Memory Minimal Predictive Models for Quantum Processes

Graeme D. Berk[1] * Jayne Thomson[2 3] Mile Gu[1 3]

[1] *Nanyang Quantum Hub, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 639673.*

[2] *Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, Singapore 138632.*

[3] *Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117543.*

Abstract. The success of increasingly enormous machine learned models begs the question: what is the minimum quantity of information a model requires to predict a process faithfully? Minimal models saturating the memory constraints imposed by physics represent the 'Occam's Razor' for observed behaviour. We extend these concepts to the quantum regime. Leveraging a novel quantum model reduction principle which we call linear dependence of future morphs, we derive a lower bound on the memory dimension of the irreducible causal mechanism that generates a quantum process. We produce an algorithm to approximately compute this bound. Using this algorithm, we numerically study the reducibility of unitary quantum processes, and observe the close link to probabilistic distinctness of model memory states.

1 Introduction

Predictive modelling is ubiquitous in quantitative science. The goal is to build a causal representation of a dynamical process, to make statistically faithful predictions about the observable future, based on the observable past. In general, a process will have many potential models, which may differ greatly in the quantity of information about the past that is stored and processed. The recent trend in machine learning has been towards larger models [10]. However, computational resources are not unlimited, motivating the desire for model reduction.

In complexity science, minimal models - and the amount of past information they require - are extensively studied in the context of modelling stochastic processes [19]. Here, each candidate predictive model is a particular sub-class of hidden Markov models (HMM) – finite-state machines with a preset encoding function that tells one which state to configure the machine in if we see a particular past. The memory requirement is then given by the past information needed to communicate the states of such a machine (e.g., the number of memory states) for optimal prediction. The provably memory minimal causal (classical) model is known as the ϵ-machine (CC in Tab. 1). Such machines operate by encoding histories of outcomes into the same memory state if they lead to equivalent statistics for the future process – a property we refer to as causal equivalence of future morphs (CEFM). The states of the ϵ-machine are known as causal states, signifying the minimal causes one must postulate to fully explain the dynamics of a process. ϵ-machines have since seen extensive use in analysing the causal relations of various complex systems [4], and have potential the to improve the performance of reinforcement-learning agents [23].

The potential benefit of quantum models came into sharp focus when it was discovered that quantum models of classical stochastic processes can be more efficient than even maximally efficient ϵ-machines [9]. Consequently, there is an ongoing effort to use quantum tools to study classical processes [5, 6, 8, 12, 14, 22] (QC in Tab. 1). It has been shown that the degree of this advantage can be unbounded for certain classes of processes [1, 7].

However, there is a growing necessity to re-equip with tools that can properly handle inherently quantum systems. Numerous classical and semi-classical methods to model quantum systems (CQ in Tab. 1) have been employed in the past [2, 11, 15, 20, 21]. However, classically grounded methods are ultimately unable to probe complex quantum systems. Thus, what is still missing is memory-efficient quantum models of quantum processes (QQ in Tab. 1). The QQ approach encompasses solutions to CC, CQ, and QC under appropriate restrictions. However, there are numerous technical hurdles that have led to a conspicuous absence of QQ results from the computational mechanics literature. At a glance, it is not even clear that memory-optimal predictive modelling of quantum processes is even possible at

*graemedean.berk@ntu.edu.sg

Table 1: The paradigms of predictive modelling of classical and quantum systems. One can use classical tools (ϵ-machines) to study classical stochastic processes, e.g. Refs. [16–18]. Alternatively, one can gain a memory advantage storing the classical causal states in a quantum memory [1, 5–9, 12, 14, 22]. Quantum processes warrant investigation in their own right, as done using classical ϵ machines in e.g. Ref. [15]. In this work we close this gap by introducing the q-transducer, and a lower bound on the memory for quantum prediction.

		Tools: Classical	Tools: Quantum
System: Classical		**CC** ϵ-Machines	**QC** Memory Advantage
System: Quantum		**CQ** HMM for QM	**QQ** q-Transducers

all, since – unlike for the classical finite state processes ϵ-machines are based on – there is a continuum of potential inputs and outputs, even when the Hilbert space is finite dimensional. We resolve these technical challenges, which allows us to quantify the fundamental limits on the memory required to predict a quantum future.

2 Observed Statistics

In our framework, an experimenter acts on a system s at a countable number of times. Between the actions of the experimenter, s evolves due to an interaction with some inaccessible environment. Thus s is influenced by the interlocking of two influences over time: the experimenter and the environment, which can be described by quantum combs [3] $\mathbf{A}$ and $\mathbf{T}$ respectively. Without loss of generality for the models we produce, we describe the control policy $\mathbf{A}$ of the experimenter as at each time t sending a pure s state into $\mathbf{T}$, and projectively measuring an outcome on s. The experimenter may also have access to an ancillary memory subsystem a. Given a policy $\mathbf{A}$, the experimenter will observe a sequence of measurement outcomes $\overleftrightarrow{x} = \ldots, x_{t-1}, x_t, x_{t+1}, \ldots$ with probability distribution $p(\overleftrightarrow{x})$. The mapping between outcomes of given a policy and probability distributions on outcome sequences defines the tomographic representation of the process tensor [13] $\mathbf{T}$.

3 Quantum Predictive Modelling

We formalise quantum predictive modelling as the problem of finding an encoding function e that takes a history of actions and observations $\mathbf{A}_{\overleftarrow{x}}$, and produces a memory state $\rho^m = e\left(\mathbf{A}_{\overleftarrow{x}}\right)$ such that ρ^m can be identified with a conditional future process tensor $\mathbf{T}^\rho$ (analogous to conditional probability distributions for stochastic processes) whose statistics are indistinguishable from the conditional future $\mathbf{T}^{\mathbf{A}_{\overleftarrow{x}}}$ obtained by contracting the process itself with the action-outcome history $\mathbf{A}_{\overleftarrow{x}}$, i.e. $\mathbf{T}^\rho = \mathbf{T}^{\mathbf{A}_{\overleftarrow{x}}}$.

We show that any valid system-memory dilation of a process tensor always can define an encoding function, and hence a quantum model $\mathscr{M}(\mathbf{T}) = (\mathcal{H}^s, \mathcal{H}^m, \mathcal{T}^{sm})$. $\mathcal{H}^s, \mathcal{H}^m$ are the system and memory Hilbert spaces, and $\mathcal{T}^{sm}$ is the biparite channel evolving the process in time. However, what are the the fundamental limits on how simple such a dilation can be? Classically, this was answered using ϵ-machines by invoking CEFM. Critically, since it is possible for quantum states to be non-identical yet still non-orthogonal, quantum causal equivalence is a global property, not a local one. This renders CEFM insufficient to specify a history to memory encoding. Realising this, we define linear independence of future morphs (LDFM) as the inner product between a given future – represented by a multitime Choi state in dilated form – against an orthonormal 'reference set' of other futures, maximised over all possible dilations.

Theorem 1 (LDFM and Causal Modelling) *Given a process* $\mathbf{T}$*, a reference set of conditional futures* $\Xi = \{|\mathbf{B}_l\rangle\}_{l=1}^{L}$*, and a trial future* $\mathbf{R}$*, the LDFM* $\mathcal{L}(\mathbf{R}|\Xi) = 1$ *iff the observable statistics of* $\mathbf{R}$ *can be expressed as a convex sum of superpositions of elements in* Ξ*.*

We define the q-transducer as the model induced by applying LDFM to the space of possible futures.

Definition 2 (q-Transducer) *Given a stationary quantum process* $\mathbf{T}$ *on a finite dimensional Hilbert space* $\mathcal{H}^s$*, its q-transducer is the model* $\mathscr{M}_q(\mathbf{T}) = (\mathcal{H}^s, \mathcal{H}^m, \mathcal{T}^{sm})$*.* $\mathcal{H}^m$ *is the lowest dimension Hilbert space such that there exists a unifilar biparite quantum channel* $\mathcal{T}^{sm}$ *such that LDFM=1 for all futures* $\mathbf{T}$ *is capable of exhibiting.*

Stationarity is a property of a process where its statistics are time-translation invariant, i.e. the results of an experiment do not depend on when it begins. Unifilarity means that one's knowledge of the hidden m state does not drift into uncertainty over time. This property is crucial to predictive modelling. We show that all unitary processes are unifilar, although the converse is not necessarily true.

3.1 Explicit Construction of Lower Bound

A necessary condition for a q-transducer of a given dimension d to exist is that an orthonormal set of memory states with size d can produce any future $\mathbf{T}$ is capable of exhibiting. We can use this fundamental fact to produce a model reduction algorithm that approximates a lower bound on q-transducer dimension.

Alg. (1) starts with an arbitrary (non-minimal) model $\mathscr{R}(\mathbf{T}) = (\mathcal{H}^s, \mathcal{H}^r, \mathcal{R}^{sr})$, and seeks to find an orthonormal reference set Ξ of dilated futures that can produce every future in some trial set of futures Θ, but is smaller than a basis for $\mathcal{H}^r$. The algorithm begins with a single element in Ξ and a single element in Θ. It proceeds by testing if Ξ can fully describe Θ as quantified by LDFM. It then iteratively adds elements to Ξ if LDFM is increased beyond a specified tolerance, until each future in Θ is fully described by Ξ. The algorithm then adds another element to Θ and repeats. The convex hull of all futures in Θ describes a convex polytope that approximates the full $\mathcal{H}^r$ space to increasing completeness. Crucially, at every step $d = |\Xi|$ is an monotonically growing lower bound on the dimension of the q-transducer. Additionally, in explicit computations the length of futures must be limited to some finite value n, meaning that our algorithm is only exact for processes with Markov order $\leq n$. A formal lower bound is obtained in the limit of $n \to \infty$ and Θ contains all pure states in $\mathcal{H}^r$ (that are realisable under $\mathbf{T}$).

This is the memory limit imposed on a quantum process by applying the principle of Occam's razor to causal equivalence, as has been done for classical processes to yield ϵ-machines. However, unlike for classical processes, for quantum processes there is an additional consideration. Does there actually exist an valid quantum channel that can evolve the minimal causal description of a quantum process from one memory state to another? Put more simply, do some quantum processes have irreconcilable memory redundancies? This remains an open question.

3.2 Reducibility of Models

A critical property to the minimality of of classical ϵ-machines is that each memory state yields distinct future probabilities compared to each other one. Intuitively, one might expect that a model not satisfying this property should not be minimal, since the 'redundant' state can be removed without changing its predictions. However, this intuition is misleading for quantum models. If two memory states yield identical futures, that does not imply a superposition of the two cannot produce a future that neither could produce independently. As such, Alg. (1) only excludes memory states from a model if those states do not contribute any new statistics under every possible superpostion. We refer to this property as weak probabilistic distinctness (WPD). On the other hand, strong probabilistic distinctness (SPD) is the property that each memory state has distinct futures compared to every other one. SPD is closely linked to minimality, because removing any basis element will necessarily lead to futures with no memory state to describe them. However, we show by explicit counterexample that minimal models need not satisfy SPD.

Which models satisfy SPD? A sufficient condition for SPD is that in the limit of time-steps $n \to \infty$, the LDFM between dilated futures approaches the fidelity between the memory states that generate those futures. We have performed numerical experiments suggesting that Haar random unitary models satisfying SPD have full measure, and correspondingly tend to only be reducible using Alg. (1) when the Markov order approximation is severely limited. On the other hand, sparse random unitaries were not observed to satisfy our fidelity condition for SPD, and are typically reducible using Alg. (1).

4 Discussion

Model reduction on quantum processes has potential to yield practical benefits for tasks including quantum compilation, simulating large quantum systems like molecules, noise reduction, and more general optimal control. To this end, we generalised foundational concepts from computational mechanics to the quantum paradigm via the process tensor formalism. From this, we devised an algorithm to find a quantum process' minimal causal explanation, which is a lower bound on the memory dimension of any model that can reproduce its statistics. It is currently unknown if this bound is tight, or whether there are processes for which quantum mechanics enforces fundamental causal redundancies. Regardless, our numerical study supports the potential for model reduction on a wide range of quantum processes.

References

[1] Aghamohammadi, C., Mahoney, J. R., and Crutchfield, J. P. (2017). Extreme quantum advantage when simulating classical systems with long-range interaction. *Sci. Rep.*, 7(1):6735.

[2] Carmichael, H. J. (1993). Quantum trajectory theory for cascaded open systems. *Phys. Rev. Lett.*, 70:2273–2276.

[3] Chiribella, G., D'Ariano, G. M., and Perinotti, P. (2009). Theoretical framework for quantum networks. *Physical Review A*, 80(2):022339.

[4] Crutchfield, J. P. (2012). Between order and chaos. *Nature Physics*, 8(1):17–24.

[5] Elliott, T. J. and Gu, M. (2018). Superior memory efficiency of quantum devices for the simulation of continuous-time stochastic processes. *npj Quantum Inf.*, 4(1):18.

[6] Elliott, T. J., Gu, M., Garner, A. J. P., and Thompson, J. (2022). Quantum adaptive agents with efficient long-term memories. *Phys. Rev. X*, 12:011007.

[7] Elliott, T. J., Yang, C., Binder, F. C., Garner, A. J. P., Thompson, J., and Gu, M. (2020). Extreme dimensionality reduction with quantum modeling. *Phys. Rev. Lett.*, 125:260501.

[8] Ghafari, F., Tischler, N., Thompson, J., Gu, M., Shalm, L. K., Verma, V. B., Nam, S. W., Patel, R. B., Wiseman, H. M., and Pryde, G. J. (2019). Dimensional quantum memory advantage in the simulation of stochastic processes. *Phys. Rev. X*, 9:041013.

[9] Gu, M., Wiesner, K., Rieper, E., and Vedral, V. (2012). Quantum mechanics can reduce the complexity of classical models. *Nat. Commun.*, 3(1):762.

[10] Kaddour, J., Harris, J., Mozes, M., Bradley, H., Raileanu, R., and McHardy, R. (2023). Challenges and applications of large language models.

[11] Kleinmann, M., Gühne, O., Portillo, J. R., Larsson, J.-Å., and Cabello, A. (2011). Memory cost of quantum contextuality. *New J. Phys.*, 13(11):113011.

[12] Mahoney, J. R., Aghamohammadi, C., and Crutchfield, J. P. (2016). Occam's quantum strop: Synchronizing and compressing classical cryptic processes via a quantum channel. *Sci. Rep.*, 6(1):20495.

[13] Milz, S., Pollock, F. A., and Modi, K. (2017). An introduction to operational quantum dynamics. *Open Systems & Information Dynamics*, 24(04):1740016.

[14] Monrà s, A. and Winter, A. (2016). Quantum learning of classical stochastic processes: The completely positive realization problem. *Journal of Mathematical Physics*, 57(1).

[15] Muñoz, R. N., Frazer, L., Yuan, G., Mulvaney, P., Pollock, F. A., and Modi, K. (2022). Memory in quantum dot blinking. *Phys. Rev. E*, 106:014127.

[16] Muñoz, R. N., Leung, A., Zecevik, A., Pollock, F. A., Cohen, D., van Swinderen, B., Tsuchiya, N., and Modi, K. (2020). General anesthesia reduces complexity and temporal asymmetry of the informational structures derived from neural recordings in drosophila. *Phys. Rev. Res.*, 2:023219.

[17] Palmer, A., Fairall, C., and Brewer, W. (2000). Complexity in the atmosphere. *IEEE Trans. Geosci. Remote Sens.*, 38(4):2056–2063.

[18] Park, J. B., Lee, J. W., Yang, J.-S., Jo, H.-H., and Moon, H.-T. (2007). Complexity analysis of the stock market. *Physica A*, 379(1):179–187.

[19] Shalizi, C. R. and Crutchfield, J. P. (2001). Computational mechanics: Pattern and prediction, structure and simplicity. *J. Stat. Phys.*, 104(3):817–879.

[20] Sivak, V. V., Eickbusch, A., Liu, H., Royer, B., Tsioutsios, I., and Devoret, M. H. (2022). Model-free quantum control with reinforcement learning. *Phys. Rev. X*, 12:011059.

[21] Wiseman, H. M. (1996). Quantum trajectories and quantum measurement theory. *J. opt., B Quantum semiclass. opt.*, 8(1):205–222.

[22] Yang, C., Garner, A., Liu, F., Tischler, N., Thompson, J., Yung, M.-H., Gu, M., and Dahlsten, O. (2021). Provable superior accuracy in machine learned quantum models.

[23] Zhang, A., Lipton, Z. C., Pineda, L., Azizzadenesheli, K., Anandkumar, A., Itti, L., Pineau, J., and Furlanello, T. (2019). Learning causal state representations of partially observable environments. *arXiv preprint arXiv:1906.10437*.