Quantum Coarse Graining for Extreme Dimension Reduction in Modeling Stochastic Temporal Dynamics

Thomas J. Elliott*

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

(Received 5 November 2020; revised 12 May 2021; accepted 13 May 2021; published 16 June 2021)

Stochastic modeling of complex systems plays an essential, yet often computationally intensive, role across the quantitative sciences. Recent advances in quantum information processing have elucidated the potential for quantum simulators to exhibit memory advantages for such tasks. Heretofore, the focus has been on lossless memory compression, wherein the advantage is typically in terms of lessening the amount of information tracked by the model, while—arguably more practical—reductions in memory dimension are not always possible. Here, we address the case of lossy compression for quantum stochastic modeling of continuous-time processes, introducing a method for coarse graining in quantum state space that drastically reduces the requisite memory dimension for modeling temporal dynamics while retaining near-exact statistics. In contrast to classical coarse graining, this compression is not based on sacrificing temporal resolution and brings memory-efficient high-fidelity stochastic modeling within reach of present quantum technologies.

DOI: 10.1103/PRXQuantum.2.020342

I. INTRODUCTION

Everywhere we look, we are surrounded by complex systems. They manifest across all scales, from the microscopic level of chemical and physical interactions, through biological processes, to geophysical and meteorological phenomena and beyond [1–8]. As the descriptor *complex* suggests, with such systems manifesting a rich tapestry of emergent behaviors it quickly becomes an insurmountable task to track their many interacting components in full. Computational tractability demands that when modeling complex systems, we keep only a partial knowledge, sufficient for predicting relevant properties of interest. Meanwhile, the remaining information that is discarded (or is not possible to observe in the first place) manifests as stochastic effects on top of this. Accordingly, stochastic modeling [9–26] is a critical part of modern science and identifying ways and means of maximizing its efficacy is a transdisciplinary endeavor.

A key bottleneck is the amount of memory available, restricting the amount of information that can be stored. Each configuration the system can take is assigned to a state in the memory; the number of states the memory can support—its *dimension*—limits the number of distinct configurations that can be tracked. A form of compression to mitigate this is coarse graining—grouping together configurations that are sufficiently close into a single combined configuration, reducing the effective dimension, at the cost of precision. This is particularly prominent for temporal information: time is a continuous parameter requiring an unbounded amount of information to specify to arbitrary precision [27]; in practice, it is coarse grained into bins of finite width [28].

For a *quantum* memory, the dimension is no longer synonymous with the number of different possible states it can support. In the context of stochastic modeling, by encoding configurations with partially overlapping features into linearly dependent quantum states, a dimensional compression can be achieved [29–33]. This quantum-compression advantage can be of significant magnitude [32], though present techniques are constrained to exact (lossless) compression, hampering widespread applicability. Nevertheless, quantum encodings have been shown to almost universally reduce the information cost of stochastic modeling [30,34–40], suggesting that many of the dimensions in the memory are barely utilized. This substantiates a strong motivation to develop lossy quantum encodings that trim down these excess dimensions while retaining high fidelity with the exact model.

Here, we introduce such a lossy compression protocol that can be applied to greatly reduce the memory dimensions devoted to tracking temporal information. Our compression is based on reconstructing approximate—yet

*physics@tjelliott.net

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

near-exact—models of a process where the quantum memory states are constrained to a low-dimensional Hilbert space, emancipating the dimension from the number and width of time bins. After reviewing the necessary background, we describe our protocol in detail for pure temporal dynamics, with examples to illustrate the high fidelities and extreme quantum advantages that can be achieved with only a few memory qubits. We then describe how the protocol can be used for compressed modeling of general continuous-time stochastic processes.

II. FRAMEWORK

A. Stochastic processes and models

Herein, we are concerned with continuous-time discrete-event stochastic processes [39,41]. These consist of a series of events described by a sequence of couples $\boldsymbol{x}_n := (x_n, t_n)$, where $x_n \in \mathcal{X}$ is the nth event in the series and $t_n \in \mathbb{R}^+$ is the time between the $(n-1)$th and nth events [42]. The sequence is probabilistic, drawn from a distribution $P(\ldots, \boldsymbol{X}_{n-1}, \boldsymbol{X}_n, \boldsymbol{X}_{n+1}, \ldots)$; throughout, we use upper case to represent random variables and lower case to represent the corresponding variates. We assume the set of possible events $\mathcal{X}$ to be finite. A contiguous block of the sequence is denoted $\boldsymbol{x}_{j:k} := \boldsymbol{x}_j \boldsymbol{x}_{j+1} \ldots \boldsymbol{x}_{k-1}$. We consider bi-infinite length sequences such that $n \in \mathbb{Z}$ and assume the process to be stationary such that $P(\boldsymbol{X}_{0:L}) = P(\boldsymbol{X}_{m:m+L}) \forall m, L \in \mathbb{Z}$. We also consider discrete-time approximations to such processes, where times are coarse grained into finite intervals of size Δt, recovering the continuous case in the limit $\Delta t \to 0$.

We can partition the process into a past and future, delineating what has happened and what is yet to happen, respectively, relative to some point in the sequence. Without loss of generality, we can set $n = 0$ to represent the present with x_0 the next event to occur, such that the past consists of $\overleftarrow{\boldsymbol{x}} := \boldsymbol{x}_{-\infty:0}(\emptyset, t_{\overleftarrow{0}})$ and the future $\overrightarrow{\boldsymbol{x}} := (x_0, t_{\overrightarrow{0}})\boldsymbol{x}_{1:\infty}$. Here, $t_{\overleftarrow{0}}$ represents the time since the last event and $t_{\overrightarrow{0}}$ the time until the next event ($t_0 = t_{\overleftarrow{0}} + t_{\overrightarrow{0}}$), and $\emptyset$ denotes that the 0th event is yet to occur [39,41].

We desire models that are able to track relevant information from the past of a process in order to faithfully replicate the corresponding future statistics [8,43]. We require the models to be causal, entailing that they can be initialized for any given past, and that they store no information about the future that could not be obtained from the past observations [29]. Such models function by means of an encoding function $f : \overleftarrow{\boldsymbol{x}} \to \mathcal{M}$ that maps pasts into memory states $\rho_m \in \mathcal{M}$ and a transition structure $\Lambda : \mathcal{M} \to \mathcal{M}, \emptyset \cup \mathcal{X}$ that produces the future statistics and updates the memory state accordingly [44]. In the continuous-time setting this transition structure is a continuous evolution, while in the discrete-time setting it acts once at each time step [27,28,37,41]. A model with a lossless encoding is able to replicate the future statistics perfectly, while a lossy one produces an approximation thereof.

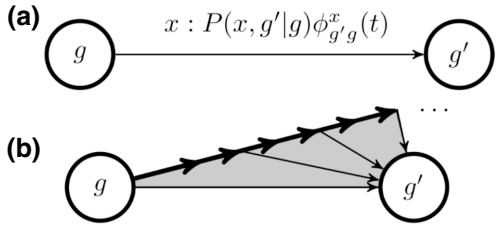


FIG. 1. (a) An HSMM representation of a continuous-time discrete-event stochastic process, showing the transition structure between modes. Each node corresponds to a mode of the model and the arrows labeled $x : p(t)$ denote transitions between modes accompanied by event x occurring at time t since the previous event, with the transition occurring with probability $p(t)$. (b) The unpacking into HMM tracking mode-occupation times. The nodes continue to represent modes and thin lines the transitions, while the thick black line indicates a continuum of states of the model, tracking both the current mode and the time since the last event.

Continuous-time stochastic processes can be represented by edge-emitting hidden semi-Markov models (HSMMs) [23,41]. A HSMM comprises of (hidden) modes $\mathcal{G}$, an event alphabet $\mathcal{X}$, and a transition dynamic Λ. Conditional on the current mode and the time for which it has been occupied, the transition dynamic describes the probability of the model emitting a symbol $x \in \mathcal{X}$ and transitioning to a new mode, with the probabilities depending on the particular process [Fig. 1(a)]. That is, the system resides in a given mode $g \in \mathcal{G}$ until an emission $x \in \mathcal{X}$ occurs, at which point it transitions to a new mode $g' \in \mathcal{G}$; the probability that a system resides in mode g for a time t before emitting symbol x and transitioning to mode g' is given by the modal wait-time distribution $\sum_{xg'} P(x, g'|g)\phi^x_{g'g}(t)$, where the probabilities $P(X, G'|G)$ describe the symbolic transition structure between modes and the dwell functions $\phi^x_{g'g}(t)$ the distribution for the time spent in a given mode before such a given transition occurs. For further details, see Refs. [39,41] and Sec. VI.

A HSMM can be unpacked [39] into an edge-emitting hidden Markov model (HMM) [45] with a continuous state space tracking the occupation time for the modes [Fig. 1(b)]. States in the HMM represent a mode and time since the last event $(g, t_{\overleftarrow{0}})$, with a transition structure taking the system to $(g, t_{\overleftarrow{0}} + dt)$ on nonevents in the next infinitesimal time interval dt, and $(g', 0)$ upon events. The corresponding emitted symbols are $\emptyset$ for nonevents and $x \in \mathcal{X}$ for each event; the transition probabilities follow from the conditional form of the modal wait-time distributions. Models of discrete-time stochastic processes can similarly be modeled by discrete-state HMMs, in which the occupation time is tracked by the corresponding coarse-grained states [28].

B. Memory and quantum advantage

A key metric of efficiency for a model is how much memory it requires to operate [46]. One way in which this can be parametrized is via the information cost—in the sense of Shannon entropy—of storing the compressed past information [34,44,46]. Another, to which we direct our focus here, is the size of the substrate into which this information is encoded—in other words, the dimension of the memory state space [30,32,44,47,48]. The choice of encoding function will impact upon the memory cost and it is ideally chosen to make it as small as possible.

For stationary stochastic processes, the optimal classical lossless memory encoding function is provided by the causal equivalence relation ($\sim_\varepsilon$) of computational mechanics [8,44,49], which partitions the entire set of semi-infinite pasts $\overleftarrow{\boldsymbol{\mathcal{X}}}$ into equivalence classes called causal states, $s \in \mathcal{S}$, such that two pasts belong to the same causal state if and only if they effect the same conditional future statistics:

$$P(\overrightarrow{\mathbf{X}} \,|\, \overleftarrow{\mathbf{x}}) = P(\overrightarrow{\mathbf{X}} \,|\, \overleftarrow{\mathbf{x}}') \Leftrightarrow \overleftarrow{\mathbf{x}} \sim_\varepsilon \overleftarrow{\mathbf{x}}'. \tag{1}$$

The memory-optimal lossless classical model (known as the ε-machine) is then constructed by designating a memory state $|s\rangle$ for each causal state s and having the causal-state encoding function f_ε assign pasts accordingly. A typical process evolving in continuous time will require an infinite-dimensional memory to record the progress through infinitesimal divisions in time [27,32,37], engendering the need for lossy approximations that evolve with discretized time steps [28,32].

With the advent of quantum information processing tools, the optimality of ε-machines has been supplanted [34]. Quantum encoding functions f_q map pasts into a set of quantum memory states; by leveraging the possibility of encoding information into an ensemble of nonorthogonal states, further compression beyond the causal-state encoding function may be attained. Prior work has centered on lowering the information cost of storing the past [30,34–40], showing that a quantum-compression advantage can almost always be procured. Recent focus has been devoted to obtaining corresponding advantages in compressing the dimension of the memory, by engineering quantum memory states with linear dependencies [29–33]. Examples have highlighted that such dimensional compression can sometimes be made arbitrarily strong with respect to the optimal classical encoding [32], though instances where it may be achieved in the lossless regime appear to be much less ubiquitous than in the case of reducing the information cost [30]. The lossy encoding protocol that we introduce seeks to remedy this present shortcoming of the quantum models in the context of tracking the temporal aspect of their dynamics, to escape the associated memory dimension divergence in the continuous limit.

C. Renewal processes

With our attention directed toward compressing the temporal information, for much of this paper we work with a special class of continuous-time stochastic processes that are purely temporal in nature: renewal processes [50]. These consist of a single mode and a single symbol, such that the resulting process is a series of identical events stochastically separated in time, with the spacing of each consecutive pair of events drawn from the same distribution. The distribution governing the time between events is called the wait-time distribution $\phi(t)$ and the survival probability $\Phi(t) := \int_t^\infty \phi(t')dt'$ is the probability that a given interval is of length t or greater [27,28,32,37].

With few exceptions, for generic renewal processes, the causal states group pasts together according to the time since the last event occurred [27,28,37]. That is, all relevant information for predicting the future of a renewal process is contained within the time since the last event—such that the causal states are in one-to-one correspondence with $t_{\overleftarrow{0}}$—and, moreover, can only provide predictive power with respect to the time $t_{\overrightarrow{0}}$ until the next event will happen.

The transition structure between the memory states of the ε-machine for a renewal process has been likened to a "conveyor belt" [27], progressing continuously along a line with time until an event occurs, whereupon the memory jumps to a "reset" state corresponding to $t_{\overleftarrow{0}} = 0$. The probability of occupying the memory state corresponding to $t_{\overleftarrow{0}}$ is given by $\pi(t_{\overleftarrow{0}}) = \mu\Phi(t_{\overleftarrow{0}})$, where $\mu := \left[\int_0^\infty t\phi(t)dt\right]^{-1}$ is the so-called mean firing rate [27,37]. The discrete-time analog consists of a linear sequence of memory states through which the system progresses, akin to the incrementation of a counter, until also resetting upon an event [28,32]. Both are illustrated in Fig. 2. The exact continuous-time version requires an infinite continuum of memory states, and thus requires a memory of unbounded dimension; when there is no maximum

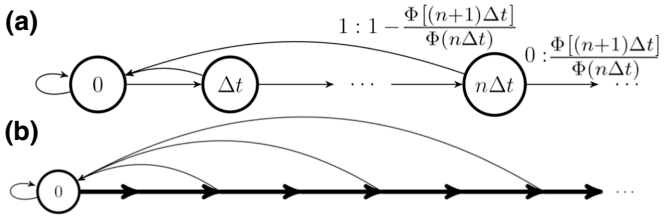


FIG. 2. (a) Discrete and (b) continuous HMM representations of ε-machines of a renewal process. The system progresses along a counter until an event occurs, upon which it transitions to the reset state. In (a), nodes correspond to states of the HMM tracking time since the last event; in (b), the thick black line represents a continuum of such states. The thin arrows represent transitions between states, with $x : p$ indicating the probability p of the transition occurring, accompanied by symbol x. Symbol 1 represents events and 0 nonevents.

value for $t_{\overleftarrow{0}}$, the discrete-time case will similarly need an infinite-dimensional memory and thus finite-dimensional approximations must also adopt a terminal state that the counter cannot exceed [32].

III. QUANTUM COARSE GRAINING

A. Quantum models of renewal processes

In previous work [37], we have established that a general renewal process with wait-time distribution $\phi(t)$ can be exactly simulated by a quantum model with a memory encoding function $f_q(\overleftarrow{x}) = |\varsigma_{t_{\overleftarrow{0}}}\rangle$, where

$$|\varsigma_{t_{\overleftarrow{0}}}\rangle := \int_0^\infty \frac{\psi(t_{\overleftarrow{0}}+t)}{\sqrt{\Phi(t_{\overleftarrow{0}})}} dt|t\rangle, \tag{2}$$

in which $\{|t\rangle\}$ is an infinite-dimensional orthogonal basis and $\psi(t) := \sqrt{\phi(t)}$ [51]. The future statistics are extracted from these memory states by means of a continuous measurement sweep that, at each infinitesimal interval δt, produces a binary outcome as to whether or not the system is found in a state $|t\rangle$ in the interval $[0, \delta t)$: if yes, then the event is deemed to have occurred and the memory is reinitialized in state $|\varsigma_0\rangle$; if no, then the event does not occur and a relabeling $t \to t - \delta t$ takes place.

A fine-grained discrete analog of this evolution with time-step interval δt can be implemented through the following unitary interaction $U_{\delta t}$ coupling the memory state to an ancillary system used to provide the measurement readout, where 0 and 1 represent nonevents and events, respectively [32]:

$$\begin{aligned} U_{\delta t}|\varsigma_t\rangle|0\rangle = &\sqrt{\frac{\Phi(t+\delta t)}{\Phi(t)}}|\varsigma_{t+\delta t}\rangle|0\rangle \\ &+ \sqrt{1-\frac{\Phi(t+\delta t)}{\Phi(t)}}|\varsigma_0\rangle|1\rangle. \end{aligned} \tag{3}$$

After measurement, the ancilla is set to $|0\rangle$ ready for the next time step. The amplitudes on the right-hand side of this equation are set such that they yield the correct probabilities for the future statistics, as $\int_t^{t+\delta t}\phi(t')dt' = \Phi(t) - \Phi(t+\delta t)$. Arbitrary complex phases can be added to these amplitudes without affecting the statistics [30,32]; on the first term, this is equivalent to appending an irrelevant phase to the quantum memory states, while on the latter it mirrors the effect of a complex phase on $\psi(t)$.

B. Quantum model memory as an integral kernel

The steady-state ρ of the quantum model memory is given by a mixture of the quantum memory states, weighted by their probability of occurrence [37]:

$$\begin{aligned} \rho :&= \int_0^\infty \pi(t_{\overleftarrow{0}})dt_{\overleftarrow{0}}|\varsigma_{t_{\overleftarrow{0}}}\rangle\langle\varsigma_{t_{\overleftarrow{0}}}| \\ &= \mu\iiint_0^\infty \psi(t_{\overleftarrow{0}}+t)\psi(t_{\overleftarrow{0}}+t')dtdt'dt_{\overleftarrow{0}}|t\rangle\langle t'|. \end{aligned} \tag{4}$$

The rank of ρ corresponds to the dimension required by the memory substrate to support the range of quantum memory states. This is given by the number of nonzero elements in the spectrum of ρ, which can be found from the characteristic equation

$$\int_0^\infty \rho(t,t')\nu(t')dt' = \lambda\nu(t). \tag{5}$$

This has the form of a homogenous Fredholm integral equation of the second kind [52], with $\rho(t,t') = \mu\int_0^\infty \psi(t_{\overleftarrow{0}}+t)\psi(t_{\overleftarrow{0}}+t')dt_{\overleftarrow{0}}$ corresponding to the kernel of the equation.

We are thus in a position to leverage results from Fredholm theory to reveal properties of the spectrum $\{\lambda\}$ of ρ. Most pertinently, if ρ represents a degenerate kernel, wherein it can be expressed as $\rho(t,t') = \sum_{j=1}^N \alpha_j(t)\beta_j(t')$ for some finite integer N and set of functions $\{\alpha_j, \beta_j\}$, then the spectrum has at most N nonzero elements [52]. Consequently, the memory states can be stored within an N-dimensional space. However, the general form of ρ as per Eq. (4) does not readily present as a degenerate kernel and, indeed, exact quantum models of renewal processes often require an infinite-dimensional memory space. Nevertheless, the amount of information retained in the memory about the past of the process typically appears to be finite [37], suggesting that many of these dimensions are barely utilized and motivating the pursuit of a lossy—yet still near-exact—compression method. A suggestive path to such compression is to truncate ρ by removing the dimensions corresponding to elements of its spectrum that are sufficiently small (as the $\{\lambda\}$ represent the occupation probabilities of the eigenstates of ρ). However, this impacts upon the transition structure of the model, rendering it nonphysical. An approach with greater finesse is needed, which we now provide.

C. Exponential sums and lossy compression

Rather than taking an existing exact model and introducing lossy distortion to effect compression, we instead construct a distortion of the underlying process that is amenable to simulation by a model with a memory of low dimension. The intent is that the exact model of the distorted process forms a near-exact compressed model of the original process.

This requires us to identify what features the wait-time distribution must possess to permit a finite-dimensional

exact model; in other words, to identify what the constraints on $\phi(t)$ are such that it will lead to $\rho(t,t')$ taking the form of a degenerate kernel. Let us begin by introducing the kernel $\kappa(t,t') := \psi(t+t')$, such that $\rho(t,t') = \mu\int_0^\infty \kappa(t,t'')\kappa(t'',t')dt''$. It then follows that the spectrum of $\kappa(t,t')$ is $\{\sqrt{\lambda/\mu}\}$, and is thus of the same rank as $\rho(t,t')$ [52]. This reduces the problem to identifying the conditions under which $\kappa(t,t')$ is a degenerate kernel. These are then the processes for which we can express $\psi(t)$ as a finite sum of functions $F_j(t)$ that satisfy $F_j(t+t') = \alpha_j(t)\beta_j(t')$. We can readily identify the appropriate functions as being (complex) exponentials, i.e., $F_j(t) = c_j\exp(z_j t)$ for some $(c_j, z_j) \in \mathbb{C}^2$. Thus, for $\psi(t) = \sum_{j=1}^N c_j\exp(z_j t)$, we correspondingly have at most N nonzero eigenvalues of the kernel $\kappa(t,t')$.

Although we begin by assuming $\psi(t)$ is real, if we allow it to be complex we instead have $\phi(t) = |\psi(t)|^2$, and $\rho(t,t') = \mu\int_0^\infty \psi(t+t_{\overleftarrow{0}})\psi^*(t'+t_{\overleftarrow{0}})dt_{\overleftarrow{0}}$. Note that even when $\psi(t) = \sum_{j=1}^N c_j\exp(z_j t)$ is complex, it can be verified through direct substitution that $\rho(t,t')$ remains a degenerate kernel of at most rank N. Thus, with an N-dimensional memory it is possible to model renewal processes for which

$$\phi(t) = \left|\sum_{j=1}^N c_j e^{z_j t}\right|^2. \tag{6}$$

Let us decompose $z_j := -\gamma_j + i\omega_j$ for $(\gamma_j, \omega_j) \in \mathbb{R}^2$. For $\phi(t)$ to be a valid distribution it must be normalizable to unity and thus we can constrain $\gamma_j \in \mathbb{R}^+$.

The complex exponentials $\exp(-zt)$ form an overcomplete basis into which any piecewise continuous function of finite exponential order can be decomposed, where the overlap of the function with the basis elements are described by its Laplace transform. Thus, for any $\psi(t)$ that is piecewise continuous and of finite exponential order, we can express the corresponding wait-time distribution in the form of Eq. (6), albeit with N not necessarily finite.

Nevertheless, this provides a constructive approach to finding lossy compressions for quantum models of renewal processes. The goal is to find exponential sums with a finite number of terms that provide a high-fidelity approximation to $\psi(t)$. In practice, it has been found that such decompositions can achieve accurate reconstructions of a function with a relatively small number of terms. Moreover, there are systematic approaches to obtaining such decompositions [53]. From the decomposition, we are then able to build an exact model of the approximate process, to effect a near-exact model of the original process.

The last step remaining is to find an explicit encoding of the memory states of the approximate model into a finite-dimensional memory space. Beginning from a (normalized) approximate decomposition $\tilde{\psi}(t) = \sum_{j=1}^N c_j\exp[(-\gamma_j + i\omega_j)t]$, we assign N "generator" states $\{|\varphi_j\rangle\}$ and a unitary operator $\tilde{U}_{\delta t}$ with the evolution [54]

$$\tilde{U}_{\delta t}|\varphi_j\rangle|0\rangle = e^{(-\gamma_j+i\omega_j)\delta t}|\varphi_j\rangle|0\rangle + \sqrt{1-e^{-2\gamma_j\delta t}}|\tilde{\varsigma}_0\rangle|1\rangle, \tag{7}$$

in analogy with Eq. (3). Here, we define

$$|\tilde{\varsigma}_0\rangle := \sum_{j=1}^N \frac{c_j}{\sqrt{2\gamma_j}}|\varphi_j\rangle, \tag{8}$$

which forms the reset state corresponding to $t_{\overleftarrow{0}} = 0$, with the rest of the quantum memory states $\{|\tilde{\varsigma}_{t_{\overleftarrow{0}}}\rangle\}$ implicitly defined by acting U with the ancilla a sufficient number of times, postselected on all measurement outcomes being 0, i.e., $|\tilde{\varsigma}_{n\delta t}\rangle = \langle 0|(I\otimes|0\rangle\langle 0|U)^n|\tilde{\varsigma}_0\rangle|0\rangle$. Nonnormalized, these states can also be expressed $|\tilde{\varsigma}_{t_{\overleftarrow{0}}}\rangle \propto \sum_{j=1}^N (c_j/\sqrt{2\gamma_j})\exp[(-\gamma_j + i\omega_j)t_{\overleftarrow{0}}]|\varphi_j\rangle$. The overlaps of the generator states can be obtained [30,38] from the recursive relations $\langle\varphi_j|\varphi_k\rangle = \langle\varphi_j|\langle 0|\tilde{U}_{\delta t}^\dagger\tilde{U}_{\delta t}|\varphi_k\rangle|0\rangle$, from which we can move from their implicit definition to expressing them explicitly in terms of an N-dimensional set of orthonormal basis states using a reverse Gram-Schmidt procedure [55]. The relevant columns of $\tilde{U}_{\delta t}$ are defined implicitly by Eq. (7) and can now readily be expressed explicitly in this basis; the remaining columns can be assigned arbitrarily provided that they preserve orthonormality of the basis states (by using e.g., a Gram-Schmidt procedure) [38].

Inputs: Renewal process wait-time distribution $\phi(t)$.
Outputs: Compressed quantum memory states $\{|\tilde{\varsigma}_{t_{\overleftarrow{0}}}\rangle\}$, evolution $\tilde{U}_{\delta t}$, approximate wait-time distribution $\tilde{\phi}(t)$.

1: Define $\psi(t) = \sqrt{\phi(t)}$.
2: Use method of choice to find an exponential sum $\tilde{\psi}(t) = \sum_{j=1}^N c_j\exp((-\gamma_j + i\omega_j)t)$ of N terms approximating $\psi(t)$. Scale weights such that $\tilde{\phi}(t)$ is normalised.
3: Implicitly define $\tilde{U}_{\delta t}$ and $\{|\varphi_j\rangle\}$ according to Eq. (7) and determine the state overlaps from the recursive relations $\langle\varphi_j|\varphi_k\rangle = \langle\varphi_j|\langle 0|\tilde{U}_{\delta t}^\dagger\tilde{U}_{\delta t}|\varphi_k\rangle|0\rangle$. Use a reverse Gram-Schmidt procedure to express the states in terms of N orthonormal basis states.
4: Assign columns of $\tilde{U}_{\delta t}$ defined in Eq. (7). Fill the remaining columns arbitrarily, using a Gram-Schmidt procedure to ensure orthonormality with existing columns.
5: Assign compressed quantum memory states $\{|\tilde{\varsigma}_{t_{\overleftarrow{0}}}\rangle\}$:

$$|\tilde{\varsigma}_{t_{\overleftarrow{0}}}\rangle \propto \sum_{j=1}^N \frac{c_j}{\sqrt{2\gamma_j}} e^{(-\gamma_j+i\omega_j)t_{\overleftarrow{0}}}|\varphi_j\rangle.$$

Algorithm 1 Quantum coarse graining for modeling renewal processes

This constructs a lossy compression of the quantum memory states, yielding a near-exact model of the process. The steps are summarized in Algorithm 1.

IV. EXAMPLES

As a demonstration of the efficacy of our quantum-compression protocol, we apply it to the modeling of two example renewal processes. For each process, we show how the quantum models quickly converge on high-fidelity approximations of the original processes with only a comparatively small memory dimension. Our approximate exponential sums are found using the method of Beylkin and Monzón [56], summarized in Appendix A.

We quantify the goodness of fit using a Kolmogorov-Smirnov (KS) statistic [57], which is defined as the maximum difference between points in the cumulative distribution functions of two probability distributions. This allows us to compare how well discrete distributions approximate continuous distributions, as the cumulative distribution function can be extended over a continuum. That is, let $C_p(t) = \int_0^t p(t')dt'$ be the cumulative distribution function of a continuous distribution $p(t)$ and $C_q(t) = \sum_{n=0}^{\text{argmax}(N|N\delta t<t)} q(n\delta t)$ the continuum form of the cumulative distribution of a discrete distribution $q(n\delta t)$. The KS statistic is then given by $\text{KS}(p,q) = \max_t |C_p(t) - C_q(t)|$. For a renewal process, the survival probability $\Phi(t) = 1 - C_\phi(t)$, and so the KS statistic here, also corresponds to the maximum difference between the survival probabilities of the exact and approximate processes at any time: $\text{KS}[\phi(t), \tilde{\phi}(t)] = \max_t |\Phi(t) - \tilde{\Phi}(t)|$, where $\tilde{\Phi}(t) := \int_t^\infty |\tilde{\psi}(t')|^2 dt'$. Thus, the KS statistic as employed here measures the largest cumulative divergence between the statistics of the approximate model and the exact process.

We compare our quantum models to approximate classical models constrained to a classical memory of the same dimension. These classical models are constructed by discretizing the process into finite-sized time steps and using gradient descent [58] to fit the parameters, taking the KS statistic as a cost function (see Appendix B). While we do not claim this to be the optimal lossy classical compression, we believe it to provide a fair indicator of the potential performance of classical compression methods for this task.

A. Alternating Poisson process

As a first example, we consider an alternating Poisson process. The output can be described by a sequential series of Poisson processes, with an event on these underlying processes alternatively coinciding with events or nonevents of the alternating Poisson process (nonevents of the Poisson processes also correspond to nonevents of the alternating Poisson process). The corresponding wait-time distribution is given by

$$\phi(t) = \gamma^2 t e^{-\gamma t}, \tag{9}$$

where the rate γ sets an arbitrary scale for units of time. This is the continuous-time analog of the so-called simple nonunifilar source process [43]. While also appearing simple to generate, it too has no finite-dimensional exact causal classical representation [28]; it is thought that an exact causal quantum model is similarly structurally complex.

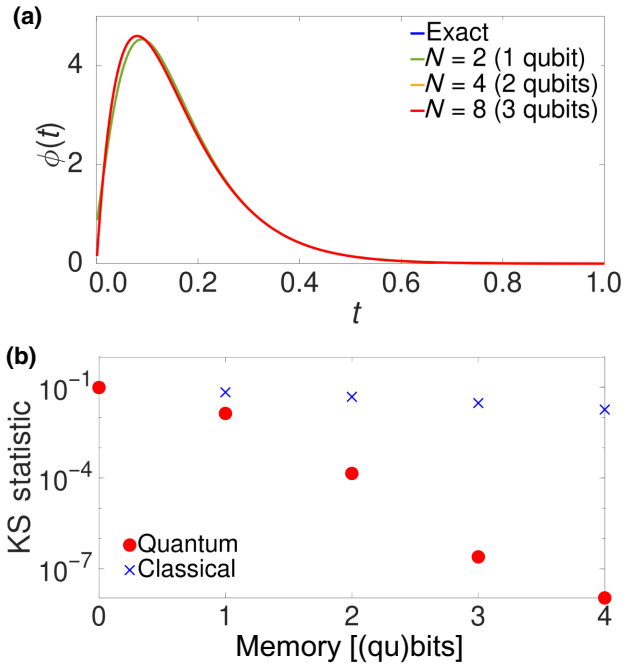


FIG. 3. (a) Wait-time distributions of compressed quantum models of an alternating Poisson process in arbitrary units. (b) KS statistics comparing the performance of compressed quantum models to compressed classical models.

Using our compression protocol, we observe excellent performance in replicating the statistics of the alternating Poisson process with low-dimensional quantum models. As can be seen in Fig. 3(a), even a single-qubit memory provides a close approximation to the exact wait-time distribution and a two-qubit memory is seemingly indistinguishable at the resolution shown. In Fig. 3(b), we compare the performance of our coarse-grained quantum models with the classical approximations, as well as a memoryless model. We see that the quantum models bear a KS statistic orders of magnitude smaller than the corresponding classical model and, moreover, appear to exhibit a more favorable scaling with increasing memory.

B. Bimodal Gaussian process

For the second example, we find compressed models of a bimodal Gaussian process. The wait-time distribution

consists of the sum of two displaced Gaussian peaks:

$$\phi(t) = p_1 e^{-(t-\mu_1)^2/\sigma_1^2} + p_2 e^{-(t-\mu_2)^2/\sigma_2^2}. \quad (10)$$

As with the previous example, the units of time are arbitrary and can be set through the σ. We consider the case where the two peaks have equal weight ($p_1 = p_2$) and equal spread ($\sigma_1 = \sigma_2$). In units where $\sigma = 1$, we then take $\mu_1 = \sqrt{5}$ and $\mu_2 = \sqrt{33.8}$. This leads to little overlap between the two peaks, requiring a model to be able to capture features at both short and long time scales in order to account for the two regions of high event probability and the low-probability trough between them.

As can be seen in Fig. 4(a), our coarse-grained models struggle to fully capture the features with one- and two-qubit memories, with the former overweighting the first peak and the latter the second. With a three-qubit memory however, the model closely follows the exact process. This is reflected in the KS statistic [Fig. 4(b)], where there is a drastic decrease when going from two qubits to three. This is possibly due to the method used here to construct the approximate exponential sum: rather than fixing the maximum allowed number of terms in advance, the method instead constructs a sum with a large number of terms and then afterward truncates to those with the largest weight. In this case, we find that the terms lost to truncation are not always negligible. This motivates future consideration of alternative methods for constructing approximate exponential sums that begin with the constraint of a maximum allowed number of terms, in order to make best use of the available memory resources. Nevertheless, we still see that our coarse-grained quantum models significantly outperform classical models with only a small number of qubits.

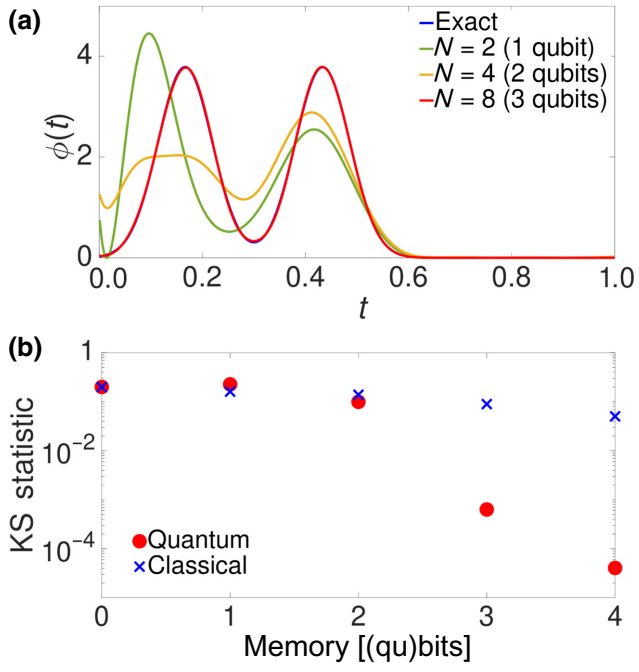


FIG. 4. (a) Wait-time distributions of compressed quantum models of a bimodal Gaussian process in arbitrary units. (b) KS statistics comparing the performance of compressed quantum models to compressed classical models.

V. COSTLY FEATURES?

We see that the quantum-compression protocol performs well on the two examples above. However, this begs the question of how well it performs in general and for which processes it will show the weakest performance. Ultimately, the accuracy of the model comes down to how good an approximation the finite exponential sum is of the wait-time distribution—or, conversely, the dimension required by the model depends on how few terms are required in the sum to reach a desired precision—as the compressed model will (experimental imperfections aside) provide an exact model of this approximation of the wait-time distribution. In this sense, the performance of our compression protocol comes down to how well the method used to construct an approximate exponential sum performs. For the particular algorithm used in our examples, see the discussion in the associated literature [56,59], also noting that the authors of these works find even better performance in practice than indicated by their bounds.

Nevertheless, we can find a useful heuristic in the information cost of the exact quantum model of the process—once the (logarithm of) the dimension drops below the information cost (i.e., once the capacity of the memory is lower than the information required for exact modeling), the compressed model must throw away useful information, limiting the accuracy it can achieve. Correspondingly, we can expect the performance of the quantum compression to be inversely correlated with the information cost of exact quantum modeling.

We can also deduce the features that would be most stubborn to compress. Consider our discussion above, comparing the exponential sum with expressing the function in the Laplace basis. Given that we want our sum to have as few terms as possible, problematic functions are those that are highly localized, as they have large spread in the Laplace basis. Indeed, the ultimate limit of this—δ functions—represent deterministic renewal processes; such processes do not allow a quantum advantage even in information cost in exact compression settings [34,37]. In Appendix C, we provide a case study of the performance of our quantum-compression protocol applied to a series of top-hat wait-time distributions of decreasing width. These processes represent increasingly accurate models of ideal clocks [48,60] and are also similarly difficult for classical compression methods. More generally, processes dominated by such sharp peaks are resistant to quantum compression in the information cost [37] and so

can be expected to also present difficulties for methods of compressing the memory dimension such as ours.

VI. DEPLOYMENT WITH GENERAL CONTINUOUS-TIME STOCHASTIC PROCESSES

A. Generalizing the protocol

Algorithm 1—our protocol for compressing quantum models of renewal processes—can be adapted to compress the temporal aspect of quantum models of general continuous-time processes with multiple modes and events [39,41].

Consider such a process with modes $g \in \mathcal{G}$, events $x \in \mathcal{X}$ and a transition dynamic Λ. The dynamic Λ effects an evolution according to $P(X, G'|G, T_0^{\leftarrow})$ describing the probability density of an event x occurring accompanied by a transition into mode g' in the next infinitesimal interval dt given that the system is currently in mode g, with time $t_0^{\leftarrow}$ since the last event. Following the corresponding literature on memory-minimal classical models [41], we assume a HSMM representation of the process where the subsequent mode is uniquely determined by (g, x)—independent of t_0. This is a slightly stronger condition than strictly necessary for the model to be causal and we discuss its relaxation later.

Along with the modal wait-time distributions $\sum_{xg'} P(x, g'|g)\phi^x_{g'g}(t)$, we can define a corresponding modal survival probability $\Phi_g(t) = \sum_{xg'} \int_t^\infty P(x, g'|g)\phi^x_{g'g}(t')dt'$ [39]. From these, one can then define a set of quantum memory states $\{|\varsigma_{gt_0^{\leftarrow}}\rangle\}$ and evolution $U_{\delta t}$ such that [61]

$$U_{\delta t}|\varsigma_{gt}\rangle|0\rangle := \sqrt{\frac{\Phi_g(t+\delta t)}{\Phi_g(t)}}|\varsigma_{gt+\delta t}\rangle|0\rangle + \sum_{xg'}\sqrt{\frac{\int_t^{t+\delta t} P(x, g'|g)\phi^x_{g'g}(t')dt'}{\Phi_g(t)}}|\varsigma_{g'0}\rangle|x\rangle. \tag{11}$$

We are now in a position to generalize Algorithm 1 for such processes. It transpires that this is, for the most part, simply a case of repeating the steps for renewal processes multiple times for each of the dwell functions.

To generalize steps 1 and 2, we define a function $\psi^x_{g'g}(t) := \sqrt{\phi^x_{g'g}(t)}$ for each of the dwell functions and, analogous to the case of renewal processes, approximate each of them by finite exponential sums $\tilde{\psi}^x_{g'g}(t)$:

$$\tilde{\psi}^x_{g'g}(t) = \sum_{j=1}^{N} c_j^{g'gx} e^{(-\gamma_j^{g'gx} + i\omega_j^{g'gx})t}. \tag{12}$$

Generalizing steps 3–5, we then similarly use these to construct a set of generator states $\{|\varphi_j^{g'gx}\rangle\}$, again defined implicitly in terms of an evolution operator:

$$\tilde{U}_{\delta t}|\varphi_j^{g'gx}\rangle|0\rangle = e^{(-\gamma_j^{g'gx} + i\omega_j^{g'gx})\delta t}|\varphi_j^{g'gx}\rangle|0\rangle + \sqrt{1 - e^{-2\gamma_j^{g'gx}\delta t}}|\tilde{\varsigma}_{g'0}\rangle|x\rangle. \tag{13}$$

Here, we analogously define memory states as linear combinations of these generator states:

$$|\tilde{\varsigma}_{gt}\rangle \propto \sum_{xg'j} \sqrt{P(x, g'|g)}\frac{c_j^{g'gx}}{\sqrt{2\gamma_j^{g'gx}}} e^{(-\gamma_j^{g'gx} + i\omega_j^{g'gx})t}|\varphi_j^{g'gx}\rangle. \tag{14}$$

This implicit definition can be used to determine the overlaps of the generator states, from which a reverse Gram-Schmidt procedure can be used to express them explicitly in terms of (at most) $N|\mathcal{X}||\mathcal{G}|$ orthonormal basis states [62]. In turn, the evolution operators and memory states may be expressed in this basis, completing the protocol.

We remark on a useful feature of this compression—that the modal wait-time distributions maintain their structure as a product of symbolic dynamics and a temporal component—with only this latter factor modified. That is, the compressed quantum models have the statistics of a process with the same transition topology but now with modal wait-time distributions $\sum_{xg'} P(x, g'|g)\tilde{\phi}^x_{g'g}(t)$, where $\tilde{\phi}^x_{g'g}(t) = |\tilde{\psi}^x_{g'g}(t)|^2$. This distortion introduces errors only in terms of the times when events occur and not the probability with which they occur. Moreover, the product structure entails that the distortion in the statistics of the compressed quantum model is no greater than the worst of the distortions of the $\phi^x_{g'g}(t)$ and that the errors in each interevent interval are independent. Thus, the performance of the protocol seen in the renewal process examples will still hold in this generalized setting. The memory of the resultant quantum model will be compressed to at most $N|\mathcal{X}||\mathcal{G}|$ dimensions.

B. Example

As an illustration of how the general case is little more than a straightforward application of Algorithm 1 for multiple times, we apply it to an example process consisting of dwell functions that are based on the examples above. Specifically, the process has two modes g_A, g_B and two possible events x, y, with the dwell function of both modes corresponding to an alternating Poisson process for event x and a bimodal Gaussian process for event y, and a transition structure such that the mode changes on event x and remains constant on event y; the probability of each event is different for the two modes. This is depicted as a HSMM in Fig. 5(a).

We measure the error in the accuracy as the average KS statistic, where the average is taken over events (for

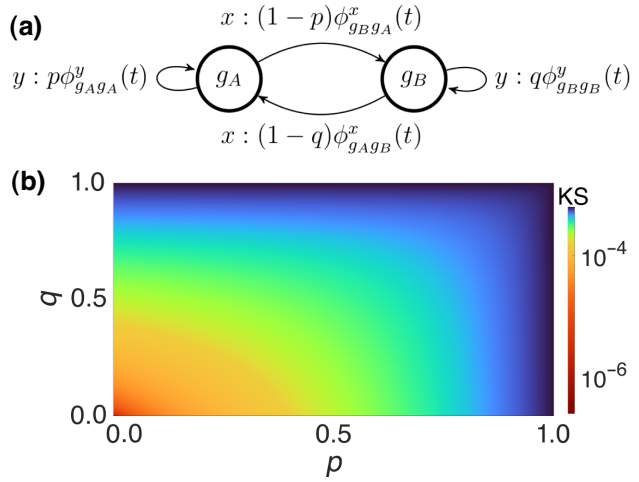


FIG. 5. (a) An HSMM representation of the example discussed in Sec. VI B. (b) The corresponding averaged KS statistic with a 32-dimensional memory for the full (p, q) parameter range.

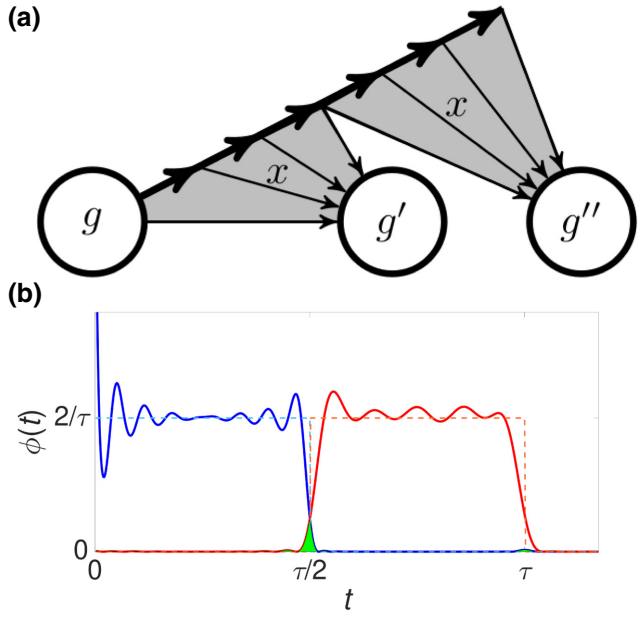


FIG. 6. (a) A HMM representation of a transition that does not satisfy the classical condition on mode update rules. (b) An example of spurious overlap of dwell functions introduced by the compression protocol. The dashed lines show the exact dwell functions, the solid lines the approximations, and the green areas the regions of overlap.

simplicity, we scale all dwell functions to have the same average mean firing rate, such that this also essentially corresponds to the average over time). That is, the average KS statistic $\overline{\mathrm{KS}} := \sum_{xgg'} \mathrm{KS}[\phi^x_{g'g}(t), \tilde{\phi}^x_{g'g}(t)]P(x,g)$. It is possible to apply the KS statistic in this way as the errors are constrained to a single interevent interval and there is no crossover of errors between the dwell functions of different events. Moreover, we need not calculate the approximations of the dwell functions anew—the approximations (and corresponding errors) found in Sec. IV are the very same approximations needed. In Fig. 5(b), we plot this for the full (p, q) parameter range for $N = 8$ (requiring 32 memory dimensions in total). Of note are the limits $p = q = 0$ (corresponding to only an alternating Poisson process) and $p = q = 1$ (corresponding to only a bimodal Gaussian process) where the errors take on their minimum and maximum, respectively, matching with those found for the renewal processes, while the error for the remainder of the parameter space interpolates between these two limits. Note that we neglect the extra dimensions made available by linear dependencies of generator states at the exceptional parameter regimes $p = 0, 1$, $q = 0, 1$, and $p = q$.

C. Scope for improvement?

Above, we follow the classical condition that the HSMM representation is such that the symbol and current mode alone determine the next mode. Yet the quantum models described in Eq. (11) still function correctly—and remain causal—with only the weaker condition on the HSMM representation that the triple (g, x, t_0) suffices to determine the subsequent mode. That is, emission of a given symbol from a given mode can result in a transition to two (or more) possible different modes, provided that also knowing the time spent in the current mode then provides sufficient information to determine the next mode. That is, the classical convention requires $H(G'|G, X) = 0$, while the exact quantum models assume only that $H(G'|G, X, T_0) = 0$ [here, $H(.)$ is the Shannon entropy [63]]. An example of such a transition satisfying only the weaker condition is illustrated in Fig. 6(a).

However, in the case where only this weaker condition holds, there can be interference between the generator states corresponding to transitions with the same symbol and initial mode but a different end mode. This manifests from errant overlaps of the approximate dwell functions $\tilde{\phi}^x_{g'g}(t)$: while $\int_0^\infty \phi^x_{g'g}(t)\phi^x_{g''g}(t)dt = 0 \forall x, g, g', g'' \neq g'$, this may not hold true for the $\tilde{\phi}^x_{g'g}(t)$. That is, there may be times t for which $\tilde{\phi}^x_{g'g}(t)$ and $\tilde{\phi}^x_{g''g}(t)$ $(g'' \neq g')$ are simultaneously nonzero, violating the condition $H(G'|G, X, T_0) = 0$. Such a violation cannot occur under the stronger classical condition, as we are already guaranteed that there is at most one g' for each pair (g, x) for which $\phi^x_{g'g}(t)$ [and thus $\tilde{\phi}^x_{g'g'}(t)$] is not zero everywhere.

As an example, consider a process where the dwell time associated with mode g and event x is uniformly distributed over the interval $[0, \tau]$, with the system transitioning into mode g' if the dwell time is less than $\tau/2$ and into g'' if it is greater than (or equal to) $\tau/2$. Then,

$\phi^x_{g'g}(t)$ is a uniform distribution over $[0, \tau/2)$ and $\phi^x_{g''g}(t)$ a uniform distribution over $[\tau/2, \tau]$. When we parse this through the compression protocol, the approximated distributions have a nonzero overlap and so have interfering probability amplitudes. This is illustrated in Fig. 6(b) for $N = 16$.

This interference requires us to modify the symbolic transition probabilities $P(X, G'|G)$ to an approximate form $\tilde{P}(X, G'|G)$ in order to appropriately normalize the memory states, which will correspondingly distort the transition structure. In particular, it can result in the model transitioning to superpositions of memory states, manifesting new (potentially infinitely many) effective modes. While these effective modes do not require additional memory dimensions to track (as they are linear combinations of existing memory states), they do allow for a gradual accumulation of errors over time, as the errors are now able to propagate across multiple interevent intervals. A further complication is presented in the freedom of choice in how to actually assign $\tilde{P}(X, G'|G)$ to enforce proper normalization—while a simple rescaling of $P(X, G'|G)$ would work, it is also possible to achieve this with an uneven rescaling, which may result in greater accuracy by offsetting the effect of the interference.

Note that the magnitude of the interference scales with the overlaps of the memory states for each mode—and hence the overlaps of their statistics: thus, the more distinguishable the statistics of the modes are, the smaller is the distortion. Further, as noted above, with the stronger condition imposed on classical models these overlaps cannot occur and thus when compressing a given such classical model we can sidestep such interference. Nevertheless, embracing this weaker condition may unlock even greater compression potential; we leave the optimization of the $\tilde{P}(X, G'|G)$ in such settings as an open question for future work.

VII. DISCUSSION AND CONCLUSIONS

We have introduced a lossy compression protocol for the quantum modeling of stochastic temporal dynamics. By harnessing nonclassical features of quantum state spaces—namely, that sets of quantum states can be at once linearly dependent and nondegenerate—an effective coarse graining of the state space inhabited by a quantum memory can be realized. This achieves a much greater compression than is possible with analogous classical methods and exact quantum compression alike. The relaxation from exact to near-exact replication naturally fits into applications where the dynamics of the system to be modeled have been inferred through observation [64,65] and are thus already an approximation of the true dynamics. This also brings the additional benefit of placing less demand on the precision of the quantum processor implementing the simulation, which in current realistic settings should not be assumed to be noiseless.

Going forward, our work encourages the development of similar lossy compression beyond tracking the temporal component of stochastic processes. For example, the framework can be applied to compress quantum clocks [48,60] and motivates the extension to other models with continuous state spaces, such as belief spaces [66,67] used in reinforcement learning [68]. Further avenues include development of analogous methods for compressed modeling of purely symbolic dynamics and input-output processes [69].

Furthermore, in spite of the significant compression advantage offered by our protocol, it is by no means optimal. Two aspects we foresee as presenting opportunities for enhancing the compression are in the choice of algorithm for constructing an approximate exponential sum and in allowing for more general complex $\psi(t)$ to be considered. Pursuing the former of these may allow for more faithful approximations of the wait-time distribution without increasing the number of allowed states. In the latter, we have a family of functions we can attempt to approximate and we need only take the one that we can most faithfully represent. In the case of general continuous-time processes, the question remains open how to best handle cases where the classical condition that the dynamics factor into a product of temporal and symbolic dynamics does not hold. Further improvements in this regime may also be found by taking a more holistic approach that coarse grains the Hilbert space in terms of symbolic and temporal dynamics simultaneously. Nevertheless, even in this initial foray, we see the potential for drastic improvement over classical techniques. Moreover, the high fidelities reached with comparatively few dimensions places it well within reach of current and near-term small-scale quantum processors with only a handful of qubits [33,70], offering exciting prospects for imminent experimental realizations.

ACKNOWLEDGMENTS

We thank Andrew Garner and Mile Gu for discussions. This work was funded by the Imperial College Borland Fellowship in Mathematics and Grant No. FQXi-RFP-1809 from the Foundational Questions Institute and the Fetzer Franklin Fund (a donor-advised fund of the Silicon Valley Community Foundation).

APPENDIX A: APPROXIMATE EXPONENTIAL SUMS

In Algorithm 1, step 2 requires that we construct an exponential sum approximating the square root of the wait-time distribution. Here, we use the method of Beylkin and Monzón [56], summarized in Algorithm 2.

Inputs: Exact function $\psi(t)$, with the domain scaled such that the region to be approximated is the interval $[0, 1]$, target precision ϵ.
Outputs: Set of triples $\{(c_j, \gamma_j, \omega_j)\}$ yielding approximate decomposition $\psi(t) \approx \sum_{j=1}^{N} c_j \exp((-\gamma_j + i\omega_j)t)$.

1: Construct M-dimensional vector $h_j := \psi(j/M)$ for $0 \leq j \leq M$ with M sufficiently large to oversample function.
2: Construct Hankel matrix $H_{jk} := h_{j+k}$ and find eigenvector σ corresponding to the eigenvalue closest to ϵ.
3: From the elements of σ, construct the polynomial $\sum_{j=0}^{M/2} \sigma_j z^j$ and solve to find the roots $\{\Gamma_k\}$.
4: Construct Vandermonde matrix $V_{jk} := \Gamma_k^j$ for $0 \leq j \leq M$. Invert to find solutions $\{c_k\}$ to $h_j := \sum_{k=1}^{M/2} c_k V_{jk}$.
5: Define $\{-\gamma_j\}$ and $\{\omega_j\}$ as the real and imaginary parts of $\{\ln(\Gamma_j)\}$ respectively.

Algorithm 2 Approximate exponential sum [56]

For our purposes, to obtain an N-term approximate sum, we keep only the triples with the N largest magnitudes for weights $\{|c_j|\}$. Prior to this, we also discard any terms with nonpositive γ_j [to ensure a valid quantum state of the form Eq. (2) can be constructed] and rescale the weights by a constant factor to ensure that the sum has unit L_2 norm. By varying the precision ϵ, we obtain different decompositions, with truncation to fewer terms favoring larger ϵ and, conversely, larger number of terms performing better with smaller ϵ. In our examples, we take $M = 1000$ and vary ϵ to find the most accurate decomposition for each N (according to the KS statistic), ultimately using values in the range 10^{-12} to 10^{-1}. We use the ROOTS function of GNU OCTAVE [71] to numerically solve the polynomials and $c = (V^T V)^{-1} V^T h$ to solve the overconstrained Vandermonde system.

APPENDIX B: LOSSY CLASSICAL COMPRESSION METHOD

As shown in Fig. 2, the transition structure between the memory states of the ε-machine of a renewal process takes the form of an incrementing counter that resets upon events. A finite-dimensional approximation must adopt the structure of Fig. 7, where the counter progresses up to a terminal state, upon which it loops back to an earlier state [32]. The variable parameters are the transition probabilities $\{p_j\}$, the time-step size Δt, and the target state of the loop R.

The optimal lossy classical compression at fixed dimension is found by minimizing the associated cost function over all possible choices of these parameters. We use a standard gradient-descent-based approach to seek the minimum of the KS statistic. For each possible choice of loop state R, we generate W seeds of random parameters for $(\{p_j\}, \Delta t)$ and run S steps of update

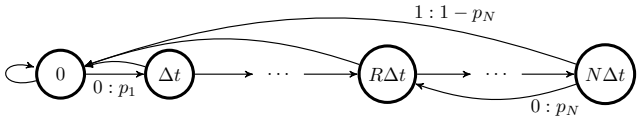


FIG. 7. The HMM topology of the most accurate $(N + 1)$-dimensional approximation of a renewal process. The variable parameters are the transition probabilities $\{p_j\}$, the time-step size Δt, and the position of the loop R.

according to $p_j \to p_j - \eta_p \nabla_j D(\{p_j\}, \Delta t)$ and $\Delta t \to \Delta t - \eta_t \nabla_t D(\{p_j\}, \Delta t)$, where D is the KS statistic (with hard constraints to ensure that the parameters remain physical). We then keep the final parameter set that reaches the minimum value of D across all choices of loop state and seeds. As with the quantum method, we rescale the wait-time distribution to the domain $[0, 1]$ and for purposes of numerical evaluation discretize it into 1000 steps. We again remark that we do not claim this method to necessarily yield the very optimal lossy classical compression at fixed dimension but simply that it should offer a ballpark figure as to its performance. That is, we believe it is reasonable to expect that the optimal classical compression will not perform significantly better than the explicit examples we find here.

We generate the initial seeds for $\{p_j\}$ uniformly in the interval $[0, 1]$ and Δt exponentially decaying. For the alternating Poisson process, we find the best performance by taking learning rates $\eta_p = 10^{-4}$ and $\eta_t = 10^{-8}$, with gradients estimated over discrete intervals $\delta p = 10^{-3}$ and $\delta t = 10^{-4}$. Empirically, the descents appear to converge on a minimum within $S = 12\,500N$ steps and running more than $W = 1000$ seeds for each loop state does not seem to yield any improved minima. For the bimodal Gaussian process, we find the best performance with much the same parameters; a slight improvement is found by increasing the learning rates by a factor of 10 for the first $1250N$ steps of descent, whereupon convergence is reached within $S = 6250N$ steps.

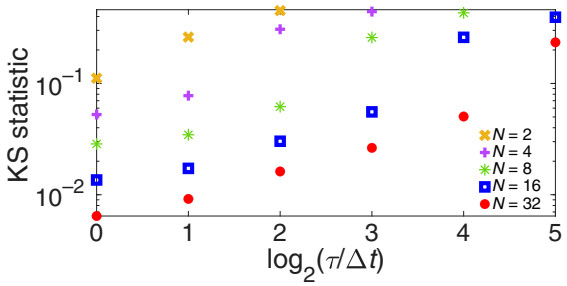


FIG. 8. KS statistics showing the performance of compressed quantum models for renewal processes with top-hat distributions of varying width.

APPENDIX C: COMPRESSION OF TOP-HAT DISTRIBUTIONS

In Sec. V, we discuss how wait-time distributions with sharp peaks are hard to compress. Here, we illustrate this with a case study of renewal processes with top-hat distributions, showing how the performance of the quantum compression degrades with narrowing of the width.

Such top-hat distributions of width Δt take on uniform values between $\tau - \Delta t$ and τ (and zero elsewhere), with τ forming a arbitrary scaling factor. We consider the cases $\Delta t/\tau = 2^0, 2^1, \ldots, 2^5$ and use Algorithm 1 to construct quantum models of between one- and five-qubit memories. In running Algorithm 2, we set $M = 6000$ and place τ at 512, with the long time scale properly accounting for the long tails of the poorer-performing models. The best performance is found for ϵ in the range of 10^{-3} to 10^{2}.

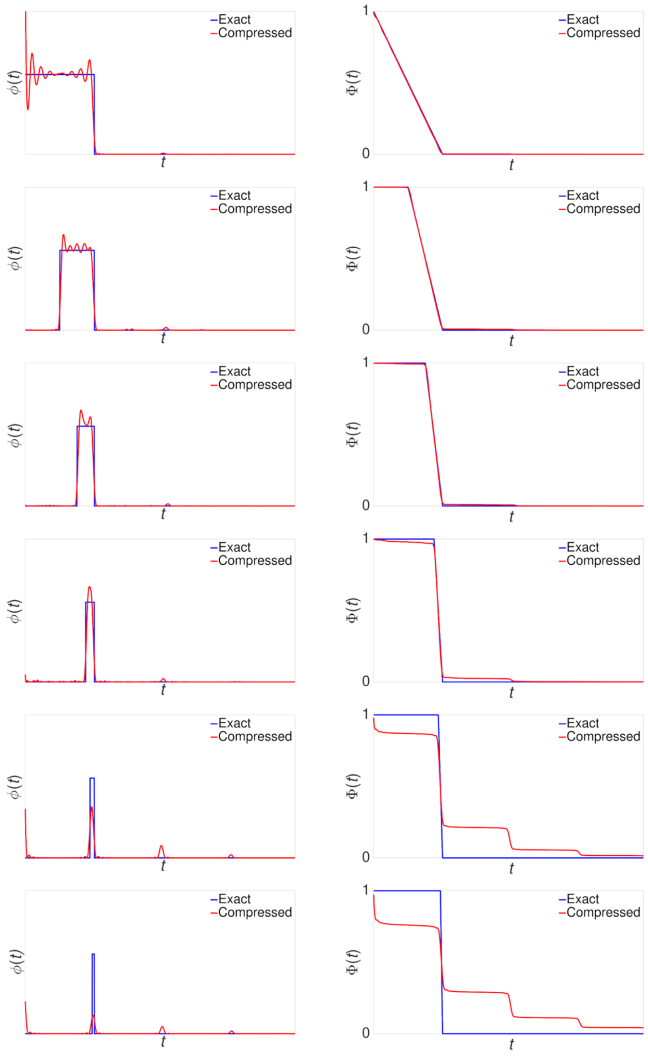


FIG. 9. The wait-time distributions (left) and survival probabilities (right) of four-qubit compressed quantum models for renewal processes with top-hat distributions of varying width (arbitrary units).

In Fig. 8, we plot the KS statistic found for the quantum models, truncated to the smallest width that the model can simulate with a KS statistic below 0.45 (note that a memoryless model of any process exists with a KS statistic no greater than 0.5). We see that each halving of the width requires an additional qubit to model with roughly the same accuracy, as one would intuitively expect—in the classical case, doubling the number of states allows a model with half the time-step size. As a point of comparison, consider that a classical model with $N < \tau/\Delta t$ cannot beat a KS statistic of 0.5; to see this, consider that the best classical model in this instance would be a deterministic counter that emits only on the last state, positioned to coincide with the time where $\Phi(t) = 0.5$. In Fig. 9, we compare wait-time distributions and survival probabilities of the compressed four-qubit quantum models to their exact counterparts for each of the widths. As the width narrows, the periodicity of the approximate distributions can be seen, due to competition between suppressing these spurious peaks with the exponential decay and the need to not suppress the modeled peak.

[1] W. H. Zurek, Complexity, entropy and the physics of information, Santa Fe Institute Studies in the Sciences of Complexity (1990).

[2] R. Badii and A. Politi, *Complexity: Hierarchical Structures and Scaling in Physics* (Cambridge University Press, Cambridge, 1999), 6.

[3] W. B. Arthur, Complexity and the economy, Science **284**, 107 (1999).

[4] M. Gell-Mann, in *Complexity and Industrial Clusters* (Springer, Heidelberg, 2002), p. 13.

[5] J. M. Ottino, Engineering complex systems, Nature **427**, 399 (2004).

[6] V. Grimm, E. Revilla, U. Berger, F. Jeltsch, W. M. Mooij, S. F. Railsback, H.-H. Thulke, J. Weiner, T. Wiegand, and D. L. DeAngelis, Pattern-oriented modeling of agent-based complex systems: Lessons from ecology, science **310**, 987 (2005).

[7] N. Boccara, *Modeling Complex Systems* (Springer Science & Business Media, New York, 2010).

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

[9] S. E. Levinson, Continuously variable duration hidden Markov models for automatic speech recognition, Comput. Speech Lang. **1**, 29 (1986).

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

[11] D. Kulp, D. Haussler, M. G. Reese, and F. H. Eeckman, in *Proceedings of the 4th International Conference on Intelligent Systems for Molecular Biology* (AAAI Press, St Louis MO, 1996), p. 134.

[12] P. Tino and M. Koteles, Extracting finite-state representations from recurrent neural networks trained on chaotic symbolic sequences, IEEE Trans. Neural Networks **10**, 284 (1999).

[13] A. J. Palmer, C. W. Fairall, and W. A. Brewer, Complexity in the atmosphere, IEEE Trans. Geosci. Remote Sens. **38**, 2056 (2000).

[14] S.-Z. Yu, Z. Liu, M. S. Squillante, C. Xia, and L. Zhang, in *Performance, Computing, and Communications Conference, 2002. 21st IEEE International* (IEEE, Phoenix AZ, 2002), p. 65.

[15] W. Gerstner and W. M. Kistler, *Spiking Neuron Models: Single Neurons, Populations, Plasticity* (Cambridge University Press, Cambridge, 2002).

[16] E. Bonabeau, Agent-based modeling: Methods and techniques for simulating human systems, Proc. Natl. Acad. Sci. **99**, 7280 (2002).

[17] R. W. Clarke, M. P. Freeman, and N. W. Watkins, Application of computational mechanics to the analysis of natural data: An example in geomagnetism, Phys. Rev. E **67**, 016203 (2003).

[18] J. Bulla and I. Bulla, Stylized facts of financial time series and hidden semi-Markov models, Comput. Stat. Data Anal. **51**, 2192 (2006).

[19] J. B. Park, J. W. Lee, J.-S. Yang, H.-H. Jo, and H.-T. Moon, Complexity analysis of the stock market, Phys. A: Stat. Mech. Appl. **379**, 179 (2007).

[20] C.-B. Li, H. Yang, and T. Komatsuzaki, Multiscale complex network of protein conformational fluctuations in single-molecule time series, Proc. Natl. Acad. Sci. **105**, 536 (2008).

[21] D. J. Wilkinson, Stochastic modelling for quantitative description of heterogeneous biological systems, Nat. Rev. Genet. **10**, 122 (2009).

[22] R. Haslinger, K. L. Klinkner, and C. R. Shalizi, The computational structure of spike trains, Neural Comput. **22**, 121 (2010).

[23] S.-Z. Yu, Hidden semi-Markov models, Artif. Intell. **174**, 215 (2010).

[24] P. E. Smouse, S. Focardi, P. R. Moorcroft, J. G. Kie, J. D. Forester, and J. M. Morales, Stochastic modelling of animal movement, Philos. Trans. R. Soc. B: Biol. Sci. **365**, 2201 (2010).

[25] E. Garavaglia and R. Pavani, About earthquake forecasting by Markov renewal processes, Methodol. Comput. Appl. Probab. **13**, 155 (2011).

[26] D. Kelly, M. Dillingham, A. Hudson, and K. Wiesner, A new method for inferring hidden Markov models from noisy time sequences, PLoS ONE **7**, e29703 (2012).

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

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

[29] 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).

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

[31] 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).

[32] 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).

[33] 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).

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

[35] 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).

[36] 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).

[37] 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).

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

[39] 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).

[40] T. J. Elliott, Memory compression and thermal efficiency of quantum implementations of non-deterministic hidden Markov models, Phys. Rev. A **103**, 052615 (2021).

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

[42] A. Khintchine, Korrelationstheorie der stationären stochastischen prozesse, Mathematische Ann. **109**, 604 (1934).

[43] J. P. Crutchfield, The calculi of emergence: Computation, dynamics and induction, Phys. D: Nonlinear Phenomena **75**, 11 (1994).

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

[45] L. Rabiner and B. Juang, An introduction to hidden Markov models, IEEE Acoustics Speech Signal Process. Mag. **3**, 4 (1986).

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

[47] Y. Yang, G. Chiribella, and M. Hayashi, Quantum stopwatch: How to store time in a quantum memory, Proc. R. Soc. A: Math. Phys. Eng. Sci. **474**, 20170773 (2018).

[48] Y. Yang and R. Renner, Ultimate limit on time signal generation, arXiv:2004.07857 (2020).

[49] J. P. Crutchfield and K. Young, Inferring Statistical Complexity, Phys. Rev. Lett. **63**, 105 (1989).
[50] W. L. Smith, Renewal theory and its ramifications, J. R. Stat. Soc. Ser. B (Methodological) **20**, 243 (1958).
[51] Note that, in principle, an arbitrary time-dependent complex phase can be added to $\psi(t)$; provided that $|\psi(t)|^2 = \phi(t)$, the model will still yield the correct statistics, albeit with a potentially different memory cost. We return to this point later.
[52] L. M. Delves and J. L. Mohamed, *Computational Methods for Integral Equations* (CUP Archive, Cambridge, 1988).
[53] We do not designate any particular such method as the optimal. In testing our protocol, we find that the algorithm of Beylkin and Monzón [56] performs well, with low computational cost.
[54] Unlike Eq. (3), there is no freedom in these amplitudes; changing their magnitude is equivalent to changing γ_j, while the phase factors correspond to different c_j and ω_j.
[55] P. Dennery and A. Krzywicki, *Mathematics for Physicists*, Dover Books on Physics Series (Dover Publications, Mineola NY, 1996).
[56] G. Beylkin and L. Monzón, On approximation of functions by exponential sums, Appl. Comput. Harmon. Anal. **19**, 17 (2005).
[57] F. J. Massey Jr, The Kolmogorov-Smirnov test for goodness of fit, J. Am. Stat. Assoc. **46**, 68 (1951).
[58] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer Science & Business Media, New York, 2009).
[59] G. Beylkin and L. Monzón, Approximation by exponential sums revisited, Appl. Comput. Harmon. Anal. **28**, 131 (2010).
[60] M. P. Woods, R. Silva, G. Pütz, S. Stupar, and R. Renner, Quantum clocks are more accurate than classical ones, arXiv:1806.00491 (2018).
[61] These exact quantum models generally achieve greater compression in information cost than analogous prior quantum models [39] as they better compensate for non-Markovianity in the transitions between modes.
[62] It is possible that some generator states may be linear combinations of those belonging to other modes. This does not break the protocol, though it will result in some memory dimensions being left unused. It may then be possible to use these dimensions to incorporate additional terms into the approximate exponential sums to increase their accuracy.
[63] M. A. Nielsen and I. Chuang, Quantum Computation and Quantum Information (2000).
[64] S. E. Marzen and J. P. Crutchfield, Inference, prediction, and entropy-rate estimation of continuous-time, discrete-event processes, arxiv:2005.03750 (2020).
[65] 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).
[66] A. R. Cassandra, L. P. Kaelbling, and M. L. Littman, in *Proceedings of the Twelfth AAAI National Conference on Artificial Intelligence* (AAAI Press, Seattle WA, 1994), p. 1023.
[67] L. P. Kaelbling, M. L. Littman, and A. W. Moore, Reinforcement learning: A survey, J. Artif. Intelligence Res. **4**, 237 (1996).
[68] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction* (MIT press, Cambridge MA, 2018).
[69] 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).
[70] V. Negnevitsky, M. Marinelli, K. Mehta, H.-Y. Lo, C. Flühmann, and J. P. Home, Repeated multi-qubit readout and feedback with a mixed-species trapped-ion register, Nature **563**, 527 (2018).
[71] *GNU Octave* https://octave.org/doc/v5.2.0/finding-roots.html.