Optimal stochastic modeling with unitary quantum dynamics

Qing Liu,[1,2,*] Thomas J. Elliott,[2,1,†] Felix C. Binder,[1,2,‡] Carlo Di Franco,[1,2] and Mile Gu[1,3,2,§]
[1]*School of Physical and Mathematical Sciences, Nanyang Technological University, 637371, Singapore*
[2]*Complexity Institute, Nanyang Technological University, 637335, Singapore*
[3]*Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, 117543, Singapore*

(Received 15 February 2019; published 18 June 2019)

Isolating past information relevant for future prediction is central to quantitative science. Quantum models offer a promising approach, enabling statistically faithful modeling while using less past information than any classical counterpart. Here we introduce a class of phase-enhanced quantum models, representing the most general means of simulating a stochastic process unitarily in causal order. The resulting constructions surpass previous state-of-art methods—both in reducing the information they need to store about the past and in the minimal memory dimension they require to store this information. Moreover, these two features are generally competing factors in optimization—leading to an ambiguity in optimal modeling that is unique to the quantum regime. Our results simultaneously offer quantum advantages for stochastic simulation and illustrate further qualitative differences between classical and quantum notions of complexity.

DOI: 10.1103/PhysRevA.99.062110

I. INTRODUCTION

Models of stochastic processes are essential to quantitative science, providing a systematic means for simulating future behavior based on past observations. Given different models exhibiting statistically identical behavior, there is a general preference for the simplest models—those which require minimal information about the past. The motivation is twofold: fundamentally, they represent a way of identifying potential causes of future events and, operationally, simulating a process using such models requires less memory—as they track less information about the past—leading to reduced resource costs.

The field of computational mechanics [1–3] provides a systematic approach to constructing the provably simplest classical causal model for any given stochastic process. These models, called ε machines, can produce statistically correct predictions using less memory than any classical alternative. The amount of past information they store has been employed as a measure of structure in diverse contexts [4–11], motivated by its interpretation as a fundamental limit on how much information from the past must be tracked in order to predict the future.

Quantum mechanics, however, enables even simpler models that bear statistically identical predictions [12–20]. This advantage, which has been observed experimentally [21,22], can scale without bound [17,23–25] and induces significant qualitative classical-quantum divergences in quantifiers of structure [26,27]. However, while presently known quantum constructions are provably optimal for specific cases [25,26], they are known not to be so in general. This motivates the search for even simpler quantum models that obtain further memory advantages in stochastic simulation, and better characterize quantum notions of structure and complexity.

In this article, we introduce phase-enhanced quantum models—a sophistication of previous quantum models—that capture all possible methods of causal simulation using unitary quantum circuits. We show that the resulting models can improve upon current state-of-the-art constructions in further reducing the amount of memory they require, according to both entropic and dimensional measures [25]. Moreover, our models reveal the origin and highlight the widespread nature of a recently discovered phenomenon [28]—which we term the ambiguity of optimality—wherein optimizing for quantum models that track minimal information about the past may sacrifice achieving minimal dimensionality of their memory (and vice versa).

II. FRAMEWORK

A. Classical models

A bi-infinite discrete-time, discrete-event stochastic process [29] is characterized by a sequence of random variables X_t that take values x_t drawn from a finite alphabet $\mathcal{A}$ at each time step $t \in \mathbb{Z}$. The process is defined by a joint probability distribution $P(\overleftarrow{X}, \vec{X})$, where $\overleftarrow{X} = \cdots X_{-2}X_{-1}$ and $\vec{X} = X_0X_1\cdots$ represent the past and future sequences of the process, respectively (we use upper case to denote random variables and lower case for their variates). A consecutive sequence of length L is denoted by $X_{0:L} = X_0 \cdots X_{L-1}$. Here, we consider stationary stochastic processes, such that $P(X_{0:L}) = P(X_{m:m+L}) \forall L, m \in \mathbb{Z}$.

An instance of a given stochastic process has a specific past $\overleftarrow{x}$, and possesses a corresponding conditional future $P(\vec{X}|\overleftarrow{x})$. A causal model of a stochastic process defines an encoding

*liuqingppk@gmail.com
†physics@tjelliott.net
‡quantum@felix-binder.net
§gumile@ntu.edu.sg

function that maps each possible $\overleftarrow{x}$ to some suitable memory state such that the same systematic action on the memory at each time step gives rise to future sequences according to this conditional future distribution. Notably, all information about the future that is stored in the memory states may be obtained from observations of the past [1,2,25].

The field of computational mechanics [1,2] offers a systematic means to construct the simplest classical causal models—ε machines. These models are defined by encoding past information into causal states $s \in \mathcal{S}$, defined by an equivalence relation on the past-future conditional distribution:

$$\overleftarrow{x}, \overleftarrow{x}' \in s \Leftrightarrow P(\vec{X}|\overleftarrow{x}) = P(\vec{X}|\overleftarrow{x}'). \tag{1}$$

A key property of ε machines is unifilarity [2]: given an initial causal state s and output symbol x, the memory transitions into a unique subsequent causal state. We may thus define an update rule $\lambda(s, x)$ to describe the new state [16].

The memory of an ε machine is often parametrized according to two metrics [1]: the statistical complexity

$$C_\mu := H(S) = -\sum_{s\in S} \pi_s \log_2(\pi_s), \tag{2}$$

which measures the amount of information stored in the memory, and the topological complexity

$$D_\mu := \log_2(\dim(\mathcal{S})), \tag{3}$$

which measures the dimension of the memory. Here, $\pi_s = \sum_{\overleftarrow{x}\in s} P(\overleftarrow{x})$ denotes the steady-state distribution of the causal states. The ε machine minimizes both these metrics over analogous measures for the memory of all other classical causal models. Nevertheless, it still stores information that is not directly relevant for simulating future statistics; C_μ can be strictly greater than the mutual information between past and future [2]. Operationally, C_μ and D_μ correspond to the size of the simulator memory (per simulator), when run in an ensemble or single-shot setting, respectively.

B. Quantum models

Quantum models can bypass these classical limits [12–18]. The present state-of-the-art systematic constructions can be expressed as a stepwise unitary circuit [15,16], where each causal state $j \in S$ is assigned to a corresponding quantum memory state $|\sigma_j\rangle$. Future sequences are manifest by sequential application of a unitary operator U satisfying

$$U|\sigma_j\rangle|0\rangle = \sum_x \sqrt{P(x|j)}|\sigma_{\lambda(j,x)}\rangle|x\rangle, \tag{4}$$

where $P(x|j) := P(X_{t+1} = x|S_t = j)$. At each time step t, the memory state (first subspace) interacts with a fresh ancilla (second subspace) initialized in $|0\rangle$ (Fig. 1). Subsequent measurement of the resulting ancilla then yields the correct conditional future statistics at each time step. Such a unitary operation exists for any stationary stochastic process [16].

We can extend the definitions of Eqs. (2) and (3) to the quantum domain:

$$C_q := -\mathrm{Tr}(\rho \log_2(\rho)), \quad D_q := \log_2(\mathrm{rank}(\rho)), \tag{5}$$

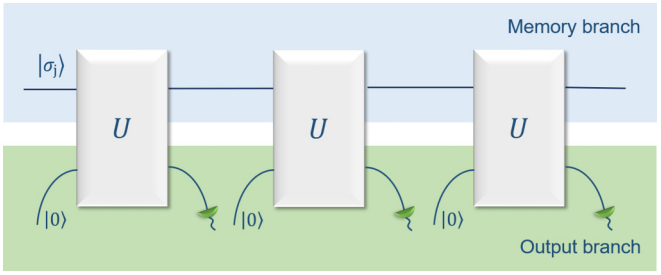

FIG. 1. Unitary quantum model produces a statistical sequence of outputs $\vec{x} = x_0x_1x_2\ldots$ by interacting a blank ancilla with the memory through U at each time step, and then measuring the state of the output branch in the computational basis.

where $\rho = \sum_j \pi_j|\sigma_j\rangle\langle\sigma_j|$. These quantities inherit the same operational significance as their corresponding classical counterparts. We refer to them as the quantum statistical memory and quantum topological memory, respectively. These quantities are model dependent [30].

As the memory states are generally not mutually orthogonal they enable memory savings in terms of both metrics [31]. In fact, the above constructions saturate bounds on pairwise memory state overlap [16,26]. That is, for any quantum model the overlap between quantum memory states $c_{jk} := \langle\sigma_j|\sigma_k\rangle$ cannot exceed the fidelities of their respective conditional future distributions $f_{jk} = \sum_{\vec{x}} \sqrt{P(\vec{x}|j)P(\vec{x}|k)}$ due to information processing inequalities. For the above construction, $c_{jk} = f_{jk}$ [13,16].

Despite this, the optimality of these models is only proven for specific processes [25,26]. C_q and D_q are thus not the true quantum analogs of statistical and topological complexity, but rather bound them from above. There is hence strong motivation to find quantum models whose memories further reduce these measures, in order to both provide a more efficient means of stochastic modeling and to capture the ultimate limits of quantum models.

III. RESULTS

A. Phase-enhanced quantum models

We construct our phase-enhanced unitary models by postulating a new set of quantum memory states $\{|\sigma_j^\varphi\rangle\}$ with a corresponding unitary interaction U^φ that generalizes Eq. (4):

$$U^\varphi|\sigma_j^\varphi\rangle|0\rangle := \sum_x \sqrt{P(x|j)}e^{i\varphi_{xj}}|\sigma^\varphi_{\lambda(j,x)}\rangle|x\rangle, \tag{6}$$

where $\{e^{i\varphi_{xj}}\}$ are the additional phase factors that depend both on the initial causal state j and the output symbol x. Given a set of memory states and unitary operator satisfying this relation, measurements of the second subspace in the computational basis $\{|x\rangle\}$ are guaranteed to produce sequences that obey the same statistics as the corresponding non-phase-enhanced model.

Theorem 1. All phase-enhanced models are valid; a corresponding unitary U^φ satisfying Eq. (6) exists for any choice of phase factors $\{e^{i\varphi_{xj}}\}$.

The proof is given in Appendix A.

Theorem 2. The set of phase-enhanced models of a given stochastic process as described above contains the unitary quantum models of the process that minimize each of the quantum statistical and topological memories.

The only possible valid modifications that can be made to Eq. (6) are refinements [2] of the memory states beyond the causal states. Modifying the transition structure between memory states in any other manner or modifying the magnitude of the terms in the action of the unitary will change the output statistics, and hence the process being modeled, ruling out such modifications. It has previously been shown that such refinements can only increase the statistical memory [26]; thus the minimal unitary quantum models must be described by Eq. (6).

As the quantum memory state overlaps $c^{\varphi}_{jk} := \langle \sigma^{\varphi}_j | \sigma^{\varphi}_k \rangle$ generally differ between different phase choices, the corresponding memory measures will also differ. In Appendix A, we show that these overlaps are given by

$$c^{\varphi}_{jk} = \sum_{\vec{x}} \sqrt{P(\vec{x}|j)P(\vec{x}|k)} e^{i(\varphi_{\vec{x}k} - \varphi_{\vec{x}j})}, \tag{7}$$

where $\varphi_{x_{m:n}j} := \sum_{t=m}^{n} \varphi_{x_t \lambda(x_{m:t}, j)}$ is shorthand for the multistep combination of phases. For any phase-enhanced model we can then compute the corresponding quantum statistical and topological memories:

$$C^{\varphi}_q := -\mathrm{Tr}(\rho^{\varphi} \log_2(\rho^{\varphi})), \quad D^{\varphi}_q := \log_2(\mathrm{rank}(\rho^{\varphi})), \tag{8}$$

where $\rho^{\varphi} = \sum_j \pi_j |\sigma^{\varphi}_j\rangle\langle\sigma^{\varphi}_j|$. Since these quantities depend on the choice of $\{\varphi_{xj}\}$, we define

$$C^{\varphi}_{q\,\mathrm{min}} := \min_{\{\varphi_{xj}\}} C^{\varphi}_q \tag{9}$$

(and similarly $D^{\varphi}_{q\,\mathrm{min}}$) as the minimal quantum statistical (topological) memory over all possible phase enhancements. Should these quantities be smaller than those without phase enhancement, i.e.,

$$C^{\varphi}_{q\,\mathrm{min}} < C_q, \quad D^{\varphi}_{q\,\mathrm{min}} < D_q, \tag{10}$$

the resulting phase-enhanced models would be more memory efficient. In the following we will demonstrate that this advantage can indeed manifest for both memory metrics.

B. Examples: Three-state Markov processes

We illustrate the power of phase enhancement by systematic study of three-state Markovian processes. The general Markov chain for such processes is given in Fig. 2, where T_{yx} is used to denote the transition probability of going from state x to state y (while emitting y). The Markov property allows us to simplify Eq. (6) to

$$U^{\varphi}|\sigma^{\varphi}_x\rangle|0\rangle = \sum_y \sqrt{T_{yx}} e^{i\varphi_{yx}} |\sigma^{\varphi}_y\rangle|y\rangle. \tag{11}$$

Theorem 3. Phase enhancements can reduce the dimension of the memory (i.e., quantum topological memory), providing advantages for single-shot stochastic modeling.

To achieve a dimensional reduction, it must be possible for the phase factors to induce a linear dependency between the quantum memory states, i.e.,

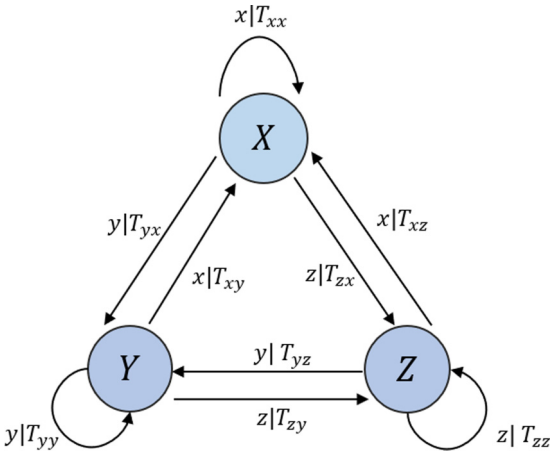

FIG. 2. General three-state Markov model: the notation $y|T_{yx}$ indicates that the transition from state x to y occurs with probability T_{yx} and the output symbol is y.

$$\exists \{\alpha_j \in \mathbb{C}\} : |\sigma^{\varphi}_k\rangle = \sum_{j \neq k} \alpha_j |\sigma^{\varphi}_j\rangle. \tag{12}$$

For three-state Markov models this can be expressed as

$$\alpha|\sigma^{\varphi}_x\rangle + \beta|\sigma^{\varphi}_y\rangle = |\sigma^{\varphi}_z\rangle \tag{13}$$

for some $\alpha, \beta \in \mathbb{R}^+$, where we are able to restrict to positive reals through freedom to add phase to the memory states $\{|\sigma^{\varphi}_j\rangle\}$. Moreover, due to global phase invariance, we can set $\varphi_{wx} = 0$ for all $w \in \{x, y, z\}$ without loss of generality.

Equation (13) can be expressed in terms of the transition probabilities:

$$\alpha\sqrt{T_{wx}} + \beta\sqrt{T_{wy}} e^{i\varphi_{wy}} = \sqrt{T_{wz}} e^{i\varphi_{wz}} \;\forall\; w. \tag{14}$$

From this we obtain the following set of inequalities:

$$|\alpha\sqrt{T_{wx}} - \beta\sqrt{T_{wy}}| \leqslant \sqrt{T_{wz}} \leqslant \alpha\sqrt{T_{wx}} + \beta\sqrt{T_{wy}} \tag{15}$$

that must be satisfied for all w. The existence of real and positive (α, β) satisfying these inequalities is a necessary and sufficient condition for a dimensional advantage.

Furthermore, given (α, β) that satisfy these conditions for a set of transition probabilities, we can determine the phases that collapse the memory to two dimensions:

$$\cos(\varphi_{wy}) = \frac{T_{wz} - \alpha^2 T_{wx} - \beta^2 T_{wy}}{2\alpha\beta\sqrt{T_{wx}T_{wy}}},$$
$$\cos(\varphi_{wz}) = \frac{\alpha\sqrt{T_{wx}} + \beta\sqrt{T_{wy}}\cos(\varphi_{wy})}{\sqrt{T_{wz}}}. \tag{16}$$

Thus, for processes satisfying these inequalities, the phase-enhanced quantum model has $D^{\varphi}_{q\,\mathrm{min}} = 1$, in contrast to the non-phase-enhanced model with $D_q = \log_2(3)$.

We performed a systematic numerical sweep with a range of α and β over the space of three-state Markov processes, and found that the inequalities Eq. (15) are satisfied by at least 17% of such processes (see Appendix B for more details). This lower bound indicates that dimensional advantages, wherein $D^{\varphi}_{q\,\mathrm{min}} < D_q \leqslant D_{\mu}$, are relatively commonplace.

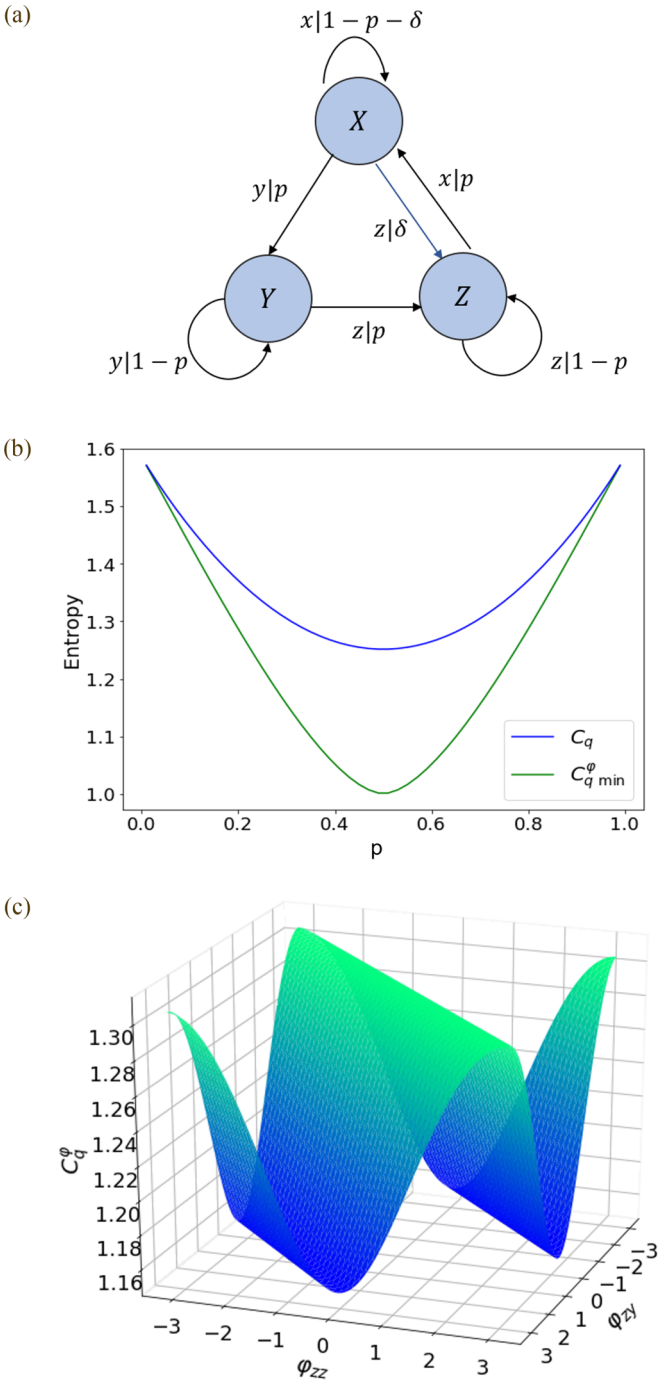

FIG. 3. (a) Symmetric three-state quasicycle with slippage δ. (b) C_q and $C_{q\,\mathrm{min}}^{\varphi}$ as a function of p for $\delta = 0$. (c) The dependence of C_q^{φ} on the phase factors when $p = 0.3$ with $\delta = 0$.

Theorem 4. Phase enhancements can reduce the quantum statistical memory.

Consider the symmetric three-state quasicycle [32], as illustrated in Fig. 3(a) with $\delta = 0$. The transition matrix T of this process is given by

$$T = \begin{pmatrix} 1-p & 0 & p \\ p & 1-p & 0 \\ 0 & p & 1-p \end{pmatrix}. \tag{17}$$

Due to certain phase symmetries such as global phase, the quantum memory states $|\sigma_j^{\varphi}\rangle$ and corresponding unitary U^{φ} can be given in their most general form as

$$\begin{aligned} U^{\varphi}|\sigma_x^{\varphi}\rangle|0\rangle &= \sqrt{1-p}|\sigma_x^{\varphi}\rangle|x\rangle + \sqrt{p}|\sigma_y^{\varphi}\rangle|y\rangle, \\ U^{\varphi}|\sigma_y^{\varphi}\rangle|0\rangle &= \sqrt{1-p}|\sigma_y^{\varphi}\rangle|y\rangle + \sqrt{p}e^{i\varphi_{zy}}|\sigma_z^{\varphi}\rangle|z\rangle, \\ U^{\varphi}|\sigma_z^{\varphi}\rangle|0\rangle &= \sqrt{p}|\sigma_x^{\varphi}\rangle|x\rangle + \sqrt{1-p}e^{i\varphi_{zz}}|\sigma_z^{\varphi}\rangle|z\rangle. \end{aligned} \tag{18}$$

This allows us to calculate the statistical memory C_q^{φ} for this model across the full range of possible phase factors. In Fig. 3(b) we compare $C_{q\,\mathrm{min}}^{\varphi}$ with C_q, observing a clear advantage with our phase-enhanced models. We also show the full dependence of C_q^{φ} on the two phase parameters in Fig. 3(c). It can be seen that $C_{q\,\mathrm{min}}^{\varphi}$ is found when $|\varphi_{zy} - \varphi_{zz}| = \pi$.

Performing a numerical search over the space of general three-state Markov processes, however, we find that entropic advantages appear to be quite rare, occurring in less than 0.5% of cases (see Appendix B for details).

Our numerical results thus indicate that for three-state Markov processes, models that admit $D_{q\,\mathrm{min}}^{\varphi} < D_q$ are much more common than those with $C_{q\,\mathrm{min}}^{\varphi} < C_q$. This begets the question, what happens to C_q^{φ} for models with dimensional advantages? We find that in many cases for which $D_{q\,\mathrm{min}}^{\varphi} < D_q$, the corresponding C_q^{φ} is strictly greater than C_q. However, since multiple choices of phases can provide a dimensional advantage, one may be tempted to think that another set of phases will show advantages in both metrics. We now study a family of processes that conclusively show that the dichotomy cannot always be resolved in this manner: unlike classical causal models, the optimal quantum model can depend on the choice of memory metric.

Theorem 5. The model that minimizes quantum topological memory and the model that minimizes quantum statistical memory do not generally coincide: there is no unique optimal quantum model, leading to an ambiguity of optimality.

A process displaying this phenomenon for models with real phases was recently highlighted [28]. Our results illustrate that this phenomenon is in fact widespread when general complex phase enhancements are introduced.

Consider a modified three-state quasicycle with slippage, as illustrated in Fig. 3(a). Our phase-enhanced models offer dimensional advantages along one line of the parameter space, while there is a large area of the space that permits models that exhibit an entropic advantage [Fig. 4(a)]. Specifically, a dimensional advantage exists iff p and δ satisfy $(1-p)\sqrt{1-p-\delta} - p\sqrt{p} = \pm\sqrt{p\delta(1-p)}$, in which case the inequalities Eq. (15) are satisfied only for a single pair of values of α and β given by

$$\begin{aligned} \alpha &= \frac{\sqrt{p}}{\sqrt{1-p-\delta}}, \\ \beta &= \frac{-p}{\sqrt{(1-p)(1-p-\delta)}}. \end{aligned} \tag{19}$$

Since a unique set of phases offer linear dependence between the memory states at each point on the aforementioned line, we can be satisfied that this gives the unique optimal model in terms of topological memory. In Fig. 4(b) we plot C_q^{φ} of this model for the parameters indicated by the red segment of the dashed line in Fig. 4(a), and compare it to C_q and $C_{q\,\mathrm{min}}^{\varphi}$.

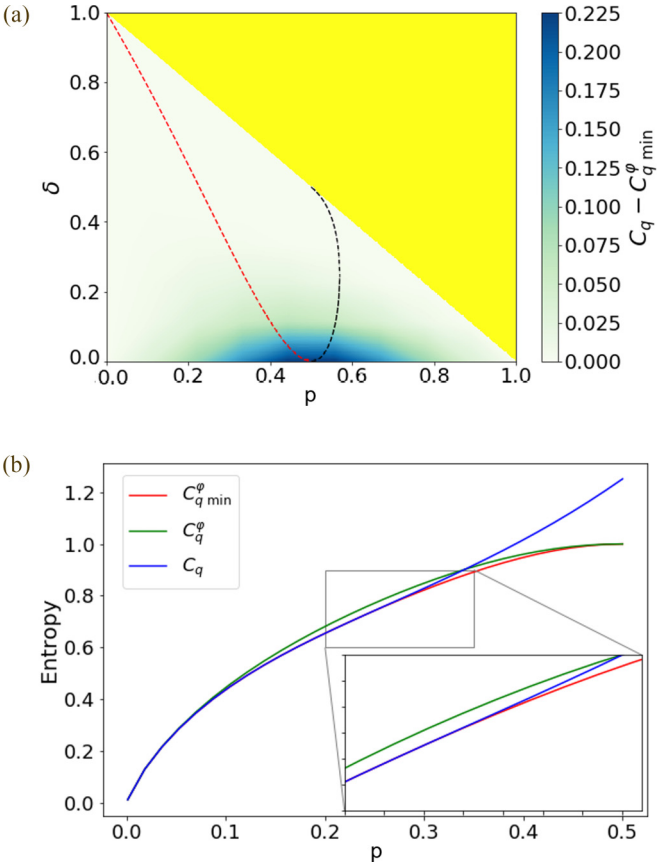

FIG. 4. In (a) we show the regions with entropic (color plot) and dimensional (dashed line) advantages. The yellow region delineates the nonphysical parameter regime. In (b) we plot C^{φ}_{q} in the parameter region denoted by the red segment of the dashed line in (a) and show that the choice of phases that lead to $C^{\varphi}_{q\,\min}$ does not always correspond to $D^{\varphi}_{q\,\min}$, giving rise to an ambiguity of optimality.

For certain parameter values $C^{\varphi}_{q} > C_q > C^{\varphi}_{q\,\min}$, confirming the ambiguity of optimality.

Geometrically, we can understand how such an ambiguity can manifest; reductions in topological memory require linear dependence between the memory states, irrespective of the distance between them, while reductions in statistical memory arise from reductions in the distance between the states. When these two factors are in competition, the ambiguity occurs.

IV. DISCUSSION

A key task in stochastic modeling is to find methods of making statistically faithful future predictions with minimal memory of the past. Here, we have introduced phase enhancement as a general means of constructing quantum models that use less memory resources than previous state-of-the-art constructions. Our analysis holds true in both ensemble and single-shot scenarios. We established that phase enhancement leads to provably optimal models in both scenarios, among all alternatives that operate through unitary quantum circuits. Through examples, we also highlighted an ambiguity in optimality—a uniquely quantum phenomenon where the optimal model for simulating a stochastic process in a single-shot scenario differs from that for simulating many such processes in parallel. We expect that such enhancements will become more typical in stochastic processes with larger numbers of causal states—the number of phase parameters that can be tweaked grows with both the number of states and alphabet size, allowing more freedom for optimization.

From Eq. (7), it is clear that the overlap between memory states is always maximized when all phase factors are zero. Moreover, for most other choices, $|c^{\varphi}_{jk}|$ is strictly less than c_{jk}; phase factors cannot increase pairwise overlaps between memory states. The potential for memory reductions through phase enhancement might thus at first blush appear counterintuitive. Nevertheless, as our examples show, such reductions are possible. Reductions in topological memory can be understood as the phase factors creating linear dependencies between the memory states. Meanwhile, its potential to reduce statistical memory nicely illustrates that increasing pairwise distinguishability between an ensemble of quantum states can still reduce higher-order distances that are captured by the von Neumann entropy [33].

ACKNOWLEDGMENTS

This work was funded by Singapore National Research Foundation Fellowship No. NRF-NRFF2016-02, the Lee Kuan Yew Endowment Fund (Postdoctoral Fellowship), Grant No. FQXi-RFP-1809 from the Foundational Questions Institute and Fetzer Franklin Fund, a donor advised fund of Silicon Valley Community Foundation, Singapore Ministry of Education Tier 1 Grant No. RG190/17, and NRF-ANR grant NRF2017-NRF-ANR004 VanQuTe. Q.L., T.J.E., and F.C.B. thank the Centre for Quantum Technologies for their hospitality.

APPENDIX A: EXISTENCE OF U^{φ} AND OVERLAP OF QUANTUM MEMORY STATES

Here we show that the unitary operator U^{φ} for our phase-encoded quantum models exists for any choice of the phases φ_{xj}, provides an expression for the overlaps of pairs of quantum memory states, and shows that the solution to this overlap converges.

Existence of U^{φ}. We introduce the notation $|1^{\varphi}_{j}\rangle$ to indicate the combined system-ancilla state after applying the unitary circuit:

$$\left|1^{\varphi}_{j}\right\rangle := U^{\varphi}\left|\sigma^{\varphi}_{j}\right\rangle|0\rangle = \sum_{x}\sqrt{P(x|j)}e^{i\varphi_{xj}}\left|\sigma^{\varphi}_{\lambda(j,x)}\right\rangle|x\rangle. \quad \text{(A1)}$$

Previous work [16] established the existence of a unitary operation U in the non-phase-encoded case if and only if

$$\langle\sigma_j|\sigma_k\rangle = \langle 1_j|1_k\rangle \;\forall\; j,k. \quad \text{(A2)}$$

Similarly, for the existence of U^{φ} in our phase-encoded models we require

$$\begin{aligned}\left\langle \sigma_j^{\varphi} \middle| \sigma_k^{\varphi} \right\rangle = \left\langle 1_j^{\varphi} \middle| 1_k^{\varphi} \right\rangle &= \sum_x \sqrt{P(x|j)} e^{-i\varphi_{xj}} \left\langle \sigma_{\lambda(j,x)}^{\varphi} \right| \langle x| \sum_{x'} \sqrt{P(x'|k)} e^{i\varphi_{x'k}} \left| \sigma_{\lambda(k,x')}^{\varphi} \right\rangle |x'\rangle \\ &= \sum_x \sqrt{P(x|j)P(x|k)} e^{i(\varphi_{xk}-\varphi_{xj})} \left\langle \sigma_{\lambda(j,x)}^{\varphi} \middle| \sigma_{\lambda(k,x)}^{\varphi} \right\rangle. \end{aligned} \tag{A3}$$

A solution for the inner product of the quantum memory states is as follows:

$$c_{jk}^{\varphi} := \left\langle \sigma_j^{\varphi} \middle| \sigma_k^{\varphi} \right\rangle = \sum_{\vec{x}} \sqrt{P(\vec{x}|j)P(\vec{x}|k)} e^{i(\varphi_{\vec{x}k}-\varphi_{\vec{x}j})}, \tag{A4}$$

which can be verified by insertion into Eq. (A3), thus proving the existence of U^{φ}.

Convergence of c_{jk}^{φ}. We must now verify that our solution to the memory state overlaps is convergent; that is, $\lim_{L\to\infty} c_{jk}^{\varphi\ [L]} = c_{jk}^{\varphi}$, where

$$c_{jk}^{\varphi\ [L]} := \sum_{x_{0:L}} \sqrt{P(X_{0:L} = x_{0:L}|S_0 = j)P(X_{0:L} = x_{0:L}|S_0 = k)} e^{i(\varphi_{x_{0:L}k}-\varphi_{x_{0:L}j})} c_{\lambda(j,x_{0:L})\lambda(k,x_{0:L})}^{\varphi}. \tag{A5}$$

Note that to avoid confusion between variables at different time steps, in this section we do not employ the shorthand $P(x|j)$ introduced in the main text.

We assume that we are dealing with synchronizable processes, such that the memory of the model can be initialized properly given the entire past. Recalling that $H(A|B) := \sum_b P(B = b)H(A|B = b)$, this condition can be expressed as

$$\lim_{L\to\infty} H(S_0|X_{-L:0}) = 0, \tag{A6}$$

and thus for large L we can express

$$\sum_{x_{-L:0}} P(X_{-L:0} = x_{-L:0})H(S_0|X_{-L:0} = x_{-L:0}) < \varepsilon(L) \tag{A7}$$

for some small $\varepsilon(L)$ that vanishes as $L \to \infty$. This allows us to divide the possible trajectories $x_{-L:0}$ into two classes: those where the memory state is (asymptotically) synchronized $[H(S_0|\overleftarrow{x}) = 0]$ and those where it is not. However, since this uncertainty is finite, the probability of such nonsynchronizing trajectories occurring must be vanishingly small for consistency with Eq. (A7) and, moreover, the total probability of such trajectories must also be vanishingly small. We can therefore devote our attention only to the former class.

For this former class, we can express

$$H(S_0|X_{-L:0} = x_{-L:0}) = -\sum_j P(S_0 = j|X_{-L:0} = x_{-L:0}) \log_2(P(S_0 = j|X_{-L:0} = x_{-L:0})) < \varepsilon'(L) \tag{A8}$$

for some $\varepsilon'(L)$ that again vanishes as $L \to \infty$. Since each term in the summation is non-negative, we can also constrain each term to satisfy the inequality individually. To satisfy this, we must have that each $P(S_0 = j|X_{-L:0} = x_{0:L})$ is either close to zero or 1. These probabilities must sum to 1, which ensures that for one value of s_0, which we shall label as m, the probability is $1 - \epsilon(L)$ for some small $\epsilon(L)$, while the others occur with probability ϵ_j that are each also small, with $\sum_{j\neq m} \epsilon_j(L) = \epsilon(L)$. In other words, after having produced a sufficiently long sequence of outputs $x_{-L:0}$ the past of the process almost certainly belongs to causal state m, and

$$\lim_{L\to\infty} P(S_0 = j|X_{-L:0} = x_{-L:0}) = \delta_{jm}. \tag{A9}$$

Now consider the expansion

$$\begin{aligned} &P(S_0 = k|X_{-L:0} = x_{-L:0}) \\ &\quad = \sum_j P(S_0 = k, S_{-L} = j|X_{-L:0} = x_{-L:0}) \\ &\quad = \sum_j P(S_{-L} = j|X_{-L:0} = x_{-L:0}) \\ &\qquad \times P(S_0 = k|S_{-L} = j, X_{-L:0} = x_{-L:0}) \\ &\quad = \sum_j P(S_{-L} = j|X_{-L:0} = x_{-L:0})\delta_{k\lambda(j,x_{-L:0})}. \end{aligned} \tag{A10}$$

For $L \to \infty$, the left-hand side becomes arbitrarily close to 1 when $k = m$, and zero otherwise.

Examining the case $k = m$, since $\sum_j P(S_{-L} = j|X_{-L:0} = x_{-L:0}) = 1$, for any j where $k \neq \lambda(j, x_{-L:0})$ we must have $P(S_{-L} = j|X_{-L:0} = x_{-L:0}) \approx 0$. Using Bayes' rule, and assuming that $P(S = j) \not\approx 0$, we have

$$\frac{P(X_{-L:0} = x_{-L:0}|S_{-L} = j)}{P(X_{-L:0} = x_{-L:0})} \approx 0, \tag{A11}$$

implying that, for any j such that $k \neq \lambda(j, x_{-L:0})$, the probability of such an output trajectory occurring given we started in a past belonging to causal state j must be vanishingly small, even relative to the probability of the trajectory occurring at all.

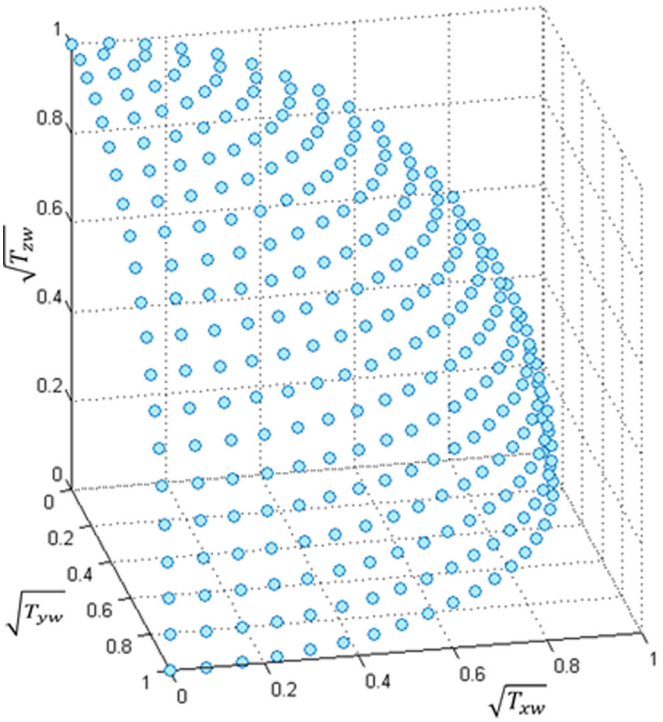

FIG. 5. Transition probabilities out of each state in a three-state Markov process can be represented as a point on the positive octant of a unit sphere. Three such points define the process. We sweep over a discretized set of such points in our numerical treatment to systematically study these processes.

Taken together, these lead us to the conclusion that

$$\lim_{L\to\infty} \lambda(j, x_{-L:0}) = \lim_{L\to\infty} \lambda(x_{-L:0}) \tag{A12}$$

for all but a set of output trajectories of vanishingly small probability; that is, for sufficiently large L the current causal state is almost certainly determined by the output sequence alone independent of the initial state prior to this sequence. Note that for processes with finite Markov order this statement is tautologically true for any trajectory once L is at least as large as the Markov order.

Returning then to Eq. (A5), we see for sufficiently large L that for all but a set of trajectories of vanishingly small probability we may replace $c^{\varphi}_{\lambda(j,x_{0:L})\lambda(k,x_{0:L})} \to c^{\varphi}_{\lambda(x_{0:L})\lambda(x_{0:L})} = 1$. Thus, for sufficiently large L, the recursive factor in the expression tends towards unity and as such $\lim_{L\to\infty} c^{\varphi\ [L]}_{jk} = c^{\varphi}_{jk}$ as required.

TABLE I. Numerical sweep search for phase enhancement.

Advantage	% of three-state processes admitting advantage
Entropic	<0.5
Dimensional ($\alpha = \beta = 1$)	~9
Dimensional [multiple (α, β)]	>17

APPENDIX B: NUMERICAL SWEEP SEARCH FOR PHASE ENHANCEMENTS

For a general three-state Markov process as depicted in Fig. 2, each state is described by the three output probabilities to each state, defined by two free parameters due to normalization of probability. These free parameters can be mapped to a point on the positive octant of a unit sphere (Fig. 5), where the square of the distance along a given axis corresponds to the probability of transitioning into the corresponding state. Each process is defined by three such points, one for each state.

In the case of searching for dimensional advantages, we systematically sweep over these surfaces, coarse grained into grids such that there are 20 evenly spaced steps along each edge of the sweep areas. For each process we then check whether the inequalities Eq. (15) are satisfied for any of the combinations of α and β given in the main text. We found that the inequalities are satisfied for approximately 9% of such processes when $\alpha = \beta = 1$. Expanding the range of α and β values to $\{1, 2, 3, 1/2, 1/4\}$ we find that the inequalities can be satisfied by at least 17% of cases. Accounting for additional values for the parameters can only increase this number.

For entropic advantages, we are unaware of any straightforwardly verifiable conditions as with Eq. (15) for the dimensional case that can be used to ascertain whether the advantage exists. Thus we must consider all possible choices of phases within a given process—four independent parameters after accounting for various symmetries—when checking for advantage. With the six additional parameters used to define the process, we found it numerically intractable to systematically sweep over all processes and phases to any reasonable resolution when searching for entropic advantages. We instead opted to randomly sample processes from the distribution used in the dimensional sweep, and then systematically sweep over all phase parameters to determine whether an advantage $C^{\varphi}_{q} < C_q$ may be found for that particular process. We found such advantages in fewer than 0.5% of processes sampled.

Our findings are summarized in Table I.

[1] J. P. Crutchfield and K. Young, Phys. Rev. Lett. **63**, 105 (1989).
[2] C. R. Shalizi and J. P. Crutchfield, J. Stat. Phys. **104**, 817 (2001).
[3] J. P. Crutchfield, Nat. Phys. **8**, 17 (2012).
[4] J. P. Crutchfield and D. P. Feldman, Phys. Rev. E **55**, R1239(R) (1997).
[5] A. J. Palmer, C. W. Fairall, and W. A. Brewer, IEEE Trans. Geosci. Remote Sensing **38**, 2056 (2000).
[6] D. P. Varn, G. S. Canright, and J. P. Crutchfield, Phys. Rev. B **66**, 174110 (2002).
[7] R. W. Clarke, M. P. Freeman, and N. W. Watkins, Phys. Rev. E **67**, 016203 (2003).
[8] J. B. Park, J. W. Lee, J.-S. Yang, H.-H. Jo, and H.-T. Moon, Physica A **379**, 179 (2007).
[9] C.-B. Li, H. Yang, and T. Komatsuzaki, Proc. Natl. Acad. Sci. USA **105**, 536 (2008).
[10] R. Haslinger, K. L. Klinkner, and C. R. Shalizi, Neural Comput. **22**, 121 (2010).

[11] D. Kelly, M. Dillingham, A. Hudson, and K. Wiesner, PloS One **7**, e29703 (2012).
[12] M. Gu, K. Wiesner, E. Rieper, and V. Vedral, Nat. Commun. **3**, 762 (2012).
[13] J. R. Mahoney, C. Aghamohammadi, and J. P. Crutchfield, Sci. Rep. **6**, 20495 (2016).
[14] J. Thompson, A. J. P. Garner, V. Vedral, and M. Gu, npj Quantum Inf. **3**, 6 (2017).
[15] C. Aghamohammadi, S. P. Loomis, J. R. Mahoney, and J. P. Crutchfield, Phys. Rev. X **8**, 011025 (2018).
[16] F. C. Binder, J. Thompson, and M. Gu, Phys. Rev. Lett. **120**, 240502 (2018).
[17] T. J. Elliott and M. Gu, npj Quantum Inf. **4**, 18 (2018).
[18] T. J. Elliott, A. J. P. Garner, and M. Gu, New J. Phys. **21**, 013021 (2019).
[19] P. M. Riechers, J. R. Mahoney, C. Aghamohammadi, and J. P. Crutchfield, Phys. Rev. A **93**, 052317 (2016).
[20] C. Yang, F. C. Binder, V. Narasimhachar, and M. Gu, Phys. Rev. Lett. **121**, 260602 (2018).
[21] M. S. Palsson, M. Gu, J. Ho, H. M. Wiseman, and G. J. Pryde, Sci. Adv. **3**, e1601302 (2017).
[22] F. Ghafari Jouneghani, M. Gu, J. Ho, J. Thompson, W. Y. Suen, H. M. Wiseman, and G. J. Pryde, arXiv:1711.03661.
[23] A. J. P. Garner, Q. Liu, J. Thompson, V. Vedral *et al.*, New J. Phys. **19**, 103009 (2017).
[24] C. Aghamohammadi, J. R. Mahoney, and J. P. Crutchfield, Sci. Rep. **7**, 6735 (2017).
[25] J. Thompson, A. J. P. Garner, J. R. Mahoney, J. P. Crutchfield, V. Vedral, and M. Gu, Phys. Rev. X **8**, 031013 (2018).
[26] W. Y. Suen, J. Thompson, A. J. P. Garner, V. Vedral, and M. Gu, Quantum **1**, 25 (2017).
[27] C. Aghamohammadi, J. R. Mahoney, and J. P. Crutchfield, Phys. Lett. A **381**, 1223 (2016).
[28] S. Loomis and J. P. Crutchfield, arXiv:1808.08639.
[29] A. Khintchine, Math. Ann. **109**, 604 (1934).
[30] Note that C_q has sometimes been alternatively referred to as the quantum statistical complexity or quantum machine complexity; we avoid such nomenclature here as the former assumes the model is minimal, while the latter invites potential confusion with D_q.
[31] M. A. Nielsen and I. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*, 10th ed. (Cambridge University Press, New York, 2011).
[32] M. Horodecki and J. Oppenheim, Nat. Commun. **4**, 2059 (2013).
[33] R. Jozsa and J. Schlienz, Phys. Rev. A **62**, 012301 (2000).