise strength of $\Lambda_n^{(M)}$, and

$$\text{tr}_E \circ \mathscr{B}_m(\rho_E \otimes \rho_S) \to \text{tr}_E \overset{m}{\underset{n=1}{\bigcirc}} \Theta_{\Lambda_n^{(M)}}(\rho_E) \otimes \frac{\mathbb{1}}{d_S} = \frac{\mathbb{1}}{d_S},$$

$$\text{(D3)}$$

which implies that $\mathcal{F}_m \to p_1 \cdots p_m \, \text{tr}[\mathcal{M} \; \Lambda_{m+1}(\rho - \mathbb{1}/d_S)] + \text{tr}[\mathcal{M} \; \Lambda_{m+1}(\mathbb{1}/d_S)]$ under Markovian noise.

APPENDIX E: FINITE NON-MARKOVIAN NOISE

Initial non-Markovian noise. Suppose a quantum noise process $\hat{\Upsilon}_m$ is non-Markovian up to some time step $\ell < m$ and almost Markovian in the remaining steps, i.e., $\hat{\Upsilon}_m \simeq \Upsilon_\ell \otimes \Upsilon_{m:\ell+1}^{(M)}$, where $\Upsilon_{m:\ell}^{(M)}$ is a Markov process from time step $\ell + 1$ to time step m. This effectively would mean that E is traced at the ℓth step and the remaining noise maps act only on S. We can describe this by replacing the action of the noise map at the ℓth step as $\Lambda_\ell(X) \to \varepsilon \otimes \text{tr}_E[\Lambda_\ell(X)]$, where X is the joint SE state at such a step, and where ε is some fiducial state of E. The remaining noise maps will be given by $\Lambda_n \to \mathcal{I}_E \otimes \Lambda_n^{(M)}$ for $\ell < n \leq m + 1$ with some CPTP maps $\Lambda_n^{(M)}$. This implies that

$$\text{tr}_E \circ \mathscr{A}_m(\rho)$$

$$\to \text{tr}_E \circ \frac{\overset{m}{\underset{n=1}{\bigcirc}}\left(\$_{\Lambda_n^{(M)}} - \Theta_{\Lambda_n^{(M)}}\right) \otimes \mathcal{I}_S}{\left(d_S{}^2 - 1\right)^m}\left(\rho - \rho_E \otimes \frac{\mathbb{1}}{d_S}\right)$$

$$= p_{\ell+1} \cdots p_m \, \text{tr}_E \circ \mathscr{A}_\ell(\rho),$$

$$\text{(E1)}$$

and also $\text{tr}_E \circ \mathscr{B}_m(\rho) = \text{tr}_E \circ \mathscr{B}_\ell(\rho) = \text{tr}\left[\mathscr{B}_\ell(\rho)\right]\mathbb{1}/d_S$. In particular, if the final noise were trace-preserving, we would have $\text{tr}\left[\mathscr{B}_\ell(\rho)\right] = 1$. In general, however, this implies

$$\mathcal{F}_m \to p_{\ell+1} \cdots p_m \, \text{tr}\left[\mathcal{M}\Lambda_{m+1}^{(M)} \circ \text{tr}_E \circ \mathscr{A}_\ell(\rho)\right]$$

$$+ \text{tr}\left[\mathscr{B}_\ell(\rho)\right] \text{tr}\left[\mathcal{M}\Lambda_{m+1}^{(M)}\left(\frac{\mathbb{1}}{d_S}\right)\right] \quad \text{with} \quad \ell < m,$$

$$\text{(E2)}$$

where here again $p_n = \text{tr}\left[\Lambda_n^{(M)}\right] - 1/d_S{}^2 - 1$ is the noise strength corresponding to $\Lambda_n^{(M)}$.

This means, as one would expect, that in such a case if non-Markovian noise cannot be resolved with a RB sequence length ℓ, it would amount to SPAM errors, with any subsequent ASF decay being Markovian. Notice, however, that for short sequence lengths, non-Markovian noise could be resolved on average with a few runs of the RB protocol; as explained in the main text, this would allow estimation of the degree of non-Markovianity in the underlying process.

Late non-Markovian noise. Now consider the opposite, where the noise process is initially Markovian but somehow E stops being superfluous after some time step $\ell < m$, i.e., $\hat{\Upsilon} \simeq \Upsilon_\ell^{(M)} \otimes \Upsilon_{m:\ell+1}$. Now we have

$$\mathscr{A}_m(\rho_E \otimes \rho_S)$$

$$\to \frac{\overset{m}{\underset{n=1}{\bigcirc}}\left(\$_{\Lambda_n} - \Theta_{\Lambda_n}\right) \otimes \mathcal{I}_S}{\left(d_S{}^2 - 1\right)^m}\left[\rho_E \otimes \left(\rho_S - \frac{\mathbb{1}}{d_S}\right)\right]$$

$$= p_1 \cdots p_\ell \, \mathscr{A}_{m:\ell+1}(\rho_E \otimes \rho_S),$$

$$\text{(E3)}$$

where here we define

$$\mathscr{A}_{m:k}(\rho) := \frac{\overset{m}{\underset{n=k}{\bigcirc}}\left(\$_{\Lambda_n} - \Theta_{\Lambda_n}\right) \otimes \mathcal{I}_S}{\left(d_S{}^2 - 1\right)^{m-k+1}}\left[\rho - \rho_E \otimes \frac{\mathbb{1}}{d_S}\right],$$

$$\text{(E4)}$$

whilst now $\mathscr{B}_m(\rho_E \otimes \rho_S) = \mathscr{B}_{m:\ell+1}(\rho_E \otimes \rho_S)$, where similarly, $\mathscr{B}_{m:k}(\rho) := \bigcirc_{n=k}^m \Theta_{\Lambda_n}(\rho_E) \otimes \mathbb{1}/d_S$. Thus

$$\mathcal{F}_m \to p_1 \cdots p_\ell \, \text{tr}[\mathcal{M} \; \text{tr}_E \circ \Lambda_{m+1} \circ \mathscr{A}_{m:\ell+1}(\rho)]$$

$$+ \text{tr}[\mathcal{M} \; \text{tr}_E \circ \Lambda_{m+1} \circ \mathscr{B}_{m:\ell+1}(\rho)] \quad \text{with} \quad \ell < m,$$

$$\text{(E5)}$$

so we get a similar behavior, but in this case, as we have seen, it would generally be harder to resolve non-Markovian effects in RB if these occur at longer sequences.

Blocks of finite non-Markovian noise. Now we may consider the case when the noise process is split in two non-Markovian processes, i.e., the first noise process somehow approximately resets the environment at step ℓ and the remaining noise process is also non-Markovian until step m, i.e., $\hat{\Upsilon}_m \simeq \Upsilon_\ell \otimes \Upsilon_{m:\ell+1}$. Now the only difference from a standard non-Markovian ASF is that at the ℓth step we have $\Lambda_\ell(X) \to \varepsilon \otimes \text{tr}_E \circ \Lambda_\ell(X)$, where again ε is some fiducial state of E and X is the state of SE at the ℓth step. This means we can write

$$\mathscr{A}_m(\rho) \to \mathscr{A}_{m:\ell+1}\left[\varepsilon \otimes \text{tr}_E \circ \mathscr{A}_\ell(\rho)\right]$$

$$= \frac{\overset{m}{\underset{n=\ell+1}{\bigcirc}}\left(\$_{\Lambda_n} - \Theta_{\Lambda_n}\right)(\varepsilon)}{\left(d_S{}^2 - 1\right)^{m-\ell}} \otimes \text{tr}_E \circ \mathscr{A}_\ell(\rho), \quad \text{(E6)}$$

whilst now,

$$\mathscr{B}_m(\rho) = \left(\underset{n=\ell+1}{\overset{m}{\bigcirc}} \Theta_{\Lambda_n} \right) \left(\underset{n=1}{\overset{\ell}{\bigcirc}} \Theta_{\Lambda_n} \right) \rho_E \otimes \frac{\mathbb{1}}{d_S}$$

$$= \left(\underset{n=\ell+1}{\overset{m}{\bigcirc}} \Theta_{\Lambda_n} \right) \mathrm{tr}_S \left[\Lambda_n \left\{ \left(\underset{n=1}{\overset{\ell-1}{\bigcirc}} \Theta_{\Lambda_n} \right) \rho_E \otimes \frac{\mathbb{1}}{d_S} \right\} \right] \rho_E \otimes \frac{\mathbb{1}}{d_S}$$

$$= \mathrm{tr} \left[\Lambda_n \left\{ \left(\underset{n=1}{\overset{\ell-1}{\bigcirc}} \Theta_{\Lambda_n} \right) \rho_E \otimes \frac{\mathbb{1}}{d_S} \right\} \right] \left(\underset{n=\ell+1}{\overset{m}{\bigcirc}} \Theta_{\Lambda_n} \right) \varepsilon \otimes \frac{\mathbb{1}}{d_S}$$

$$= \mathrm{tr} \left[\mathscr{B}_\ell(\rho) \right] \mathscr{B}_{m:\ell+1}(\varepsilon \otimes \mathbb{1}/d_S), \tag{E7}$$

so we may write

$$\mathcal{F}_m \to \mathrm{tr} \left\{ \mathcal{M} \ \mathrm{tr}_E \circ \Lambda_{m+1} \circ \mathscr{A}_{m:\ell+1} \left[\varepsilon \otimes \mathrm{tr}_E \circ \mathscr{A}_\ell(\rho) \right] \right\} + \mathrm{tr} \left[\mathscr{B}_\ell(\rho) \right] \mathrm{tr} \left\{ \mathcal{M} \ \mathrm{tr}_E \circ \Lambda_{m+1} \circ \mathscr{B}_{m:\ell+1}(\varepsilon \otimes \mathbb{1}/d_S) \right\} \quad \text{with } \ell < m. \tag{E8}$$

This is a much more complicated behavior, but notice that similarly now after a sequence length ℓ, the first block of non-Markovian noise will be manifest only as SPAM errors. Also, now in essence any other possible mixture of Markovian and non-Markovian noise can be considered, e.g., if there is Markovian noise in-between this would give rise to p factors within the first summand of Eq. (E8) containing $\mathscr{A}$, and $\mathrm{tr}[\mathscr{B}(\rho)]$ factors in the second summand.

In particular, suppose we have two blocks of finite non-Markovian noise, first one of length $k < \ell$, and then a second block of length $\ell < m$. Then we get a recursive expression for the ASF of the form

$$\mathcal{F}_m \to \mathrm{tr} \left\{ \mathcal{M} \ \mathrm{tr}_E \circ \Lambda_{m+1} \circ \mathscr{A}_{m:\ell+k+1} \left[\varepsilon_\ell \otimes \mathrm{tr}_E \circ \mathscr{A}_{\ell:k+1} \left[\varepsilon_k \otimes \mathrm{tr}_E \circ \mathscr{A}_k(\rho) \right] \right] \right\}$$

$$+ \mathrm{tr} \left[\mathscr{B}_k(\rho) \right] \mathrm{tr} \left[\mathscr{B}_{\ell:k+1}(\varepsilon_k \otimes \mathbb{1}/d_S) \right] \mathrm{tr} \left\{ \mathcal{M} \ \mathrm{tr}_E \circ \Lambda_{m+1} \circ \mathscr{B}_{m:\ell+k+1}(\varepsilon_\ell \otimes \mathbb{1}/d_S) \right\} \quad \text{with} \quad k < \ell < m. \tag{E9}$$

If moreover the initial state is uncorrelated, $\rho = \rho_E \otimes \rho_S$, we get

$$\mathcal{F}_m \to \frac{\mathrm{tr} \left[\underset{n=1}{\overset{k}{\bigcirc}} \left(\$_{\Lambda_n} - \Theta_{\Lambda_n} \right) (\rho_E) \right] \mathrm{tr} \left[\underset{n=k+1}{\overset{\ell}{\bigcirc}} \left(\$_{\Lambda_n} - \Theta_{\Lambda_n} \right) (\varepsilon_k) \right]}{(d_S{}^2 - 1)^\ell} \mathrm{tr} \left\{ \mathcal{M} \ \mathrm{tr}_E \circ \Lambda_{m+1} \circ \mathscr{A}_{m:\ell+k+1} \left[\varepsilon_\ell \otimes \rho_S \right] \right\}$$

$$+ \mathrm{tr} \left[\mathscr{B}_k(\rho_E \otimes \rho_S) \right] \mathrm{tr} \left[\mathscr{B}_{\ell:k+1}(\varepsilon_k \otimes \mathbb{1}/d_S) \right] \mathrm{tr} \left\{ \mathcal{M} \ \mathrm{tr}_E \circ \Lambda_{m+1} \circ \mathscr{B}_{m:\ell+k+1}(\varepsilon_\ell \otimes \mathbb{1}/d_S) \right\} \quad \text{with} \quad k < \ell < m. \tag{E10}$$

This then generalizes to blocks with finite non-Markovianity $\Delta\ell_n = \ell_n - (\ell_{n-1} + \ell_{n-2} - \cdots - \ell_1)$, where $\ell_1 < \ell_2 < \cdots < \ell_n < m$ are sequence lengths.

APPENDIX F: CLASSICAL NON-MARKOVIAN NOISE

Classical dephasing noise. For the case of classical correlations we now exemplify how we may describe these through a classical memory specified by an external classical stochastic process whose outputs control the noise Λ_i at every step i. We can depict a circuit for the RB sequence as in Fig. 11(a). Here we focus on the model by Ref. [30] and verify that we obtain the same behavior for the ASF.

In particular, such a model takes a qubit system with errors $\Lambda_i^{(C)}(\cdot) = \lambda_i(\cdot)\lambda_i^\dagger$, where $\lambda_i = \exp(-i\delta_i \otimes Z) = \exp(-i\delta_i Z)$, where $Z = \mathrm{diag}(1, -1)$ and with δ_i is a random variable determined by the classical external control; such variables are randomly sampled but then fixed for the whole RB experiment. Thus a sequence of length m can be treated as a Markovian time-dependent decay $\mathbb{E}[\mathcal{S}_m(\rho)] = p_1 \cdots p_m \Lambda_{m+1}^{(C)}(\rho)$, where knowing the Kraus operators λ,

we can compute for small δ

$$p_i = \frac{|\mathrm{tr}(e^{-i\delta_i \hat{Z}})|^2 - 1}{d_S{}^2 - 1} = \frac{4\cos^2 \delta_i - 1}{3}. \tag{F1}$$

Letting the classical memory be a normally distributed discrete stochastic process $X_i \sim \mathcal{N}(\mu = 0, \sigma^2)$ with mean $\mu = 0$ and variance σ^2, the so-called Markovian scenario considers the control operations at step i giving a realization $X_i = x_i$ and setting $\delta_i = x_i$. That is, all errors being independent of each other. Ignoring SPAM errors, suppose $\rho = \mathcal{M} = |0\rangle\langle 0|$, so that

$$\mathcal{F}_m^{\mathrm{C-Mark}} = \mathrm{tr}\{\mathcal{M}\mathbb{E}[\mathcal{S}_m(\rho)]\} = \prod_{i=1}^m \left(\frac{4\cos^2 \delta_i - 1}{3} \right). \tag{F2}$$
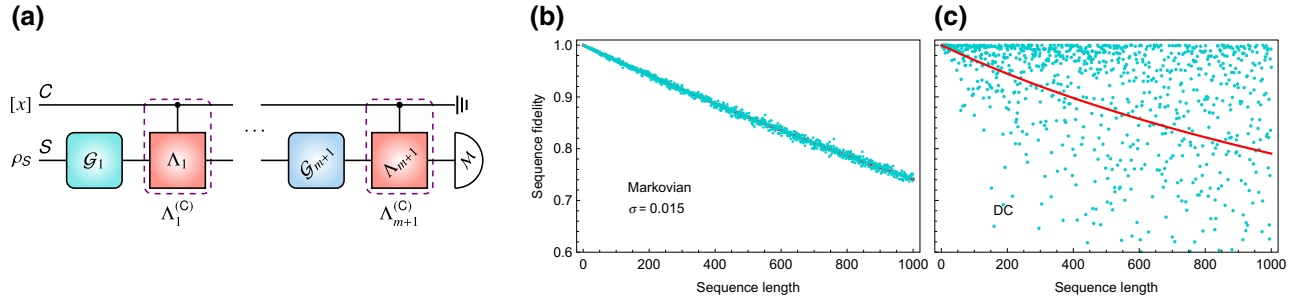
FIG. 11. **Classical non-Markovian noise**. (a) A RB sequence with classical non-Markovian noise for an initial state ρ_S, gates $\{\mathcal{G}_i\}$ and final measurement $\mathcal{M}$. Correlations are determined by a classical stochastic process $[x]$, whose output at each step i controls the corresponding error map Λ_i on S giving some other map $\Lambda_i^{(C)}$. We reproduce the ASF in Ref. [30] for classical dephasing noise in the cases of (b) classical Markovian noise and (c) classical dc noise with a standard deviation $\sigma = 0.015$; in both cases (teal) dots represent numerical averages over 100 samples and the solid (red) curve is the analytical ASF.

The other extreme scenario is when all noise random variables are identical so that $\delta_i = \delta$, so-called dc noise; here the control should measure a realization with probability p and update the memory with a PDF of the form $\sum_i \mathbb{P}(X = x_i)\Delta(\delta - x_i)$, where here Δ is a Dirac delta distribution. Then $\mathbb{E}[\mathcal{S}_m(\rho)] = p^m\, e^{-i\delta\hat{Z}} \rho\, e^{i\delta\hat{Z}}$, which similarly for $\rho = \mathcal{M} = |0\rangle\langle 0|$ becomes simply

$$\mathcal{F}_m^{\text{C-dc}} = \text{tr}\{\mathcal{M}\mathbb{E}[\mathcal{S}_m(\rho)]\} = \left(\frac{4\cos^2\delta - 1}{3}\right)^m. \quad (\text{F3})$$

For both extreme cases we see that we effectively reproduce the behavior described in Ref. [30], in particular, for the average sequence fidelity. Here we still have to average over the classical random variable:

$$\langle\mathcal{F}_m^{\text{C-Mark}}\rangle = \left[\int_{-\infty}^{\infty}\left(\frac{\exp\left(\delta^2/2\sigma^2\right)}{\sigma\sqrt{2\pi}}\right)\left(\frac{4\cos^2\delta - 1}{3}\right)d\delta\right]^m,$$

$$\langle\mathcal{F}_m^{\text{C-dc}}\rangle = \int_{-\infty}^{\infty}\left(\frac{\exp\left(\delta^2/2\sigma^2\right)}{\sigma\sqrt{2\pi}}\right)\left(\frac{4\cos^2\delta - 1}{3}\right)^m d\delta. \quad (\text{F4})$$

For the Markovian case, the average can be carried out to obtain a decay $\langle\mathcal{F}_m^{\text{C-Mark}}\rangle = P^m$, where here P is the true error rate together with the classical noise. For a standard deviation of $\sigma = 0.015$, this gives $\langle\mathcal{F}_m^{\text{C-Mark}}\rangle \approx (0.9997)^m$. The dc case, as expected is more complicated, and one possibility is to expand the cosine function around $\delta = 0$ to analyze the average fidelity, similar to how it is done in Ref. [30] with contributions up to δ^2. The final behavior of $\langle\mathcal{F}_m^{\text{C-dc}}\rangle$ differs both from an exponential and a simple product of noise-strengths. We show plots for the average fidelities in both cases with a standard deviation of $\sigma = 0.015$ in Figs. 11(b) and 11(c).

The shallow pocket model. We now consider a similar model for a qubit S coupled to degree of freedom (DOF) on a real line, which acts as an environment. This is labeled a shallow pocket model because such DOF cannot store energy internally. This is an interesting model for several reasons, but here mainly because it leads to completely positive and divisible dynamics of S but it is nevertheless non-Markovian [34,62]. For RB, however, the nature of classical correlations is what leads to a treatment of the ASF as a time-dependent Markovian one.

The shallow pocket model now considers $\Lambda_n^{(C)}(\cdot) = \lambda_n(\cdot)\lambda_n^\dagger$ with $\lambda_n = \exp(-i\tau_n\, \hat{x}_n \otimes Z) = \exp(-i\tau_n\, x_n\, Z)$, where $\hat{x}_n$ is a position operator at time step n and τ_n are time intervals representing evolution time of the nth step. This immediately implies that the average sequence is of the form $\mathbb{E}[\mathcal{S}_m(\rho)] = p_1 \cdots p_m\, \Lambda_{m+1}^{(C)}(\rho)$, where $\rho = \rho_S \otimes |\psi\rangle\langle\psi|$. The initial state of the environment DOF is taken as $|\psi\rangle$ such that $\langle\psi|x_1\rangle = \sqrt{\gamma/\pi}\,/(x_1 + i\gamma)$. Now tracing out the environment at the end of the process is equivalent to integrating x over the reals with a factor $\langle\psi|x\rangle\langle x_m|\psi\rangle\delta_{xx_2}\delta_{x_2 x_3}\cdots\delta_{x_m x_{m+1}}$. Thus we can think of the external DOF as a classical dc noise distributed with a probability density function $|\langle\psi|x\rangle|^2$.

That is, now we have

$$p_{\tau_n}(x_n) = \frac{|\text{tr}\left(e^{-i\tau_n\, x_n\, \hat{Z}}\right)|^2 - 1}{d_S^2 - 1} = \frac{4\cos^2(\tau_n\, x_n) - 1}{3}. \quad (\text{F5})$$

Notice that all p's have to be constrained to $[0, 1]$, so to have a meaningful ASF the equivalent of our distribution, namely $|\langle\psi|x\rangle|^2$, has to contain a low enough equivalent of a variance, which amounts to choosing an appropriate value for γ. Hence, now taking $\rho = |0\rangle\langle 0| \otimes |\psi\rangle\langle\psi|$ and $\mathcal{M} = |0\rangle\langle 0|$, we get

$$\mathcal{F}_m^{\text{shallow}} = \text{tr}\{\mathcal{M}\,\mathbb{E}[\mathcal{S}_m(\rho)]\} = \frac{\gamma}{\pi}\int_{-\infty}^{\infty}\frac{p_{\tau_1 x}\cdots p_{\tau_m x}}{x^2 + \gamma^2}\, dx, \quad (\text{F6})$$

which is somewhat harder to evaluate given that expanding around small x is not a viable option. Regardless, the point we make here is that classical correlations such as the one before of dephasing noise or the shallow pocket model can be treated on RB with a standard Markovian time-dependent approach.

APPENDIX G: NUMERICAL CALCULATIONS

SPAM errors. As in the main text, here we consider a qubit subject to time-independent unitary noise $\Lambda(\cdot) = \lambda(\cdot)\lambda^\dagger$ on a full N-qubit system, where $\lambda = \exp(-i\delta H)$ with H given by the N-site Ising spin chain

$$
\begin{aligned}
H &= \sum_{i=1}^{N} \left(\frac{J}{2} X_i X_{i+1} + h_x X_i + h_y Y_i \right) \\
&= \begin{pmatrix}
0 & h_x - ih_y & h_x - ih_y & J \\
h_x + ih_y & 0 & J & h_x - ih_y \\
h_x + ih_y & J & 0 & h_x - ih_y \\
J & h_x + ih_y & h_x + ih_y & 0
\end{pmatrix}, \quad \text{(G1)}
\end{aligned}
$$

with X_i, Y_i being Pauli matrices acting on the ith site. We take a closed chain so that $X_{N+1} := X_1$. In particular, in the main text we take only $N = 2$ qubits, with site $i = 1$ being system S.

Here we pick the values $J = 1.7$, $h_x = 1.47$, and $h_y = -1.05$, fixing $\delta = 0.029\,475$. To take into account SPAM errors numerically, suppose the initial state ρ is previously affected by the same Λ error for some small $\delta = \Delta_1$, and that $\mathcal{M}$ is slightly rotated via $\exp(-i\Delta_2 Y)$ for a small Δ_2. In Fig. 12 we show examples for both mild, $\Delta_1 = 0.04232$

and $\Delta_2 = 0.09321$, and much worse, $\Delta_1 = 0.2932$ and $\Delta_2 = 0.10321$. We also consider the case where the preparation affects only S by some rotation $\exp(-i\gamma X)$ with a small γ, but does not generate correlations with E.

In all cases SPAM makes it harder to numerically resolve non-Markovian effects. Similar to the Markovian case, SPAM errors generate an offset of the ASF, but in general they also affect the decay rate of the errors. This can be argued to be mainly due to the correlating effect of errors but changes in the decay rates can also be seen when the preparation does not generate correlations with E. The impact of SPAM in the characterization of non-Markovian noise with RB is thus an issue that still has to be studied in greater detail.

Absence of nonexponential behavior. We notice that for a similar noise model for a couple of qubits,

$$
\begin{aligned}
H &= J_x X_1 X_2 + J_y Y_1 Y_2 \\
&= \begin{pmatrix}
0 & 0 & 0 & J_x - J_y \\
0 & 0 & J_x + J_y & 0 \\
0 & J_x + J_y & 0 & 0 \\
J_x - J_y & 0 & 0 & 0
\end{pmatrix}, \quad \text{(G2)}
\end{aligned}
$$

essentially no deviation from an exponential is seen. We look again at time-independent noise given by $\lambda = \exp(-iH\delta)$ with small $\delta = 0.029\,475$ and take $\rho = |00\rangle\langle 00|$, where one of the qubits is identified as system S and the other one as the environment E, and take $\mathcal{M} = |0\rangle\langle 0|$. We show the corresponding ASF in Fig. 13 for the arbitrary choices $J_x = 1.2$, $J_y = -2.7$.
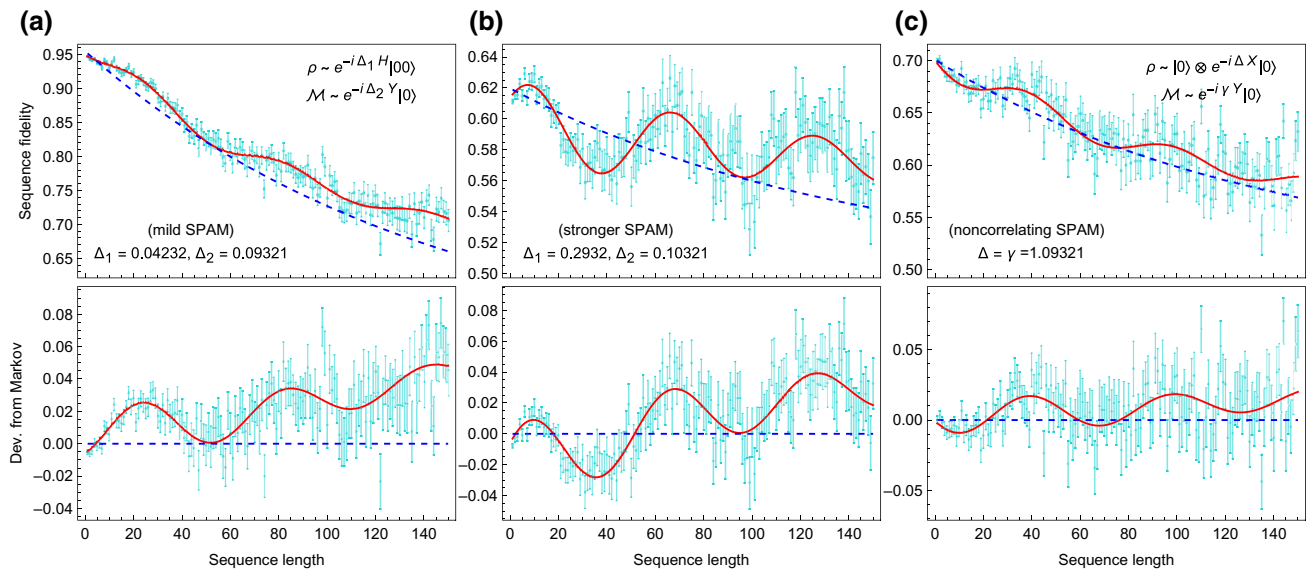

FIG. 12. **Effect of** SPAM **errors in the two-qubit spin noise in Eq.** (24). In the non-Markovian case, SPAM errors result in an offset but also appear to affect the error rates. In (a) and (b) the initial state is affected by the same sequence noise with $\Lambda \exp(-i\Delta_1 H)$ for different values of Δ_1 and $\mathcal{M}$ is slightly rotated via $\exp(-i\Delta_2 Y)$ with a small Δ_2. In (c) the initial state is only affected on system S with a rotation $\exp(-i\gamma X)$ and a relatively large γ. In all cases the sample size remains fixed at 100 samples.

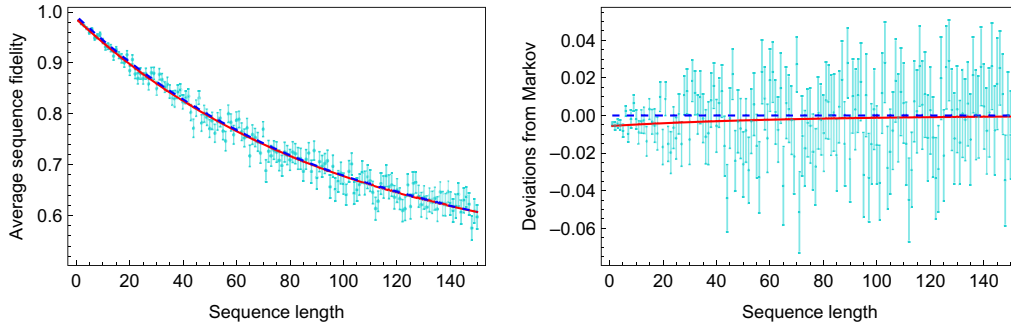FIG. 13. RB non-Markovianity blindness on an *XX*-spin chain. Despite being generically non-Markovian, the noise model of Eq. (G2) displays virtually no deviations from a Markovian noise model when $J = 1.2$, $J_y = 2.7$.

Notice that small deviations do occur at very short sequence lengths, although they are practically negligible. While of course, we are not quantifying the non-Markovianity of the model, and also different choices of the couplings might display larger deviations, the point we want to make is that there are going to be models that are blind, or at least myopic, to non-Markovianity when employing RB, and the circumstances when this occurs are still to be better understood.

Increasing environment dimension. We now look at the effect of increasing the number of qubits in E; noticeably the environment dimension does not show up explicitly in the main ASF in Eq. (7). We now employ similar conditions on the Hamiltonian in Eq. (G1) for a changing value of N. In Fig. 14 we show the deviations from RB non-Markovianity for up to five environmental qubits, and notice that the nonexponential deviations get effectively damped, albeit slowly and for longer sequence lengths first. This is expected behavior, but nevertheless it is still a question what is exactly the dependence of the general non-Markovian ASF in environment dimension.

Finite non-Markovian sequence lengths and non-Markovian deviations. Whenever we have finite non-Markovian noise, say over an initial sequence length approximately ℓ, described by the CP maps $\Lambda_1, \Lambda_2, \ldots, \Lambda_\ell$, and an uncorrelated input state, by choosing to fix $\ell - 1$ Cliffords after the first one to be identities, by Eq. (E2), we get a Markovian decay with $\mathcal{F}_m \simeq p_{\ell:1} p_{\ell+1} \cdots p_m A + B$, where

$$p_{\ell:1} := \frac{\text{tr}\left[\Lambda_{\ell:1}^{(M)}\right] - 1}{d_S^2 - 1}, \tag{G3}$$

with

$$\Lambda_{\ell:1}^{(M)}(\cdot) := \text{tr}_E[\Lambda_\ell \circ \Lambda_{\ell-1} \circ \cdots \circ \Lambda_1(\varepsilon \otimes \cdot)], \tag{G4}$$

with each Λ_n in terms of Kraus operators λ_{i_n} acting on SE spaces and co-spaces as $\Lambda_n := \sum_i \lambda_{i_n} \otimes \lambda_{i_n}^\dagger$.

That is, the initial block of finite non-Markovian noise looks like a single noise map $\Lambda_{\ell:1}$ if we randomize over a
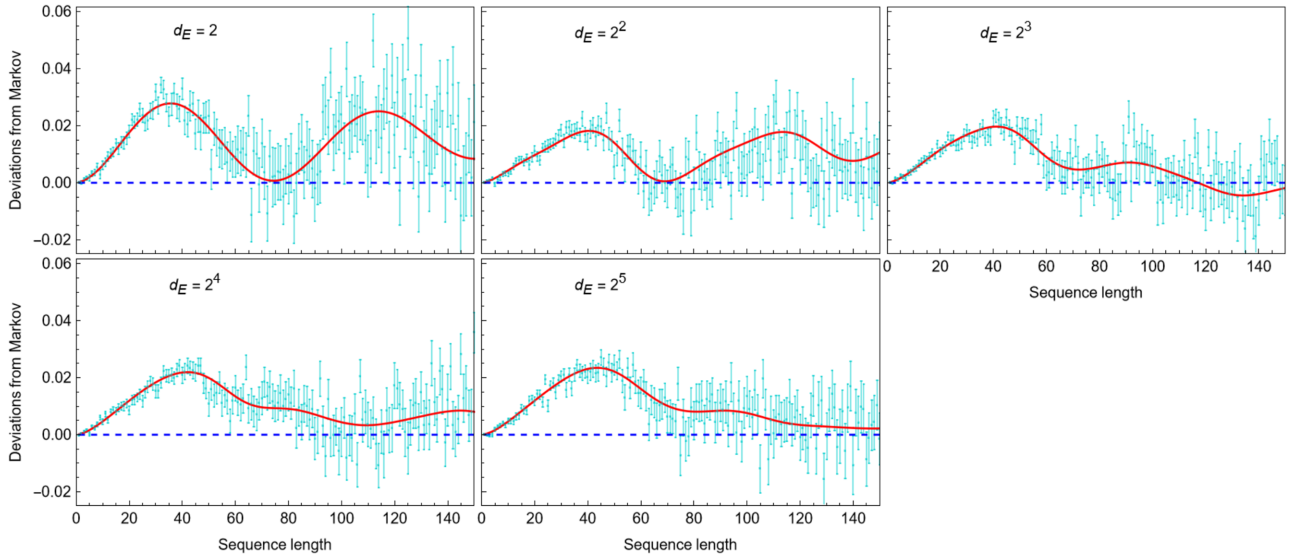

FIG. 14.   **Behavior of the** ASF **in environment dimension for an Ising spin chain as time-independent noise.** We take as time-independent noise Eq. (G1) for a single qubit in S and a variable number of qubits in E, with site $i = 1$ being system S. For all we pick $J = 1.7$, $h_x = 0.9$, and $h_y = -1.05$, fixing $\delta = 0.029\,475$ and take $\rho = |0\rangle\langle 0|^{\otimes 3}$ and $\mathcal{M} = |0\rangle\langle 0|$, ignoring SPAM errors.

single Clifford within this block, with the remaining ones set to identities.

While this is an idealized scenario, we can use it to estimate sequence lengths at which non-Markovian noise effects are relevant in a RB experiment. In the main text we model such a noise process with a noise map at the nth step given by

$$\Lambda_n^{(\ell)} := q_{n-\ell}\Lambda + (1 - q_{n-\ell})\Lambda^{(M)}, \quad \text{where}$$

$$q_n := \frac{1}{1 + \exp(n - \ell)}, \tag{G5}$$

where here again $\Lambda = \lambda \otimes \lambda^\dagger$ with $\lambda = \exp(-i\delta H)$, where H is given by Eq. (G1) and $\Lambda^{(M)}$ acts on S as $\Lambda^{(M)}(\cdot) = \varepsilon \otimes \text{tr}_E[\Lambda(\cdot)]$. In the main text we also fix the values $J = 1.7$, $h_x = 0.5$, and $h_y = -1.05$ and set $\delta \approx 0.03$, although we now pick a $\delta^{(M)} = 2.5\delta$ for $\Lambda_n^{(M)}$. This implies that the noise acts jointly over the whole SE throughout the full process, but it acts almost fully as Λ for $m < \ell$, whilst it turns almost to act solely on S with $\Lambda^{(M)}$ for $m > \ell$.

In the top figure of Fig. 6 in the main text, we display the ASFs for a set of RB experiments with $\rho = |00\rangle\langle 00|$ and $\mathcal{M} = |0\rangle\langle 0|$ for different sets of fixed identities at sequence lengths $1, 2, \ldots, 8$. We fix $\ell = 9$ and we describe the way in which an experimenter can estimate this value of ℓ from the data of the experiments alone, as well as construct a sensible time-independent Markovian ASF with which they can quantify the amount of non-Markovian deviations; this is shown in the bottom panel of the same figure.

The procedure is the following given a single ASF, $\mathcal{F}_m$, displaying a nonexponential decay over a finite sequence length:

1. Fix identities at sequence lengths of $\mathcal{F}_m$ manifestly displaying deviations from an exponential decay and run RB experiments for each of them, obtaining corresponding ASFs $\mathcal{F}_{m/\{i,\ldots,j\}}$, where $i, \ldots, j$ are sequence lengths at which identities are fixed.

2. Identify the section of the original $\mathcal{F}_m$ manifestly displaying exponential behavior and extract the noise rate p at such a section.

3. Fit an exponential to each $\mathcal{F}_{m/\{i,\ldots,j\}}$; Eq. (E2) implies that the curve with an exponential rate $p_{m/\{i',\ldots,j'\}}$ closest to p will indicate the length at which the noise turns almost Markovian (or where non-Markovian effects become negligible).

4. Finally, a Markovianized ASF can be constructed with rate $p_{m/\{i',\ldots,j'\}}$ and at least two reasonable constraints for the SPAM constants, such as $A + B = 1$ and $A \approx B$ if the SPAM errors are assumed low and the decay rate is not too high, $p \approx 1$.

For the particular example in the main text, step 1 is displayed in the top panel of Fig. 6, each over 150 samples.

For step 2, we took points $\{m, \mathcal{F}_m\}$ from $m = 12$ to $m = 30$, which more manifestly display an exponential decay. These are fitted to an exponential $f_m \approx (0.7847)(0.9325)^m + 0.4915$, i.e., we extract $p \simeq 0.9325$.

For step 3, we identify the closest decay rate to p occurred for $\mathcal{F}_{m/\{1,\ldots,8\}}$, with $p_{m/\{1,\ldots,8\}} \simeq 0.9278$. This indicates that $\ell \approx 9$. Since we fix $\ell = 9$, this procedure is essentially identifying that the non-Markovian effects of the noise on the ASF become negligible at sequence length $m = 9$; notice that at this length $q_0 = 1/2$, i.e., $\Lambda_9^{(9)} = \frac{1}{2}\left(\Lambda + \Lambda^{(M)}\right)$ so that the noise will still act jointly on SE with at least half probability. In this sense ℓ is just approximated numerically. In the bottom panel of Fig. 6, the dot-dashed line displays the curve given by $\tilde{f}_m \simeq (0.7847)p_{m/\{1,\ldots,8\}}^m + 0.4915$, showing the slight offset due to this numerical estimation.

Finally, at step 4 we simply fix $A \approx B$ in $\mathcal{F}^{(M)} = Ap_{m/\{1,\ldots,8\}}^m + B$ assuming low spam errors; in Fig. 6 we specifically take $A = 0.5085$ and $B = 0.4915$ with the demand that B converges to the same value as in f_m and $\tilde{f}_m$ for $m \to \infty$. As is the case for RB, this Markovianized ASF curve at most informs us about the gate fidelity with respect to the identity of the Markovianized noise through $p_{m/\{1,\ldots,8\}}$.

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

[2] B. Lévi, C. C. López, J. Emerson, and D. G. Cory, Efficient error characterization in quantum information processing, Phys. Rev. A **75**, 022314 (2007).

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

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

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

[6] J. Helsen, I. Roth, E. Onorati, A. H. Werner, and J. Eisert, A general framework for randomized benchmarking (2020), ArXiv:2010.07974.

[7] I. L. Chuang and M. A. Nielsen, Prescription for experimental determination of the dynamics of a quantum black box, J. Mod. Optic **44**, 2455 (1997).

[8] I. Roth, R. Kueng, S. Kimmel, Y.-K. Liu, D. Gross, J. Eisert, and M. Kliesch, Recovering Quantum Gates from Few Average Gate Fidelities, Phys. Rev. Lett. **121**, 170502 (2018).

[9] J. J. Wallman and S. T. Flammia, Randomized benchmarking with confidence, New J. Phys. **16**, 103032 (2014).

[10] M. Mohseni, A. T. Rezakhani, and D. A. Lidar, Quantum-process tomography: Resource analysis of different strategies, Phys. Rev. A **77**, 032322 (2008).

[11] E. Nielsen, J. K. Gamble, K. Rudinger, T. Scholten, K. Young, and R. Blume-Kohout, Gate set tomography (2020), ArXiv:2009.07301.

[12] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, Quantum State Tomography via Compressed Sensing, Phys. Rev. Lett. **105**, 150401 (2010).

[13] S. T. Flammia, D. Gross, Y.-K. Liu, and J. Eisert, Quantum tomography via compressed sensing: Error bounds, sample complexity and efficient estimators, New J. Phys. **14**, 095022 (2012).

[14] S. T. Flammia and Y.-K. Liu, Direct Fidelity Estimation from Few Pauli Measurements, Phys. Rev. Lett. **106**, 230501 (2011).

[15] M. P. da Silva, O. Landon-Cardinal, and D. Poulin, Practical Characterization of Quantum Devices Without Tomography, Phys. Rev. Lett. **107**, 210404 (2011).

[16] O. Moussa, M. P. da Silva, C. A. Ryan, and R. Laflamme, Practical Experimental Certification of Computational Quantum Gates Using a Twirling Procedure, Phys. Rev. Lett. **109**, 070504 (2012).

[17] In Ref. [18], it was pointed out that this relation between the ASF and the gate fidelity of the noise is not unique due to gauge invariance. An in-depth discussion can be seen in Refs. [19,80] with an overview and generalization in Ref. [6].

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

[19] J. J. Wallman, Randomized benchmarking with gate-dependent noise, Quantum **2**, 47 (2018).

[20] J. Helsen, X. Xue, L. M. K. Vandersypen, and S. Wehner, A new class of efficient randomized benchmarking protocols, npj Quantum Inf. **5**, 71 (2019).

[21] A. Carignan-Dugas, J. J. Wallman, and J. Emerson, Characterizing universal gate sets via dihedral benchmarking, Phys. Rev. A **92**, 060302 (2015).

[22] A. K. Hashagen, S. T. Flammia, D. Gross, and J. J. Wallman, Real randomized benchmarking, Quantum **2**, 85 (2018).

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

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

[25] C. A. Ryan, M. Laforest, and R. Laflamme, Randomized benchmarking of single- and multi-qubit control in liquid-state NMR quantum information processing, New J. Phys. **11**, 013034 (2009).

[26] D. K. Park, G. Feng, R. Rahimi, J. Baugh, and R. Laflamme, Randomized benchmarking of quantum gates implemented by electron spin resonance, J. Magn. Reson. **267**, 68 (2016).

[27] J. M. Epstein, A. W. Cross, E. Magesan, and J. M. Gambetta, Investigating the limits of randomized benchmarking protocols, Phys. Rev. A **89**, 062321 (2014).

[28] M. A. Fogarty, M. Veldhorst, R. Harper, C. H. Yang, S. D. Bartlett, S. T. Flammia, and A. S. Dzurak, Nonexponential

fidelity decay in randomized benchmarking with low-frequency noise, Phys. Rev. A **92**, 022326 (2015).

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

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

[31] B. H. Fong and S. T. Merkel, Randomized benchmarking, correlated noise, and ising models (2017), ArXiv:1703.09747.

[32] J. Qi and H. K. Ng, Randomized benchmarking in the presence of time-correlated dephasing noise, Phys. Rev. A **103**, 022607 (2021).

[33] J. M. Gambetta, A. D. Córcoles, S. T. Merkel, B. R. Johnson, J. A. Smolin, J. M. Chow, C. A. Ryan, C. Rigetti, S. Poletto, T. A. Ohki, M. B. Ketchen, and M. Steffen, Characterization of Addressability by Simultaneous Randomized Benchmarking, Phys. Rev. Lett. **109**, 240504 (2012).

[34] S. Milz and K. Modi, Quantum stochastic processes and quantum non-markovian phenomena, PRX Quantum **2**, 030201 (2021).

[35] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Quantum Circuit Architecture, Phys. Rev. Lett. **101**, 060401 (2008).

[36] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Theoretical framework for quantum networks, Phys. Rev. A **80**, 022339 (2009).

[37] F. Costa and S. Shrapnel, Quantum causal modelling, New J. Phys. **18**, 063032 (2016).

[38] F. A. Pollock, C. Rodríguez-Rosario, T. Frauenheim, M. Paternostro, and K. Modi, Operational Markov Condition for Quantum Processes, Phys. Rev. Lett. **120**, 040405 (2018).

[39] F. A. Pollock, C. Rodríguez-Rosario, T. Frauenheim, M. Paternostro, and K. Modi, Non-markovian quantum processes: Complete framework and efficient characterization, Phys. Rev. A **97**, 012127 (2018).

[40] D. Gottesman, Stabilizer codes and quantum error correction (1997), ArXiv:quant-ph/9705052.

[41] D. Gottesman, The heisenberg representation of quantum computers (1998), ArXiv:quant-ph/9807006.

[42] S. Aaronson and D. Gottesman, Improved simulation of stabilizer circuits, Phys. Rev. A **70**, 052328 (2004).

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

[44] Y. Nakata, D. Zhao, T. Okuda, E. Bannai, Y. Suzuki, S. Tamiya, K. Heya, Z. Yan, K. Zuo, S. Tamate, Y. Tabuchi, and Y. Nakamura, Quantum circuits for exact unitary *t*-designs and applications to higher-order randomized benchmarking (2021), ArXiv:2102.12617.

[45] P. Taranto, Memory effects in quantum processes, Int. J. Quantum Inf. **18**, 1941002 (2020).

[46] S. Milz, F. A. Pollock, and K. Modi, An introduction to operational quantum dynamics, Open Syst. Inf. Dyn. **24**, 1740016 (2017).

[47] J. Watrous, *The Theory of Quantum Information* (Cambridge University Press, Cambridge, 2018).

[48] Strictly speaking, here we are defining $\mathfrak{C}_m = \mathbb{1} \otimes \mathfrak{A}_m$, where $\mathfrak{A}_m$ is the Choi state of the sequence of gates $\mathcal{G}_i$, and $\mathcal{G}_i$ can be defined to act on either auxiliary space A_i or B_i, the choice depends only on what auxiliary space the swap $\mathscr{S}_i$ on the definition of Υ_m swaps with, so that the contraction $\Upsilon_m \mathfrak{C}_m^T$ contracts the correct spaces.

[49] S. Milz, F. Sakuldee, F. A. Pollock, and K. Modi, Kolmogorov extension theorem for (quantum) causal modelling and general probabilistic theories, Quantum **4**, 255 (2020).

[50] H. I. Nurdin and J. E. Gough, From the Heisenberg to the Schrödinger picture: Quantum stochastic processes and process tensors (2021), ArXiv:2109.09256.

[51] S. Milz, M. S. Kim, F. A. Pollock, and K. Modi, Completely Positive Divisibility Does not Mean Markovianity, Phys. Rev. Lett. **123**, 040401 (2019).

[52] P. Figueroa-Romero, K. Modi, and F. A. Pollock, Almost markovian processes from closed dynamics, Quantum **3**, 136 (2019).

[53] P. Figueroa-Romero, F. A. Pollock, and K. Modi, Markovianization with approximate unitary designs, Commun. Phys. **4**, 127 (2021).

[54] A. Yuvaraj, The action of the map $\$_{\Lambda_n}$ can alternatively be written as $\$_\Lambda(\cdot) = \sum_\mu \operatorname{tr}_S(\lambda_\mu)(\cdot)\operatorname{tr}_S(\lambda_\mu^\dagger)$ with λ_μ the Kraus operators of Λ.

[55] K. Modi, Operational approach to open dynamics and quantifying initial correlations, Sci. Rep. **2**, 581 (2012).

[56] M. Ringbauer, C. J. Wood, K. Modi, A. Gilchrist, A. G. White, and A. Fedrizzi, Characterizing Quantum Dynamics with Initial System-Environment Correlations, Phys. Rev. Lett. **114**, 090402 (2015).

[57] K. Rudinger, T. Proctor, D. Langharst, M. Sarovar, K. Young, and R. Blume-Kohout, Probing Context-Dependent Errors in Quantum Processors, Phys. Rev. X **9**, 021045 (2019).

[58] C. Giarmatzi and F. Costa, Witnessing quantum memory in non-markovian processes, Quantum **5**, 440 (2021).

[59] S. Milz, C. Spee, Z.-P. Xu, F. A. Pollock, K. Modi, and O. Gühne, Genuine multipartite entanglement in time, SciPost Phys. **10**, 141 (2021).

[60] C. Arenz, R. Hillier, M. Fraas, and D. Burgarth, Distinguishing decoherence from alternative quantum theories by dynamical decoupling, Phys. Rev. A **92**, 022102 (2015).

[61] C. Arenz, D. Burgarth, P. Facchi, and R. Hillier, Dynamical decoupling of unbounded hamiltonians, J. Math. Phys. **59**, 032203 (2018).

[62] P. Taranto, F. A. Pollock, and K. Modi, Memory strength and recoverability of non-markovian quantum stochastic processes (2019), ArXiv:1907.12583.

[63] This observation can be seen to follow, e.g., because we may upper bound the total RB non-Markovianity of a sequence length m experiment, $\mathcal{N}_q^{\mathcal{F}_m}$ in Eq. (15), as $\mathcal{N}_q^{\mathcal{F}_m} \le \sum \mathcal{N}_q$, where $\mathcal{N}_q$ is general non-Markovianity for each intermediate step up to m as in Eq. (5) with D being a Schatten q norm, $D = \|X\|_q := (\operatorname{tr}|X|^q)^1/q$.

[64] P. Taranto, F. A. Pollock, S. Milz, M. Tomamichel, and K. Modi, Quantum Markov Order, Phys. Rev. Lett. **122**, 140401 (2019).

[65] P. Taranto, S. Milz, F. A. Pollock, and K. Modi, Structure of quantum stochastic processes with finite markov order, Phys. Rev. A **99**, 042108 (2019).

[66] G. A. L. White, F. A. Pollock, L. C. L. Hollenberg, K. Modi, and C. D. Hill, Non-markovian quantum process tomography (2021), ArXiv:2106.11722.

[67] I. Heinz and G. Burkard, Crosstalk analysis for single-qubit and two-qubit gates in spin qubit arrays (2021), ArXiv:2105.10221.

[68] D. Tamascelli, A. Smirne, S. F. Huelga, and M. B. Plenio, Nonperturbative Treatment of Non-Markovian Dynamics of Open Quantum Systems, Phys. Rev. Lett. **120**, 030402 (2018).

[69] I. A. Luchnikov, S. V. Vintskevich, H. Ouerdane, and S. N. Filippov, Simulation Complexity of Open Quantum Dynamics: Connection with Tensor Networks, Phys. Rev. Lett. **122**, 160401 (2019).

[70] B. Dirkse, J. Helsen, and S. Wehner, Efficient unitarity randomized benchmarking of few-qubit clifford gates, Phys. Rev. A **99**, 012315 (2019).

[71] M. Girling, C. Cirstoiu, and D. Jennings, Estimation of correlations and non-separability in quantum channels via unitarity benchmarking (2021), ArXiv:2104.04352.

[72] N. C. Brown and K. R. Brown, Leakage mitigation for quantum error correction using a mixed qubit scheme, Phys. Rev. A **100**, 032325 (2019).

[73] A. Strikis, D. Qin, Y. Chen, S. C. Benjamin, and Y. Li, Learning-based quantum error mitigation (2020), ArXiv:2005.07601.

[74] P. Parrado-Rodríguez, C. Ryan-Anderson, A. Bermudez, and M. Müller, Crosstalk suppression for fault-tolerant quantum error correction with trapped ions, Quantum **5**, 487 (2021).

[75] G. D. Berk, S. Milz, F. A. Pollock, and K. Modi, Extracting quantum dynamical resources: Consumption of non-markovianity for noise reduction (2021), ArXiv:2110.02613.

[76] I. Bengtsson and K. Życzkowski, *Geometry of Quantum States: An Introduction to Quantum Entanglement* (Cambridge University Press, Cambridge, 2017).

[77] P. Figueroa-Romero, Equilibration and typicality in quantum processes (2021), ArXiv:2102.02289.

[78] B. Collins, Moments and cumulants of polynomial random variables on unitary groups, the itzykson-zuber integral, and free probability, Int. Math. Res. Notices **2003**, 953 (2003).

[79] B. Collins and P. Śniady, Integration with respect to the haar measure on unitary, orthogonal and symplectic group, Commun. Math. Phys. **264**, 773 (2006).

[80] S. T. Merkel, E. J. Pritchett, and B. H. Fong, Randomized benchmarking as convolution: Fourier analysis of gate dependent errors (2018), ArXiv:1804.05951.