Randomised benchmarking for non-Markovian noise

Pedro Figueroa-Romero,1,* Kavan Modi,2,† Thomas M. Stace,3,§ and Min-Hsiu Hsieh1,‡

1Hon Hai Quantum Computing Research Center, Taipei, Taiwan
2ARC Centre for Engineered Quantum Systems and School of Physics and Astronomy, Monash University, Victoria, Australia
3ARC Centre for Engineered Quantum Systems, University of Queensland, Brisbane, Australia

(Dated: July 13, 2021)

Estimating the features of noise is the first step in a chain of protocols that will someday lead to fault tolerant quantum computers. The randomised 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 characterisation 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 an 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 pathway to better understanding 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 randomised benchmarking (RB) [1–5] has become the gold standard to certify the performance of gate-sets and characterise the noise in computations involving these sets. RB generally refers to an experimental protocol allowing to estimate 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 to fully reconstruct noise 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 characterise 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 [17], 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, behaves as a decaying exponential in 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 characterisation 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 Ref. [29, 30] (and similarly in Ref. [31]), the noise is modelled 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 generalised to correla-
tions being mediated by a bath, modelled as a multi-mode bosonic field interacting with the qubit. Similarly, correlations arising as interaction between neighbouring qubits, so-called crosstalk [33] have been addressed in multi-qubit RB protocols, generally noticing that averaging over a single qubit generally leads to a non-exponential 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 characterise non-Markovian processes [39].

In this manuscript, we derive an analytical expression for the ASF of an RB experiment with the Clifford group under non-Markovian gate-independent noise. This allows to study the behaviour 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 manuscript is structured as follows. In Section II we introduce the RB protocol and discuss the theoretical setting employed in the remainder of the paper. In Section III we introduce the process tensor framework and elaborate on how it is a natural framework for non-Markovian RB. In Section 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 Section 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 Section 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 Section 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 Section 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 Section IX with an overview of our results and a perspective for future work.

II. RANDOMISED BENCHMARKING

While there are many variants of RB, and a general framework encompassing these can be established [6], for concreteness here we will consider an 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 realised efficiently on a quantum processor [40–42]. The RB protocol is then as follows:

1. Prepare an initial state \( \rho \).

2. Sample \( m \) distinct elements, \( G_1, G_2, \ldots, G_m \), uniformly at random from the Clifford group. Let \( G_{m+1} := \circlearrowleft_{i=1}^{m+1} G_i \) denote composition of maps and \( G^{i}(\cdot) = G^{i}(\cdot)G \), where \( G \) is unknown. In practice, this amounts to applying a noisy sequence \( S_m := \circlearrowleft_{i=1}^{m+1} G_i \) of length \( m \) on \( \rho \), where \( G_i \) are the physical noisy gates associated to \( G \).

3. Apply the composition \( \circlearrowleft_{i=1}^{m+1} G_i \) on \( \rho \). In practice, this amounts to applying a noisy sequence \( S_m := \circlearrowleft_{i=1}^{m+1} G_i \) of length \( m \) on \( \rho \), where \( G_i \) are the physical noisy gates associated to \( G \).

4. Estimate the probability \( f_m = \text{tr}[MS_m(\rho)] \) via a POVM element \( M \).

5. Repeat \( n \) times the steps 1 to 4 for the same initial state \( \rho \), same POVM element \( M \), and different sets of gates chosen uniformly at random \( \{G^{(1)}_{i1}, G^{(2)}_{i2}, \ldots, G^{(n)}_{in}\} \) from the Clifford group to obtain the probabilities \( f_{m}^{(1)}, f_{m}^{(2)}, \ldots, f_{m}^{(n)} \). Compute the average \( \bar{f}_m = 1/n \sum_{i=1}^{n} f_{m}^{(i)} \).

6. Examine the behaviour of the ASF \( \bar{f}_m \) 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 analysing 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

\[
\bar{f}_m = Ap^m + B, \tag{1}
\]

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 (SPAM) errors [5]. This implies that having performed an 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 to the gate fidelity of the noise with respect to the identity [43], and hence the labelling of \( \bar{f}_m \) 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 modelling each noisy gate as $\hat{G}_t = \Lambda \circ \hat{G}_t$, for some completely positive trace preserving (CPTP) map $\Lambda$; then the analytical average of the survival probabilities is given by the average over gates $\hat{G}_t$. 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 [44], such as the Clifford group. This implies that averaging over gates can be replaced with that over the Haar measure to obtain $T_m$, and similarly the use of higher unitary designs could serve to characterise higher-order statistical properties of noise in RB [45]. 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. RB has been studied within a time-independence assumption and a weak gate dependence, both generally render a linear combination of exponential decays for the noise is given by $\rho = \Lambda_1 \circ \hat{G}_1 \circ \ldots \circ \Lambda_{m+1} \circ \hat{G}_{m+1} \circ \rho_\text{anc}$. Here we are interested in temporal correlations 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, labelled $SE$, composed of a $d_S$-dimensional system $S$ and a $d_E$-dimensional environment $E$. An experimenter, in principle, would apply the sequence $S_m = \otimes_{i=1}^{m+1} G_i$ of Clifford gates $\hat{G}_t$ solely on $S$ (they may or may not have access to $E$). We consider different scenarios for the initial state on $S$, which can be correlated with $E$. We now can model the noisy gates as $\hat{G}_t = \Lambda_t \circ (I_E \otimes \hat{G}_t)$, where $\Lambda_t$ acts on the full $SE$ system and $I_E$ is an identity map on $E$.

Such a sequence 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, 46] 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, say $\Upsilon_m$, and contract it with the set of Clifford gates, which we can also incorporate in a tensor, say $\xi_m$. This can be depicted as in the circuit of Fig. 1.

These tensors, $\Upsilon_m$ and $\xi_m$, just as any quantum map, can have different representations [47]. Here we will employ the Choi-state representation, as detailed in Appendix A, which simply is a generalisation of the Choi-Jamiolkowski isomorphism for quantum channels [48]. The tensor Choi-state for the noise is given by

$$\Upsilon_m = \text{tr}_E \left\{ \left( \otimes_{i=1}^{m+1} (\Lambda_i \otimes I_\text{anc}) \circ \mathcal{I}_t \right) \rho \otimes \psi^{\otimes m+1} \right\},$$

where $I_\text{anc}$ is an identity map on an ancillary space $\text{anc} = A_1 B_1 \ldots A_{m+1} B_{m+1} \simeq S^{2(m+1)}$ composed of $m + 1$ pairs of $S$ systems, $\mathcal{I}_t$ is a swap gate between $S$ and one of these pairs in the $i$th ancillary space, say $A_i$, and $\rho \otimes \psi^{\otimes m+1} = \sum_i \otimes \{ ii \otimes jj \}$ is an unnormalized maximally entangled state. The tensor state for the gate sequence, on the other hand, is simply $\xi_m$.

$$\xi_m = I_S \otimes \otimes_{i=1}^{m+1} (I_A \otimes \mathcal{G}_i) \rho \otimes \psi^{\otimes m+1}.$$  

The notion of Markovianity is formalised 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, 50]. 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 $\xi_m$ in Eq. (3). That is, temporal correlations in the process tensor correspond to spatial correlations in the Choi-state representation, and more precisely then, a process tensor $\Upsilon_m^{\text{MB}}$ will be Markovian if and only if there are noise maps $\Lambda_t^{\text{MB}}$ acting solely on $S$ such that

$$\Upsilon_m^{\text{MB}} = \rho_S \otimes \otimes_{i=1}^{m+1} \left( \Lambda_i^{\text{MB}} \otimes I_B \right) \psi^{\otimes m+1}.$$  

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

$$N := \min_{\Upsilon_m^{\text{MB}}} D(\Upsilon_m, \Upsilon_m^{\text{MB}}).$$
where the choice of such distance measure is rather a matter of practicality, as the minimisation over Markovian processes will often make the computation of $N$ 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, $\gamma_m^{(j)} = \rho_S \otimes \mu_{m+1}^{(j)} \tr_{E_{j-1}}[\gamma^m]$, where $\tr_{E_{j-1}}$ means trace over all except between steps $i$ to $j$, or otherwise placing relevant bounds on $N$ for Schatten-norm measures, as done in Ref. [51, 52] to study some statistical properties of non-Markovian processes. Here, we will care about quantifying how non-Markovian an 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 $f_m = \tr[M\Sigma_m(\rho)]$ in a run with $m$ noisy gates in terms of the process tensor with

$$S_m(\rho) = \tr_E \left\{ \otimes_{i=1}^{m+1} \Lambda_i \circ (I_S \otimes \mathcal{G}_i) \rho \right\}$$

$$= \tr_E \left( \gamma_m^{(E)} \right),$$

where $\tr_E$ means a partial trace over all intermediate input and output systems except the final $S$ and $T$ denotes a transpose. Computing the ASF, $F_m$, then amounts to computing the average $E \left[ \otimes_{i=1}^{m+1} (I_S \otimes \mathcal{G}_i)^{\otimes m+1} \right] \phi^{\otimes m+1}$. This is a big simplification allowing 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 Section II, we can replace averaging over Clifford gates with averaging over the unitary group with the uniform Haar measure. To finally obtain $F_m$, we have to contract the average gate sequence tensor with the noise tensor $\gamma_m$, which will contain the noise inherent to the RB sequence, and in particular can be labelled as non-Markovian if the individual noise is correlated between time-steps or Markovian otherwise.

We now present a general expression for the ASF $F_m$ 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 an 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 (ASF). This yields the following:

**Theorem.** Let $\rho$ be the initial state of a system-environment, $SE$, composite with $d_SD_E = \dim(SE)$. Let $S_m(\rho)$ describe a randomised benchmarking (RB) sequence of length $m$ over Clifford gates with the $CP$ map $\Lambda_n$ acting on $SE$ being the associated noise at the $n^{th}$ time-step. Then, the average sequence fidelity (ASF) $F_m$ with a POVM element $M$ is given by

$$F_m = \tr[M \Sigma_m(\rho)]$$

$$= \tr[ M \tr_E \circ \Lambda_{m+1} \circ (\mathcal{A}_m + \mathcal{B}_m) \rho],$$

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

$$\mathcal{A}_m(\rho) := \sum_{n=1}^{m} \langle s | \Lambda_n(\epsilon) \otimes s | s' \rangle \langle s' | \rangle,$$

$$\mathcal{B}_m(\rho) := \tr_E \circ \Lambda_{m+1}(\rho) \otimes \mathbb{1},$$

with $\rho_E := \tr_S(\rho)$ being the reduced initial state in $E$; here $\Lambda_n, \Theta_n$ are maps acting solely on $E$ as defined by

$$\Lambda_n(\epsilon) := \sum_{s,s'} \langle s | \Lambda_n(\epsilon) \otimes s | s' \rangle \langle s' | \rangle,$$

$$\Theta_n(\epsilon) := \tr_E \left[ \Lambda_n(\epsilon \otimes \mathbb{1} \otimes \frac{1}{d_S}) \right],$$

for any operator $\epsilon$ 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 $F_m = \tr[ M \tr_E \circ \Lambda_{m+1} \circ (\mathcal{A}_m + \mathcal{B}_m) \rho]$, where the average $E \left[ \langle \mathcal{T}_m \rangle \right]$ 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.

Crucially, we should first notice that in the noiseless limit, $\Lambda_1 = \Lambda_2 = \ldots = \Lambda_n = I$, we recover $F_m \rightarrow \tr[M\rho_S]$, so that indeed Eq. (7) is bounded by one. For the ideal case of $SE$ being a closed system, each $\Lambda_n$ 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 $\mathcal{A}_m$ and $\mathcal{B}_m$ in Eqs. (8) and (9), resp., where the depolarising effect of the noise on $S$ is manifest, with $\mathcal{A}_m$ being partially depolarising in $S$ and $\mathcal{B}_m$ completely depolarising in $S$. The action of $\mathcal{B}_m$, 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 $\mathcal{A}_m$ and $\mathcal{B}_m$ give a product state, with $E$ carrying all the noise factors. Finally upon applying $\tr_E \circ \Lambda_{m+1} \circ \mathcal{A}$, this would render a factor analogous to a product of noise-strengths $p_1p_2\cdots p_m$.

The notation we use for $\mathcal{A}$ and $\mathcal{B}$, which here are quantum maps, is suggestive in that these reduce to the corresponding $AP^n$ and $B$, resp., in the static Markovian limit. In a Markovian scenario the environment is superfluous and we would have $\Lambda_n \rightarrow I_E \otimes \Lambda_n^{(m)}$ together with $\rho \rightarrow \rho_E \otimes \rho_S$, 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],

$$F_{m}^{(M)} = p_1 \cdots p_m A + B,$$

where,

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

$$A = \text{tr} \left[ \Lambda_{m+1}^{(M)} \left( \rho_S - \frac{1}{d_S} \right) \right],$$

$$B = \text{tr} \left[ \Lambda_{m+1}^{(M)} \left( \frac{1}{d_S} \right) \right].$$

That is, we get $\mathcal{A}_n(\rho) \rightarrow p_1 \cdots p_m (\rho - \frac{1}{d_S})$ and $\mathcal{B}_n(\rho) \rightarrow \frac{1}{d_S}$ in this limit, which makes it clear that $\mathcal{B}$ renders only SPAM and non-Markovian noise contributions. Here $\text{tr} \left[ \Lambda_{m}^{(M)} \right] = \sum_i \text{tr} \left( \Lambda_{m}^{(M)} \right)_i$ where $\Lambda_{m}^{(M)}$ are the Kraus operators of $\Lambda_n$. Furthermore, despite being complicated in the general case, the map $S_{\Lambda_n}^{(M)}$ simply picks up a noise multiplicative factor, $S_{\Lambda_n}^{(M)}(\varepsilon) = \text{tr} \left[ \Lambda_{m}^{(M)} \right] \varepsilon$ 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 [54, 55]. These could be particularly relevant in a non-Markovian RB experiment because the averaging over $S$ gates only depolarises 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 $F_m \rightarrow \text{tr} \left[ \sum_{\alpha} \left( \Lambda_{\alpha}^{(M)} \right) \right]$, 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 $F_m$, as we exemplify numerically in Appendix G.

Finally, non-exponential 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 $C_{\Lambda_n}^{(M)} \left( \Lambda_{\alpha} - \Theta_{\Lambda_n} \right) = (S_{\Lambda} - \Theta_{\Lambda})^{(M)}$, which will generally not render an exponential decay in the ASF. It is important to point out that while non-Markovianity generally leads to non-exponential decays, there can also be other contextual factors [56], such as gate-dependence or other rather arbitrary external time-dependence leading to such behaviour.

V. QUANTIFYING NON-MARKOVIANITY IN RANDOMISED 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 [57, 58]. As examples of classical correlations, in Appendix F we reproduce the ASF of the model in Ref. [30], where classical temporal correlations are modelled via dephasing noise determined by a classical stochastic process; this effectively renders an ASF analogous to one that is Markovian time-dependent or static with the noise parameter being a random variable. We also illustrate this via a shallow pocket model [34, 59–61], 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 $N$ 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.

An 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 static non-Markovian processes leading to exponential or almost exponential behaviour, although as mentioned above, in general a static noise does not lead to an exponential behaviour unless the environment is superfluous. In Appendix G we exemplify this numerically with a spin interaction as source of non-Markovian noise. If in general there exists a whole class of non-Markovian processes can be classified as RB blind, it 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 non-Markovianity $N$ with Eq. (5), we really only care about how much its associated ASF for the Clifford group deviates from a Markovian one. Ideally, we would look for a direct Markovian counterpart of the original non-Markovian noise process that we have. This means we would look at deviations from the ASF generated by the Markovianised process $\Upsilon^{(M)}$, where each noise map in the original non-Markovian noise $\Lambda_n$ 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 \( \Lambda_m^{(\text{SM})} \) on system \( S \) acting as \( \Lambda_m^{(\text{SM})}(\rho) = \text{tr}_E \circ \Lambda_m(\rho \otimes \sigma) \) for an arbitrary pure state \( \sigma \). This is depicted in Fig. 2.

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

\[
N_{q_m}^{\text{SM}} := \left\| \mathcal{F}_m - \mathcal{F}_m^{(\text{SM})} \right\|_q = \left( \sum_{n=1}^{m} \left| \text{tr} \left[ M \text{tr}_E \left[ (\mathcal{Y}_n - \mathcal{Y}_n^{(\text{SM})}) (E_1^n) \right] \right] \right|^q \right)^{1/q},
\]

where \( \mathcal{F}_m^{(\text{SM})} \) is the ASF of the Markovian noise process associated to \( \mathcal{Y}_m \), given by \( \mathcal{Y}_m^{(\text{SM})} := \rho_S \otimes \left( \otimes_{i=1}^{m+1} \left( \Lambda_i^{(\text{SM})} \otimes I_B \right) \right) \rho_S^{m+1} \).

where \( \mathcal{Y}_m \) is the ASF of the Markovian noise process associated to \( \gamma_m \), given by \( \gamma_m := \rho_S \otimes \left( \otimes_{i=1}^{m+1} \left( \Lambda_i^{(\text{SM})} \otimes I_B \right) \right) \rho_S^{m+1} \), where

\[
\Lambda_m^{(\text{SM})(\sigma)} := \text{tr}_E \circ \Lambda_m(\varepsilon \otimes \sigma),
\]

for any \( \sigma \) acting on \( S \) and an arbitrary pure state \( \varepsilon \) on \( E \).

The measure \( N_{q_m}^{\text{SM}} \) boils down to how well the POVM element \( M \) can tell \( \rho_S \circ \Lambda_m(\rho) \) from \( \rho_S \circ \Lambda_m(\rho_S \otimes \mathbb{I}/d_S) \), as well as \( \rho_S \circ \Lambda_m(\rho) \) from \( \Lambda_m^{(\text{SM})}(\rho_S \otimes \mathbb{I}/d_S) \) for CPTP noise. Generic bounds can potentially become possible with this non-Markovian state. One could thus know the expression for the non-Markovianity. Of course, the RB non-Markovianity measure in Eq. (15) also already makes it manifest that if an underlying noise process in an RB sequence is Markovian, then \( N_{q_m}^{\text{SM}} = 0 \). The converse, however, might not necessarily be true or deviations could be negligible in practice.

There could be several scenarios where Eq. (15) could be computed or estimated. One might be either the full Markov process \( \gamma_m \), or just an error rate is known, but once the RB experiment is run, deviations from \( \gamma_m \) are observed which most plausibly could be explained by non-Markovianity. This means we could actually compute \( N_{q_m}^{\text{SM}} \) directly from the experimental data and e.g. analyse 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 \( \gamma_m \), and thus know the expression for the non-Markovian ASF \( \mathcal{F}_m \) in Eq. (17). Then we may construct the Markovian counterpart \( \nu_m^{(\text{SM})} = p_1 \cdots p_m A + B \) of the ASF, compute \( N_{q_m}^{\text{SM}} \) in Eq. (15) and compare with the actual RB data.

Perhaps the most common case, however, will be that an RB experiment is run without a-priori knowledge of a model for the noise and a non-exponential 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 an RB experiment, as well as to operationally construct an analogue of a Markovianised ASF, \( \mathcal{F}_m^{(\text{SM})} \), 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 [46, 62, 63], 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 \( \rho \) be an initial state on a system-environment, SE, composite and let \( S_m^{(\text{SM})}(\rho) \) describe an RB sequence of length \( m \) with noise described by CP maps \( \Lambda_n \) on SE for all \( n \) up to a sequence length \( \ell < m \), with the rest of the sequence having noise CPTP maps \( \Lambda_n^{(\text{SM})} \) on S and associated noise-strengths \( p_n \). Then the average sequence fidelity (ASF) upon acting with a POVM element \( M \) is given by

\[
\mathcal{F}_m = \text{tr} \left[ ME[S_m^{(\text{SM})}(\rho)] \right] = p_{\ell+1} \cdots p_m \text{tr} \left[ \Lambda_m^{(\text{SM})} \circ \text{tr}_E(\mathcal{F}_m) \right] + B \text{tr} \left[ \mathcal{F}_m(\rho) \right],
\]

where \( B = \text{tr} \left[ \Lambda_m^{(\text{SM})}(\mathbb{I}/d_S) \right] \) and \( \mathcal{A}_n, \mathcal{B}_n \) are defined in Eq. (8) and Eq. (9), resp.

This implies that after a sequence length \( \ell \), non-Markovian noise will be manifest in an RB experiment as SPAM errors and not affect the subsequent decay, which for static 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 \( \ell \).

The main reason why this is important is twofold: first, detecting non-Markovian effects with an RB experiment will be 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 \( \ell \) 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 \( \ell \) inherent in the noise process, can then potentially be determined through an RB experiment. This would also be related to determining the order of a finite quantum Markov order process [64].

In section VIII we show one such example where the sequence length \( \ell \) of non-Markovian noise can be estimated from an RB experiment’s data alone, and where a sensible static Markovianised ASF, \( \mathcal{F}_m^{(\text{SM})} \), 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 \( \ell \), described by CPTP maps \( \Lambda_n \) and an initial uncorrelated state, by choosing to fix \( \ell + 1 \) Clifford
Determind time-scales of finite non-Markovian noise. A noise process with initial finite non-Markovian noise over a sequence length \( \ell \approx 5 \) will decay as described by a Markovian ASF after such step, with the non-Markovian part contributing as SPAM error factors. By fixing to identify the gates of at least time-steps step 2 and 3, the decay of the ASF corresponding to such sequence becomes entirely Markovian; this allows 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 Section VIII.

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

Corollary 2 (Late non-Markovian noise). Let \( \rho \) be an initial state on a system-environment, SE, composite and let \( S_{m;\ell+1}^{n}(\rho) \) describe an RB sequence of length \( m \) with noise given by CP maps \( \Lambda_{n} \) on SE for all \( n \) up to a sequence length \( \ell < m \), then at the \( \ell \)th step by \( \Lambda_{\ell}(\cdot) \to \varepsilon \otimes \text{tr}_{E} \circ \Lambda_{\ell}(\cdot) \) for some \( E \) state \( \varepsilon \), and with the rest of the sequence having noise CP maps \( \Lambda_{n} \) on SE. Then the average sequence fidelity (ASF) upon acting with a POM element \( M \) is given by

\[
\mathcal{F}_{m} = \text{tr} \left[ M \left[ S_{m;\ell+1}^{n}(\rho) \right] \right] = \text{tr} \left[ M \circ \Lambda_{m+1} \circ \mathcal{A}_{m;\ell+1}(\rho) \right] + \text{tr} \left[ M \circ \Lambda_{\ell} \circ \mathcal{B}_{m;\ell+1}(\rho) \right],
\]

where

\[
\mathcal{A}_{m;\ell}(\rho) := \frac{\sum_{n=1}^{m} (\Lambda_{n} - \Theta_{\Lambda_{n}}) \otimes I_{S}}{d_{S}^{2} - 1} \left( \rho - \rho_{E} \otimes \frac{1}{d_{S}} \right),
\]

\[
\mathcal{B}_{m;\ell}(\rho) := \sum_{n=m+1}^{\infty} \Theta_{\Lambda_{n}}(\rho_{E}) \otimes \frac{1}{d_{S}},
\]

with \( \Theta_{\Lambda_{n}} \) defined in Eq. (10) and Eq. (11), resp.

This case might be relevant in practice whenever the sequence length \( \ell \) 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 an RB experiment.

Furthermore, in the middle of this 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

Corollary 3 (Blocks of finite non-Markovian noise). Let \( \rho \) be an initial state on a system-environment, SE, composite and let \( S_{m;\ell+1;\ell}^{n}(\rho) \) describe an RB sequence of length \( m \) with noise given by CP maps \( \Lambda_{n} \) on SE for all \( n \) up to a sequence length \( \ell < m \), then at the \( \ell \)th step by \( \Lambda_{\ell}(\cdot) \to \varepsilon \otimes \text{tr}_{E} \circ \Lambda_{\ell}(\cdot) \) for some \( E \) state \( \varepsilon \), and with the rest of the sequence having noise CP maps \( \Lambda_{n} \) on SE. Then the average sequence fidelity (ASF) upon acting with a POM element \( M \) is given by

\[
\mathcal{F}_{m} = \text{tr} \left[ M \left[ S_{m;\ell+1;\ell}^{n}(\rho) \right] \right] = \text{tr} \left[ M \circ \Lambda_{m+1} \circ \mathcal{A}_{m;\ell} \left( \varepsilon \otimes \frac{1}{d_{S}} \right) \right] + \text{tr} \left[ M \circ \Lambda_{n+1} \circ \mathcal{B}_{m;\ell+1}(\rho) \right],
\]

with \( \mathcal{A}_{m;\ell} \) and \( \mathcal{B}_{m;\ell} \) 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 Corollary 1, 2, 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 \( \mathcal{S} \) subject to static unitary noise \( \Lambda(\cdot) = \Lambda(\cdot), \) where \( \lambda = \)}
FIG. 4. Average sequence fidelity (ASF) for static unitary non-Markovian noise and deviations from its Markovianised counterpart. We consider the noise model described by the two-spin interaction of Eq. (24) with \( J = 1.7 \), \( h_x = 1.47 \) and \( h_y = -1.05 \), for a single qubit as system \( S \). We take \( \rho = |00⟩⟨00| \) and \( M = |0⟩⟨0| \).

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 Markovianised process with static noise \( \Lambda_0^\text{SM}(\cdot) = \text{tr}_E \circ \Lambda(\varepsilon \otimes \cdot) \), here with \( \varepsilon = |00⟩⟨00| \). Bottom: Deviations of the ASF by both the analytical data (continuous red line) and numerical data produced by Eq. (7), from the Markovianised ASF, \( F^\text{SM}_m \) (dashed blue line).

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 \( J = 1.7 \), \( h_x = 1.47 \) 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 Markovianised process. Top: the initial state \( \rho = |00⟩⟨00| \) is affected by the sequence noise \( \Lambda \sim \exp(-i\Delta_1 H) \) for a small \( \Delta_1 \approx 0.04 \) and \( M = |0⟩⟨0| \) is slightly rotated via \( \Lambda \sim \exp(-i\Delta_2 Y) \) with a small \( \Delta_2 \approx 0.09 \). Bottom: \( \Delta_1 \approx 0.29 \) and \( \Delta_2 \approx 0.10 \) are increased considerably, amounting to large SPAM errors. In all cases the sample size is 100.

\[
\exp(-i\delta H), \text{ due to interaction with another qubit, identified as } E, \text{ where } H \text{ given by the two-spin interaction}
\]
\[
H = J X_1X_2 + h_x(X_1 + X_2) + h_y(Y_1 + Y_2), \tag{24}
\]
with \( X_i, Y_i \) being Pauli matrices acting on the \( i \)-th 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 [65].

We take \( J = 1.7 \), \( h_x = 1.47 \) and \( h_y = -1.05 \) arbitrarily, for which we compute the ASF \( F^\text{SM}_m \) as a function of \( m \), both by numerical averaging and employing Eq. (7) with \( \delta = 0.03 \). We take \( \rho = |00⟩⟨00| \) and \( M = |0⟩⟨0| \) and ignore SPAM errors. We display the results in Fig. 4 together with its Markovianised ASF, \( F^\text{SM}_m \), whereby the static noise is modelled as a CP map given by \( \Lambda_0^\text{SM}(\cdot) = \text{tr}_E \circ \Lambda(\varepsilon \otimes \cdot) \); 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 non-exponential. 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 \( m \leq 50 \). 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, \( N^\text{SM}_m \) with respect to the Markovian counterpart can also be swiftly computed through the sum of absolute values of the differences between \( F^\text{SM}_m \) and
In particular in Fig. 4 the RB non-Markovianity is not particularly high (between $\mathcal{N}_m^{-100} \approx 2.1$ and $\mathcal{N}_m^{-100} \approx 0.04$) 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 $\rho$ is affected by the same $\Lambda$ error for some small $\delta = \Delta_t$, and that $\mathcal{M}$ is slightly rotated via $\exp(-i\Delta_2 Y)$ for a small $\Delta_2$. If Fig. 5 we show examples for both mild, $\Delta_1 = 0.04232$ and $\Delta_2 = 0.09321$, and much stronger noise with, $\Delta_1 = 0.2932$ and $\Delta_2 = 0.10321$.

In Appendix G, we also show the case where the preparation affects only $S$ by some rotation $\exp(-iyX)$ with a small $y$, 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 Section 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 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 [66, 67]. 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, $\mathcal{N}_q^{-1} \approx 0$, and the behaviour 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 analyse a plausible realistic scenario for an 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, MARKOVIANISED AVERAGE SEQUENCE FIDELITY AND COHERENT NOISE**

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

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

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

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 $\ell$ is. Given Corollary 1, however, we know that whenever we have finite static 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 static noise (i.e. with almost equal

![FIG. 6. Determining the sequence length of finite non-Markovian noise. Top: An 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 $\mathcal{F}_{m[i,...,j]}$, where $[i,...,j]$ denotes steps taken to identity (shown joined in the plot for clarity), fitting exponentials to these, and approximately matching their decay rates $p_{m[i,...,j]}$ with the one in the manifestly exponential part in the original data. Bottom: The non-Markovian noise sequence length was determined to be $\ell = 9$; the dot-dash purple line denotes the curve with the decay rate $p_{m[1,...,9]}$ and constants $A, B$ of the fitted exponential of the original data starting at $m = 9$. Once $\ell$ is determined, a sensible Markovianised ASF, shown as a dashed blue line, can be taken with $p = p_{m[1,...,9]}$ and reasonable criteria for fixing $A, B$; here we choose $A = B$ assuming low SPAM errors. The analytical Markovianised static ASF of the form we proposed in Section V is shown with orange dots just as a comparison.](image-url)
noise-strengths $p_{t+1} \approx \ldots \approx p_m$) after such sequence length $\ell$. 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 to identify a Markovianised static 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 non-exponential; 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, $F_m$, both analytical and numerical, for a finite noise memory process with noise modelled by Eq. (25), again taking $\rho = |00 \rangle \langle 00|$ and $M = |00 \rangle \langle 00|$. We also display numerical ASFs, denoted $F_m^{(1)}, \ldots, F_m^{(9)}$, with fixed identities at sequence lengths $i, \ldots, j$. The corresponding ASFs $F_m^{(1)}, \ldots, F_m^{(9)}$ will normally be decreasing as $F_m > F_{m+1} > \cdots > F_{m+10}$ 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 $p_m^{(1)}, \ldots, p_m^{(9)}$ of one of these sequences $F_m^{(1)}, \ldots, F_m^{(9)}$ with the corresponding one of the manifestly Markovian part in the full sequence. Once the decay rate has been determined, a sensible static Markovianised ASF, $\hat{F}_m^{(1)}, \ldots, \hat{F}_m^{(9)}$, 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 was determined to be $\ell \approx 9$ by approximately matching $p_m^{(1)}, \ldots, p_m^{(9)}$ with the corresponding one for the exponential fit between sequence lengths $15 \leq m \leq 30$ 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 Markovianised ASF with $\hat{F}_m = A p_m^{(1)}, \ldots, p_m^{(9)} + B$ with $A \approx B$ supposing SPAM errors to be small; we compared this with a Markovianised construction as proposed in Section V, with static noise given throughout by $A^{\text{SM}}$.

This practical approach can work reasonably well, as we show in this example, and allow 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 static Markovianised 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 $F_m^{(1)}, \ldots, F_m^{(9)}$ 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.

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 Markovianised ASF, $\hat{F}_m$, 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 characterised, e.g., with via unitarity measures [8, 24, 68, 69] or otherwise, in principle it could be addressed and calibrated if we have access to the SE qubits. Precisely then, we may tell if the noise is unitary over the whole SE if we get a general non-exponential behaviour 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.
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 non-exponential behaviour, 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 exemplified 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.

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 manuscript 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, and/or to study context-dependent errors, with one possible way being the one we propose in Fig. 8. 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-Markovian noise will display non-Markovian effects in state preparation and measurement (SPAM) 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 for our main result makes the depolarising 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

IX. CONCLUSIONS AND DISCUSSION

We have, i) derived a general analytical expression for the average sequence fidelity (ASF) of a randomised benchmarking (RB) 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 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 manuscript we also discuss the effect of state preparation and measurement (SPAM) errors in RB with non-Markovian noise, as well as the case of classical correlations, which we can argue be treated as a Markovian time-dependent problem, and more generally the idea of RB blindness to a subclass of non-Markovian noise processes.

ACKNOWLEDGMENTS

We thank Rob J. Harris and Felix A. Pollock for conversations. KM is supported through Australian Research Council Future Fellowship FT160100073 and Discovery Project grant DP210100597.

[1] J. Emerson, R. Alicki, and K. Życzkowski, J. Opt. B-Quantum S.O. 7, S347 (2005).
[2] B. Lévi, C. C. López, J. Emerson, and D. G. Cory, 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, Phys. Rev. A 77, 012307 (2008).
[4] E. Magesan, J. M. Gambetta, and J. Emerson, Phys. Rev. Lett. 106, 180504 (2011).
[5] E. Magesan, J. M. Gambetta, and J. Emerson, 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 [quant-ph].
[7] I. L. Chuang and M. A. Nielsen, J. Mod. Optic 44, 2455–2467 (1997).
[8] I. Roth, R. Kueng, S. Kimmel, Y.-K. Liu, D. Gross, J. Eisert, and M. Kliesch, Phys. Rev. Lett. 121, 170502 (2018).
[9] J. J. Wallman and S. T. Flammia, New J. Phys. 16, 103032 (2014).
[10] M. Mohseni, A. T. Rezakhani, and D. A. Lidar, 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 [quant-ph].
[12] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert,
Appendix A: The process tensor of the noise and gate sequences

The process tensor is a multi-linear map taking completely positive (CP) maps as input and giving a single quantum state as output. The operational scenario is the following: an initial quantum state $\rho$ on the joint $SE$ composite is acted on with an operation $G_i$ solely on system $S$, which in general is given by a CP map: subsequently the whole composite evolves unitarily through a unitary map $U_i$, after which an operation $G_k$ is performed on $S$, then the whole evolves unitarily under a unitary map $U_k$, and so on, until an intervention $G_k$, followed finally by a unitary map $U_k$. This means the final state in system $S$ will be given by

$$\rho_S^{(k)} = \text{tr}_E \left[ \bigotimes_{i=1}^k U_i \circ G_i \right] \rho,$$

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

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

where $\mathcal{G}_{k:1} = (G_1, G_2, \ldots, G_k)$. Such operations $G_i$ are said to form an intervention and belong to an instrument, which can be understood as a generalisation of a POVM, and the particular outcomes of each intervention yield a joint probability distribution describing a stochastic process.

The generalisation of a Choi-state, as given by the Choi-Jamiołkowski isomorphism [48, 70], for a $k$-step process tensor follows by introducing $k$ maximally entangled states $\psi_{A_iB_i} \in \mathcal{B} (\mathcal{H}_A \otimes \mathcal{H}_B)$, where $\mathcal{H}_A \equiv \mathcal{H}_S$ and similarly for $B$, and letting half of each act as an input at every step by swapping the input spaces with the corresponding ancilla. This is more clearly illustrated in Fig. 9.

![Fig. 9. The Choi-state representation of a $k$-step process tensor.](image)

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

$$\mathcal{T}_{k:1} = \text{tr}_E \left[ \bigotimes_{i=1}^k U_i \circ \mathcal{J}_i \right] \rho \otimes \psi^{\otimes k},$$

where here we are implicitly writing $U_i$ for $U_i \otimes I_{A_1B_1\cdots A_{k-1}B_{k-1}}$, and $\psi^{\otimes k} = \psi_{A_1B_1} \otimes \cdots \otimes \psi_{A_kB_k}$. The generalised swap $\mathcal{J}_i$ between system $S$ and an ancilla $A_i$ at time-step $i$ is defined by $\mathcal{J}_i(\cdot) := \mathcal{G}_i(\cdot) \mathcal{G}_i$, where

$$\mathcal{G}_i := \sum_{j=1}^d I_E \otimes |f \rangle \langle f| \otimes I_{A_1B_1\cdots A_{i-1}B_{i-1}} \otimes | j \rangle \langle j| \otimes I_{B_{i-1}A_iB_{i+1} \cdots A_kB_k},$$

where $d$ is the dimension of $A_i$ and $d_E$ is the dimension of $E$.
Similar to the case of quantum channels, the isomorphism between the action of the process tensor and its Choi representation is manifest [71] through the relationship

\[ \mathcal{T}_k \left[ \tilde{G}_{k:1} \right] = \text{tr}_\Sigma \left[ \Upsilon_{k:1} \left( I_S \otimes \mathcal{Y}^T_{k:1} \right) \right], \tag{A5} \]

where here

\[ \mathcal{Y}_{k:1} = \left( \bigotimes_{i=1}^k I_{A_i} \otimes \tilde{G}_i \right) \psi^{\otimes k}, \tag{A6} \]

is the Choi-state for the operations \( \tilde{G}_{k:1} \), the notation \( \text{tr}_\Sigma \) stands for partial trace over all except output in \( S \), and \( T \) denotes a transpose.

The tensor \( \mathcal{Y}_{k:1} \) 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, \( \Upsilon_{k:1}^{\text{Markov}} = \bigotimes_{i=1}^k \Phi_{i:i-1} \otimes \rho_s, \) the Choi-states \( \Phi_{i:j} \) can either correspond to a closed system dynamics between steps \( i \) and \( j \), or in general to a \( \mathcal{CP} \) 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 Figure 10. The order of the spaces will be relevant whenever two tensors are contracted and can be written generally through swaps with ancillas and half maximally entangled states.

In this manuscript we describe a noise RB sequence as a process tensor with the dynamics being described by the noise at each step, \( U_i \rightarrow \Lambda_i \).

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, \( E \left( \psi^T_m \right) \).

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 \( \mu \), over the \( d_S \)-dimensional unitary group, which we denote \( U(d_S) \). The Haar measure is the unique measure on \( U(d) \) satisfying invariance under left and right multiplication, i.e. it is invariant under arbitrary rotations. Specifically, given a subset \( V \subseteq U(d) \), we have \( \mu(V) = \int_W d\mu(U) \) for the Haar measure, \( \mu \), with the (left-right invariance) property

\[ \mu(W) = \mu(VW) = \int_W d\mu(VU) = \int_W d\mu(UV) = \mu(WV), \tag{B1} \]

for any fixed \( V \in U(d) \). For any quantity \( f \) depending on a unitary \( U \in U(d) \), we denote integration over such unitary by \( E[f(U)] \) and refer to it as the Haar or uniform average of \( f \).
Let the action of the unitary maps $G_i$ be given by $G_i(\cdot) = G_i(\cdot) G_i^\dagger$, then let us compute

$$
E \left[ \mathcal{C}_m \right] = \mathbb{I}_S \otimes \sum_{\ell = 1}^{d_S} \int_{\ell, k \in \mathbb{C}(\ell_k)} \left( |\ell_i k_i| \otimes G_i^\dagger |\ell_j k_j| G_j \right) d\mu(G_1) \cdots d\mu(G_m),
$$

(B2)

where crucially, $G_{m+1} = G_i^\dagger G_i^\dagger \cdots G_1^\dagger$. 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 the Weingarten function, which here takes the values

$$
W_g[(1), (2), d] = \frac{1}{d^2 - 1}, \quad W_g[(1), (1), d] = -\frac{1}{d(d^2 - 1)},
$$

on the two possible permutations $\tau\sigma^{-1} \in S_2$.

Then we can let $G_{\ell} = \sum G_i |\ell_i \rangle \langle \ell_i| \rho_i |\ell_i \rangle \langle \ell_i| \rho_i^{-1}$ for each $G_{\ell}$ and employ the 2-moment above; let us take the integral over $G_1$ first,

$$
E \left[ \mathcal{C}_m \right] = \mathbb{I}_S \otimes \sum_{\ell = 1}^{d_S} \sum_{|\ell\rangle \in \mathbb{C}(\ell_2)} W_g(\tau\sigma^{-1}, d_S) |\ell_1 \rangle \langle \ell_1| \tau_{\ell_1(1)}(\sigma_{\ell_1(1)}(1)) |\ell_2 \rangle \langle \ell_2| \tau_{\ell_2(2)}(\sigma_{\ell_2(2)}(2)) \otimes \mathbb{I} \cdots \otimes \mathbb{I} dG_1 \cdots dG_m d\mu(G_2) \cdots d\mu(G_m),
$$

(B5)

then we can do similarly with all remaining unitaries by also labelling each permutation mapping $\sigma_\ell$ and $\tau_\ell$ for the corresponding integral over each $G_\ell$, i.e.

$$
E \left[ \mathcal{C}_m \right] = \mathbb{I}_S \otimes \sum_{\ell = 1}^{d_S} \sum_{|\ell\rangle \in \mathbb{C}(\ell_2)} W_g(\tau\sigma^{-1}, d_S) |\ell_1 \rangle \langle \ell_1| \tau_{\ell_1(1)}(\sigma_{\ell_1(1)}(1)) |\ell_2 \rangle \langle \ell_2| \tau_{\ell_2(2)}(\sigma_{\ell_2(2)}(2)) \otimes \mathbb{I} \cdots \otimes \mathbb{I} dG_1 \cdots dG_m d\mu(G_2) \cdots d\mu(G_m),
$$

(B6)

where for easiness of notation we dropped the primes and denoted $W_g(\tau\sigma^{-1}, d_S)$, where a sum is implicit over basis vectors and permutations on $S_2$, and where $\delta(a, b)$ stands for the usual Kronecker $\delta_{ab}$. We finally notice that $E \left[ \mathcal{C}_m \right] = E \left[ \mathcal{C}_m^\dagger \right]$.

**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

$$
E \left[ \mathcal{C}_m^\dagger \right] = \mathbb{I}_S \otimes \sum_{|\ell\rangle \in \mathbb{C}(\ell_2)} W_g |\ell_1 \rangle \langle \ell_1| \tau_{\ell_1(1)}(\sigma_{\ell_1(1)}(1)) |\ell_2 \rangle \langle \ell_2| \tau_{\ell_2(2)}(\sigma_{\ell_2(2)}(2)) \otimes \mathbb{I} \cdots \otimes \mathbb{I},
$$

(C1)

where $W_g$ implicitly depends on $\tau\sigma^{-1}$, with each $\tau$ and $\sigma$ being summed over the symmetric group $S_2$, and with an implicit sum over each $\mu$ and $v$.

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

$$
\gamma_1^M = (\Lambda_{1}^M \otimes I) \circ \mathcal{J}_{\sigma} \circ (\Lambda_{1}^M \otimes I) \circ \mathcal{J}_{\rho} \left( \rho \otimes \rho_{\otimes 2} \right)
$$

$$
= \sum_{|\ell_i\rangle} |\ell_i \rangle \langle \ell_i| \kappa_{\sigma}(\ell_i) \otimes 1 \kappa_{\rho}(\rho_{\otimes 2}) \kappa_{\sigma}(\ell_i) \otimes 1 \kappa_{\rho}(\rho_{\otimes 2})
$$

$$
= \sum_{|\ell_i\rangle} |\ell_i \rangle \langle \ell_i| \alpha_1 \beta_\sigma \beta_\sigma |\ell_i \rangle \langle \ell_i| \beta_\rho \beta_\rho |\ell_i \rangle \langle \ell_i| \beta_\rho \beta_\rho |\ell_i \rangle \langle \ell_i|.\n$$

(C2)
Let us simply denote $d_3$ as $d$, as there is no environment to care about. Then we obtain the average sequence

$$E[S_1(\rho)] = \text{tr}_S \left[T_{m}^{\lambda_1} E \left[ C_{m}^{T} \right] \right]$$

$$= \mathcal{A}_2^{\lambda_1} \left[ \sum W_{g}[\psi](\lambda_{i_1}^{(m)})[u_{\sigma(1)}^M\chi_{\sigma(1)}^M]|\psi_i^M\chi_{\sigma(1)}^M[u_{\sigma(2)}^M\chi_{\sigma(2)}^M]\right]$$

$$= \mathcal{A}_2 \left[ \frac{1}{d^2-1} \left( \text{tr} \left( \sum_{\sigma=1} \lambda_{i_1}^{(m)} \lambda_{i_1}^{(m)} \right) \mathbb{I} \right) + \frac{1}{d(d^2-1)} \left( \frac{1}{d} \right) \right]$$

$$= \mathcal{A}_2 \left[ \frac{d \text{tr} \left( \sum_{\sigma=1} \lambda_{i_1}^{(m)} \lambda_{i_1}^{(m)} \right) - \sum_{\sigma=1} \text{tr} (\lambda_{i_1}^{(m)})^2}{d^2-1} \right]$$

$$\quad \quad \quad + \frac{d \sum_{\sigma=1} \text{tr} (\lambda_{i_1}^{(m)})^2 - \text{tr} \left( \sum_{\sigma=1} \lambda_{i_1}^{(m)} \lambda_{i_1}^{(m)} \right)}{d(d^2-1)}$$

$$= \mathcal{A}_2 \left[ \frac{d \text{tr} \left( \sum_{\sigma=1} \lambda_{i_1}^{(m)} \lambda_{i_1}^{(m)} \right) - \sum_{\sigma=1} \text{tr} (\lambda_{i_1}^{(m)})^2}{d^2-1} \right]$$

Now, if the noise is trace-preserving as well, we have $\text{tr} \left( \sum_{\sigma=1} \lambda_{i_1}^{(m)} \lambda_{i_1}^{(m)} \right) = \text{tr}(\mathbb{I}) = d$. Then we get

$$E[S_1(\rho)] = \mathcal{A}_2^{\lambda_1} \circ \mathcal{D}_p(\rho),$$

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

$$p := \frac{\sum_{\sigma=1} \text{tr} (\lambda_{i_1}^{(m)})^2 - 1}{d^2 - 1} \in [0,1],$$

which has to be constrained to $[0,1]$. If we denote the noise map $\Lambda_1^{\lambda_1} = \sum_{\sigma=1} \lambda_{i_1}^{(m)} \otimes \lambda_{i_1}^{(m)}$, with the Kraus operators acting on the respective system $S$ Hilbert space and conjugate space, resp., we can simply write

$$p = \frac{\text{tr} \left( \Lambda_1^{\lambda_1} \right) - 1}{d^2 - 1}. \quad (C6)$$

The noise-strength can be shown to be related to the gate fidelity of $\Lambda_1$ with respect to the identity $[43]$, i.e. $f_{\Lambda_1,\mathbb{I}} = \int d\psi \langle \psi | \Lambda_1^{\lambda_1} | \psi \rangle \langle \psi | \psi \rangle$, as $p = \frac{df_{\Lambda_1,\mathbb{I}} - 1}{d^2 - 1}$. This is the relevant parameter which in practice can be recovered by running several sequences for different lengths and averaging the resulting probabilities.

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

$$\mathcal{D}_p(X) = \sum W_{g}[\psi](\lambda_{i_1}^{(m)})[u_{\sigma(1)}^M\chi_{\sigma(1)}^M]|\psi_i^M\chi_{\sigma(1)}^M[u_{\sigma(2)}^M\chi_{\sigma(2)}^M],$$

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

$$E[S_m(\rho)] = \text{tr}_S \left[T_{m}^{\lambda_1} E \left[ C_{m}^{T} \right] \right]$$

$$= \mathcal{A}_{m+1} \circ \mathcal{D}_{p_m} \circ \cdots \circ \mathcal{D}_{p_1}(\rho)$$

$$= p_1 p_2 \cdots p_m \Lambda_{m+1}^{\lambda_1} \left( \rho - \frac{1}{d} \right) + \Lambda_{m+1}^{\lambda_1} \left( \frac{1}{d} \right).$$

(C8)

where here now $p_m := \frac{n(\Lambda_1^{\lambda_1} - 1)}{d^2 - 1}$, 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

$$F_m = p^m \text{tr} \left[ M \Lambda_{m+1}^{\lambda_1} \left( \rho - \frac{1}{d} \right) \right] + \text{tr} \left[ M \Lambda_{m+1}^{\lambda_1} \left( \frac{1}{d} \right) \right] := Ap^m + B,$$

with $A := \text{tr} \left[ M \Lambda_{m+1}^{\lambda_1} \left( \rho - \frac{1}{d} \right) \right]$ and $B := \text{tr} \left[ M \Lambda_1^{\lambda_1} \left( \frac{1}{d} \right) \right]$, 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

\[
\begin{align*}
\mathbf{T}_1 & = \text{tr}_E \left[ \Lambda_2 \otimes I \circ \aleph_2 \circ (\Lambda_1 \otimes I) \circ \aleph_1 (\rho \otimes \psi^{(2)}) \right] \\
& = \text{tr}_E \left[ (\Lambda_2 \otimes I) \aleph_2 (\Lambda_1 \otimes I) \aleph_1 (\rho \otimes \psi^{(2)}) \right] \\
& = \text{tr}_E \left[ (\Lambda_2 \otimes I) \aleph_2 (\Lambda_1 \otimes I) \aleph_1 (\rho \otimes \psi^{(2)}) \right] \\
& \quad \otimes |\beta_1, \alpha_1, \beta_2, \alpha_2, \gamma_1, \delta_1, \gamma_2, \delta_2|,
\end{align*}
\]

where here \(\xi_{ab} := 1_E \otimes |a, b|\), hence we get

\[
\begin{align*}
E[S_1(\rho)] & = \text{tr}_E \left[ \mathbf{T}_1 E(\xi^{(1)}) \right] \\
& = \text{tr}_E \left[ \Lambda_2 \left[ \sum Wg \xi_{\aleph_2 \aleph_1}, \rho \xi_{\aleph_2 \aleph_1} \otimes \xi_{\aleph_2 \aleph_1} \right] \right] .
\end{align*}
\]

Let us write the Kraus operators of the \(n\)th noise map, \(\Lambda_n\), as

\[
\Lambda_n := \sum_{e, s = 1}^{d_E} \sum_{i, j = 1}^{d} L_{i, j}^e X s |e_n s_n \rangle \langle e_n s_n|,
\]

where the \(e\) and \(s\) indices refer to systems \(E\) and \(S\), resp.; the subindex \(n\) is simply a label for the \(n\)th Kraus operator. Then

\[
\begin{align*}
E[S_1(\rho)] & = \sum L_{i, j}^e X s |e_n s_n \rangle \langle e_n s_n| \text{tr}_E \left[ \Lambda_2 \left[ \langle e_1 | e_1' \rangle \otimes |e_2 X u_2 | s_1 X s_1' | u_{\tau_1(1)} | u_{\tau_1(2)} \rangle \rho \langle e_1' | e_1 \rangle \otimes |e_1 X u_1 | s_1' X s_1' | u_{\tau_2(1)} | u_{\tau_2(2)} \rangle \right] \right],
\end{align*}
\]

and let us now similarly write the initial state as

\[
\rho = \sum_{e, s = 1}^{d_E} \chi^{e e'} |e s \rangle \langle e s'|,
\]

where \(\sum \chi^{e e'} = 1, \)

then also

\[
\begin{align*}
E[S_1(\rho)] & = \sum L_{i, j}^e X s |e_n s_n \rangle \langle e_n s_n| \text{tr}_E \left[ \Lambda_2 \left[ \langle e_1 | e_1' \rangle \otimes |e_2 X u_2 | s_1 X s_1' | u_{\tau_1(1)} | u_{\tau_1(2)} \rangle \rho \langle e_1' | e_1 \rangle \otimes |e_1 X u_1 | s_1' X s_1' | u_{\tau_2(1)} | u_{\tau_2(2)} \rangle \right] \right].
\end{align*}
\]

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

\[
\Phi_{\delta \Delta}(X) := \frac{d_d \delta \Delta \delta \xi_{\Delta \Delta}}{d_S (d_E^2 - 1)} X + \frac{d_d \delta \Xi \delta \xi_{\Delta \Delta}}{d_S (d_E^2 - 1)} \left( \mathbb{I}_S \right),
\]

Now let

\[
\alpha_{\delta \xi_{\Delta \Delta}}, \beta_{\delta \xi_{\Delta \Delta}} := \frac{d_d \delta \Delta \delta \xi_{\Delta \Delta}}{d_S (d_E^2 - 1)},
\]

so that

\[
\Phi_{\delta \Delta}(X) = \alpha_{\delta \xi_{\Delta \Delta}} X + \beta_{\delta \xi_{\Delta \Delta}} \left( \mathbb{I}_S \right) = \alpha_{\delta \xi_{\Delta \Delta}} \left( X - \mathbb{I}_S \right) + \beta_{\delta \xi_{\Delta \Delta}} \left( \mathbb{I}_S \right). \]
Now, we can also define

\[
\mathcal{E}_{\Phi^{(2)}}^{(2)} := \sum_{i_1, i_2} \sum_{e, e'} e_{i_1} e_{i_2} e_{e_1} e_{e_2} \sum_{s, s'} s_{i_1} s_{i_2}^* \sum_{s, s'} s_{e_1} s_{e_2}^* \sum_{s, s'} s_{e_1} s_{e_2}^* \\
= \left\{ I_E \otimes \langle s_2 | s_1 \rangle \right\} \left( E \otimes | s_1 \rangle \langle s_1 | \right) \left( E \otimes | s_2 \rangle \langle s_2 | \right).
\]

where summation is over all \( i_1, i_2, \ldots \) and \( e_0, e_1, \ldots \), 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 \( \mathcal{E}_{\Phi^{(2)}}^{(2)} = \left\{ I_E \otimes \langle s_2 | s_1 \rangle \right\} \left( E \otimes | s_1 \rangle \langle s_1 | \right) \left( E \otimes | s_2 \rangle \langle s_2 | \right) \) as in the main text, where an identity on \( E \) is implicit. With this we can write Eq. (C15) as

\[
\mathbb{E} [S_1(\rho)] = \sum_{i, i'} \mathcal{E}_{\Phi^{(2)}}^{(2)} \left[ \alpha_{i,i'} \left( \frac{\delta_{xx'} \delta_{xx'} \delta_{xx'} \delta_{xx'} \delta_{xx'}}{dS} \right) + \frac{\delta_{xx'} \delta_{xx'} \delta_{xx'} \delta_{xx'} \delta_{xx'}}{dS} \left( \frac{\delta_{xx'} \delta_{xx'} \delta_{xx'} \delta_{xx'} \delta_{xx'}}{dS} \right) \right] |s_2 \rangle |\xi_2 \rangle.
\]

In general, for an arbitrary sequence length, we have

\[
\mathcal{E}_{\Phi^{(m)}}^{(m)} := \sum_{e, e'} \chi_{s}^{s'} \left\langle \prod_{n=m}^{m} L_{\eta_n}^{\epsilon_{\eta_n}} |e_n \rangle \langle e'_n| \prod_{n=1}^{m} L_{\eta_n}^{\epsilon_{\eta_n}} |e_n \rangle \langle e'_n| \left\rangle \right\rangle \right. \\
= \sum_{i} \left\langle \prod_{n=m}^{m} \langle s_n | \lambda_{i_n} | s'_n \rangle \right\rangle \left\langle \prod_{n=1}^{m} | \zeta_{i_n} \rangle \langle \zeta_{i_n}| \right\rangle \\
\]

so that,

\[
\mathbb{E} [S_m(\rho)] = \sum_{i, i'} \mathcal{E}_{\Phi^{(m+1)}}^{(m+1)} \left[ \alpha_{i,i'} \left( \frac{\delta_{xx'} \delta_{xx'} \delta_{xx'} \delta_{xx'} \delta_{xx'}}{dS} \right) + \frac{\delta_{xx'} \delta_{xx'} \delta_{xx'} \delta_{xx'} \delta_{xx'}}{dS} \left( \frac{\delta_{xx'} \delta_{xx'} \delta_{xx'} \delta_{xx'} \delta_{xx'}}{dS} \right) \right] |s_{m+1} \rangle |\xi_{m+1} \rangle.
\]

The sequential application of \( \Phi \) maps is given by

\[
\Phi_{s_m \xi_m} \circ \cdots \circ \Phi_{s_1 \xi_1}(X) := \alpha_{s_{m} \xi_{m}}^{(m)} \left( X - \frac{1}{dS} \right) + \Delta_{s_{m} \xi_{m}}^{(m)} \frac{1}{dS},
\]

where

\[
\alpha_{s_{m} \xi_{m}}^{(m)} := \prod_{n=1}^{m} \alpha_{s_n \xi_n}^{(n)},
\]

and \( \Delta_{s_{m} \xi_{m}}^{(m)} \) is a sum of all \( m \)-term product combinations of \( \alpha_{s_1 \xi_1}^{(1)}, \ldots, \alpha_{s_m \xi_m}^{(m)} \) and \( \beta_{s_1 \xi_1}^{(1)}, \ldots, \beta_{s_m \xi_m}^{(m)} \), that is,

\[
\Delta_{s_{m} \xi_{m}}^{(1)} \sim \alpha_{1} + \beta_{1} \quad \Delta_{s_{m} \xi_{m}}^{(2)} \sim \alpha_{1,2} + \alpha_{1,2} + \beta_{1,2} + \beta_{1,2} \quad \Delta_{s_{m} \xi_{m}}^{(3)} \sim \alpha_{1,2,3} + \alpha_{1,2,3} + \alpha_{1,2,3} + \beta_{1,2,3} + \beta_{1,2,3} + \beta_{1,2,3} \quad \Delta_{s_{m} \xi_{m}}^{(m+1)} = \prod_{i=1}^{m} \delta_{s_i \xi_i} \delta_{s_i \xi_i} \delta_{s_i \xi_i} \delta_{s_i \xi_i} \delta_{s_i \xi_i}.
\]

where \( \alpha_i = \alpha_{s_i \xi_i}^{(i)} \) and similarly for \( \beta_i \); in general there are \( 2^m \) of these summands on \( \Delta_{s_{m} \xi_{m}}^{(m)} \). However, notice that as \( \alpha_{1} \sim \frac{1}{dS} \delta_{s_{1} \xi_{1}} \delta_{s_{1} \xi_{1}} \delta_{s_{1} \xi_{1}} \delta_{s_{1} \xi_{1}} \delta_{s_{1} \xi_{1}} \), every term simplifies to products of deltas, i.e.

\[
\Delta_{s_{m} \xi_{m}}^{(1)} \sim \frac{\delta_{s_{1} \xi_{1}} \delta_{s_{1} \xi_{1}} \delta_{s_{1} \xi_{1}} \delta_{s_{1} \xi_{1}} \delta_{s_{1} \xi_{1}}}{dS}, \quad \Delta_{s_{m} \xi_{m}}^{(2)} \sim \frac{\delta_{s_{1} \xi_{1}} \delta_{s_{1} \xi_{1}} \delta_{s_{1} \xi_{1}} \delta_{s_{1} \xi_{1}} \delta_{s_{1} \xi_{1}}}{dS}, \quad \Delta_{s_{m} \xi_{m}}^{(m+1)} = \frac{\prod_{i=1}^{m} \delta_{s_i \xi_i} \delta_{s_i \xi_i} \delta_{s_i \xi_i} \delta_{s_i \xi_i} \delta_{s_i \xi_i}}{dS}.
\]

Thus we can rewrite Eq. (C22) as

\[
\mathbb{E} [S_m(\rho)] = \sum_{i, i'} \mathcal{E}_{\Phi^{(m+1)}}^{(m+1)} \left[ \alpha_{s_{m} \xi_{m}}^{(m)} \left( \frac{\delta_{xx'} \delta_{xx'} \delta_{xx'} \delta_{xx'} \delta_{xx'}}{dS} \right) + \frac{\delta_{xx'} \delta_{xx'} \delta_{xx'} \delta_{xx'} \delta_{xx'}}{dS} \left( \frac{\delta_{xx'} \delta_{xx'} \delta_{xx'} \delta_{xx'} \delta_{xx'}}{dS} \right) \right] |s_{m+1} \rangle |\xi_{m+1} \rangle.
\]
and so for a measurement $M$, on average,

$$F_m = \sum_{s_d, s', \zeta} \mathcal{E}_{s_d|s', \zeta|\zeta}^{(m+1)} \left( \mathcal{A}_{s_d|s', \zeta|\zeta}^{(m+1)} + \mathcal{B}_{s_d|s', \zeta|\zeta}^{(m+1)} \right),$$

where

$$\mathcal{A}_{s_d|s', \zeta|\zeta}^{(m+1)} := \alpha_{s_d|s', \zeta|\zeta}^{(m)} \left( \delta_{s_d|s', \zeta|\zeta}^{(m+1)} - \frac{\delta_{s_d|s', \zeta|\zeta}^{(m+1)}}{d_S^2} \right) \langle \zeta_{m+1} | M | s_{m+1} \rangle,$$

$$\mathcal{B}_{s_d|s', \zeta|\zeta}^{(m+1)} := \Delta_{s_d|s', \zeta|\zeta}^{(m)} \left( \frac{\delta_{s_d|s', \zeta|\zeta}^{(m+1)}}{d_S} \right) \langle \zeta_{m+1} | M | s_{m+1} \rangle.$$  

This expression contains $4d_S^2(m+1)$ 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 $\mathcal{E}_{s_d|s', \zeta|\zeta}^{(m+1)}$.

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

$$\sum_{s_d, s', \zeta} \mathcal{E}_{s_d|s', \zeta|\zeta}^{(m+1)} \mathcal{A}_{s_d|s', \zeta|\zeta}^{(m+1)} \mathcal{B}_{s_d|s', \zeta|\zeta}^{(m+1)}$$

$$= \sum \text{tr} \left[ \prod_{n=1}^{m+1} \langle s_n | \rho | s_n \rangle \left( \prod_{n=1}^{m+1} \langle \zeta_n | \Lambda_n^i | \zeta_n \rangle \right) \right] \left( \prod_{n=1}^{m+1} \langle \zeta_n | \Lambda_n^i | \zeta_n \rangle \right) \left( \frac{\prod_{n=1}^{m+1} \left( d_S \delta_{s_n|s_n'} \delta_{\zeta_n|\zeta_n'} - \delta_{s_n|\zeta_n} \delta_{s_n'|\zeta_n'} \right)}{d_S^2(d_S^2 - 1)^m} \right) \left( \delta_{s_n|s_n'} \delta_{\zeta_n|\zeta_n'} - \frac{\delta_{s_n|\zeta_n} \delta_{s_n'|\zeta_n'}}{d_S} \right)$$

so now let us define the following. Let

$$S_{\Lambda_n}(\varepsilon) := \sum_{s_i} \text{tr}_S(\Lambda_n) \varepsilon \text{ tr}_S(\Lambda_n^i),$$

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

for any operator $\varepsilon$ acting on $E$. Then we notice that

$$\frac{1}{d_S^2} \sum \left( \prod_{n=m}^{m+1} \langle s_n | \rho | s_n \rangle \left( \prod_{n=1}^{m} \langle \zeta_n | \Lambda_n^i | \zeta_n \rangle \right) \right) \left( d_S \delta_{s_n|s_n'} \delta_{\zeta_n|\zeta_n'} - \delta_{s_n|\zeta_n} \delta_{s_n'|\zeta_n'} \right)$$

$$= \frac{1}{d_S^2} \sum \left( \prod_{n=m}^{m+1} \langle s_n | \rho | s_n \rangle \left( \text{tr}_S(\Lambda_n) \langle \varepsilon | \rho | s_n \rangle \right) - \text{tr}_S[\Lambda_n(\langle \varepsilon | \rho | s_n \rangle \otimes I)] \right) \left( \prod_{n=2}^{m} \langle \zeta_n | \Lambda_n^i | \zeta_n \rangle \right) \left( d_S \delta_{s_n|s_n'} \delta_{\zeta_n|\zeta_n'} - \delta_{s_n|\zeta_n} \delta_{s_n'|\zeta_n'} \right)$$

$$= \frac{1}{d_S^2} \sum \left( \prod_{n=m}^{m+1} \langle s_n | \rho | s_n \rangle \left( S_{\Lambda_n} - \Theta_{\Lambda_n} \right) \langle \varepsilon | \rho | s_n \rangle \right) \left( \prod_{n=2}^{m} \langle \zeta_n | \Lambda_n^i | \zeta_n \rangle \right) \left( d_S \delta_{s_n|s_n'} \delta_{\zeta_n|\zeta_n'} - \delta_{s_n|\zeta_n} \delta_{s_n'|\zeta_n'} \right)$$

$$= \sum_{n=1}^{m} \left( S_{\Lambda_n} - \Theta_{\Lambda_n} \right) \langle \varepsilon | \rho | s_n \rangle,$$

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

$$\sum_{s_d, s', \zeta} \mathcal{E}_{s_d|s', \zeta|\zeta}^{(m+1)} \mathcal{A}_{s_d|s', \zeta|\zeta}^{(m+1)} \mathcal{B}_{s_d|s', \zeta|\zeta}^{(m+1)} = \text{tr} \left[ M \text{ tr}_E \circ \Lambda_{m+1} \left( \prod_{n=1}^{m} (S_{\Lambda_n} - \Theta_{\Lambda_n}) \otimes I_S \right) \left( \rho - \rho_E \otimes \frac{1}{d_S} \right) \right] \left( d_S^2 - 1 \right)^m.$$
where $\rho_E := \text{tr}_S(\rho)$. Now for the second term, similarly (again we omit implicit identity operators),

\[
\sum_{s,s',t'=1}^{d_S} E^{(m+1)}_{s,s',t'} \xi^{(n+1)}_{s',t'} = \frac{1}{d_S^{m+1}} \sum_{s,s',t'=1}^{d_S} \text{tr} \left( \langle s_{m+1}| s'_{m+1} \rangle \left( \prod_{n=1}^{m} \langle s_n| s_n' \rangle \right) \langle s| \rho |s \rangle \right) \left( \langle s'|_{m+1} | s'_{m+1} \rangle \right) \left( \langle s_{m+1}| s_{m+1} \rangle \right) \left( \langle s'|_{m+1} | s'_{m+1} \rangle \right)
\]

Thus we can finally write

\[
\mathcal{F}_m = \text{tr} \left[ M \text{tr}_E \circ \Lambda_{m+1} \circ (\mathcal{A}_m + \mathcal{B}_m) \rho \right],
\]

where

\[
\mathcal{A}_m(\rho) := \left( \sum_{n=1}^{m} (S_{n} - \Theta_{n}) \right) \otimes I_S \left( \rho - \rho_E \otimes \frac{1}{d_S} \right)
\]

\[
\mathcal{B}_m(\rho) := \left( \sum_{n=1}^{m} \Theta_{n} \right) \rho_E \otimes \frac{1}{d_S},
\]

with $S_n$ and $\Theta_n$ defined in Eq. (C33) and Eq. (C34), resp.

**Appendix D: Markovian limit**

For the Markovian limit we take $\Lambda_n \rightarrow I_E \otimes \Lambda_{n}^\text{M}$ and $\rho = \rho_E \otimes \rho_S$. First, let us notice that, assuming $\Lambda_{n}^\text{M}$ are CPTP,

\[
S_{\Lambda_{n}}(\varepsilon) = \text{tr} \left[ \Lambda_{n}^\text{M} \right] \varepsilon, \quad \Theta_{n}(\varepsilon) = \text{tr} \left[ \Lambda_{n}^\text{M} \left( \frac{1}{d_S} \right) \right] \varepsilon = \varepsilon,
\]

for any operators $\varepsilon$ acting on $E$ and $\sigma$ on $S$. Then this implies that

\[
\text{tr}_E \circ \mathcal{A}_m(\rho_E \otimes \rho_S) \rightarrow \text{tr}_E \circ \left( \sum_{n=1}^{m} (S_{n} - \Theta_{n}) \right) \left( \rho_E - \frac{1}{d_S} \right)
\]

\[
= \frac{1}{(d_S^2 - 1)^m} \text{tr} \left[ \Lambda_{n}^\text{M} \right] \left( \rho_E - \frac{1}{d_S} \right)
\]

\[
= \prod_{n=2}^{m} \left( \text{tr} \left[ \Lambda_{n}^\text{M} \right] - 1 \right) \left( \rho_E - \frac{1}{d_S} \right)
\]

\[
= p_1 \cdots p_m \left( \rho_E - \frac{1}{d_S} \right),
\]

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

\[
\text{tr}_E \circ \mathcal{B}_m(\rho_E \otimes \rho_S) \rightarrow \text{tr}_E \circ \left( \sum_{n=1}^{m} \Theta_{n} \right) \rho_E \otimes \frac{1}{d_S} = \frac{1}{d_S},
\]

which implies that $\mathcal{F}_m \rightarrow p_1 \cdots p_m \text{tr}[M \Lambda_{m+1}(\rho - 1/d_S)] + \text{tr}[M \Lambda_{m+1}(1/d_S)]$ under Markovian noise.
Appendix E: Finite non-Markovian noise

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

\[
\text{tr}_E \circ \mathcal{A}_m(\rho) \rightarrow \text{tr}_E \circ \sum_{n=1}^{m} \frac{(S_{\Lambda_n} - \Theta_{\Lambda_n}) \otimes \mathcal{I}_S}{(d_S^2 - 1)^m} \left[ \rho_E \otimes \left( \rho_S - \frac{1}{d_S} \right) \right] = p_{\ell+1} \cdots p_m \text{tr}_E \circ \mathcal{A}(\rho),
\]

(E1)

and also \( \text{tr}_E \circ \mathcal{B}_m(\rho) = \text{tr}_E \circ \mathcal{B}_m(\rho) = \text{tr}[\mathcal{B}_m(\rho)] \frac{1}{d_S} \). In particular if the final noise were trace-preserving, we would have \( \text{tr}[\mathcal{B}_m(\rho)] = 1 \). In general, however, this implies

\[
\mathcal{F}_m \rightarrow p_{\ell+1} \cdots p_m \text{tr} \left[ M \Lambda_{m+1}^{(m)} \circ \mathcal{A}(\rho) \right] + \text{tr}[\mathcal{B}_m(\rho)] \text{tr} \left[ M \Lambda_{m+1}^{(m)} \left( \frac{1}{d_S} \right) \right] \quad \text{with} \quad \ell < m,
\]

(E2)

where here again \( p_m = \frac{\text{tr}[\Lambda_{m+1}^{(m)} - 1]}{d_S^2 - 1} \) is the noise-strength corresponding to \( \Lambda_m^{(n)} \).

This means, as one would expect, that in such a case if non-Markovian noise cannot be resolved with an RB sequence length \( \ell \), 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 to estimate 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 \( \mathcal{E} \) stops being superfluous after some time-step \( \ell < m \), i.e. \( \hat{\mathcal{T}} \approx \mathcal{T}_\ell \otimes \mathcal{T}_{m,\ell+1}^{(m)} \). Now we have

\[
\mathcal{A}_m(\rho_E \otimes \rho_S) \rightarrow \sum_{n=1}^{m} \frac{(S_{\Lambda_n} - \Theta_{\Lambda_n}) \otimes \mathcal{I}_S}{(d_S^2 - 1)^m} \left[ \rho_E \otimes \left( \rho_S - \frac{1}{d_S} \right) \right] = p_{\ell+1} \cdots p_\ell \mathcal{A}_{m+1}(\rho_E \otimes \rho_S),
\]

(E3)

where here we defined

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

(E4)

whilst now \( \mathcal{B}_m(\rho_E \otimes \rho_S) = \mathcal{B}_{m+1}(\rho_E \otimes \rho_S) \), where similarly, \( \mathcal{B}_{m,k}(\rho) := \sum_{n=k}^{m} \Theta_{\Lambda_n}(\rho_E) \otimes \frac{1}{d_S} \). Thus

\[
\mathcal{F}_m \rightarrow p_{\ell+1} \cdots p_\ell \text{tr} \left[ M \mathcal{A}_{m+1}(\rho_E \otimes \Lambda_{m+1} \circ \mathcal{A}_{m+1}(\rho_E \otimes \rho_S)) \right] + \text{tr}[M \mathcal{I}_E \otimes \mathcal{A}_{m+1} \circ \mathcal{B}_{m+1}(\rho)] \quad \text{with} \quad \ell < m,
\]

(E5)

so we get a similar behaviour, 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 \( \ell \) and the remaining noise process is also non-Markovian until step \( m \), i.e. \( \hat{\mathcal{T}}_m \approx \mathcal{T}_\ell \otimes \mathcal{T}_{m,\ell+1}^{(m)} \). Now the only difference from a standard non-Markovian ASF is that at the \( \ell \)-th step we have \( \Lambda_{\ell}(X) \rightarrow \varepsilon \otimes \text{tr}_E[\Lambda_{\ell}(X)] \), where again \( \varepsilon \) is some fiducial state of \( \mathcal{E} \) and \( X \) is the state of \( \mathcal{SE} \) at the \( \ell \)-th step. This means we can write

\[
\mathcal{A}_m(\rho) \rightarrow \mathcal{A}_{m+1} \left[ \varepsilon \otimes \text{tr}_E \circ \mathcal{A}(\rho) \right] = \sum_{n=1}^{m} \frac{(S_{\Lambda_n} - \Theta_{\Lambda_n})(\epsilon)}{(d_S^2 - 1)^{m-n+1}} \otimes \text{tr}_E \circ \mathcal{A}(\rho),
\]

(E6)
whilst now,

\[
\mathcal{B}_m(\rho) = \left( \sum_{n=0}^{m} \Theta_{\Lambda_n} \right) \left( \sum_{n=1}^{\ell} \Theta_{\Lambda_n} \right) \rho_{\text{E}} \otimes \frac{\mathbb{I}}{d_S} \\
= \left( \sum_{n=0}^{m} \Theta_{\Lambda_n} \right) \text{tr}_{S} \left[ \Lambda_n \left( \left( \sum_{n=1}^{\ell} \Theta_{\Lambda_n} \right) \rho_{\text{E}} \otimes \frac{\mathbb{I}}{d_S} \right) \right] \rho_{\text{E}} \otimes \frac{\mathbb{I}}{d_S} \\
= \text{tr} \left[ \Lambda_n \left( \left( \sum_{n=1}^{\ell} \Theta_{\Lambda_n} \right) \rho_{\text{E}} \otimes \frac{\mathbb{I}}{d_S} \right) \left( \sum_{n=0}^{m} \Theta_{\Lambda_n} \right) \varepsilon \otimes \frac{\mathbb{I}}{d_S} \right] \\
= \text{tr} \left[ \mathcal{B}_\ell(\rho) \right] \mathcal{B}_{m,\ell+1}(\varepsilon \otimes \frac{\mathbb{I}}{d_S}),
\]

so we may write

\[
\mathcal{F}_m \rightarrow \text{tr} \left[ M \text{ tr}_{\text{E}} \circ \Lambda_{m+1} \circ \mathcal{A}_{m,\ell+1} \left[ \varepsilon \otimes \text{ tr}_{\text{E}} \circ \mathcal{A}_{\ell}(\rho) \right] \right] + \text{tr} \left[ \mathcal{B}_\ell(\rho) \right] \text{tr} \left[ M \text{ tr}_{\text{E}} \circ \Lambda_{m+1} \circ \mathcal{B}_{m,\ell+1}(\varepsilon \otimes \frac{\mathbb{I}}{d_S}) \right]
\]

with \( \ell < m \). \((E8)\)

This is a much more complicated behaviour, but notice that similarly now after a sequence length \( \ell \), 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 \( \mathcal{A} \), and \( \text{tr}[\mathcal{B}(\rho)] \) 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

\[
\mathcal{F}_m \rightarrow \text{tr} \left[ M \text{ tr}_{\text{E}} \circ \Lambda_{m+1} \circ \mathcal{A}_{m,\ell+1} \left[ \varepsilon \otimes \text{ tr}_{\text{E}} \circ \mathcal{A}_{\ell}(\rho) \right] \right] + \text{tr} \left[ \mathcal{B}_\ell(\rho) \right] \text{tr} \left[ M \text{ tr}_{\text{E}} \circ \Lambda_{m+1} \circ \mathcal{B}_{m,\ell+1}(\varepsilon \otimes \frac{\mathbb{I}}{d_S}) \right]
\]

with \( k < \ell < m \). \((E9)\)

If moreover the initial state is uncorrelated, \( \rho = \rho_{\text{E}} \otimes \rho_{S} \), we get

\[
\mathcal{F}_m \rightarrow \text{tr} \left[ \frac{1}{(d_S^2 - 1)^{\ell}} \sum_{n=1}^{k} (S_{\Lambda_n} - \Theta_{\Lambda_n}) (\rho_{\text{E}}) \right] \text{tr} \left[ \frac{1}{(d_S^2 - 1)^{\ell}} \sum_{n=1}^{\ell} (S_{\Lambda_n} - \Theta_{\Lambda_n}) (\varepsilon_{k}) \right] \text{tr} \left[ M \text{ tr}_{\text{E}} \circ \Lambda_{m+1} \circ \mathcal{A}_{m,\ell+1}(\varepsilon_{\ell} \otimes \frac{\mathbb{I}}{d_S}) \right]
\]

with \( k < \ell < m \). \((E10)\)

This then generalises 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 no exemplify how we may describe these through a classical memory specified by an external classical stochastic process whose outputs control the noise \( \Lambda_i \) 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 behaviour for the ASF.

In particular, such model takes a qubit system with errors \( \Lambda^{(C)}(\cdot) = \lambda_{\ell}(\cdot) \delta_{\ell} \), where \( \lambda_{\ell} = \exp(-i\delta_{\ell} \otimes Z) = \exp(-i\delta_{\ell} Z) \) where \( Z = \text{diag}(1,-1) \) and with \( \delta_{\ell} \) is a random variable determined by the classical external control. Thus a sequence of length \( m \) can be treated as a Markovian time-dependent decay \( E[S_{m}(\rho)] = p_1 \cdots p_m \Lambda^{(C)}_{m+1}(\rho) \), where knowing the Kraus operators \( \lambda_i \), we can compute for small \( \delta \)

\[
p_i = \frac{\text{tr}(e^{-i\delta_{\ell} Z})^2 - 1}{d_S^2 - 1} = \frac{4 \cos^2 \delta_{\ell} - 1}{3},
\]

\((F1)\)
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 \( \sigma^2 \), the so-called Markovian scenario considers the control operations at step \( i \) giving a realisation \( X_i = x_i \) 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

\[
\mathcal{F}_m^{\text{C-Mark}} = \text{tr} \{ \mathcal{M} \mathbb{E} \{ S_m(\rho) \} \} = \prod_{i=1}^m \left( \frac{4 \cos^2 \delta_i - 1}{3} \right).
\]

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 realisation with probability \( p \) and update the memory with a PDF of the form \( \sum_i P(X = x_i) \delta(\delta - x_i) \), where here \( \delta \) is a Dirac delta distribution. Then \( \mathbb{E} \{ S_m(\rho) \} = p^m e^{-\mu^2} e^{\delta^2} \), which similarly for \( \rho = \mathcal{M} = |0\rangle\langle 0| \) becomes simply

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

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 analyse the average fidelity, similar to how it is done in [30] with contributions up to \( \delta^2 \). The final behaviour 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 Fig. 11 (b), 11 (c).

**The shallow pocket model.** We now consider a similar model for a qubit \( S \) coupled to degree of freedom (d.o.f.) on a real line, which acts as an environment. This is labelled a shallow pocket model because such d.o.f. 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, 61]. 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 \( \Lambda^{(C)}_m(\cdot) = \Lambda_m(\cdot) \Lambda^{(C)}_{m+1} \) with \( \Lambda_n = \exp(-i \tau_n \sigma_n \otimes Z) = \exp(-i \tau_n \hat{x}_n Z) \), where \( \hat{x}_n \) is a position operator at time-step \( n \) and \( \tau_n \) are time-intervals representing evolution time of the \( n^{\text{th}} \) step. This immediately implies that the average sequence is of the form \( \mathbb{E} \{ S_m(\rho) \} = p_1 \cdots p_m \Lambda^{(C)}_{m+1} \rho \), where \( \rho = \rho_S \otimes |\psi\rangle\langle \psi| \). The initial state of the environment d.o.f. is taken as \( |\psi\rangle \) such that \( \langle \psi | x_{1} \rangle = \sqrt{\frac{\tau}{\pi}} / (x_{1} + iy) \). Now tracing out the environment at the end of the process is equivalent to integrating \( x \) over the reals with a factor \( \langle \psi | x_m \hat{x}_m | \psi \rangle \delta_{x_m x_{m+1}} \cdots \delta_{x_2 x_1} \delta_{x_1 x_{0}} \). Thus we can think of the external d.o.f. as a classical DC noise distributed with a probability density function \( |\langle \psi | x \rangle|^2 \).
That is, now we have
\[ p_{\tau_n}(x_n) = \frac{1}{d_S^2 - 1} \left| \langle e^{-i\tau_n Z} \rangle \right|^2 - 1 = \frac{4\cos^2(\tau_n x_n) - 1}{3}. \] (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 \(|\langle \psi|\chi\rangle|^2\), has to contain a low enough equivalent of a variance, which amounts to choosing an appropriate value for \( \gamma \). Hence, now taking \( \rho = |0\rangle\langle 0| \otimes |\psi\rangle\langle \psi| \) and \( M = |0\rangle\langle 0| \), we get
\[ \mathcal{F}_m^{\text{shallow}} = \text{tr}[M \mathcal{E}(S_m(\rho))] = \frac{\gamma}{\pi} \int_{-\infty}^{\infty} P_{\tau_1,\ldots,\tau_N} d \tau, \] (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 static unitary noise \( \Lambda(\cdot) = \Lambda(\cdot) \hat{\lambda} \) on a full \( N \)-qubit system, where \( \lambda = \exp(-i\delta H) \) with \( H \) given by the \( N \)-site Ising spin chain
\[ H = \frac{\sum_{i=1}^{N} \left( \frac{1}{2} X_i X_{i+1} + h_x X_i + h_y Y_i \right)}{J} = \begin{pmatrix} 0 & h_x - ih_y & h_x - ih_y & J \\ h_x + ih_y & 0 & J & h_x - ih_y \\ h_x + ih_y & J & 0 & h_x - ih_y \\ J & h_x + ih_y & h_x + ih_y & 0 \end{pmatrix}, \] (G1)
with \( X_i, Y_i \) being Pauli matrices acting on the \( i \)th 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 \( \delta \approx 0.03 \). To take into account SPAM errors numerically, suppose the initial state \( \rho \) is previously affected by the same \( \Lambda \) error for some small \( \delta = \Delta_1 \), and that \( M \) is slightly rotated via \( \exp(-i\Delta_2 Y) \) for a small \( \Delta_2 \). If 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 \( \gamma \), but somehow 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 characterisation of non-Markovian noise with RB is thus an issue that still has to be studied in greater detail.

**Absence of non-exponential behaviour.** We notice that for a similar noise model for a couple of qubits,
\[ H = J_x X_1 X_2 + J_y Y_1 Y_2 = \begin{pmatrix} 0 & 0 & 0 & J_x - J_y \\ 0 & 0 & J_x + J_y & 0 \\ 0 & J_x + J_y & 0 & 0 \\ J_x - J_y & 0 & 0 & 0 \end{pmatrix}, \] (G2)
essentially no deviation from an exponential is seen. We look again at static noise given by \( \lambda = \exp(-i\hat{\delta} H) \) with small \( \delta \approx 0.03 \) 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 \( M = |0\rangle\langle 0| \). We show the corresponding ASF in Fig. 13 for the arbitrary choices \( J_x = 1.2, J_y = -2.7 \).

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 5 environmental qubits, and notice that the non-exponential deviations get effectively damped, albeit slowly and for longer sequence lengths first. This is expected behaviour, 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 $\sim \ell$, described by the CP maps $A_1, A_2, \ldots, A_\ell$, 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 $F_m = P_{\ell:1} P_{\ell+1} \cdots P_m A + B$, where

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

\[ \Lambda_i^{\text{SM}}(\cdot) := \text{tr}_E[\Lambda_{\ell} \circ \Lambda_{\ell-1} \circ \cdots \circ \Lambda_{1}(\epsilon \otimes \cdot)], \]

(G4)

with each \( \Lambda_n \) in terms of Kraus operators \( \lambda_n \) acting on SE spaces and co-spaces as \( \Lambda_n := \sum_{j} \lambda_{i_n} \otimes \lambda_{i_n}^\dagger \).

That is, the initial block of finite non-Markovian noise looks like a single noise map \( \Lambda_{\ell-1} \) if we randomise over a single Clifford within this block, with the remaining ones set to identities.

While this is an idealised scenario, we can use it to estimate sequence lengths at which non-Markovian noise effects are relevant in an RB experiment. In the main text we model such a noise process with a noise map at the \( n^{th} \) step given by

\[ \Lambda_n^{(f)} := q_{n-\ell}\Lambda + (1-q_{n-\ell})\Lambda^{\text{SM}}, \quad \text{where} \quad q_n := \frac{1}{1 - \exp(n - \ell)} \]

(G5)

where here again \( \Lambda = \lambda \otimes \lambda^\dagger \) with \( \lambda = \exp(-i\delta H) \) where \( H \) is given by Eq. (G1) and \( \Lambda^{\text{SM}} \) acts on \( S \) as \( \Lambda^{\text{SM}}(\cdot) = \epsilon \otimes \text{tr}_E[\Lambda(\cdot)] \).

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 \( \delta^{(f)} = 2.5\delta \) for \( \Lambda_n^{(f)} \). This implies that the noise acts jointly over the whole SE throughout the full process, but it acts almost fully as \( \Lambda \) for \( m < \ell \), whilst it turns almost to act solely on \( S \) with \( \Lambda^{\text{SM}} \) for \( m \geq \ell \).

In the top figure of Fig. 6 in the main text, we display the ASFs for a set of RB experiments with \( \rho = |0\rangle|0\rangle \) and \( M = |0\rangle|0\rangle \) 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 \( \ell \) from the data of the experiments alone, as well as construct a sensible static 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, \( F_m \), displaying a non-exponential decay over a finite sequence length:

1. Fix identities at sequence lengths of \( F_m \) manifestly displaying deviations from an exponential decay and run RB experiments for each of them, obtaining corresponding ASFs \( F_m^{(i,\ldots,j)} \), where \( i, \ldots, j \) are sequence lengths at which identities were fixed.
2. Identify the section of the original \( F_m \) manifestly displaying exponential behaviour and extract the noise rate \( p \) at such section.
3. Fit an exponential to each $F_{m[i,\ldots,j]}$; Eq. (E2) implies that the curve with an exponential rate $p_{m[i',\ldots,j']} \approx p$ will indicate the length at which the noise turns almost Markovian (or where non-Markovian effects become negligible).

4. Finally, a Markovianised 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 $\{m, F_m\}$ from $m = 12$ to $m = 30$, which more manifestly display an exponential decay. These were fitted to an exponential $f_m \approx (0.7847)(0.9325)^m + 0.4915$, i.e. we extract $p \approx 0.9325$.

For step 3, we identified the closest decay rate to $p$ occurred for $F_{m[1,\ldots,8]}$, with $p_{m[1,\ldots,8]} \approx 0.9278$. This indicates that $\ell \approx 9$. Since we fixed $\ell = 9$, 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. $\Lambda_0^{(9)} = \frac{1}{2} (\Lambda + \Lambda^{\text{SE}})$ so that the noise will still act jointly on SE with at least half probability. In this sense is $\ell$ just approximated numerically. In the bottom panel of Fig. 6, the dot-dashed line displays the curve given by $\tilde{f}_m \approx (0.7847)p_{m[1,\ldots,8]}^m + 0.4915$, showing the slight offset due to this numerical estimation.

Finally, at step 4 we simply fix $A = B$ in $F_{m[i,\ldots,8]} = Ap_{m[i,\ldots,8]}^m + B$ 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 $m \rightarrow \infty$. As is the case for RB, this Markovianised ASF curve at most informs us about the gate fidelity with respect to the identity of the Markovianised noise through $p_{m[1,\ldots,8]}$. 