

## Markovianization with approximate unitary designs

Pedro Figueroa-Romero <sup>1✉</sup>, Felix A. Pollock<sup>1</sup> & Kavan Modi <sup>1✉</sup>

Memoryless processes are ubiquitous in nature, in contrast with the mathematics of open systems theory, which states that non-Markovian processes should be the norm. This discrepancy is usually addressed by subjectively making the environment forgetful. Here we prove that there are physical non-Markovian processes that with high probability look highly Markovian for all orders of correlations; we call this phenomenon Markovianization. Formally, we show that when a quantum process has dynamics given by an approximate unitary design, a large deviation bound on the size of non-Markovian memory is implied. We exemplify our result employing an efficient construction of an approximate unitary circuit design using two-qubit interactions only, showing how seemingly simple systems can speedily become forgetful. Conversely, since the process is closed, it should be possible to detect the underlying non-Markovian effects. However, for these processes, observing non-Markovian signatures would require highly entangling resources and hence be a difficult task.

<sup>1</sup>School of Physics & Astronomy, Monash University, Clayton, VIC, Australia. ✉email: [pedro.figueroaromero@monash.edu](mailto:pedro.figueroaromero@monash.edu); [kavan.modi@monash.edu](mailto:kavan.modi@monash.edu)

A foundational question of modern physics is to understand the origins of irreversibility<sup>1</sup>. In particular, to determine whether fundamental laws, which are fully reversible, are consistent with phenomena like equilibration and thermalization. The dynamical version of this conundrum concerns the emergence of forgetful processes from isolated ones. In quantum mechanics, an isolated process is unitary, and cannot lose information; past behavior in one part of the system will always be remembered, eventually returning to influence the future.

However, there are many ways in which nature manifests forgetful processes, where a system's evolution is determined with a seeming disregard to its previous interactions with its surroundings. For example, a carbon atom does not typically remember its past and behaves like any other carbon atom. Such processes are not isolated, and the general intuition is that the dynamics of a system, in contact with a large environment, can be approximately described as memoryless<sup>2</sup>. Yet, formal derivations of memory-less quantum processes require several assumptions about the coupling strength with the environment, the timescales of dynamical correlations, and an infinite-dimensional reservoir. For finite-sized environments, this can only be achieved exactly by continually refreshing (discarding and replacing) the environment's state, i.e., artificially throwing away information from the environment. The problem this poses is akin to the one made by the Fundamental Postulate of Statistical Mechanics<sup>1</sup>, which a-priori sets the probabilities of a closed system to be in any of its accessible microstates as equal.

Thus the foundational question remains open: can forgetful processes arise from isolated processes without any artificial discarding of information? Because forgetful processes are often called Markovian, we refer to the mechanism for forgetting as *Markovianization*, in the same spirit as the terms equilibration and thermalization<sup>1,3–8</sup>. Indeed, Markovianization is likely to come about through mechanisms intimately related to these other processes. For instance, dissipative Markov processes have fixed points to which the system relaxes; this is a mechanism for equilibration, and also possibly for thermalization. We have previously argued for the emergence of Markovianization for mathematically typical processes, using averages with respect to the Haar measure<sup>9</sup>; however, such processes are far from physically typical<sup>1</sup>.

In this paper, we identify a class of isolated physical processes which approximately Markovianize in a strong sense, where even the multi-time quantum correlations vanish. To do so, we employ large deviation bounds for approximate unitary designs derived by R. Low<sup>10</sup>, and apply them to the process tensor formalism<sup>11–13</sup>, which describes quantum stochastic processes. We show that, similar to the way in which quantum states thermalize, quantum processes can Markovianize in the sense that they can converge to a class of typical processes, satisfying a meaningful large deviation principle whenever they are undergone within a large environment and under complex enough—but not necessarily fully random—dynamics. As a proof of principle, we employ a recent efficient construction of approximate unitary designs with quantum circuits<sup>14</sup> to illustrate how a dilute gas would quickly Markovianize. These results directly impose bounds on complexity and timescales for standard master equations employed in the theory of open systems. Finally, we discuss possible extensions of our results to many-body systems with time-independent Hamiltonians. Our results are timely given the ever-increasing interest and relevance in determining the breakdown of the Markovian approximation in modern experiments<sup>15–18</sup>.

## Results

**Quantum stochastic processes.** A classical stochastic process on a discrete set of times is the joint probability distribution of a

time-ordered random variable,  $\mathbb{P}(x_k, \dots, x_0)$ . A process is said to have finite memory whenever the state of the system at a given time is only conditionally dependent on its previous  $m$  states:  $\mathbb{P}(x_k|x_{k-1}, \dots, x_0) = \mathbb{P}(x_k|x_{k-1}, \dots, x_{k-m})$ . Here,  $m$  is the Markov order; when  $m = 1$  the process is called Markovian, and when  $m = 0$  the process is called random. Finite memory processes, and in particular Markov processes, have garnered significant attention in the sciences for two principal reasons. First, the complexity of a process grows with the Markov order and thus it is easier to work with finite memory processes. Second, many physical processes tend to be well approximated by those with finite memory.

Generalizations of Markov processes and Markov order to the quantum realm have been plagued with technical difficulties<sup>19</sup>, which have their origin in the fundamentally invasive nature of quantum measurement. However, recently, a generalized and unambiguous characterization of quantum stochastic processes within the process tensor framework<sup>11,20</sup> has paved the way to alleviating these difficulties. The success of this framework lies in generalizing the notion of time-ordered events in the quantum realm.

Consider a system-environment composite SE of dimension  $d_{SE} = d_S d_E$  with an initial state  $\rho^{(0)}$  that undergoes an evolution  $\mathcal{U}_0$ . An intervention  $\mathcal{A}_0$  is then made on the system S alone, followed by evolution  $\mathcal{U}_1$ . For concreteness, onward we will consider  $\mathcal{U}_i \neq \mathcal{U}_j$ . Then a second intervention  $\mathcal{A}_1$  on S alone. This continues until a final intervention  $\mathcal{A}_k$  is performed following  $\mathcal{U}_k$ . A quantum event  $x_i$  at the  $i$ th time step corresponds to an outcome of the corresponding intervention, and is represented by a completely positive (CP) map  $\mathcal{A}_{x_i}(\cdot) := \sum_{\nu} A_{x_i}^{\nu}(\cdot) A_{x_i}^{\nu\dagger}$  with Kraus operators  $\{A^{\nu}\}$  satisfying  $\sum A^{\nu\dagger} A^{\nu} \leq \mathbb{1}$ . In other words, an intervention is the action of an instrument  $\mathcal{J} = \{\mathcal{A}_{x_i}\}_{x_i}$  where  $A_i = \sum_{x_i} \mathcal{A}_{x_i}$  is a completely positive trace preserving (CPTP) map. This is depicted schematically in Fig. 1. In general, the evolution  $\mathcal{U}$  is allowed to be a CPTP map on SE. In this paper, however, we are interested in an isolated SE, where the  $\mathcal{U}$ s are unitary transformations:  $\mathcal{U}(\cdot) := U(\cdot)U^{\dagger}$ , with  $U$  a unitary operator.

The probability to observe a sequence of quantum events is given by

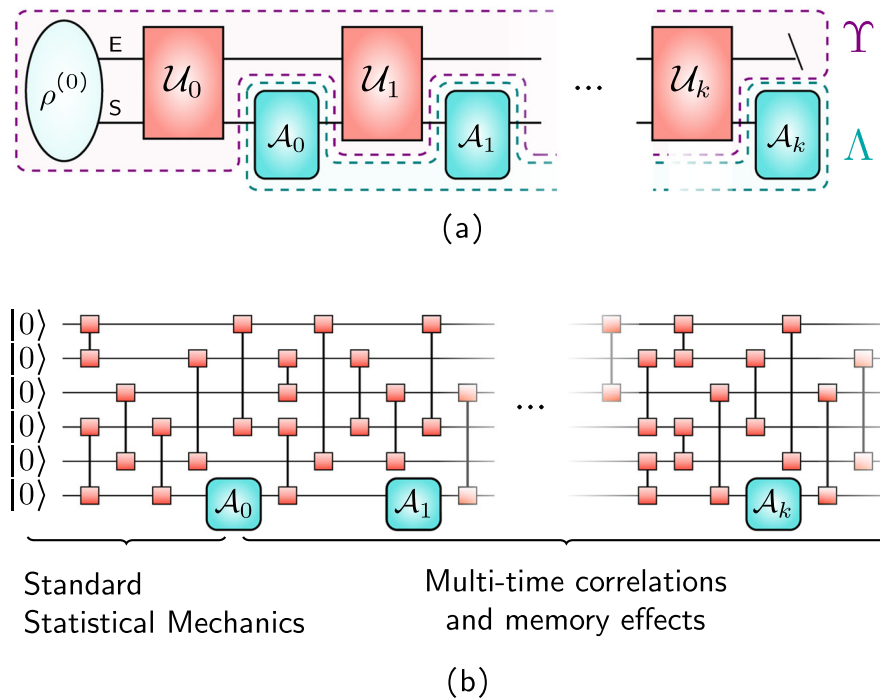
$$\mathbb{P}(x_k, \dots, x_0 | \mathcal{J}_k, \dots, \mathcal{J}_0) = \text{tr} \left[ \mathcal{A}_{x_k} \mathcal{U}_{k-1} \dots \mathcal{A}_{x_0} \mathcal{U}_0 \rho^{(0)} \right].$$

This can be rewritten, clearly separating the influence of the environment from that of the interventions, in a multi-time generalization of the Born rule<sup>21–23</sup>:

$$\mathbb{P}(x_k, \dots, x_0 | \mathcal{J}_k, \dots, \mathcal{J}_0) = \text{tr} \left[ \Upsilon \Lambda^T \right], \quad (1)$$

where T denotes transpose,  $\Lambda := \mathcal{A}_{x_0} \otimes \dots \otimes \mathcal{A}_{x_k}$ , and the effects on the system due to interaction with the environment have been isolated in the so-called process tensor  $\Upsilon$ . We have depicted  $\Upsilon$  and  $\Lambda$  in Fig. 1a as the red and green comb-like regions, respectively. A circuit depiction of the same process  $\Upsilon$ , along with the instruments  $\Lambda$  is given in Fig. 1b.

Maps like the process tensor are abstract objects with many different representations<sup>12</sup>. In this manuscript, for convenience, we work with the Choi state representation<sup>12,24</sup> of the process tensor, shown in Eq. (10) of the Methods section. The process tensor  $\Upsilon$  is a complete representation of the stochastic quantum process, containing all accessible multi-time correlations<sup>25–28</sup>. Similarly, the tensor  $\Lambda$  contains all of the details of the instruments and their outcomes. This tensor, in general, is also a quantum comb, where the bond represents information fed forward through an ancillary system. Finally, the process tensor can be formally shown to be the quantum generalization of a



**Fig. 1 Quantum processes and the process tensor.** **a** A  $k$ -step quantum process  $\Upsilon$  on system  $S$  alone is due to the time evolution of an initial system-environment (SE) state  $\rho^{(0)}$  with distinct unitary transformations  $\mathcal{U}_i$  with  $i = 0, 1, \dots, k$ . In between each pair of unitaries, an external operation (e.g., a measurement)  $\mathcal{A}_i$  for  $i = 0, 1, \dots, k$  is applied; this can also be described by a tensor  $\Lambda$ . **b** An  $n$ -qubit SE-system ( $|0\rangle$ ) depicting a single qubit) with two-qubit gate interactions (depicted by vertical lines between squares) only; a subsystem qubit is probed at the  $i^{\text{th}}$  step through  $\mathcal{A}_i$ . While the standard approach towards typicality or equilibrium properties concerns the whole SE dynamics and/or a single measurement on system  $S$  as in Standard Statistical Mechanics, we show that complex—not necessarily uniformly random—dynamics within large environments will be highly Markovian with high probability.

classical stochastic process, satisfying a generalized extension theorem with consistency conditions for a family of joint probabilities to guarantee the existence of an underlying continuous quantum stochastic process<sup>13</sup>, and reducing to classical stochastic process in the correct limit<sup>29,30</sup>.

**Measuring non-Markovianity.** The convenience of using the Choi state  $\Upsilon$  is that it translates temporal correlations between timesteps into spatial correlations. Furthermore, as detailed in the Methods section on the process tensor,  $\Upsilon$  can be efficiently described when written as a matrix product operator<sup>11,31</sup>, whose bond dimension represents the dimension of a quantum environment that could mediate the non-Markovian correlations. In particular, when the bond dimension is one, the process is Markovian. Specifically, a process  $\Upsilon^{(M)}$  is Markovian if and only if it has the form

$$\Upsilon^{(M)} = \mathcal{E}_{1:0} \otimes \dots \otimes \mathcal{E}_{k:k-1}, \tag{2}$$

with  $\mathcal{E}_{j:i}$  a CPTP map on the system connecting the  $i^{\text{th}}$  to the  $i + 1^{\text{th}}$  time<sup>12,20</sup>. This quantum Markov condition in Eq. (2) allows for a precise quantification of memory effects; it is fully consistent with the classical Markov condition, and contains all of the popular witnesses of quantum non-Markovianity<sup>19</sup>. Importantly, it allows for operationally meaningful measures of non-Markovianity: for instance, the relative entropy of the process tensor with respect to its marginals, which happen to be the closest Markovian process tensor, i.e.  $\mathcal{N}_S := \min_{\Upsilon^{(M)}} S(\Upsilon \parallel \Upsilon^{(M)})$ , quantifies the probability of mistaking  $\Upsilon$  and  $\Upsilon^{(M)}$ , which decreases in the number of realizations of the process  $n$  as  $\exp(-n\mathcal{N}_S)$ .

For the current considerations, a natural choice is the so-called diamond norm. Just as trace distance is a natural metric for

differentiating two quantum states, in the sense of having a clear operational definition, the natural distance for differentiating two quantum channels is the diamond norm, which allows for the use of additional ancillas<sup>32</sup>. We are interested in optimally differentiating between a non-Markovian process from a Markovian one, which leads to the multi-time diamond distance:

$$\mathcal{N}_\diamond := \frac{1}{2} \min_{\Upsilon^{(M)}} \|\Upsilon - \Upsilon^{(M)}\|_\diamond, \tag{3}$$

where  $\|X\|_\diamond := \sup_{\{\mathcal{O}_i\}, i} \|\sum_i \text{tr}[\mathcal{O}_i X \otimes \mathbb{1}] |i\rangle \langle i|\|_1$  is a generalized diamond norm<sup>27,33</sup>, with the supremum over  $i \geq 1$  and a set of CP maps  $\{\mathcal{O}_i\}$ . This definition generalizes the diamond norm for quantum channel distinguishability<sup>34</sup> (also called cb-norm<sup>35</sup> or completely bounded trace norm<sup>24</sup>), reducing to it for a single step process tensor, and similarly being interpreted as the optimal probability to discriminate a process from the closest Markovian one in a single shot, given any set of measurements, which can be made together with an ancilla.

Vanishing non-Markovianity in Eq. (3) would imply that the process must have the form of Eq. (2). The derivations of such processes make ad-hoc assumptions such as artificially refreshing the environment between time-steps (i.e., assumption of an infinite bath) that render approximations such as Born-Markov. Classical processes additionally require randomness injection by hand for stochasticity. Here, we show that a class of underlying quantum mechanisms lead to the emergence of Markovianity without ad-hoc assumptions. Namely, We show that the above measure of non-Markovianity in Eq. (3) vanishes as the global SE dynamics becomes more complex. This is entirely analogous to entanglement being the underlying mechanism explaining the emergence of statistical mechanics from quantum dynamics alone and accounting for the artificial postulate of equal a-priori probabilities<sup>3</sup>.

**Markovianization with unitary designs.** The generic form of open quantum dynamics is non-Markovian, but, despite this, it is often very well approximated by simpler Markovian dynamics. How this memorylessness emerges is not dissimilar to questions, regarding the emergence of thermodynamic behavior, which have pervaded quantum mechanics since its conception. Indeed, it can be shown that canonical quantum states are typical<sup>36–39</sup>, and we now know that the fundamental postulate of equal a-priori probabilities of statistical mechanics can be traced back to the entanglement between subsystems and their environment<sup>3</sup>. It turns out that, very similarly, if we sample a generic quantum process occurring in a large finite environment at random, it will be almost Markovian with very high probability<sup>9</sup>.

This sampling procedure can be formalized through the so-called Haar probability measure,  $\mu_h$ , over the  $d$ -dimensional unitary group  $\mathbb{U}(d)$ , which is the unique (up to a multiplicative constant) measure with the property that, if  $U \in \mathbb{U}(d)$  is distributed according to the Haar measure, then so is any composition  $UV$  or  $VU$ , with a fixed  $V \in \mathbb{U}(d)$ . It can be normalized to one, so as to constitute a legitimate probability measure<sup>40</sup>. The Haar measure allows one to swiftly obtain statistical properties of uniformly distributed quantities<sup>40–45</sup> and, furthermore, to prove concentration of measure results<sup>46–48</sup>; these somewhat surprisingly imply that, when drawn from the right distribution, certain quantities will become overwhelmingly likely to be close to another fixed quantity as the Hilbert space dimension is increased. Henceforth, we write  $U \sim \mu_h$  to refer to  $U$  as distributed according to the Haar measure and, similarly, we use  $\mathbb{P}_h$  and  $\mathbb{E}_h$  to denote probabilities and expectations with respect to the Haar measure.

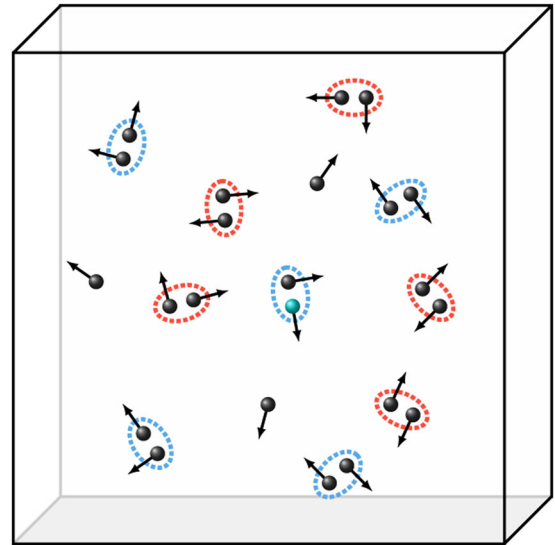
The Result by Modi et al. on Markovian Typicality<sup>9</sup>, which is reproduced in detail in the Methods section on, gives a mathematically sound result of concentration of measure around Markovian processes. However, it assumes a Haar-distributed uniform sampling of unitary dynamics, and we know that nature seldom behaves randomly<sup>49,50</sup>. The dynamics of a vast number of physically relevant models can be approximated as Markovian<sup>51</sup>, so can we say that these also satisfy a concentration of measure with respect to Markovianity?

In some circumstances, sets of physical processes can approximate some of the statistical features of the Haar measure<sup>1,52–54</sup>, for example, consider the toy model depicted in Fig. 2, comprising a dilute gas of  $n$  particles evolving autonomously in a closed box. The gas particles interact with each other in one of two ways as they randomly move inside the box. Following and intervening on a special impurity particle, taken to be the system, this model can be approximately thought to be described by a circuit such as the one in Fig. 1b. The simplicity of this system suggests that it can only uniformly randomize after a large number of random two-qubit interactions, progressively resembling genuine Haar random dynamics.

One possible way to quantify this progressive resemblance of the Haar measure is given by the concept of unitary designs. In general an  $\epsilon$ -approximate  $t$ -design, which we denote  $\mu_t$ , can be defined through

$$\left\| \mathbb{E}_{\mu_t} [\mathcal{V}^{\otimes s}(X)] - \mathbb{E}_h [\mathcal{U}^{\otimes s}(X)] \right\| \leq \epsilon, \quad \forall s \leq t \quad (4)$$

for a suitable metric  $\|\cdot\|$ , where  $\mathcal{U}(\cdot) := U(\cdot)U^\dagger$  and  $\mathcal{V}(\cdot) := V(\cdot)V^\dagger$  are unitary maps with  $U, V \in \mathbb{U}(d)$ . Here, as above, the notation  $\mathbb{E}_\Omega$  indicates the expectation value with respect to a given probability measure  $\mu_\Omega$ , i.e.  $V \sim \mu_t$  and  $U \sim \mu_h$ . That is,  $\mu_t$  approximates the Haar measure up to the  $t^{\text{th}}$  moment with a small error  $\epsilon$ . In the case we are interested in, the unitary maps will correspond to SE unitaries, as depicted in Fig. 1(a),



**Fig. 2 A toy model analogous to a system with dynamics given by an approximate unitary design with two kinds of two-qubit interactions only.** An impurity particle (teal) immersed in a gas of  $n_E$  particles (arrows depicting direction of motion) within a closed box, where all particles interact in pairs in one of two ways (dashed circles) at random, can be similarly described by an approximate unitary design. The result of Theorem 1 ensures that for a large enough  $n_E$  and number of interactions, most processes analogous to this one with approximate unitary designs will be almost Markovian.

according to the either the Haar measure or a unitary design. We also do not assume anything about the parameter  $t$  other than it is a positive non-zero integer.

Notice what this would mean for a model similar to that of Fig. 2: as individual random two-body interactions of each kind accumulate, what we expect is for the dynamics to start scrambling their information across the whole gas in the box, progressively becoming more complex and uniformly random<sup>55</sup>. Unitary designs give us this finite quantification of the approximation to uniform Haar randomness and, in this case, it can give us a precise way to account for the progressive emergence of complexity from seemingly simple individual two-body interactions.

Unitary designs for  $t = 2, 3$  have been widely studied<sup>56–65</sup> and efficient constructions are known for larger values of  $t$ <sup>14,59,66</sup>. The latter are of particular relevance, precisely as designs for large  $t$ , i.e., those with a higher complexity<sup>55</sup>, are expected to satisfy tighter large deviation bounds, approaching concentration of measure as the level and quality of the design increases.

Such large deviation bounds over approximate unitary designs were derived in a general form by R. Low<sup>10</sup> for a polynomial function satisfying a concentration of measure bound, and we now use them to demonstrate the phenomenon of Markovianization for corresponding classes of processes.

**Theorem 1**

Given a  $k$ -step process  $\Upsilon$  on a  $d_S$  dimensional subsystem, generated from global unitary  $d_{SE}$  dimensional SE dynamics distributed according to an  $\epsilon$ -approximate unitary  $t$ -design  $\mu_t$ , the likelihood that its non-Markovianity exceeds any  $\delta > 0$  is bounded as

$$\mathbb{P}_t [\mathcal{N}_\diamond \geq \delta] \leq B, \quad (5)$$

where  $B$  is defined as

$$B := \frac{d_S^{3m(2k+1)}}{\delta^{2m}} \left[ \binom{m}{C}^m + (2B)^{2m} + \frac{\epsilon}{d_{SE}^2} \eta^{2m} \right], \quad (6)$$

for any  $m \in (0, t/4]$  and

$$\eta := \left( d_{SE}^4 d_S^{2k} + d_S^{-(2k+1)} \right) / 4, \tag{7}$$

where  $C$  is defined in Eq. (14) and  $B$  an upper bound on the expected norm-1 non-Markovianity  $\mathbb{E}_h[\mathcal{N}_1]$ , defined in Eq. (15).

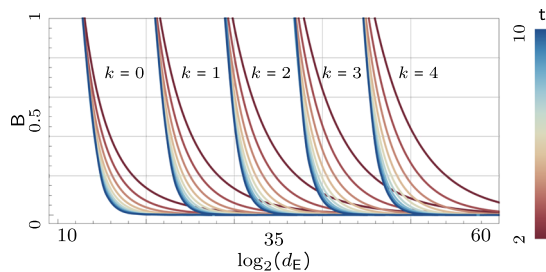
The proof is displayed in full in the Methods section. The overall strategy is as done by R. Low<sup>10</sup>: a bound on the moments  $\mathbb{E}_t[\mathcal{N}_\diamond^m]$  is given in terms of  $B$ ,  $C$  and  $\eta$ , followed by Markov’s inequality. The quantity  $\eta$  is related to the  $\epsilon$ -approximate unitary  $t$ -design  $\mu_\epsilon$  through

$$\mathbb{E}_t[\mathcal{N}_2^{2m}] \leq \mathbb{E}_h[\mathcal{N}_2^{2m}] + \frac{\epsilon}{d_{SE}^t} \eta^{2m}, \tag{8}$$

for any  $m > 0$  and corresponds to the sum of the moduli of the coefficients of  $\mathcal{N}_2^2$ . We explicitly determine a bound on this quantity within the proof of Theorem 1 in the Methods section, which is the one we take as definition in Eq. (7).

The choice of  $0 < m \leq t/4$  can be made to optimize the right-hand-side of the inequality, which ideally should be small whenever  $\delta$  is. The term  $d_S^{3(2k+1)}/\delta^2$  arises from bounding  $\mathcal{N}_\diamond$  and Markov’s inequality, while the three summands within square brackets will be small provided *i*)  $C$  is large, *ii*)  $B$  is small and *iii*) the unitary design sufficiently small  $\epsilon$  and large  $t$  is well-approximate and high enough. For conditions *i*) and *ii*), we require a fixed  $k$  such that  $d_E \gg d_S^{2k+1}$ : this implies  $B \approx 0$ , so that ignoring subleading terms, we require  $\epsilon \ll \delta^{2m} (2d_E^{-2} d_S^{-(10k+11)/4})^{4m} d_{SE}^t$  for a meaningful bound, as detailed in the Methods section on Convergence towards Markovianity.

Overall, the bound in Eq. (5) approaches concentration whenever  $d_E$  is large relative to  $d_S$  and  $k$ , together with large enough  $t$ , as shown in Fig. 3. Generally, it can be seen by inspection that the scaling in these cases will be polynomially vanishing in  $d_E$ , exponentially vanishing in  $t$  (upon appropriate choice of parameter  $m$ ), and becomes loose, polynomially in  $d_S$  and exponentially in  $k$ . Therefore, the vast majority of processes sampled from such a  $t$ -design are indistinguishable from Markovian ones in this limit. This can be intuitively understood as that for processes of small subsystems in large environments ( $d_E \gg d_S^{2k+1}$ ) undergoing complex enough dynamics (large enough  $t$ ) will look almost Markovian with high probability if the system is probed not too many times (small  $k$ ). We will now show how these processes can be modeled in terms of random circuits.



**Fig. 3 Upper bound on the probability for non-Markovianity to exceed a small amount for processes with distinct number of interventions and design dynamics against environment size.** Upper bound  $B$ , defined by Eq. (5), on  $\mathbb{P}_t[\mathcal{N} \geq 0.1]$ , the probability  $\mathbb{P}_t$  over an  $\epsilon$ -approximate  $t$ -design for the non-Markovianity  $\mathcal{N}$  to exceed  $\delta = 0.1$ , against  $\log_2(d_E)$ , where  $d_E$  is environment dimension, for a subsystem qubit undergoing a joint closed approximate unitary design interaction between a given number of interventions  $k$ . We fix an  $\epsilon = 10^{-12}$  approximate unitary  $t$ -design for different values  $2 \leq t \leq 10$  and fixed values of timesteps  $k$ , optimizing  $m$  for each case.

**Markovianization by circuit design.** While no explicit sets forming unitary  $t$ -designs for  $t \geq 4$  are known to date, several efficient constructions generating approximate unitary designs by quantum circuits are known. Using these constructions we can highlight the physical implications of the theorem above. We begin by discussing the details of one such construction. As suggested in Fig. 1b, this construction only requires simple two-qubit interactions and, under certain conditions, yields an approximate unitary design, from which we can use Eq. (5) in our main Theorem to verify that Markovianization emerges.

We focus specifically on Result 2 by Winter et al.<sup>14</sup>, reproduced in the Methods section on efficient unitary designs, where a circuit with interactions mediated by two-qubit diagonal gates with three random parameters is introduced. The intuition behind such construction is that repeated alternate applications of these diagonal gates quickly randomizes the system. Notice that this idea now fully captures the gas scenario depicted in Fig. 2, where we only have two types of random two-body interactions repeatedly occurring, and we focus on one of the particles of the gas. The detail of this construction is reproduced in the Methods section on efficient circuit unitary designs.

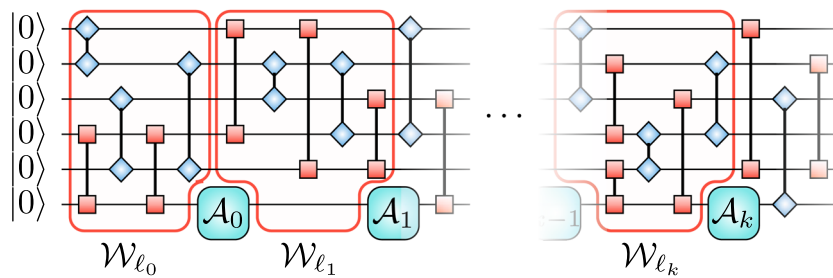
We can illustrate this idea in Fig. 4, where we depict an  $n$ -qubit SE composite with  $k$  interventions on one of the qubits, with the interactions within the circuit being only between pairs of qubits and of only two kinds; these form blocks of unitaries between each time-step  $i$  that we label  $\mathcal{W}_\ell$ , where  $\ell$  is related to the amount of two-qubit interactions as explicitly defined in Eq. (42). The main Result 2 of by Winter et al.<sup>14</sup> states that for an  $n$ -qubit system, when  $t$  is of order  $\sqrt{n}$ , a circuit  $\mathcal{W}_\ell$  yields an  $\epsilon$ -approximate unitary  $t$ -design if  $\ell \geq t - \log_2(\epsilon)/n$ , up to leading order in  $n$  and  $t$ .

Furthermore, of great relevance in this result is the fact that almost all 2-qubit gates in each repetition of  $\mathcal{W}_\ell$  can be applied simultaneously because they commute<sup>64,67</sup>. Therefore, if  $\mathcal{W}_\ell$  yields an approximate unitary design as above, the order of the non-commuting gate depth  $\mathfrak{D}$ , defined by Winter et al.<sup>67</sup> as the circuit depth when each commuting part of the circuit is counted as a single part, will coincide with the bound on the order of the number of repetitions  $\ell$ . That is, the non-commuting gate depth asymptotes to

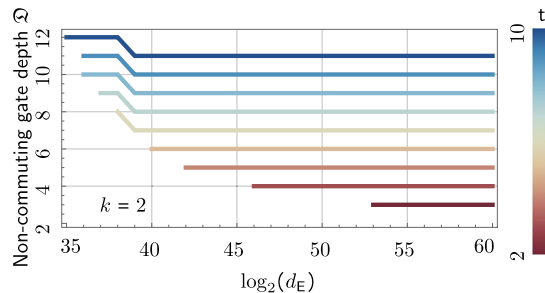
$$\mathfrak{D} \sim t - \log_2(\epsilon)/n. \tag{9}$$

We can now think of the system from the toy model of Fig. 2 as given by a spin locally interacting with a large,  $n_E$ -qubit environment, via a random time-independent Hamiltonian, with Eq. (5) statistically predicting under which conditions memory effects can be neglected. Notice that this is only a physical picture evoked by the  $\mathcal{W}_\ell$  circuits rather than exactly being the model described by it. In Fig. 5 we take such a system for a single qubit and demand a bound  $B \leq 0.01$  on the probability  $\mathbb{P}_t[\mathcal{N}_\diamond \geq 0.1]$  for a  $k = 2$  timestep process; with this, we plot the scaling of the non-commuting gate depth  $\mathfrak{D}$  required to achieve an  $\epsilon = 10^{-12}$  approximate unitary  $t$ -design using  $\mathcal{W}_\ell$  circuits for different values of  $2 \leq t \leq 10$ . While the number of 2-qubit gates is on the order of  $10^4$ , the number of repetitions  $\ell$  is at most 12 for an approximate 10-design and stays mostly constant as the number of environment qubits increases.

This construction naturally accommodates the cartoon example in Fig. 2. As long as the two interactions in the example together generate the necessary level of complexity, Markovianization will emerge. This shows, in principle, how simple dynamics described by approximate unitary designs can Markovianize under the right conditions. Moreover, taking the physical interpretation of a qubit locally interacting through two-qubit diagonal unitaries with a large environment, it also hints at how



**Fig. 4** Circuit diagram for a quantum process which can Markovianize under only two different types of 2-qubit interaction dynamics. For an  $n$ -qubit system (where each  $|0\rangle$  is a single qubit), the unitaries  $\mathcal{W}_{\ell}$ , composed of  $\ell$  alternate repetitions only two distinct types of random interactions (depicted by diamonds and squares joined by the interacting qubits), and defined by Eq. (42), generate an  $\epsilon$ -approximate unitary  $t$ -design whenever  $\ell \geq t - \log_2(\epsilon)/n$ , as shown by Winter et al.<sup>14</sup>. This can be thought of as stemming from repeated alternate applications of random 2-qubit gates diagonal in only two Pauli bases. A qubit probed with a set of operations  $\{A_i\}$  on a system undergoing  $\epsilon$ -approximate unitary  $t$ -design dynamics  $\mathcal{W}_{\ell}$  on a large environment will Markovianize for small design error  $\epsilon$  and large complexity  $t$  as specified in the main text.



**Fig. 5** Scaling of the non-commuting gate depth of the approximate unitary design by Winter et al.<sup>14</sup> for a 2-step process on a single qubit to Markovianize with respect to environment size. Scaling of the non-commuting gate depth  $D$ , given by the minimum amount of alternate repetitions  $\ell$  of the two kinds of random two-qubit diagonal gates within the unitary  $\mathcal{W}_{\ell}$ , plotted against the environment qubits  $n_E = \log_2(d_E)$ , to generate an  $\epsilon = 10^{-12}$  approximate unitary  $t$ -design for  $2 \leq t \leq 10$ . This is such that for a single-qubit system undergoing a process with  $k = 2$  timesteps, the probability  $\mathbb{P}_t$  for the non-Markovianity  $\mathcal{N}$  exceeding 0.1 is less or equal than 0.01, i.e.  $\mathbb{P}_t[\mathcal{N} \geq 0.1] \leq B \leq 0.01$ .

macroscopic systems can display Markovianization of small subsystem dynamics in circuits requiring just a small gate depth. Furthermore, for macroscopic systems with coarse observables, the same Markovianization behavior would remain resilient to a much larger number of interventions.

## Discussion

We have shown how physical quantum processes Markovianize, i.e., forget the past, for a class of physically motivated systems that can finitely approximate random ones. Forgetfulness is indeed a common feature of the world around us, and one that is crucial for doing science. Without forgetfulness, repeatability would be impossible. After all, if each carbon atom remembered its own past then it will be unique and there would be no sense in classifying atoms and molecules. Beyond these foundational considerations, our results have direct consequences for the study of open systems using standard tools, such as master equations and dynamical maps. The latter of which can be seen as a family of one-step process tensors (with initial SE correlations a minimum of two steps must be considered<sup>16,68</sup>). Specifically, our results, for the case of  $k \leq 2$ , can be used to estimate the time scale, using gate depth as a proxy, on which an approximate unitary design's open dynamics can be described (with high probability) with a truncated memory kernel<sup>2,69,70</sup>, or even a Markovian master equation.

Conversely, for larger  $k$ , our results would have implications for approximations made in computing higher order correlation functions, such as the quantum regression theorem<sup>71</sup>. These higher order approximations are independent of those at the level of dynamical maps, which can, e.g., be divisible, even when the process is non-Markovian<sup>72</sup>. This is reflected in the loosening behavior of the bound in Eq. (5) as the number of timesteps increases, which can be interpreted as a growing potential for temporal correlations to become relevant when more information about the process is accessible.

This breadth of applicability is in contrast with the results of Modi et al.<sup>9</sup>, where it was shown that quantum processes satisfy a concentration of measure with respect to Haar measure around Markovian ones, which has two main drawbacks: first, as stated above, Haar random interactions do not exist in nature and hence the relevance of the result is limited. Second, the rate of Markovianization is far too strong. Almost all processes, sampled according to the Haar measure, will simply look random, i.e., Markov order  $m = 0$  even for a large  $k$ . This, unlike our current result, misses almost all interesting physical dynamical processes. While the behaviour of our large deviation bound is polynomial, rather than exponential, thus not exhibiting concentration per-se, we have nevertheless exemplified how, with modestly large environments and relatively simple interactions, almost Markovian processes can come about with high probability. Physical macroscopic environments will be far larger than the scale shown in Figs. 3 and 5.

Despite the fundamental relevance of our result, it is well known that typicality arguments can have limited reach. For instance, the exotic Hamiltonians, introduced by Gemmer et al.<sup>73</sup>, which lead to strange relaxation, may not Markovianize even though the SE process is highly complex with a large  $E$ . There is also still significant scope for further addressing physical aspects, such as the question of whether, and how, a time-independent Hamiltonian can give rise to an approximate unitary design<sup>14</sup>, the relevant time scales of Markovianization, or the potential role of different approaches to pseudo-randomness such as that by Kastoryano et al.<sup>74</sup>, where it is shown that driven quantum systems can converge rapidly to the uniform distribution. Furthermore, a renewed wave of interest in thermalization has come along with the so-called Eigenstate Thermalization Hypothesis (ETH), which is a stronger and seemingly more fundamental condition on thermalization<sup>75–80</sup>, and we would thus expect a deep connection in the sense of ETH between Markovianization and thermalization to be forthcoming. In any case, it is clear that many physical systems Markovianize at some scale, and it only remains to discover how.

**Methods**

**The process tensor.** The Choi state representation of the process tensor is given by

$$\Upsilon = \text{tr}_E \left[ \mathbf{U}_{k;0} (\rho \otimes \Psi^{\otimes k}) \mathbf{U}_{k;0}^\dagger \right], \tag{10}$$

where each  $\Psi$  is a maximally entangled state on an ancillary space of dimension  $d_S^2$ , and where

$$\mathbf{U}_{k;0} := (U_k \otimes \mathbb{1}) S_k \cdots (U_1 \otimes \mathbb{1}) S_1 (U_0 \otimes \mathbb{1}), \tag{11}$$

is a unitary operator acting on the whole SE together with the  $2k$  ancillas. All identities act on the ancillary system, the  $U_i$  are SE unitary operators at step  $i$ , and  $S_i$  is a SWAP operator between system S and half of the  $i^{\text{th}}$  ancillary space at the  $i^{\text{th}}$  time-step of the process.

The definition in Eq. (10) is a generalization of the standard Choi state for quantum channels, as given by the Choi-Jamiolkowski isomorphism (CJI)<sup>24</sup>. The CJI for quantum channels establishes a one to one correspondence with a quantum state on a larger Hilbert space, given as the action of the channel onto half a maximally entangled state. The standard definition uses unnormalized maximally entangled states; however, here we are concerned with distinguishability of Choi states through the diamond norm in Eq. (3) and the Schatten norms in Eq. (IV B), so we avoid a normalization factor in these by normalizing the Choi states by definition. A discussion in full depth about the process tensor, its different representations and its properties and relevance is given by Modi et al.<sup>12</sup>.

As stated in the main text,  $\Upsilon$  can be efficiently described when written as a matrix product operator<sup>11,31</sup>. A matrix product operator (MPO) gets its name from the representation of an  $n$ -body operator  $\hat{O}$  as  $\hat{O} = \sum_{\{p_i, q_i\}} \hat{O}_{p_1 \dots p_n}^{q_1 \dots q_n} |p_1 \dots p_n\rangle \langle q_1 \dots q_n|$ , where the coefficients can be represented as a product of matrices,  $\hat{O}_{p_1 \dots p_n}^{q_1 \dots q_n} = \text{tr}[M_1^{p_1, q_1} M_2^{p_2, q_2} \dots M_n^{p_n, q_n}]$ . In particular, a matrix product density operator is a MPO with  $M_i^{p_i, q_i} = \sum_{\ell} A_i^{p_i, \ell} \otimes (A_i^{q_i, \ell})^\dagger$ , with the dimension of the matrices  $A_i^{p_i, \ell}$  known as the bond dimension. For the process tensor  $\Upsilon$ , these matrices generically correspond to  $M_i^{r_i, s_i} = \langle r_i | U_i | s_{i-1} \rangle \otimes \langle s_i | U_i^\dagger | s_{i-1} \rangle$  where  $|r_i\rangle$  and  $|s_i\rangle$  are subsystem S basis vectors and  $U_i$  is an SE unitary at timestep  $i$ <sup>11</sup>. This means the bond dimension of  $\Upsilon$  is  $d_E$ , which in practice should be much smaller, given that only part of the environment interacts with the system at any given time.

**A non-ambiguous measure of non-Markovianity.** As with any distinguishability measure, the non-Markovianity metric of Eq. (3) is not unique, and we choose the diamond norm for its mentioned operational significance. However, more generally, for any Schatten  $p$ -norm  $\|X\|_p := \text{tr}(|X|^p)^{1/p}$ , a similar quantity can be defined  $\mathcal{N}_p := \frac{1}{2} \min_{\Upsilon^{(M)}} \|\Upsilon - \Upsilon^{(M)}\|_p$ , as done with  $p=1$  in by Modi et al.<sup>9</sup>, whenever  $\Upsilon$  is normalized such that  $\text{tr}[\Upsilon] = \text{tr}[\Upsilon^{(M)}] = 1$ . Then, we have the hierarchy  $\mathcal{N}_1 \geq \mathcal{N}_2 \geq \dots$ , induced by that of the Schatten norms. As the black diamond norm is generally difficult to compute exactly, a particularly useful relation to Eq. (3) is  $d_E^{2k-1} \mathcal{N}_\diamond \leq \mathcal{N}_1 \leq \mathcal{N}_\diamond$ , in the sense that once any Schatten norm is known, the black diamond norm is automatically bounded.

Nevertheless, we highlight that, in general, any distinguishability measure  $\mathcal{N}$  between a process  $\Upsilon$  and the closest Markovian one  $\Upsilon^{(M)}$  will capture all non-Markovian features across multiple time steps, i.e., all multi-time phenomena and memory effects<sup>20</sup>. This is in contrast to other measures of non-Markovianity, e.g. trace-distance based measure<sup>19</sup> and other based on divisibility<sup>81</sup>, that have been proposed in recent years. In particular, all other measures relying on completely positive divisibility are only able to account for temporal correlations across at most three time-steps and are not sufficient to enforce the multi-time Markov condition<sup>82</sup>. This is even true in the classical case. Concretely, there are explicit examples of multi-time non-Markovian processes that are shown to be completely positive divisible processes, thus also deemed to be Markovian by the trace-distance based measure<sup>20,82</sup>. On the other hand, if a process satisfies the multi-time Markov condition, then it will be completely positive divisible.

In other words, the multi-time Markov condition is a stronger one that contains Markov conditions based on completely positive divisibility. This is why we consider the multi-time Markov condition in this manuscript.

**Markovian typicality.** In general, we say that a function  $f$  from a metric space  $\mathfrak{E}$  with metric  $\Delta_\mathfrak{E}$  and probability measure  $\mu_\sigma$ , to the real numbers, satisfies a concentration of measure around its mean if, for any point  $x \in \mathfrak{E}$  and any  $\delta > 0$ ,

$$\mathbb{P}_\sigma[f(x) \geq \mathbb{E}_\sigma(f) + \delta] \leq \alpha_\sigma(\delta/L), \tag{12}$$

where as done in the remainder of this manuscript,  $\mathbb{P}_\sigma$  and  $\mathbb{E}_\sigma$  explicitly refer to the probability and expectation with  $x \sim \mu_\sigma$ , and where  $L > 0$  is the so-called Lipschitz constant of  $f$ , which can be determined according to  $|f(x) - f(y)| \leq L \Delta_\mathfrak{E}(x, y)$  for any two points  $x, y \in \mathfrak{E}$ . Whenever  $L$  is small, intuitively this implies that  $f$  varies slowly in such space. Finally, the function  $\alpha_\sigma$  is called a concentration rate; it generally must be vanishing in increasing  $\delta$  in order for (12) to constitute concentration of measure, and it intuitively tells us how strong such concentration is.

Particularly well-known is the example of concentration of measure in the hypersphere of a high dimension, where for all functions that do not change too rapidly, i.e. with a small Lipschitz constant  $L$ , the function evaluated on a point picked uniformly at random will be close to its mean value with high probability, i.e. specifically  $\alpha_\sigma$  decays exponentially with  $-\delta^2$ . This is also known as Levy's lemma<sup>46</sup> and it has, remarkably, also been used by Winter et al.<sup>3</sup> to show that the fundamental theorem of statistical mechanics arises from entanglement.

Similarly, Modi et al.<sup>9</sup> showed that quantum processes satisfy a concentration of measure around Markovian ones, explaining the emergence of Markovianity without a-priori assumptions. In particular, there, the trace distance  $\mathcal{N}_1$  was used as a measure of non-Markovianity, which strictly speaking gives the distinguishability between explicitly constructed Choi states of corresponding process tensors and has no operational meaning; however, we can use the relation  $d_E^{2k-1} \mathcal{N}_\diamond \leq \mathcal{N}_1 \leq \mathcal{N}_\diamond$  to relate this to the stricter notion of non-Markovianity defined in terms of the diamond norm in Eq. (3). This implies that the main result by Modi et al.<sup>9</sup>, where all SE unitaries of Eq. (10) were randomly sampled according to the Haar measure, can be written equivalently as

$$\mathbb{P}_h[\mathcal{N}_\diamond \geq d_E^{2k+1} \mathcal{B} + \delta] \leq \exp\left\{-4 C \delta^2 d_E^{-2(2k+1)}\right\}, \tag{13}$$

where

$$C = \frac{d_{SE}(k+1)}{16} \left( \frac{d_S - 1}{d_S^{k+1} - 1} \right)^2, \tag{14}$$

is the Lipschitz constant of  $\mathcal{N}_1$ , and

$$\mathcal{B} = \begin{cases} \frac{1}{2} \sqrt{d_E \mathbb{E}_h[\text{tr}(\Upsilon^2)]} - x + \frac{y}{2} & \text{if } d_E < d_S^{2k+1} \\ \frac{1}{2} \sqrt{d_S^{2k+1} \mathbb{E}_h[\text{tr}(\Upsilon^2)]} - 1 & \text{otherwise,} \end{cases} \tag{15}$$

is an upper bound on  $\mathbb{E}_h[\mathcal{N}_1]$ , the expected non-Markovianity over the Haar measure, with  $x := d_E d_S^{-2(2k+1)}(1+y)$  and  $y := 1 - d_E d_S^{-2(2k+1)}$ , and

$$\mathbb{E}_h[\text{tr}(\Upsilon^2)] = \frac{d_E^2 - 1}{d_E(d_{SE} + 1)} \left( \frac{d_E^2 - 1}{d_E^2 - 1} \right)^k + \frac{1}{d_E}, \tag{16}$$

the expected purity, i.e., the noisiness of the process  $\Upsilon$ , over the Haar measure. Holding everything else constant, the bound  $\mathcal{B} \geq \mathbb{E}_h[\mathcal{N}_1]$  satisfies

$$\lim_{d_E \rightarrow \infty} \mathcal{B} = 0 \text{ and } \lim_{k \rightarrow \infty} \mathcal{B} = 1, \tag{17}$$

so that the expected non-Markovianity vanishes in the  $d_E \rightarrow \infty$  limit and becomes loosest in the  $k \rightarrow \infty$  limit case.

The significance of Eq. (15) is thus that quantum processes with not too many interventions in high dimensional environments will look to be almost Markovian with high probability. This means that, even when processes generically carry temporal correlations, these are typically low, explaining the emergence of Markovian processes without ad-hoc assumptions such as the Born-Markov approximation of weak coupling<sup>51</sup>.

**Unitary designs.** The result in Eq. (15) assumes that the dynamics are Haar distributed; however, implementing a Haar random unitary requires an exponential number of two-qubit gates and random bits<sup>83</sup>, thus Haar random dynamics cannot be obtained efficiently in a physical setting.

An exact unitary  $t$ -design is defined<sup>10</sup> as a probability measure  $\mu_t$  on  $\mathbb{U}(d)$  such that for all positive  $s \leq t$ , and all  $d^s \times d^s$  complex matrices  $X$ ,

$$\mathbb{E}_t[\Upsilon^{\otimes s}(X)] = \mathbb{E}_h[\Upsilon^{\otimes s}(X)], \forall s \leq t. \tag{18}$$

As per the definition in Eq. (18), a unitary  $t$ -design reproduces up to the  $t^{\text{th}}$  moment over the uniform distribution given by the Haar measure. In particular,  $\mu_t$  can consist of a finite ensemble  $\{V_i, p_i\}_{i=1}^N$  of unitaries  $V_i$  and probabilities  $p_i$ , as is now common in applications such as so-called randomized benchmarking of error rates in quantum gates<sup>60,62</sup>.

Moreover, this definition can be relaxed by letting a unitary design approximate the Haar measure with a small error  $\epsilon$ . In this manuscript we specifically employ the definition by R. Low<sup>10</sup> for unitary designs. It uses the fact that the definition of an exact  $t$ -design,  $\mu_t$ , can be written in terms of a balanced monomial  $\Theta$  of degree less or equal to  $t$  in the elements of the unitaries  $U$ . A balanced monomial of degree  $t$  is a monomial in the unitary elements with precisely  $t$  conjugated and  $t$  unconjugated elements: for example,  $U_{ab} U_{cd} U_{ef}^* U_{hg}^*$  is a balanced monomial of degree 2. Thus, writing Eq. (18) in terms of matrix elements, this can be seen to be equivalent to requiring  $\mathbb{E}_t[\Theta(V)] = \mathbb{E}_h[\Theta(U)]$  for all monomials  $\Theta$  of degree  $s \leq t$ . Similarly, for an  $\epsilon$ -approximate  $t$ -design we adopt the definition by R. Low<sup>10</sup> with Eq. (4) implying

$$\left| \mathbb{E}_t[\Theta(V)] - \mathbb{E}_h[\Theta(U)] \right| \leq \frac{\epsilon}{d^t}, \tag{19}$$

for monomials  $\Theta$  of degree  $s \leq t$ . From now on, we will focus on the more general approximate designs. We will see below that the degree  $\epsilon$  to which the distribution

of the unitary dynamics on  $\mu_t$  differs from an exact design for given  $t$  depends on the complexity of the model.

**Large deviation bounds for  $t$ -designs.** The general idea for the main result by R. Low<sup>10</sup> (similarly applied before by Horodecki et al.<sup>66</sup>) is that given a  $\mu_t$  distribution as an  $\epsilon$ -approximate unitary  $t$ -design and a concentration result for a polynomial  $\mathcal{X}$  of degree  $p$ , then one can compute the last term  $f_t$  in

$$\mathbb{E}_t \mathcal{X}^m = \mathbb{E}_h \mathcal{X}^m + f_t, \tag{20}$$

with  $m \leq t/2p$ , which will generally have a dependence  $f_t = f_t(\epsilon, t, \mathcal{X})$ . Using Markov's inequality

$$\begin{aligned} \mathbb{P}_t(\mathcal{X} \geq \delta) &= \mathbb{P}_t(\mathcal{X}^m \geq \delta^m) \\ &\leq \frac{\mathbb{E}_t \mathcal{X}^m}{\delta^m} \\ &= \frac{1}{\delta^m} [\mathbb{E}_h \mathcal{X}^m + f_t], \end{aligned} \tag{21}$$

which is the form of the main large-deviation bound.

Specifically, the results that we employ are the following, proved R. Low<sup>10</sup>.

**Theorem 2**

(Large deviation bounds for  $t$ -designs by R. Low<sup>10</sup>) Let  $\mathcal{X}$  be a polynomial of degree  $T$ . Let  $f(U) = \sum_i \alpha_i \Theta_i(U)$  where  $\Theta_i(U)$  are monomials and let  $\alpha(f) = \sum_i |\alpha_i|$ . Suppose that  $f$  has probability concentration

$$\mathbb{P}_h[|f - \zeta| \geq \delta] \leq C \exp(-C\delta^2), \tag{22}$$

and let  $\mu_t$  be an  $\epsilon$ -approximate unitary  $t$ -design, then

$$\mathbb{P}_{\mu_t}[|f - \zeta| \geq \delta] \leq \frac{1}{\delta^{2m}} \left( C \left(\frac{m}{C}\right)^m + \frac{\epsilon}{d^T} (\alpha + |\zeta|)^{2m} \right), \tag{23}$$

for any integer  $m$  with  $2mT \leq t$ .

This is the most general result providing a large-deviations bound on approximate unitary designs, where  $\zeta$  can be any quantity, in particular the expectation of  $f$ . The main idea from this result (similarly applied before by Horodecki et al.<sup>66</sup>) is that given a  $\mu_t$  distribution as an  $\epsilon$ -approximate unitary  $t$ -design and a concentration result for a polynomial  $f$  of degree  $T$ , then one can compute

$$\mathbb{E}_t[f^m] = \mathbb{E}_h[f^m] + g(\epsilon, t, f), \tag{24}$$

where  $m \leq t/2T$ . Using Markov's inequality we have

$$\mathbb{P}_t(f \geq \delta) = \mathbb{P}_t(\delta^m [\mathbb{E}_h[f^m] + g(\epsilon, t, f)]), \tag{25}$$

which is the form of the main large deviations bound in Eq. (23). More precisely, the other two main results that come along with the proof of Theorem 2 by R. Low<sup>10</sup>, and allowing to compute the right hand-side of Eq. (25) are the following.

**Lemma 3**

(3.4 of by R. Low<sup>10</sup>) Let  $\mathcal{X}$  be a polynomial of degree  $T$  and  $\zeta$  any constant. Let  $f(U) = \sum_i \alpha_i \Theta_i(U)$  where  $\Theta_i(U)$  are monomials and let  $\alpha(f) = \sum_i |\alpha_i|$ . Then for an integer  $m$  such that  $2mT \leq t$  and  $\mu_t$  an  $\epsilon$ -approximate unitary  $t$ -design,

$$\mathbb{E}_t[|f - \zeta|^{2m}] \leq \mathbb{E}_h[|f - \zeta|^{2m}] + \frac{\epsilon}{d^T} (\alpha + |\zeta|)^{2m}. \tag{26}$$

**Lemma 4**

(5.2 of by R. Low<sup>10</sup>) Let  $X$  be any non-negative random variable with probability concentration

$$\mathbb{P}(X \geq \delta + \gamma) \leq C \exp(-\mathfrak{C} \delta^2), \tag{27}$$

where  $\gamma \geq 0$ , then

$$\mathbb{E}[X^m] \leq C \left(\frac{2m}{\mathfrak{C}}\right)^{m/2} + (2\gamma)^m, \tag{28}$$

for any  $m > 0$ .

So, in essence, given these results, we determine the right-hand sides of Eq. (26) and Eq. (28) through the measure of non-Markovianity in Eq. (3) and all the other relevant quantities in such terms.

**Proof of Theorem 1**

*A bound on the Haar moments of  $\mathcal{N}_2$ .* Let us start by noticing that  $\|X\|_1 \geq \|X\|_2$ , so a concentration for  $\mathcal{N}_1$  given by  $\mathbb{P}_h[\mathcal{N}_1 \geq B + \delta] \leq \exp(-C\delta^2)$  where here  $C =$

$$\frac{d_{SE}^{k+1}}{4} \left(\frac{d_S - 1}{d_S^{k+1} - 1}\right)^2 \text{ (here 4 times the one defined in Eq. (14) in the main text), and } B$$

is defined in Eq. (15), also implies

$$\mathbb{P}_h[\mathcal{N}_2 \geq B + \delta] \leq e^{-C\delta^2}, \tag{29}$$

so that in turn Lemma 4 through Eq. (28) implies that

$$\begin{aligned} \mathbb{E}_h[\mathcal{N}_2^{2m}] &\leq \left(\frac{4m}{C}\right)^m + (2B)^{2m} \\ &= \left[\frac{16m}{(k+1)d_{SE}} \left(\frac{d_S^{k+1} - 1}{d_S - 1}\right)^2\right]^m + (2B)^{2m}, \end{aligned} \tag{30}$$

for any  $m > 0$ .

*A bound on the design moments of  $\mathcal{N}_2$ .* For the case of all unitaries at each step being independently sampled,  $\mathcal{N}_2^2$  is a polynomial of degree  $p = 2$  when the unitaries are all distinct (random interaction type). We can thus take  $\mathcal{N}_2^2$  and apply Lemma 3 for a unitary  $t$ -design  $\mu_t$  with  $t \geq 4m$ , which actually holds for real  $m > 0$ , as

$$\mathbb{E}_t[\mathcal{N}_2^{2m}] \leq \mathbb{E}_h[\mathcal{N}_2^{2m}] + \frac{\epsilon}{d_{SE}^T} \eta^{2m} \tag{31}$$

where  $\eta$  is the sum of the moduli of the coefficients of

$$\begin{aligned} \mathcal{N}_2^2 &= \left(\frac{1}{2} \min_{\Upsilon^{(M)}} \|\Upsilon - \Upsilon^{(M)}\|_2\right)^2 \\ &\leq \frac{1}{4} \|\Upsilon - \frac{1}{d_S^{2k+1}}\|_2^2 \\ &= \frac{1}{4} [\text{tr}(\Upsilon^2) - d_S^{-(2k+1)}]. \end{aligned} \tag{32}$$

The proof of Lemma 3.4 by R. Low<sup>10</sup> requires  $m$  to be an integer through the multinomial theorem; in the notation of the cited paper, this can be relaxed to be a real number by applying the multinomial theorem for a real power: convergence requires an ordering such that  $|\alpha_i \mathbb{E} M_i| > 2^{1-n} |\alpha_{t-n} \mathbb{E} M_{t-n}|$  for each  $n = 1, \dots, t - 1$  for both the approximate design and Haar expectations.

Let us explicitly write the process  $\Upsilon$ , defined in Eq. (10) in the main text, as a function of the set of unitaries  $\mathbf{U} := \{U_i\}_{i=0}^k$ , i.e.

$$\Upsilon[\mathbf{U}] = \text{tr}_E[U_k S_k \cdots U_1 S_1 U_0 (\rho \otimes \Psi^{\otimes k}) U_0^\dagger S_1 U_1^\dagger \cdots S_k U_k^\dagger], \tag{33}$$

where here implicitly  $U_\ell$  stands for  $U_\ell \otimes \mathbb{1}_{2k-\text{ancillas}}$  and the maximally entangled states  $\Psi$  are taken to be normalized. As the swaps between the system and the  $i^{\text{th}}$  half ancillary system are given by  $S_i = \sum \mathfrak{E}_{\alpha\beta} \otimes \mathbb{1} \otimes |\beta\rangle \langle \alpha| \otimes \mathbb{1}$  where  $\mathfrak{E}_{\alpha\beta} := \mathbb{1}_E \otimes |\alpha\rangle \langle \beta|_S$ , this can be written as

$$\begin{aligned} \Upsilon[\mathbf{U}] &= d_S^{-k} \sum \text{tr}_E \left[ U_k \mathfrak{E}_{\alpha_k \beta_k} \cdots U_1 \mathfrak{E}_{\alpha_1 \beta_1} U_0 \rho U_0^\dagger \mathfrak{E}_{\delta_1 \gamma_1} U_1^\dagger \cdots \mathfrak{E}_{\delta_k \gamma_k} U_k^\dagger \right] \\ &\quad \otimes |\beta_1 \alpha_1 \cdots \beta_k \alpha_k\rangle \langle \delta_1 \gamma_1 \cdots \delta_k \gamma_k|. \end{aligned} \tag{34}$$

Now, the standard approach to compute the sum of the moduli of the coefficients of a given polynomial is to evaluate on an argument (here a  $d_{SE} \times d_{SE}$  matrix) full of ones (so that all single monomials equal to one) and take each summand to the corresponding modulus. We follow this approach, however, we first notice that the environment part in Eq. (34) is just a product of the environment parts of all unitaries and initial state. To see this, let  $U = \sum U_{e's'}^{es} |es\rangle \langle e's'|$  where  $|e\rangle$  and  $|s\rangle$  are  $E$  and  $S$  bases. Unitarity then implies  $\sum \overline{U_{es}^{ab}} U_{e's'}^{ab} = \delta_{ee'} \delta_{ss'}$ , where the overline denotes complex conjugate, and so this means that  $\text{tr}_E[V \mathfrak{E}_{\alpha\beta} U \rho U^\dagger \mathfrak{E}_{\gamma\delta} V^\dagger] = \sum V_{e's'}^{es} \overline{V_{e's'}^{es}} U_{bs_2}^{e's_2} \overline{U_{bs_2}^{e's_2}} \rho_{b_1 b_1} \phi(S)$  where  $\phi(S)$  stands for the system  $S$  part; for each  $b$  index the rest of the terms are summed over  $e$ ; this generalizes similarly for any number of unitaries. This implies that at most  $d_E$  terms need to be set to one and we can evaluate  $\Upsilon$  in a set of matrices  $\mathcal{J} = \{\mathbb{1}_E \otimes J_S, \dots, \mathbb{1}_E \otimes J_S, J_E \otimes J_S\}$  with  $J$  a matrix with each element equal to one in the respective  $E$  or  $S$  systems: let  $\rho = \sum \rho_{ese's'} |es\rangle \langle e's'|$ , then

$$\begin{aligned} \Upsilon[\mathcal{J}] &= d_S^{-k} \sum \rho_{ese's'} \text{tr}[d_E J_E |e\rangle \langle e'| J_S |\alpha_k\rangle \langle \beta_k| \cdots |\alpha_1\rangle \langle \beta_1| J_S |s\rangle \langle s'| J_S |\delta_1\rangle \langle \gamma_1| \cdots |\delta_k\rangle \langle \gamma_k|] \\ &= \frac{d_E}{d_S^k} \sum \rho_{ese's'} J_S |\alpha_k\rangle \langle \beta_k| \cdots |\alpha_1\rangle \langle \beta_1| J_S |s\rangle \langle s'| J_S |\delta_1\rangle \langle \gamma_1| \cdots |\delta_k\rangle \langle \gamma_k| \\ &\quad \otimes |\beta_1 \alpha_1 \cdots \beta_k \alpha_k\rangle \langle \delta_1 \gamma_1 \cdots \delta_k \gamma_k|, \end{aligned} \tag{35}$$



and hence (we now omit the subindex **S** on the  $J$  matrices for simplicity),

$$\begin{aligned} \left(\frac{d_S^t}{d_E}\right)^2 \text{tr}[\Upsilon^2(\mathcal{J})] &= \sum_{\rho_{\text{disc}}^t \rho_{\text{disc}}^t} \text{tr}[J|\alpha_k\rangle \langle \beta_k| \cdots |\alpha_1\rangle \langle \beta_1| J|\sigma\rangle \langle \sigma'| J|\delta_1\rangle \langle \gamma_1| \cdots J|\delta_k\rangle \\ &\quad \langle \gamma_k| J^2 |\gamma_k\rangle \langle \delta_k| J \cdots |\gamma_1\rangle \langle \delta_1| J|\sigma\rangle \langle \sigma'| J|\beta_1\rangle \langle \alpha_1| J \cdots |\beta_k\rangle \langle \alpha_k| J] \\ &= d_S^2 \sum_{\rho_{\text{disc}}^t \rho_{\text{disc}}^t} \text{tr}[J|\alpha_k\rangle \langle \beta_k| \cdots |\alpha_1\rangle \langle \beta_1| J|\sigma\rangle \langle \sigma'| J|\delta_1\rangle \langle \gamma_1| \cdots \\ &\quad \langle \gamma_{k-1}| J|\delta_k\rangle \langle \delta_k| J|\gamma_{k-1}\rangle \cdots |\gamma_1\rangle \langle \delta_1| J|\sigma\rangle \langle \sigma'| J|\beta_1\rangle \langle \alpha_1| J \cdots |\beta_k\rangle \langle \alpha_k| J] \\ &= d_S^{2k+1} \sum_{\rho_{\text{disc}}^t \rho_{\text{disc}}^t} \text{tr}[J|\alpha_k\rangle \langle \beta_k| \cdots |\alpha_1\rangle \langle \beta_1| J|\sigma\rangle \langle \sigma'| J|\beta_1\rangle \langle \alpha_1| J \cdots |\beta_k\rangle \langle \alpha_k| J] \\ &= d_S^{2k+3} \sum_{\rho_{\text{disc}}^t \rho_{\text{disc}}^t} \langle \beta_k| J|\alpha_{k-1}\rangle \cdots \langle \alpha_2| J|\alpha_1\rangle \langle \beta_1| J|\sigma\rangle \langle \sigma'| J|\beta_1\rangle \langle \alpha_1| J \cdots \langle \alpha_{k-1}| J|\beta_k\rangle \\ &= d_S^{2k+5} \sum_{\rho_{\text{disc}}^t \rho_{\text{disc}}^t} \langle \beta_{k-1}| J|\alpha_{k-2}\rangle \cdots \langle \alpha_2| J|\alpha_1\rangle \langle \beta_1| J|\sigma\rangle \langle \sigma'| J|\beta_1\rangle \langle \alpha_1| J \cdots \langle \alpha_{k-1}| J|\beta_{k-1}\rangle \\ &= d_S^{2(2k+1)} \sum_{\rho_{\text{disc}}^t \rho_{\text{disc}}^t} \end{aligned} \tag{36}$$

where to obtain the second line we used the fact that  $J^n = d^{n-1}J$  for positive integers  $n$ , here applied for  $n=2$ , together with the trace over system **S** given by  $\sum \langle \gamma_k | \cdot | \gamma_k \rangle$ . This is similarly done to get the third line by  $\sum |\delta_k\rangle \langle \delta_k| = \mathbb{1}_S$ , and taking the trace summing over  $|\gamma_{k-1}\rangle$ , which can subsequently be done for all  $|\gamma_i\rangle$  and  $|\delta_i\rangle$ . For the fourth line, the cyclicity of the trace was used, followed by an identity taken by summing up over  $|\alpha_k\rangle$ , using  $J^2 = dJ$ , and taking the trace. This can be done through all remaining steps, giving the last equality. This, together with Eq. (32), implies that (now writing simply  $i, j$  for **SE** indices),

$$\begin{aligned} 4\eta &\leq d_E^2 d_S^{2(k+1)} \left(\sum |\rho_{ij}|\right)^2 + \frac{1}{d_S^{2k+1}} \\ &\leq d_E^4 d_S^{2(k+2)} \sum |\rho_{ij}|^2 + \frac{1}{d_S^{2k+1}} \\ &\leq d_E^4 d_S^{2(k+2)} + \frac{1}{d_S^{2k+1}}, \end{aligned} \tag{37}$$

where in the second line we used  $\|X\|_1^2 \leq d \|X\|_2^2$  for element-wise norms  $\|X\|_p^p = (\sum |x_{ij}|^p)$  and in the third line we used  $\|\rho\|_2^2 \leq 1$ .

**Markov's inequality on  $\mathcal{N}_\diamond$ .** As  $d_S^{2k+1} \mathcal{N}_\diamond \leq \mathcal{N}_1 \leq \sqrt{d_S^{2k+1}} \mathcal{N}_2$ , also for  $0 < m \leq t/4$ ,

$$\begin{aligned} \mathbb{P}_t[\mathcal{N}_\diamond \geq \delta] &\leq \mathbb{P}_t\left[\sqrt{d_S^{2(2k+1)}} \mathcal{N}_2 \geq \delta\right] = \mathbb{P}_t\left[\mathcal{N}_2^{2m} \geq \frac{\delta^{2m}}{d_S^{2m(2k+1)}}\right] \\ &\leq \frac{d_S^{3m(2k+1)} \mathbb{E}_t \mathcal{N}_2^{2m}}{\delta^{2m}} \leq \left(\frac{d_S^{2(2k+1)}}{\delta^2}\right)^m \left[\left(\frac{4m}{C}\right)^m + (2B)^{2m} + \frac{\epsilon}{d_{SE}^t} \eta^{2m}\right] \\ &= \left(\frac{d_S^{2(2k+1)}}{\delta^2}\right)^m \left\{ \left[\frac{16m}{(k+1)d_{SE}} \left(\frac{d_S^{k+1}-1}{d_S-1}\right)\right]^{2m} + (2B)^{2m} + \frac{\epsilon}{16^m d_{SE}^t} \left(d_E^4 d_S^{2(k+2)} + \frac{1}{d_S^{2k+1}}\right)^{2m} \right\}, \end{aligned} \tag{38}$$

where in the third line we used Markov's inequality. This concludes the proof of Theorem 1.

**Convergence towards Markovianity.** We may first examine the third and penultimate lines leading to Eq. (38) for meaningful bounds  $\mathbb{P}_t[\mathcal{N}_\diamond \geq \delta]$ . The term  $d_S^{3(2k+1)}/\delta^2$  arises from bounding the diamond norm and Markov's inequality; while  $\delta$  is arbitrary, the  $d_S^{3(2k+1)}$  could still be relevant when multiplied with  $\mathbb{E}_t \mathcal{N}_2^{2m}$ . This latter term will be small provided 1)  $C$  is large, 2)  $B$  is small and 3) the unitary design is approximate and high enough.

For 1) and 2), as detailed by Modi et al.<sup>9</sup>, we require a fixed  $k$  such that  $d_E \gg d_S^{2k+1}$ . This implies  $B \approx 0$ , so that

$$\begin{aligned} \mathbb{P}_t[\mathcal{N}_\diamond \geq \delta] &\leq \left(\frac{d_S^{3(2k+1)}}{\delta^2}\right)^m \left\{ \left[\frac{16m}{(k+1)d_{SE}} \left(\frac{d_S^{k+1}-1}{d_S-1}\right)\right]^{2m} + \frac{\epsilon}{16^m d_{SE}^t} \left(d_E^4 d_S^{2(k+2)} + \frac{1}{d_S^{2k+1}}\right)^{2m} \right\} \\ &\approx \left\{ \left[\frac{16m}{\delta^2(k+1)} \frac{d_S^{2(2k+1)}}{d_E}\right]^m + \frac{\epsilon d_E^{2m-1} d_S^{m(10k+11)-1}}{\delta^{2m} 16^m} \right\}. \end{aligned} \tag{39}$$

Now, supposing the  $t$ -design is exact, i.e.  $\epsilon=0$ , we require  $m \leq \delta^2 \frac{(k+1)d_E}{16 d_S^k}$ , together with  $m \leq t/4$ . On the other hand if  $\epsilon$  is non-zero, we require

$$\epsilon \ll \left[\delta^2 \left(\frac{2}{d_E^2 d_S^{(10k+11)/4}}\right)\right]^{4m} d_E^t d_S^t. \tag{40}$$

The choice of real  $m$  is only restricted by  $0 < m \leq t/4$ , but otherwise is arbitrary. The right-hand side of Eq. (38) is not monotonic in  $m$  over all the remaining parameters, so it won't always be optimal for some fixed choice. One may thus optimize the choice of  $m$  numerically for each particular case.

**Efficient circuit unitary designs.** As mentioned in the main text, we focus on Result 2 of Winter et al.<sup>14</sup>. To begin with, an efficient approximation for a unitary design on a system composed of  $n$ -qubits is shown by Winter et al.<sup>14</sup> for a circuit labeled  $\text{RDC}(\mathcal{I}_2)$ , where the name stands for Random Diagonal Circuit, and refers to a circuit where  $\mathcal{I}_2 = \{I_i\}$  is a set of subsets of qubit labels  $I_i \subset \{1, \dots, n\}$ ,

such that  $|I_i| = 2$ , i.e., at step  $i$ ,  $I_i$  picks a pair of qubits, to which a Pauli-Z-diagonal gate with three random parameters is applied. This construction can already be seen in the results of Winter et al.<sup>64</sup> as arising from only two types of random diagonal interactions, which can be simplified into a product of Z-diagonal ones.

A particular case which further simplifies things is then denoted by  $\text{RDC}_{\text{disc}}^{(t)}(\mathcal{I}_2)$ , where the subscript disc and the superscript  $t$  refer to discrete sets from which the parameters of the diagonal gates will be sampled, and which are determined by a given natural number  $t$ . Specifically, all gates in  $\text{RDC}_{\text{disc}}^{(t)}(\mathcal{I}_2)$  have the simplified form

$$\text{diag}\{1, e^{i\phi_1}\} \otimes \text{diag}\{1, e^{i\phi_2}\} \text{diag}\{1, 1, 1, e^{i\theta}\}, \tag{41}$$

where  $\text{diag}$  denotes Pauli-Z diagonal, and with  $\phi_1, \phi_2$  chosen independently from the discrete set  $\{2\pi m/(t+1) : m \in \{0, \dots, t\}\}$  and  $\theta$  chosen from  $\{2\pi m/(\lfloor t/2 \rfloor + 1) : m \in \{0, \dots, \lfloor t/2 \rfloor\}\}$ . We emphasise that this is still a circuit with 2-qubit diagonal gates with only three random parameters each, and therein lies its simplicity.

Now let  $H_n = H^{\otimes n}$  be  $n$  copies of the Hadamard gate, then the main Result 2 by Winter et al.<sup>14</sup> states that for an  $n$ -qubit system, when  $t$  is of order  $\sqrt{n}$ , a circuit of the form

$$\mathcal{W}_\ell := \left(\text{RDC}_{\text{disc}}^{(t)}(\mathcal{I}_2) H_n\right)^{\ell} \text{RDC}_{\text{disc}}^{(t)}(\mathcal{I}_2), \tag{42}$$

yields an  $\epsilon$ -approximate unitary  $t$ -design if

$$\ell \geq t - \log_2(\epsilon)/n, \tag{43}$$

up to leading order in  $n$  and  $t$ .

All the 2-qubit gates in each repetition of  $\mathcal{W}_\ell$ , except those in  $H_n$ , can be applied simultaneously because they commute<sup>64,67</sup>. Thus, as explained in the main text, if  $\mathcal{W}_\ell$  yields an approximate unitary design, the order of the non-commuting gate depth  $\mathfrak{D}$  will coincide with the bound on the order of the number of repetitions  $\ell$ .

### Data availability

No datasets were generated or analyzed during the current study

### Code availability

The code used in the analysis of the datasets is available from the corresponding authors on reasonable request.

Received: 25 November 2020; Accepted: 27 April 2021;

Published online: 09 June 2021

### References

- Gogolin, C. & Eisert, J. Equilibration, thermalisation, and the emergence of statistical mechanics in closed quantum systems. *Rep. Prog. Phys.* **79**, 056001 (2016).
- Breuer, H., Petruccione, F. & Petruccione, S. *The Theory of Open Quantum Systems* (Oxford University Press, 2002). <https://books.google.com.au/books?id=0Yx5VzaMYm8C>.
- Popescu, S., Short, A. J. & Winter, A. Entanglement and the foundations of statistical mechanics. *Nat. Phys.* **2**, 754–758 (2006).
- Linden, N., Popescu, S., Short, A. J. & Winter, A. Quantum mechanical evolution towards thermal equilibrium. *Phys. Rev. E* **79**, 061103 (2009).
- Gemmer, J., Michel, M. & Mahler, G. *Quantum Thermodynamics: Emergence of Thermodynamic Behavior Within Composite Quantum Systems*. Lecture Notes in Physics (Springer Berlin Heidelberg, 2009). <https://books.google.com.au/books?id=Ua5tCQAAQBAJ>.
- Vinayak & Žnidarič, M. Subsystem dynamics under random hamiltonian evolution. *J. Phys. A-Math. Theor.* **45**, 125204 (2012).
- Brandão, F. G. S. L. et al. Convergence to equilibrium under a random hamiltonian. *Phys. Rev. E* **86**, 031101 (2012).
- Masanes, L., Roncaglia, A. J. & Acín, A. Complexity of energy eigenstates as a mechanism for equilibration. *Phys. Rev. E* **87**, 032137 (2013).
- Figueroa-Romero, P., Modi, K. & Pollock, F. A. Almost Markovian processes from closed dynamics. *Quantum* **3**, 136 (2019).
- Low, R. A. Large deviation bounds for  $k$ -designs. *P. Roy. Soc. A-Math.Phys.* **465**, 3289–3308 (2009).
- Pollock, F. A., Rodríguez-Rosario, C., Frauenheim, T., Paternostro, M. & Modi, K. Non-markovian quantum processes: complete framework and efficient characterization. *Phys. Rev. A* **97**, 012127 (2018).
- Milz, S., Pollock, F. A. & Modi, K. An introduction to operational quantum dynamics. *Open Syst. Inf. Dyn.* **24**, 1740016 (2017).
- Milz, S., Sakuldee, F., Pollock, F. A. & Modi, K. Kolmogorov extension theorem for (quantum) causal modelling and general probabilistic theories. *Quantum* **4**, 255 (2020).

14. Nakata, Y., Hirche, C., Koashi, M. & Winter, A. Efficient quantum pseudorandomness with nearly time-independent hamiltonian dynamics. *Phys. Rev. X* **7**, 021006 (2017).
15. Gessner, M. et al. Local detection of quantum correlations with a single trapped ion. *Nat. Phys.* **10**, 105–109 (2014).
16. Ringbauer, M. et al. Characterizing quantum dynamics with initial system-environment correlations. *Phys. Rev. Lett.* **114**, 090402 (2015).
17. Morris, J., Pollock, F. A. & Modi, K. Non-markovian memory in IBMQX4. *arXiv:1902.07980* (2019).
18. Winick, A., Wallman, J. J. & Emerson, J. Phenomenological measure of quantum non-markovianity. *arXiv:1901.00267* (2019).
19. Breuer, H.-P., Laine, E.-M., Piilo, J. & Vacchini, B. Colloquium: Non-markovian dynamics in open quantum systems. *Rev. Mod. Phys.* **88**, 021002 (2016).
20. Pollock, F. A., Rodríguez-Rosario, C., Frauenheim, T., Paternostro, M. & Modi, K. Operational markov condition for quantum processes. *Phys. Rev. Lett.* **120**, 040405 (2018).
21. Oreshkov, O., Costa, F. & Brukner, C. Quantum correlations with no causal order. *Nat. Commun.* **3**, 1092 (2012).
22. Costa, F. & Shrapnel, S. Quantum causal modelling. *New J. Phys.* **18**, 063032 (2016).
23. Shrapnel, S., Costa, F. & Milburn, G. Updating the born rule. *New J. Phys.* **20**, 053010 (2018).
24. Watrous, J. *The Theory of Quantum Information* (Cambridge University Press, 2018).
25. Taranto, P., Pollock, F. A., Milz, S., Tomamichel, M. & Modi, K. Quantum markov order. *Phys. Rev. Lett.* **122**, 140401 (2019).
26. Taranto, P., Milz, S., Pollock, F. A. & Modi, K. Structure of quantum stochastic processes with finite markov order. *Phys. Rev. A* **99**, 042108 (2019).
27. Taranto, P., Pollock, F. A. & Modi, K. Memory strength and recoverability of non-markovian quantum stochastic processes. *arXiv:1907.12583* (2019).
28. Giarmatzi, C. & Costa, F. Witnessing quantum memory in non-markovian processes. *Quantum* **5**, 440 (2021).
29. Strasberg, P. & Diaz, M. G. Classical quantum stochastic processes. *Phys. Rev. A* **100**, 022120 (2019).
30. Milz, S. et al. When is a non-markovian quantum process classical? *Phys. Rev. X* **10** (2020). <https://doi.org/10.1103/PhysRevX.10.041049>.
31. Verstraete, F., Garcia-Ripoll, J. J. & Cirac, J. I. Matrix product density operators: simulation of finite-temperature and dissipative systems. *Phys. Rev. Lett.* **93**, 207204 (2004).
32. Gilchrist, A., Langford, N. K. & Nielsen, M. A. Distance measures to compare real and ideal quantum processes. *Phys. Rev. A* **71**, 062310 (2005).
33. Chiribella, G., D'Ariano, G. M. & Perinotti, P. Theoretical framework for quantum networks. *Phys. Rev. A* **80**, 022339 (2009).
34. Aharonov, D., Kitaev, A. & Nisan, N. Quantum circuits with mixed states. In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, STOC '98, 20–30 (Association for Computing Machinery, New York, NY, USA, 1998). <https://doi.org/10.1145/276698.276708>.
35. Paulsen, V. *Completely Bounded Maps and Operator Algebras*. Cambridge Studies in Advanced Mathematics (Cambridge University Press, 2003).
36. Goldstein, S., Lebowitz, J. L., Mastrodonato, C., Tumulka, R. & Zanghi, N. Normal typicality and von neumann's quantum ergodic theorem. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **466**, 3203–3224 (2010).
37. Lloyd, S. *Black Holes, Demons and the Loss of Coherence: how Complex Systems Get Information, and what They Do with it* (Rockefeller University, 1988). <https://books.google.com.au/books?id=rCt8NwAACAAJ>.
38. Goldstein, S., Lebowitz, J. L., Tumulka, R. & Zanghi, N. Canonical typicality. *Phys. Rev. Lett.* **96**, 050403 (2006).
39. Gemmer, J., Michel, M. & Mahler, G. *Quantum Thermodynamics: Emergence of Thermodynamic Behavior Within Composite Quantum Systems*. Lecture Notes in Physics (Springer Berlin Heidelberg, 2009). [https://books.google.com.au/books?id=NaCjwZCnI\\_gc](https://books.google.com.au/books?id=NaCjwZCnI_gc).
40. Gu, Y. *Moments of Random Matrices and Weingarten Functions*. Master's thesis, Queen's University, Ontario, Canada (2013). <http://hdl.handle.net/1974/8241>.
41. Collins, B. & Sniady, P. Integration with respect to the haar measure on unitary, orthogonal and symplectic group. *Commun. Math. Phys.* **264**, 773–795 (2006).
42. Puchała, Z. & Miszczyk, J. Symbolic integration with respect to the haar measure on the unitary groups. *B. Pol. Acad. Sci. Tech.* **65**, 21–27 (2017).
43. Page, D. N. Average entropy of a subsystem. *Phys. Rev. Lett.* **71**, 1291–1294 (1993).
44. Giraud, O. Purity distribution for bipartite random pure states. *J. Phys. A-Math. Theor.* **40**, F1053–F1062 (2007).
45. Pasquale, A. D. et al. Statistical distribution of the local purity in a large quantum system. *J. Phys. A-Math. Theor.* **45**, 015308 (2011).
46. Ledoux, M. *The Concentration of Measure Phenomenon*. Mathematical surveys and monographs (American Mathematical Society, 2005). [https://books.google.com.au/books?id=mCX\\_cWL6rqrwC](https://books.google.com.au/books?id=mCX_cWL6rqrwC).
47. Milman, V. & Schechtman, G. *Asymptotic Theory of Finite Dimensional Normed Spaces*. No. no. 1200 in Lecture Notes in Mathematics (Springer-Verlag, 1986). <https://books.google.com.au/books?id=cTnvAAAAAAAJ>.
48. Boucheron, S., Lugosi, G. & Massart, P. *Concentration Inequalities: A Nonasymptotic Theory of Independence* (OUP Oxford, 2013). <https://books.google.com.au/books?id=koNqWRlUhp0C>.
49. Garnerone, S., de Oliveira, T. R. & Zanardi, P. Typicality in random matrix product states. *Phys. Rev. A* **81**, 032336 (2010).
50. Garnerone, S., de Oliveira, T. R., Haas, S. & Zanardi, P. Statistical properties of random matrix product states. *Phys. Rev. A* **82**, 052312 (2010).
51. Schlosshauer, M. *Decoherence: and the Quantum-To-Classical Transition*. The Frontiers Collection (Springer Berlin Heidelberg, 2007). <https://books.google.com.au/books?id=URAchzQlTgC>.
52. Guhr, T., Müller-Groeling, A. & Weidenmüller, H. A. Random-matrix theories in quantum physics: common concepts. *Phys Rep* **299**, 189 – 425 (1998).
53. D'Alessio, L., Kafri, Y., Polkovnikov, A. & Rigol, M. From quantum chaos and eigenstate thermalization to statistical mechanics and thermodynamics. *Adv Phys* **65**, 239–362 (2016).
54. Mehta, M. *Random Matrices*. ISSN (Elsevier Science, 2004). [https://books.google.com.au/books?id=Kp3Nxx03\\_gmWC](https://books.google.com.au/books?id=Kp3Nxx03_gmWC).
55. Roberts, D. A. & Yoshida, B. Chaos and complexity by design. *J. High Energy Phys.* **2017**, 121 (2017).
56. Emerson, J., Alicki, R. & Życzkowski, K. Scalable noise estimation with random unitary operators. *J. Opt. B: Quantum S. O.* **7**, S347–S352 (2005).
57. Dankert, C., Cleve, R., Emerson, J. & Livine, E. Exact and approximate unitary 2-designs and their application to fidelity estimation. *Phys. Rev. A* **80**, 012304 (2009).
58. Gross, D., Audenaert, K. & Eisert, J. Evenly distributed unitaries: on the structure of unitary designs. *J. Math. Phys.* **48**, 052104 (2007).
59. Harrow, A. W. & Low, R. A. Random quantum circuits are approximate 2-designs. *Commun. Math. Phys.* **291**, 257–302 (2009).
60. Dankert, C., Cleve, R., Emerson, J. & Livine, E. Exact and approximate unitary 2-designs and their application to fidelity estimation. *Phys. Rev. A* **80**, 012304 (2009).
61. Nakata, Y. & Murao, M. Diagonal unitary 2-designs and their implementations by quantum circuits. *Int. J. Quantum Inf.* **11**, 1350062 (2013).
62. Wallman, J. J. & Flammia, S. T. Randomized benchmarking with confidence. *N. J. Phys.* **16**, 103032 (2014).
63. Webb, Z. The clifford group forms a unitary 3-design (2015).
64. Nakata, Y., Hirche, C., Morgan, C. & Winter, A. Unitary 2-designs from random X- and Z-diagonal unitaries. *J. Math. Phys.* **58**, 052203 (2017).
65. Zhu, H. Multiqubit clifford groups are unitary 3-designs. *Phys. Rev. A* **96**, 062336 (2017).
66. Brandão, F. G. S. L., Harrow, A. W. & Horodecki, M. Local random quantum circuits are approximate polynomial-designs. *Commun. Math. Phys.* **346**, 397–434 (2016).
67. Nakata, Y., Hirche, C., Morgan, C. & Winter, A. Decoupling with random diagonal unitaries. *Quantum* **1**, 18 (2017).
68. de Vega, I. The quantum dynamical map of the spin boson model (2020).
69. Cerrillo, J. & Cao, J. Non-markovian dynamical maps: numerical processing of open quantum trajectories. *Phys. Rev. Lett.* **112**, 110401 (2014).
70. Pollock, F. A. & Modi, K. Tomographically reconstructed master equations for any open quantum dynamics. *Quantum* **2**, 76 (2018).
71. Guarnieri, G., Smirne, A. & Vacchini, B. Quantum regression theorem and non-markovianity of quantum dynamics. *Phys. Rev. A* **90**, 022110 (2014).
72. Milz, S., Kim, M. S., Pollock, F. A. & Modi, K. Completely positive divisibility does not mean markovianity. *Phys. Rev. Lett.* **123**, 040401 (2019).
73. Knipschild, L. & Gemmer, J. Modern concepts of quantum equilibration do not rule out strange relaxation dynamics. *Phys. Rev. E* **101**, 062205 (2020).
74. Banchi, L., Burgarth, D. & Kastoryano, M. J. Driven quantum dynamics: Will it blend? *Phys. Rev. X* **7**, 041015 (2017).
75. Deutsch, J. M. Quantum statistical mechanics in a closed system. *Phys. Rev. A* **43**, 2046–2049 (1991).
76. Srednicki, M. Chaos and quantum thermalization. *Phys. Rev. E* **50**, 888–901 (1994).
77. Srednicki, M. The approach to thermal equilibrium in quantized chaotic systems. *J. Phys. A Math. General* **32**, 1163–1175 (1999).
78. D'Alessio, L., Kafri, Y., Polkovnikov, A. & Rigol, M. From quantum chaos and eigenstate thermalization to statistical mechanics and thermodynamics. *Adv. Phys.* **65**, 239–362 (2016).
79. Murthy, C. & Srednicki, M. Bounds on chaos from the eigenstate thermalization hypothesis. *Phys. Rev. Lett.* **123**, 230606 (2019).

80. Brenes, M., Pappalardi, S., Goold, J. & Silva, A. Multipartite entanglement structure in the eigenstate thermalization hypothesis. *Phys. Rev. Lett.* **124**, 040605 (2020).
81. Rivas, Á., Huelga, S. F. & Plenio, M. B. Quantum non-markovianity: characterization, quantification and detection. *Rep. Prog. Phys.* **77**, 094001 (2014).
82. Milz, S., Kim, M. S., Pollock, F. A. & Modi, K. Completely positive divisibility does not mean markovianity. *Phys. Rev. Lett.* **123**, 040401 (2019).
83. Knill, E. Approximation by quantum circuits (1995).

### Acknowledgements

P.F.R. is supported by the Monash Graduate Scholarship (MGS) and the Monash International Postgraduate Research Scholarship (MIPRS). K.M. is supported through Australian Research Council Future Fellowship FT160100073 and Discovery Project DP210100597. K.M. thanks Andreas Winter for an insightful discussion and especially Austin Gleeson for highlighting the forgetfulness of nature some two decades ago.

### Author contributions

KM and FAP defined the research topic and supervised the research. PFR performed the analytical calculations and plot visualizations. All authors contributed to the theoretical aspects, interpretation of results and writing of the article.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42005-021-00629-w>.

**Correspondence** and requests for materials should be addressed to P.F.–R. or K.M.

**Peer review information** *Communications Physics* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021