Memory compression and thermal efficiency of quantum implementations of nondeterministic hidden Markov models

Thomas J. Elliott ![ORCID]*

Department of Mathematics, Imperial College London, London SW7 2AZ, United Kingdom

Stochastic modeling is an essential component of the quantitative sciences, with hidden Markov models (HMMs) often playing a central role. Concurrently, the rise of quantum technologies promises a host of advantages in computational problems, typically in terms of the scaling of requisite resources such as time and memory. HMMs are no exception to this, with recent results highlighting quantum implementations of deterministic HMMs exhibiting superior memory and thermal efficiency relative to their classical counterparts. In many contexts, however, nondeterministic HMMs are viable alternatives; compared to them, the advantages of current quantum implementations do not always hold. Here, we provide a systematic prescription for constructing quantum implementations of nondeterministic HMMs that reestablish the quantum advantages against this broader class. Crucially, we show that whenever the classical implementation suffers from thermal dissipation due to its need to process information in a time-local manner, our quantum implementations will both mitigate some of this dissipation and achieve an advantage in memory compression.

I. INTRODUCTION

Hidden Markov models (HMMs) provide a powerful representative tool for modeling stochastic systems [1]. They are able to generate a diverse range of complex, non-Markovian stochastic processes, finding applications in a broad spectrum of fields, including speech recognition [2], time-series analysis [3,4], cryptanalysis [5], machine learning [6–8], bioinformatics [9–11], economics [12], and statistical physics [13]. Given their near-universal deployment across the quantitative sciences, they form an essential topic of research in their own right.

Akin to how quantum mechanics revolutionized physics in the early 20th century, quantum information processing promises to do the same for computational science [14]. Technologies based on this paradigm offer a means to implement better, faster, more efficient algorithms and protocols [15]. Naturally, this has spurred investigations into quantum extensions of HMMs, including characterization of their expressivity [16–21], how they can be inferred [22–24], and how they can outperform classical automata [25,26].

Two key areas where quantum HMMs manifest advantages are in stochastic simulation and information engines. *Quantum implementations* of deterministic HMMs offer memory compression in the former [27–34] and greater thermal efficiency in the latter [35], compared to the corresponding classical implementations. The restriction of determinism is sometimes a necessity: nondeterministic consumers of information reservoirs have been shown to be capable of violating the second law of information thermodynamics [36]; and adaptive systems must typically act in a causal man-

ner [37,38]. In many other contexts, however, the broader class of nondeterministic HMMs can be employed, against which the current quantum advantages do not hold. Here, we generalize quantum implementations to encompass them and recover the advantages.

We introduce a systematic approach to constructing quantum implementations of nondeterministic HMMs, showing that memory and thermodynamical advantages arise for almost any HMM. Our findings strengthen, extend, and supersede analogous prior results [33,35,39], which were limited in scope by considering an incomplete suite of possible quantum implementations.

II. PROCESSES, GENERATORS, AND IMPLEMENTATIONS

A bi-infinite, discrete-time, discrete-event stochastic process [40] is characterized by a sequence of random variables X_t taking on values $x_t \in \mathcal{X}$, with the index labeling the time step $t \in \mathbb{Z}$. Throughout, we will use uppercase to denote random variables and lowercase for the corresponding variates. A contiguous block of the sequence of length L is denoted $X_{t:t+L} := X_t \ldots X_{t+L-1}$; a process $\mathcal{P}$ is then defined by the distribution $P(X_{-\infty:\infty})$. We restrict our attention to stationary stochastic processes, wherein $P(X_{0:L}) = P(X_{t:t+L}) \forall L, t \in \mathbb{Z}$. We take the index $t = 0$ to represent the present time step, such that $X_{-\infty:0}$ is the past of the process and $X_{0:\infty}$ is its future.

We consider HMMs [41] that give rise to such processes. Here we will focus on edge-emitting (or Mealy) HMMs, while noting that our results can be readily adapted to state-emitting (or Moore) HMMs [42]. An edge-emitting HMM $(\mathcal{S}, \mathcal{X}, \{T_{s's}^x\})$ consists of a (potentially infinite) set of states $s \in \mathcal{S}$, an alphabet of symbols $\mathcal{X}$, and a transition structure $\{T_{s's}^x := P(s', x|s)\}$ describing the probability of a system in

*physics@tjelliott.net

state s transitioning to state s' while emitting symbol x. We deal with irreducible HMMs, such that any state $s' \in \mathcal{S}$ can be reached by any state $s \in \mathcal{S}$ with nonzero probability after a sufficient number of transitions; this can be diagnosed by verifying that the matrix $T_{s's} = \sum_x T^x_{s's}$ is irreducible. An irreducible HMM has a unique stationary state $\pi(s)$.

The emitted symbols of a HMM generate a stochastic process, described by the distribution

$$P(x_{0:L}) := \sum_{s_0 \dots s_L} T^{x_{L-1}}_{s_L s_{L-1}} \dots T^{x_0}_{s_1 s_0} \pi(s_0). \qquad (1)$$

We refer to the HMM as a *generator* of the process. An important class of generators are for *deterministic* (sometimes referred to as *unifilar*) HMMs [42], for which the end state of a transition is uniquely determined by the start state and emitted symbol, i.e., for any given s and x, $T^x_{s's}$ is nonzero for, at most, one s'. When the HMM is deterministic, we say that the generator is *predictive*. The counterparts to this are *co-unifilar* HMMs and *retrodictive* generators, for which the start state is uniquely determined by the end state and emitted symbol [43].

With classical information processing, a physical instantiation of a HMM must encode $\{s\}$ as distinguishable states $\{|s\rangle\}$ of a system, while with quantum information processing, we can map them to (generally) nonorthogonal quantum states $\{s\} \rightarrow \{|\sigma_s\rangle\}$ [33].

Definition 1. (Quantum implementations). A quantum implementation $(\{|\sigma_s\rangle\}, \Phi)$ of a generator $(\mathcal{S}, \mathcal{X}, \{T^x_{s's}\})$ is a set of quantum states $\{|\sigma_s\rangle\}$ and a quantum channel Φ which satisfy

$$\Phi(|\sigma_s\rangle\langle\sigma_s|) = \sum_{s'x} T^x_{s's} |\sigma_{s'}\rangle\langle\sigma_{s'}| \otimes |x\rangle\langle x|. \qquad (2)$$

Being defined with respect to a generator [35], this is a subtly different notion to that of a quantum model [27], which is defined with respect to a process, though a quantum implementation also represents a quantum model of the process generated by the generator it implements.

III. MEMORY COMPRESSION

One way the performance of a generator and its implementation can be evaluated is in terms of the amount of memory it requires. Two key quantifiers of this are the size of the memory state space and the amount of information it stores. With respect to a generator g, we denote [33,44–46]

$$\begin{aligned} D_g &:= \log_2[\mathrm{rank}(\rho)], \\ C_g &:= -\mathrm{Tr}(\rho \log_2[\rho]), \end{aligned} \qquad (3)$$

where $\rho = \sum_s \pi(s)|\sigma_s\rangle\langle\sigma_s|$ is the stationary state of the memory. For classical implementations, these reduce to the (logarithm of) the number of states and the Shannon entropy of the stationary distribution $\pi(s)$, respectively.

A particular privileged generator of a process is the ε-machine of computational mechanics [42,44,47], representing the predictive generator that can be classically implemented with the minimal amount of memory according to both measures of Eq. (3). The corresponding measures D_μ and C_μ are known as the topological and statistical complexity. Re-

cently, a growing body of work has established that quantum implementations of ε-machines (with memory costs D_q and C_q) can generally undercut this minimality [27–34]: $D_q \leqslant D_\mu$ and $C_q \leqslant C_\mu$, with the compression advantage sometimes able to grow unboundedly large [30,32,46,48–51]. The present state-of-the-art quantum implementations [33] are defined implicitly through a unitary interaction,

$$U|\sigma_s\rangle|0\rangle = \sum_x \sqrt{P(x|s)}e^{i\varphi_{sx}}|\sigma_{\lambda(s,x)}\rangle|x\rangle, \qquad (4)$$

where $\{\varphi_{sx} \in \mathbb{R}\}$ are free parameters (to be optimized over) and $\lambda(s, x)$ is the predictive update rule that determines the subsequent memory state. Decoherence of the output register in the computational basis (by, e.g., a nonselective measurement or coupling to an external ancilla) results in a channel of the form of Eq. (2). However, this construction is valid only for predictive generators.

While there are many scenarios where one would require predictive behavior, such a restriction is not necessary when one is purely concerned with generating a stochastic process. Pertinently, for some processes, generators based on nondeterministic HMMs can be classically implemented with less memory than the ε-machine [3,45,52]: $D_g < D_\mu$ and $C_g < C_\mu$. It behooves us to ask whether analogous quantum implementations of nonpredictive generators can be constructed and whether they exhibit similar memory compression advantages. We now answer both these questions in the affirmative.

We begin by assuming that preprocessing has been performed on the classical generator to prune any redundancy due to states with equivalent future morphs. That is, if two states $s, s' \in \mathcal{S}$ satisfy $P(X_{0:\infty}|S_0 = s) = P(X_{0:\infty}|S_0 = s')$, then we merge one of the states into the other. This is repeated until each of the remaining states possesses unique conditional future distributions. Such redundancies are fully classical in nature and can readily be remedied by such classical means.

The next step is to consider *why* the construction given by Eq. (4) does not work for nonpredictive generators. Simply replacing the predictive update rule with a sum over all possible end states weighted by their conditional probabilities will not yield the correct statistics. This is due to the lack of decoherence between different memory states in the outcome of the interaction: while the output register undergoes decoherence, superpositions of different end memory states given this output and the initial memory state are preserved; this leads to interference between the amplitudes of their future morphs, in turn corrupting the statistics. To remedy this defect, we must introduce a mechanism to break the coherence.

This can be achieved through the introduction of an auxiliary system that is imprinted with information about the end state into which the system transitions and is subsequently discarded into the environment. The information to be encoded into this system has some freedom, and different encodings will yield different levels of memory compression. Here we will take a direct approach and encode the label of the end memory state; we discuss other possibilities later. The analog of the interaction given by Eq. (4) for our implementation becomes

$$U|\sigma_s\rangle|0\rangle|0\rangle = \sum_{s'x} \sqrt{T^x_{s's}}|\sigma_{s'}\rangle|x\rangle|s'\rangle, \qquad (5)$$

where we have set all $\{\varphi\}$ to zero. After discarding the auxiliary space and decohering the output space, we obtain a channel of the form Eq. (2). We emphasize that while the information discarded into the auxiliary system is classical, the system itself must nevertheless be quantum, as it becomes entangled with the memory states. Intuitively, one can understand why the introduction of this system works by analogy with purification [14]; the evolution essentially creates the purified form of the target memory state distribution, such that when the auxiliary system is traced out, the corresponding marginal in the memory system is as desired. That is, the auxiliary system provides a precise mechanism for breaking apart superpositions of memory states.

We can express the channel Φ_{qg} in terms of Kraus operators [14] $\{L_s^x\}$,

$$\Phi_{qg}(\rho) := \sum_{s'x} L_{s'}^x \rho L_{s'}^{x\,\dagger}, \qquad (6)$$

where $L_{s'}^x |\sigma_s\rangle := \sqrt{T_{s's}^x} |\sigma_{s'}\rangle |x\rangle$. We can also break this down into a set of Kraus operators $\{K_s^x\}$ acting purely on the memory subspace, defined through $L_s^x := K_s^x \otimes |x\rangle$. Armed with a viable quantum implementation of a nondeterministic HMM, we can now evaluate its memory compression advantage. We use C_{qg} and D_{qg} to represent the memory cost of the quantum implementation of a generator g, and, similarly, C_{cg} and D_{cg} for the classical implementation.

Theorem 1. (Memory compression advantage). The quantum implementation $(\{|\sigma_s\rangle\}, \Phi_{qg})$ of Eq. (6) achieves a memory compression advantage relative to the classical implementation $C_{qg} < C_{cg}$ whenever the generator $(\mathcal{S}, \mathcal{X}, \{T_{s's}^x\})$ is nonretrodictive.

From Eq. (5), we can use the unitarity of U to find the overlaps of the quantum memory states,

$$
\begin{aligned}
\langle \sigma_j | \sigma_k \rangle &= \langle \sigma_j | U^\dagger U | \sigma_k \rangle \\
&= \sum_{j'k'xy} \sqrt{T_{j'j}^x T_{k'k}^y} \langle \sigma_{j'} | \sigma_{k'} \rangle \langle x|y \rangle \langle j'|k' \rangle \\
&= \sum_{j'x} \sqrt{T_{j'j}^x T_{j'k}^x}. \qquad (7)
\end{aligned}
$$

Thus, there is a nonzero overlap between any pair of quantum states iff both states have at least one possible outgoing transition for which the symbol and end state are identical. With these overlaps, we can then use the Gram matrix [53] of the stationary state to determine the memory costs of the implementation. The Gram matrix of the stationary state ρ is given by $\rho_G := \sum_{ss'} \sqrt{\pi(s)\pi(s')} \langle \sigma_s | \sigma_{s'} \rangle |s\rangle \langle s'|$ and possesses the same spectrum as ρ. Thus, by determining the spectrum of ρ_G, we are able to calculate the memory costs given by Eq. (3). Further, it is evident that the diagonal elements of ρ_G are given by $\pi(s)$, and so the classical stationary state is obtained from the Gram matrix by a projection in the classical memory state basis. Since projective measurements can never decrease any Rényi entropy [54], we can immediately conclude that $D_{qg} \leqslant D_{cg}$ and $C_{qg} \leqslant C_{cg}$: the quantum implementation never performs worse than the classical in terms of memory compression.

Moreover, for any Rényi entropy other than $\alpha = 0$ (corresponding to D_g), projective measurements *strictly* increase the

entropy unless the projection leaves the state unchanged. Such a projection will preserve the state iff the off-diagonals in the projection basis are all zero. Thus, for the stationary states of our quantum implementations, there is a strict memory compression advantage $C_{qg} < C_{cg}$ iff there is at least one pair of states $(s, s' \neq s) \in \mathcal{S}^2$ for which $\langle \sigma_s | \sigma_{s'} \rangle \neq 0$. For no such pair to exist, there must be no two states in the generator that share a common transition in which the symbol and end state are the same. If there are no such states, then, given the symbol and end state of a transition, it is possible to specify the start state with certainty; this is precisely the definition of a retrodictive generator. Thus, there is no memory compression advantage to a quantum implementation of a generator iff it is retrodictive.

From the overlaps given by Eq. (7), it is possible to decompose the quantum memory states $\{|\sigma_s\rangle\}$ and U into an orthogonal basis through a reverse Gram-Schmidt procedure [55], analogous to methods in the case of deterministic generators [31].

IV. THERMAL EFFICIENCY

There is a deep physical connection between information theory and thermodynamics [56], first hinted at by Maxwell's demon [57] and Szilard's engine [58], and captured more formally by Landauer's work on the dissipative costs of irreversible computation [59]. Central to this is the information processing second law [60,61], which—mirroring its thermodynamical namesake—places bounds on the entropic costs of changing a system's configuration. In the present context, the structured pattern produced by a generator outputting a stochastic process forms an information reservoir that can be harvested as a work source [62–64]. The work cost involved in producing the pattern depends on the generator and its implementation; we now show that our quantum implementations can also offer enhanced thermal efficiency. This complements and extends recent results on the thermodynamics of quantum implementations of predictive generators [35,39].

Fundamentally, the information processing second law mandates that the work cost for any generator to produce a pattern is bounded by the entropy of the pattern. That is, no generator can achieve a work cost $W < W_{\min} := -k_B T \ln[2] h_\mu$, where $h_\mu := \lim_{t \to \infty} H(X_{0:t})/t$ is the process' entropy rate. However, typical generators incur additional dissipative costs beyond this bound—a so-called *modularity* or *locality dissipation*—due to the generator acting in a time-local manner to produce the pattern [65]. An implementation of a generator is said to be *thermally inefficient* if this additional cost is nonzero [39].

In Appendix A, we show that results previously derived for the cost of quantum implementations of predictive generators [35] can be directly applied to our quantum implementations of general generators. In particular, for an asymptotically large ensemble of identical implementations of a generator g acting independently in parallel, the average work cost per time step per implementation is given by

$$W_{qg} = k_B T \ln[2][I(S';X) - H(X)], \qquad (8)$$

where $I(S';X)$ is the (quantum) mutual information between the two subspaces of the state, $\sum_{s,s',x} \pi(s) T_{s's}^x |\sigma_{s'}\rangle \langle \sigma_{s'}| \otimes$

$|x\rangle\langle x|$. An equivalent expression holds for the work cost W_{cg} of the classical implementation. With this, we are able to establish a clear link between memory compression and thermal efficiency.

Theorem 2. (Thermal inefficiency allows quantum advantages). Iff the classical implementation of a generator $(\mathcal{S}, \mathcal{X}, \{T^x_{s's}\})$ is thermally inefficient, the quantum implementation $(\{|\sigma_s\rangle\}, \Phi_{qg})$ of Eq. (6) will both exhibit a memory compression advantage $C_{qg} < C_{cg}$ and mitigate at least some of the thermal inefficiency, $W_{qg} < W_{cg}$.

For both the classical and quantum implementations, the second term of Eq. (8) is identical, as it is a function of the process only. Thus, any thermodynamical advantage must be found by a reduction in the first term alone. The data processing inequality [66] informs us that the mutual information between two variables is nonincreasing under local transformations of one of the variables, and leaves it unchanged iff the transformation is reversible. The mutual information terms of the two implementations are linked by the mapping $\{|s\rangle \to |\sigma_s\rangle\}$ on the memory subsystem; crucially, this mapping is reversible iff the quantum memory states are all mutually orthogonal—i.e., they are all nonoverlapping. This is identical to the condition for when memory compression cannot be achieved. Thus, iff $C_{qg} = C_{cg}$, then $W_{qg} = W_{cg}$, and

$$C_{qg} < C_{cg} \Leftrightarrow W_{qg} < W_{cg}. \tag{9}$$

Finally, it has previously been shown that a classical implementation has no thermal inefficiency iff the generator is retrodictive [36,39,65]. Since the quantum implementation of any nonretrodictive generator achieves memory compression, they must also reduce the thermal inefficiency, by Eq. (9).

It is natural to then ask whether quantum implementations can bypass the locality dissipation entirely and, if so, under what conditions. The answer provides a complementary view to the previous theorem.

Theorem 3. (Noneradication of dissipation). A quantum implementation $(\{|\sigma_s\rangle\}, \Phi_{qg})$ of a generator $(\mathcal{S}, \mathcal{X}, \{T^x_{s's}\})$ that operates with no locality dissipation can be simulated by a classical implementation at no additional memory cost. Any reduction of thermal inefficiency by a quantum implementation is only mitigation, not eradication; any thermally efficient quantum implementation yields no memory compression advantage.

The proof is given in Appendix B; it utilizes a framework developed in proving a similar statement for quantum implementations of predictive generators [39], effectively condensing and generalizing this earlier derivation. The crux of this generalized theorem is that *no* quantum implementation of *any* generator can simultaneously achieve an advantage in memory compression and perfect thermal efficiency. This highlights that optimizing efficiency in the modeling of stochastic processes involves a trade-off: generators with minimally dissipative classical implementations are often not the most memory efficient—this frustration extends into the quantum domain.

V. EXAMPLE: SIMPLE NONUNIFILAR SOURCE

As an example, we consider the so-called simple nonunifilar source (SNS) process, which allows us to compare and
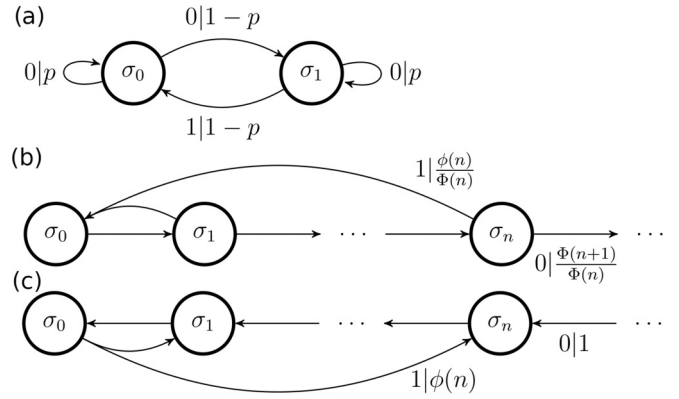

FIG. 1. Different possible HMM generators of the Simple Nonunifilar Source process. (a) Two state nondeterministic generator, (b) minimal memory predictive generator, and (c) minimal retrodictive generator. Notation $x|p$ denotes that with probability p the indicated transition occurs accompanied by the emission of symbol x.

contrast memory and work costs of classical and quantum implementation deterministic and nondeterministic generators, and explore some of these trade-offs between thermal efficiency and compression. Depicted in Fig. 1, we will look at (A) a two-state nondeterministic generator of the process, (B) the minimal memory predictive generator (i.e., the ε-machine), and (C) the minimal retrodictive generator. As a measure of memory, we will consider the Shannon–von Neumann entropy of the steady state, which we denote by C_{MG}, where $G = \{A, B, C\}$ indicates which generator the measure is with respect to, and $M = \{c, q\}$ indicates whether the implementation is classical or quantum. We will also consider the dimension of the implementations' memories and the work costs W_{MG}. We will predominantly focus on the case where $p = 1/2$, and later remark on the general case.

For the two-state generator A, it is clear from inspection that the two memory states have equal steady-state occupation probabilities. Thus, we can immediately conclude $C_{cA} = 1$. For the quantum implementation, we have

$$U|\sigma_0\rangle|0\rangle|0\rangle = \frac{1}{\sqrt{2}}|\sigma_0\rangle|0\rangle|0\rangle + \frac{1}{\sqrt{2}}|\sigma_1\rangle|0\rangle|1\rangle,$$

$$U|\sigma_1\rangle|0\rangle|0\rangle = \frac{1}{\sqrt{2}}|\sigma_0\rangle|1\rangle|0\rangle + \frac{1}{\sqrt{2}}|\sigma_1\rangle|0\rangle|1\rangle. \tag{10}$$

We can use that $U^\dagger U = 1$ to determine that $\langle\sigma_0|\sigma_1\rangle = 1/2$. Without loss of generality, we can then assign $|\sigma_0\rangle = |0\rangle$ and $|\sigma_1\rangle = (1/2)|0\rangle + (\sqrt{3}/2)|1\rangle$. The steady state of the memory is given by $(1/2)(|\sigma_0\rangle\langle\sigma_0| + |\sigma_1\rangle\langle\sigma_1|)$, from which we determine $C_{qA} = 0.811$, yielding a quantum compression advantage. Note that here and throughout the example, all numerical values are given to three significant figures.

From Eq. (A4), the work cost is given by $W = H(S') - H(S'X)$, where X is the emitted symbol, S' is the state of the system after emission, and we have normalized $k_B T \ln[2] = 1$. The first term corresponds to the entropic memory C, and so we need only calculate the entropy of the combined final state of a transition with the output. For generator A, it can be straightforwardly deduced that $P(0, \sigma_0) = 1/4, P(1, \sigma_0) =$

1/4, and $P(0, \sigma_1) = 1/2$. This allows us to calculate that $W_{cA} = -0.5$ and $W_{qA} = -0.558$, thus also yielding a quantum thermal advantage. Note that the work costs are negative as we are transforming a blank tape containing no information to a stochastic pattern with nonzero entropy.

Next, we look at the minimal memory predictive generator B. Let us first remark on the drastic difference in the number of states in the model—an infinite number—compared to the two of the nondeterministic generator A. This emphasizes the potential power of turning to nondeterministic generators generally for compressing the memory dimension. To aid us in calculating the memory and work costs of generator B, we note that the SNS is a renewal process, allowing results from Refs. [30,67] to be put to use. Let us introduce some quantities from the modeling of renewal processes. The wait-time distribution $\phi(n)$ is defined as the probability that after a 1 is seen, there are n consecutive 0's before the next 1. The survival probability $\Phi(n) := \sum_{n'=n}^{\infty} \phi(n')$ describes the probability that the number of 0's between a consecutive pair of 1's is at least n. Finally, the mean firing rate is defined as $\mu := 1/\sum_{n=0}^{\infty} \Phi(n)$. In previous works, it has been established that the steady-state probability distribution of the states σ_n is given by $P(\sigma_n) = \mu\Phi(n)$, and that the states of the corresponding quantum implementation can be encoded as

$$|\sigma_n\rangle = \sum_{n'=0}^{\infty} \frac{\sqrt{\phi(n'+n)}}{\sqrt{\Phi(n)}} |n'\rangle.$$

Moreover, it has been shown that for the SNS, $\phi(n) = np^{n-1}(1-p)^2$.

With all this at hand, we can calculate $C_{cB} = 2.71$ and $C_{qB} = 0.386$. We observe that this quantum implementation has a lower entropic memory cost than that of generator A, exemplifying that different quantum implementations would be favored depending on whether the dimension or the entropic cost of memory is prioritized [33]. To calculate the work cost, we use that $P(1, \sigma_0) = \sum_n P(\sigma_n)P(1, \sigma_0|\sigma_n) = \mu \sum_n \phi(n) = \mu$ and $P(0, \sigma_n) = P(\sigma_{n-1})P(0|\sigma_{n-1}) = \mu\Phi(n)$ for $n = 1 : \infty$. Then, $W_{cB} = 0$ and $W_{qB} = -0.468$. Thus, while the quantum implementation of B has a lower entropic cost than the implementations of A, it has a higher work cost than both—highlighting that though quantum compression and thermal advantages go hand in hand relative to a classical implementation of the same generator, there is no such definitive hierarchy between quantum implementations of different generators.

Finally, for completeness, we consider the retrodictive generator C, which should exhibit no locality dissipation. As the SNS is a renewal process, the classical implementation of generator C bears the same entropic memory cost as that of the minimal deterministic generator B [67]. Moreover, Theorem 1 tells us that there is no quantum compression advantage for a retrodictive generator. Thus, $C_{cC} = C_{qC} = 2.71$. From Theorems 2 and 3, we can also deduce $W_{qC} = W_{cC}$; we now only need to calculate this latter quantity. We have that $P(1, \sigma_n) = P(\sigma_n|1)P(1) = \mu\phi(n)$ and $P(0, \sigma_n) = P(\sigma_{n+1}) = \mu\Phi(n+1)$. Thus, we find $W_{cC} = W_{qC} = -0.678$. We can verify that this saturates the bound set by the information processing second law by checking that it is equal to the (negative of the) entropy rate of the process. This can be calculated from the deterministic generator B through $h_\mu = \sum_n P(\sigma_n)H[P(X|\sigma_n)]$, which

TABLE I. Summary of memory and thermal costs of classical and quantum implementations of different generators of the simple nonunifilar source process.

	c	q
A	$C_{cA} = 1, W_{cA} = -0.5$	$C_{qA} = 0.811, W_{qA} = -0.558$
B	$C_{cB} = 2.71, W_{cB} = 0$	$C_{qB} = 0.386, W_{qB} = -0.468$
C	$C_{cC} = 2.71, W_{cC} = -0.678$	$C_{qC} = 2.71, W_{qC} = -0.678$

indeed yields $h_\mu = 0.678$. Thus, we see that this example also highlights the general trade-off between compression and work cost, as the implementations with the lowest work cost are not those that achieve the best compression, by either measure of memory. That is, the efficient retrodictive generator C requires an infinite number of states and has a higher entropic cost than generator A, which may be compressed down to two dimensions, at the cost of some additional dissipation. We summarize these results in Table I.

Finally, let us discuss the case where $p \neq 1/2$. By varying p, the trade-off between thermal efficiency and compression can become even more marked. As $p \to 1$, the entropic cost of implementing the thermally efficient generator C_{cC} (and C_{qC}) will diverge. Meanwhile, both the classical and quantum implementations of A have entropic costs bounded from above by 1. Indeed, a more detailed calculation finds that $C_{qA} \to 1$. Moreover, for all generators and implementations considered, it can be seen that the work costs all tend to 0 in this limit. That is, as $p \to 1$, the increased compression offered by the two-state nondeterministic generator implementations grows while their thermal inefficiency diminishes.

VI. DISCUSSION

Our work extends a series of results on the power and limitations of using quantum technologies to implement generators of stochastic processes by generalizing to a much broader class of HMMs. At the core of this, we have shown that if a classical generator of a stochastic process has any thermal inefficiency due to its time-local operative nature, then it can be implemented with greater thermal and memory efficiency by a quantum device. We also provided a systematic method for determining the architecture of such an implementation. When perfect thermal efficiency is required, however, the best classical and quantum implementations are one and the same.

To achieve the necessary decoherence required to break apart superpositions of memory states for nonpredictive generators, we introduced an additional auxiliary system that carries information about the transition. To provide a universal, systematic protocol, we set this information to be the end state of the transition. This choice is not unique; for any transition s to s' with emitted symbol x, any state $|\psi(s, s', x)\rangle$ will suffice, provided

$$\langle \psi(s, s', x)|\psi(s, s'', x)\rangle = \delta_{s's''} \tag{11}$$

for all pairs of possible transitions. The choice of $\{|\psi(s, s', x)\rangle\}$ affects the overlaps of the quantum memory states through Eq. (7), potentially allowing for further thermal and memory advantages. Particularly, for generators of

highly non-Markovian processes in which most transitions have only a small number of possible end states given the initial state and symbol, setting $|\psi(s, s', x)\rangle = |s'\rangle$ will likely be rather inefficient [28,31,50]. We leave the optimization of this freedom to future work. The use of complex phases to improve memory efficiency in quantum implementations of predictive generators [33] is a special case of this auxiliary space, restricted to one dimension. The theorems above hold for any choice of $\{|\psi(s, s', x)\rangle\}$, provided that they satisfy Eq. (11) and do not remove all possible nonorthogonalities that can be engineered between quantum memory states. We also remark that the structure of this additional system corroborates the interpretation that nonpredictive generators contain "sideband" information about the future of the process they generate [36]: were an external party to retrieve this system from the environment after it is discarded, they would be able to better anticipate the future of the process than they could from observing its past alone.

As with quantum implementations of predictive generators, proof-of-principle demonstrations of these advantages are within reach of current experiments [68–70]. In taking quantum memory and thermodynamical advantages in stochastic modeling to a more general medium, we open up a number of future research avenues, some paralleling developments in quantum implementations of predictive generators and some unique to the wider spectrum. For example, our results can be further extended to encompass input-output [37,38] and continuous-time processes [30,32,50,51,67,71,72] as well as inference protocols [24], and the trade-off between memory efficiency and thermal efficiency can be explored to determine the generator that best compromises the two—and whether classical and quantum implementations agree on which generator this should be. This latter direction will also require developments even in the purely classical setting: while the optimal predictive generators can be systematically found, this remains an open question across all generators in general. Nevertheless, our results show that no matter how good a classical implementation is, a quantum counterpart is almost certainly better.

APPENDIX A: WORK COST FOR QUANTUM IMPLEMENTATIONS

In Ref. [35], it was shown that for N independent copies of a system R and blank ancilla undergoing unitary evolutions to correlated states of the joint system ancillas $R'A$, there is a procedure of erasing the A to obtain R' alone that is successful with probability $1 - 2^{-\sqrt{N}}$, bearing a work cost per copy W of

$$\frac{W}{k_\mathrm{B} T \ln[2]} = H(R) - H(R'), \qquad (A1)$$

as $N \to \infty$, where $H(\cdot)$ is the von Neumann entropy [14]. We do not reproduce this derivation here and refer the interested reader to Ref. [35]. We will, however, show how it can be deployed in our context.

Recall from the main text that we have the evolution

$$U |\sigma_s\rangle |0\rangle |0\rangle = \sum_{s'x} \sqrt{T_{s's}^x} |\sigma_{s'}\rangle |x\rangle |s'\rangle. \qquad (A2)$$

With a further unitary step that copies the output onto another auxiliary space such that we obtain a final state,

$$\sum_{s'x} \sqrt{T_{s's}^x} |\sigma_{s'}\rangle |x\rangle |s'\rangle |x\rangle, \qquad (A3)$$

we can then trace out the two auxiliary spaces to obtain the desired evolution according to Eq. (2). In terms of Eq. (A1), the initial system state corresponds to the initial memory state S, the final system state is the combined final memory state and output $S'X$, and the ancillae to be reset are the two auxiliary spaces. Thus, the work cost for our evolution is given by

$$\frac{W_{qg}}{k_\mathrm{B} T \ln[2]} = H(S) - H(S'X), \qquad (A4)$$

where here, for shorthand, we use S and S' to represent the states corresponding to the initial and final memory states. Noting that $H(S) = H(S')$ for a memory initialized in the stationary state of the evolution, we can recast this as

$$W_{qg} = k_\mathrm{B} T \ln[2][I(S'; X) - H(X)], \qquad (A5)$$

as given in the main text. This matches the result for quantum implementations of deterministic generators, as the principal difference is the extra auxiliary space added to the ancilla—which does not appear directly in the work cost given by Eq. (A1). Moreover, by including a further auxiliary space as part of the ancilla that after the evolution also outputs the initial memory state, we recover the classical implementation where all memory states are orthogonal—and, therefore, see that Eq. (A4) similarly holds classically. Note that these work costs are negative; positive work can be extracted by the generator implementations [35].

APPENDIX B: PROOF OF THEOREM 3

In Ref. [39], it was shown that a quantum implementation of a deterministic generator can operate with zero modularity cost only if the generator is also retrodictive, and hence does not allow for any quantum memory advantage. Theorem 3 declares that this result holds for quantum implementations in general. Our proof mirrors many aspects of that for the deterministic case [39], with appropriate generalization. Before proceeding with our proof, we introduce a definition and result involved in the proof of the deterministic case.

Definition 2. (Maximal local commuting measurement) [39]. Given a bipartite state ρ_{AB}, a maximal local commuting measurement (MLCM) of A for B is a local

measurement X on A such that

$$\rho_{AB} = \bigoplus_x P(x)\rho_{AB}^x, \tag{B1}$$

where $\rho_{AB}^x := (\Pi_X^x \otimes I_B)\rho_{AB}(\Pi_X^x \otimes I_B)$ and Π_X^x are the projection operators corresponding to the measurement, and no further nontrivial local measurement Y on A can be performed without disturbing $\{\rho_{AB}^x\}$.

The MLCM is proven to be unique [39].

Proposition 1. (Reversible local operations) [39]. Given a bipartite state ρ_{AB} and a local operation Φ_A on A such that $\rho_{CB} := (\Phi_A \otimes I_B)[\rho_{AB}]$, let X be the MLCM of A for B, and Y the MLCM of C for B. Then, $I(A;B) = I(C;B)$ iff Φ_A can be expressed by Kraus operators $\{K^\alpha\}$ of the form

$$K^\alpha = \bigoplus_{xy} e^{i\varphi_{xy\alpha}}\sqrt{P(y,\alpha|x)}U^{y|x}, \tag{B2}$$

where $\{\varphi_{xy\alpha} \in \mathbb{R}\}$ are arbitrary and $P(Y,\alpha|X)$ is a stochastic channel that is nonzero only when ρ_{AB}^x and ρ_{CB}^y are equivalent up to a local unitary operation $U^{y|x}$.

The proof is given in Ref. [39].

We are now in a position to prove Theorem 3, repeated here for convenience.

Theorem 3. (Noneradication of dissipation). A quantum implementation $(\{|\sigma_s\rangle\}, \Phi_{qg})$ of a generator $(\mathcal{S}, \mathcal{X}, \{T_{s's}^x\})$ operates with no locality dissipation iff it can be simulated by a classical implementation at no additional memory cost. Any reduction of thermal inefficiency by a quantum implementation is only mitigation, not eradication; any thermally efficient quantum implementation yields no memory compression advantage.

Define $\rho_g(t)$ be the combined state of the system $X_{0:t}S_t$ formed from the memory state and previous t outputs. Let Θ be the MLCM of S_t for $X_{0:t}$ such that $\rho_g(t) = \bigoplus_\theta P(\theta)\rho_g^\theta(t)$, where $\rho_g^\theta(t) := (I_{X_{0:t}} \otimes \Pi^\theta)\rho_g(t)(I_{X_{0:t}} \otimes \Pi^\theta)$, and, similarly, let Θ' be the MLCM of S_{t+1} for $X_{0:t+1}$.

Recall from Eq. (6) that we have

$$\Phi_{qg}(\rho) := \sum_{s'x} L_{s'}^x \rho L_{s'}^{x\,\dagger}, \tag{B3}$$

where we can further decompose $L_{s'}^x := K_{s'}^x \otimes |x\rangle$ to obtain Kraus operators $\{K_{s'}^x\}$ acting on the memory space alone. From Proposition 1, if there is to be no locality dissipation, we require these to take the form

$$K_{s'}^x = \bigoplus_{\theta\theta'} \sqrt{P(s',x,\theta'|\theta)}U^{\theta'x|\theta}. \tag{B4}$$

Consider now that a further unitary operation is used to imprint the label of the end state of the transition as an additional subspace of the output pattern. The implementation is now of the generator $(\{\theta\}, \mathcal{SX}, \{T_{\theta'\theta}^{sx}\})$ that produces a joint stochastic process combining the outputs of the original stochastic process and the trajectory of future memory states. Crucially though, this shares the same memory cost as the implementation of the original generator and has the same form for the $\{K_{s'}^x\}$ as Eq. (B4). However, for Eq. (B4) to hold, the new generator must be retrodictive, i.e., given (x, s', θ'), θ is uniquely determined [39]. Since the generator is retrodictive, the quantum implementation achieves no memory compression relative to a classical implementation of the same generator. Therefore, there exists a classical implementation of the retrodictive generator of the joint process that bears the same memory cost as the quantum implementation of the original generator. This classical implementation can then be used to generate the process associated with the original generator by coarse graining its output to discard the parts associated with the memory state trajectory.

Thus, whenever a quantum implementation of a generator achieves no locality dissipation, a classical implementation of the same process can be constructed that has the same memory cost. This yields the content of Theorem 3.

---

[1] L. Rabiner and B. Juang, An introduction to hidden Markov models, IEEE Acoust. Speech Signal Proc. **3**, 4 (1986).

[2] L. R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, Proc. IEEE **77**, 257 (1989).

[3] J. P. Crutchfield, The calculi of emergence: Computation, dynamics and induction, Physica D **75**, 11 (1994).

[4] C. Yang, F. C. Binder, M. Gu, and T. J. Elliott, Measures of distinguishability between stochastic processes, Phys. Rev. E **101**, 062137 (2020).

[5] C. Karlof and D. Wagner, Hidden Markov model cryptanalysis, in *International Workshop on Cryptographic Hardware and Embedded Systems*, edited by C. D. Walter, Ç K. Koç, and C. Paar (Springer, Berlin, Heidelberg, 2003), pp. 17–34.

[6] Z. Ghahramani and M. I. Jordan, Factorial hidden Markov models, in *Advances in Neural Information Processing Systems*, edited by D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo (MIT Press, Cambridge, MA, 1996), pp. 472–478.

[7] S. Fine, Y. Singer, and N. Tishby, The hierarchical hidden Markov model: Analysis and applications, Mach. Learn. **32**, 41 (1998).

[8] K. Seymore, A. McCallum, and R. Rosenfeld, Learning hidden Markov model structure for information extraction, in *AAAI-99 Workshop on Machine Learning for Information Extraction*, edited by M. E. Califf (AAAI, Menlo Park, California, 1999), pp. 37–42.

[9] P. Baldi, Y. Chauvin, T. Hunkapiller, and M. A. McClure, Hidden Markov models of biological primary sequence information., Proc. Natl. Acad. Sci. USA **91**, 1059 (1994).

[10] A. Krogh, B. Larsson, G. Von Heijne, and E. L. Sonnhammer, Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes, J. Mol. Biol. **305**, 567 (2001).

[11] M. Stanke and S. Waack, Gene prediction with a hidden Markov model and a new intron submodel, Bioinformatics **19**, ii215 (2003).

[12] R. Bhar and S. Hamori, *Hidden Markov Models: Applications to Financial Economics* (Springer Science & Business Media, New York, 2004), Vol. 40.

[13] S. Gammelmark, K. Mølmer, W. Alt, T. Kampschulte, and D. Meschede, Hidden Markov model of atomic quantum jump

dynamics in an optically probed cavity, Phys. Rev. A **89**, 043839 (2014).

[14] M. A. Nielsen and I. Chuang, *Quantum Computation and Quantum Information* (Cambridge Universty Press, Cambridge UK, 2000).

[15] A. Montanaro, Quantum algorithms: An overview, npj Quantum Inf. **2**, 15023 (2016).

[16] K. Wiesner and J. P. Crutchfield, Computation in finitary stochastic and quantum processes, Physica D **237**, 1173 (2008).

[17] A. Monras, A. Beige, and K. Wiesner, Hidden quantum Markov models and nonadaptive read-out of many-body states, Appl. Math. Comput. Sci. **3**, 93 (2011).

[18] B. O'Neill, T. M. Barlow, D. Šafránek, and A. Beige, *Hidden Quantum Markov Models with One Qubit*, AIP Conference Proceedings Vol. 1479 (American Institute of Physics, Melville, New York, 2012), pp. 667–669.

[19] L. A. Clark, W. Huang, T. M. Barlow, and A. Beige, Hidden quantum Markov models and open quantum systems with instantaneous feedback, in *ISCS 2014: Interdisciplinary Symposium on Complex Systems*, edited by A. Sanayei, O. E. Rössler, and I. Zelinka (Springer, Switzerland, 2015), pp. 143–151.

[20] M. Cholewa, P. Gawron, P. Głomb, and D. Kurzyk, Quantum hidden Markov models based on transition operation matrices, Quantum Inf. Proc. **16**, 101 (2017).

[21] S. Adhikary, S. Srinivasan, G. Gordon, and B. Boots, Expressiveness and learning of hidden quantum Markov models, in *International Conference on Artificial Intelligence and Statistics* (PMLR, 2020), pp. 4151–4161.

[22] A. Monras and A. Winter, Quantum learning of classical stochastic processes: The completely positive realization problem, J. Math. Phys. **57**, 015219 (2016).

[23] S. Srinivasan, G. Gordon, and B. Boots, Learning hidden quantum Markov models, in *International Conference on Artificial Intelligence and Statistics* (PMLR, 2018), pp. 1979–1987.

[24] M. Ho, M. Gu, and T. J. Elliott, Robust inference of memory structure for efficient quantum modeling of stochastic processes, Phys. Rev. A **101**, 032327 (2020).

[25] J. Gruska, D. Qiu, and S. Zheng, Potential of quantum finite automata with exact acceptance, Intl. J. Foundations Comput. Sci. **26**, 381 (2015).

[26] Y. Tian, T. Feng, M. Luo, S. Zheng, and X. Zhou, Experimental demonstration of quantum finite automaton, npj Quantum Inf. **5**, 56 (2019).

[27] M. Gu, K. Wiesner, E. Rieper, and V. Vedral, Quantum mechanics can reduce the complexity of classical models, Nat. Commun. **3**, 762 (2012).

[28] J. R. Mahoney, C. Aghamohammadi, and J. P. Crutchfield, Occam's quantum strop: Synchronizing and compressing classical cryptic processes via a quantum channel, Sci. Rep. **6**, 20495 (2016).

[29] C. Aghamohammadi, S. P. Loomis, J. R. Mahoney, and J. P. Crutchfield, Extreme Quantum Memory Advantage for Rare-Event Sampling, Phys. Rev. X **8**, 011025 (2018).

[30] T. J. Elliott and M. Gu, Superior memory efficiency of quantum devices for the simulation of continuous-time stochastic processes, npj Quantum Inf. **4**, 18 (2018).

[31] F. C. Binder, J. Thompson, and M. Gu, Practical Unitary Simulator for Non-Markovian Complex Processes, Phys. Rev. Lett. **120**, 240502 (2018).

[32] T. J. Elliott, A. J. P. Garner, and M. Gu, Memory-efficient tracking of complex temporal and symbolic dynamics with quantum simulators, New J. Phys. **21**, 013021 (2019).

[33] Q. Liu, T. J. Elliott, F. C. Binder, C. D. Franco, and M. Gu, Optimal stochastic modeling with unitary quantum dynamics, Phys. Rev. A **99**, 062110 (2019).

[34] S. P. Loomis and J. P. Crutchfield, Strong and weak optimizations in classical and quantum models of stochastic processes, J. Stat. Phys. **176**, 1317 (2019).

[35] S. P. Loomis and J. P. Crutchfield, Thermal Efficiency of Quantum Memory Compression, Phys. Rev. Lett. **125**, 020601 (2020).

[36] A. J. P. Garner, Oracular information and the second law of thermodynamics, arXiv:1912.03217.

[37] N. Barnett and J. P. Crutchfield, Computational mechanics of input–output processes: Structured transformations and the ϵ-transducer, J. Stat. Phys. **161**, 404 (2015).

[38] J. Thompson, A. J. P. Garner, V. Vedral, and M. Gu, Using quantum theory to simplify input-output processes, npj Quantum Inf. **3**, 6 (2017).

[39] S. P. Loomis and J. P. Crutchfield, Thermodynamically efficient local computation and the inefficiency of quantum memory compression, Phys. Rev. Research **2**, 023039 (2020).

[40] A. Khintchine, Correlation theory of stationary stochastic processes, Math. Ann. **109**, 604 (1934).

[41] D. R. Upper, Theory and algorithms for hidden Markov models and generalized hidden Markov models, Ph.D. thesis, University of California, Berkeley, 1997.

[42] C. R. Shalizi and J. P. Crutchfield, Computational mechanics: Pattern and prediction, structure and simplicity, J. Stat. Phys. **104**, 817 (2001).

[43] J. P. Crutchfield, C. J. Ellison, and J. R. Mahoney, Time's Barbed Arrow: Irreversibility, Crypticity, and Stored Information, Phys. Rev. Lett. **103**, 094101 (2009).

[44] J. P. Crutchfield and K. Young, Inferring Statistical Complexity, Phys. Rev. Lett. **63**, 105 (1989).

[45] J. B. Ruebeck, R. G. James, J. R. Mahoney, and J. P. Crutchfield, Prediction and generation of binary Markov processes: Can a finite-state fox catch a Markov mouse?, Chaos: Interdis. J. Nonlin. Sci. **28**, 013109 (2018).

[46] J. Thompson, A. J. P. Garner, J. R. Mahoney, J. P. Crutchfield, V. Vedral, and M. Gu, Causal Asymmetry in a Quantum World, Phys. Rev. X **8**, 031013 (2018).

[47] J. P. Crutchfield, Between order and chaos, Nat. Phys. **8**, 17 (2012).

[48] C. Aghamohammadi, J. R. Mahoney, and J. P. Crutchfield, Extreme quantum advantage when simulating classical systems with long-range interaction, Sci. Rep. **7**, 6735 (2017).

[49] A. J. P. Garner, Q. Liu, J. Thompson, V. Vedral *et al.*, Provably unbounded memory advantage in stochastic simulation using quantum mechanics, New J. Phys. **19**, 103009 (2017).

[50] T. J. Elliott, C. Yang, F. C. Binder, A. J. P. Garner, J. Thompson, and M. Gu, Extreme Dimensionality Reduction With Quantum Modeling, Phys. Rev. Lett. **125**, 260501 (2020).

[51] T. J. Elliott, Quantum coarse-graining for extreme dimension reduction in modelling stochastic temporal dynamics, PRX Quantum (to be published) [arXiv:2105.06831].

[52] W. Löhr and N. Ay, Non-sufficient memories that are sufficient for prediction, in *International Conference on Complex*

Sciences, edited by J. Zhou (Springer, Berlin, Heidelberg, 2009), pp. 265–276.

[53] R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, 1990).

[54] This follows from the Schur-Horn theorem (that the spectrum of a Hermitian matrix majorizes its diagonal elements [73]) and the Schur concavity of Rényi entropies [74].

[55] P. Dennery and A. Krzywicki, *Mathematics for Physicists*, Dover Books on Physics Series (Dover, New York, 1996).

[56] H. Leff and A. F. Rex, *Maxwell's Demon 2 Entropy, Classical and Quantum Information, Computing* (CRC Press, Boca Raton, FL, 2002).

[57] J. C. Maxwell, *Theory of Heat* (Longmans, Green, London, 1872).

[58] L. Szilard, On the decrease in entropy in a thermodynamic system when intelligent beings intervene, Z. Phys. **53**, 840 (1929).

[59] R. Landauer, Irreversibility and heat generation in the computing process, IBM J. Res. Dev. **5**, 183 (1961).

[60] S. Deffner and C. Jarzynski, Information Processing and the Second Law of Thermodynamics: An Inclusive, Hamiltonian Approach, Phys. Rev. X **3**, 041003 (2013).

[61] J. M. R. Parrondo, J. M. Horowitz, and T. Sagawa, Thermodynamics of information, Nat. Phys. **11**, 131 (2015).

[62] D. Mandal and C. Jarzynskib, Work and information processing in a solvable model of Maxwell's demon, Proc. Natl. Acad. Sci. USA **109**, 11641 (2012).

[63] A. J. P. Garner, J. Thompson, V. Vedral, and M. Gu, Thermodynamics of complexity and pattern manipulation, Phys. Rev. E **95**, 042140 (2017).

[64] A. B. Boyd, D. Mandal, and J. P. Crutchfield, Leveraging environmental correlations: The thermodynamics of requisite variety, J. Stat. Phys. **167**, 1555 (2017).

[65] A. B. Boyd, D. Mandal, and J. P. Crutchfield, Thermodynamics of Modularity: Structural Costs Beyond the Landauer Bound, Phys. Rev. X **8**, 031036 (2018).

[66] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 2012).

[67] S. E. Marzen and J. P. Crutchfield, Informational and causal architecture of discrete-time renewal processes, Entropy **17**, 4891 (2015).

[68] M. S. Palsson, M. Gu, J. Ho, H. M. Wiseman, and G. J. Pryde, Experimentally modeling stochastic processes with less memory by the use of a quantum processor, Sci. Adv. **3**, e1601302 (2017).

[69] F. Ghafari, N. Tischler, J. Thompson, M. Gu, L. K. Shalm, V. B. Verma, S. W. Nam, R. B. Patel, H. M. Wiseman, and G. J. Pryde, Dimensional Quantum Memory Advantage in the Simulation of Stochastic Processes, Phys. Rev. X **9**, 041013 (2019).

[70] F. Ghafari, N. Tischler, C. Di Franco, J. Thompson, M. Gu, and G. J. Pryde, Interfering trajectories in experimental quantum-enhanced stochastic simulation, Nat. Commun. **10**, 1630 (2019).

[71] S. Marzen and J. P. Crutchfield, Informational and causal architecture of continuous-time renewal processes, J. Stat. Phys. **168**, 109 (2017).

[72] S. E. Marzen and J. P. Crutchfield, Structure and randomness of continuous-time, discrete-event processes, J. Stat. Phys. **169**, 303 (2017).

[73] M. A. Nielsen, Characterizing mixing and measurement in quantum mechanics, Phys. Rev. A **63**, 022114 (2001).

[74] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and its Applications* (Springer, New York, 1979), Vol. 143.