Extreme dimensional compression with quantum modelling

Thomas J. Elliott§, 1, 2, Chengran Yang§, 2, 1, †, Felix C. Binder‡, 2, 1, Andrew J. P. Garner, ‡, 4, Jayne Thompson, ‡, 4 and Mile Gu§, 1, 4, ¶

1 Complexity Institute, Nanyang Technological University, Singapore 637335
2 School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371
3 Institute for Quantum Optics and Quantum Information, Austrian Academy of Sciences, Boltzmanngasse 3, Vienna 1090, Austria
4 Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117543

(Dated: September 9, 2019)

Predicting the future based on past events is a cornerstone of life. From meteorologists forecasting the weather, through investors trading on stock markets, to a predator chasing its prey, the ability to identify causes and accurately anticipate effects is central to survival and success. To carry out these essential tasks, models must be formulated, and information about past observations must be stored within memory.

In this context, processes with long historical dependence typically require models that store extensive information about past observations. This is because a model must ascribe each set of past causes that can give rise to distinct future effects to distinct configurations in its memory. When there are many such causes, the memory must support many configurations. Classically, the number of configurations is synonymous with the dimension of the memory – tracking a process with causes reaching far into the past typically requires a large memory with many dimensions.

In contrast, the number of configurations a quantum memory can take is separate from its dimension. This has lead to quantum encodings with reduced memory dimension for several Markovian processes – where each output is conditional only on its immediate predecessor [1, 9–13]. Here, we demonstrate that not only do these quantum advantages persist for non-Markovian processes, but that they become even more pronounced in this regime. We consider a family of such processes whose minimal classical models must devolve unboundedly many bits to tracking the past. For this family, we identify quantum models of equal accuracy that can store all relevant information within a single two-dimensional quantum system (qubit). This represents the ultimate limit of quantum compression and highlights an immense practical advantage of quantum technologies for the forecasting and simulation of complex systems.

This allows for all relevant history to be stored in a single two-state quantum system (qubit), evincing an extreme quantum advantage that scales without bound. Moreover, our protocol requires only a single probe qubit to extract the future statistics. This turns a problem from the converse scenario – that tracking a finite quantum system can require infinite classical resources [5–8] – into a useful tool.

This complements recent advances at the interface of complexity and quantum science, where it has been found that quantum models can drastically reduce the amount of past information – as measured by information entropy – that must be stored in memory to replicate the future behaviour of a process [1, 9–15]. Our work indicates that this advantage (along with its quantitative scaling divergences) also persists for the memory dimension. Crucially, this brings practical, verifiable, and significant quantum memory advantages within the reach of present technologies.

**Framework and tools.** A stochastic process \( \mathcal{X} \) can be characterised by an observation sequence \( \overrightarrow{X} \), detailing what happens and when [16]. We can partition this sequence in two: a past \( \overleftarrow{X} \) that describes everything that has happened up to the present; and a future \( \overrightarrow{X} \) describing everything yet to come (we use upper case to denote random variables, and lower case for their corresponding variates). The goal of causal modelling is to use the past (and only the past) to simulate the future [11, 17–19]. Specifically, a *causal model* \( M \) stores in its memory states \( m \in M \) determined from an encoding function of the past \( f : \{\overrightarrow{X}\} \to M \), such that it can produce futures \( \overrightarrow{X} \) according to \( P(\overrightarrow{X} | m = f(\overleftarrow{X})) = P(\overrightarrow{X} | \overleftarrow{X}) \).

Two widely-used metrics for a causal model’s memory efficiency are [17]:

- \( C_M := - \sum_{m \in M} P(m) \log_2[P(m)]; \)
- \( D_M := \log_2[\dim(M)]; \)

---

*Electronic address: physics@tjelliott.net
†Electronic address: yangchengran92@gmail.com
‡Electronic address: mgu@quantumcomplexity.org
§These authors contributed equally to the results.
where \( P(m) = \sum_{\mathcal{X} \in m} P(\mathcal{X}) \) is the probability of finding the memory in state \( m \) in the process’ steady-state. These measures respectively characterise the information stored by the memory and the dimension of the substrate into which it is encoded. Operationally, they represent the memory required to implement the model in an asymptotic ensemble (\( C_M \)) or single-shot (\( D_M \)) setting.

When \( \mathcal{X} \) is a bi-infinite, stationary sequence with discrete events, the \( \varepsilon \)-machine of computational mechanics \[17\]–[19] is the provably most efficient classical causal model according to both these metrics. The corresponding minimal measures are labelled as \( C_\mu \) and \( D_\mu \), and referred to as the statistical and topological complexity respectively \[17\]. The key elements of these models are causal states \( s \in \mathcal{S} \), a set of equivalence classes defined such that if two pasts have identical future predictions, the (causal state) memory encoding function \( f_s : \{\mathcal{X}\} \to \mathcal{S} \) assigns them to the same state: \( f_s(\mathcal{X}) = f_s(\mathcal{X}') \Leftrightarrow P(\mathcal{X}|\mathcal{X}') = P(\mathcal{X}'|\mathcal{X}') \). Causal states are in essence a state of knowledge, minimally encapsulating all information relevant to future prediction that can be obtained from observations of the past; they closely mirror the belief states of reinforcement learning \[20\]–[21]. The \( \varepsilon \)-machine describes a stochastic transition structure between causal states, with transitions accompanied by the output of a symbol. This can be represented by a hidden Markov model \[18\]. These complexity measures have been applied to study structure in systems from a variety of fields, including neuroscience \[22\]–[23], biology \[24\]–[25], economics \[26\], geophysics \[27\], meteorology \[28\], and condensed matter physics \[29\].

These optimality results do not hold within the quantum domain \[9\]. For quantum causal models \[1\]–[4]–[14]–[30]–[37], each past \( \mathcal{X} \) is assigned a quantum state \( |f(\mathcal{X})\rangle \) to be stored in the model memory. The efficiency metrics become \( C_q := -\text{Tr}[\rho \log_2(p)] \) and \( D_q := \text{log}_2(\text{Tr}[\rho]) \), where \( \rho = \sum_{\mathcal{X}} P(\mathcal{X}) |f(\mathcal{X})\rangle \langle f(\mathcal{X})| \). We refer to these as the quantum statistical memory and quantum topological memory of a model respectively; they inherit the same operational significance in the quantum regime as the corresponding classical quantities \[9\]. As with classical causal models, these quantum memory states encode information from the past of the process, and must not contain any information that can only be obtained from its future; the full description of a quantum model then includes the means by which its memory is probed to produce a sample of the future statistics given the observed past, which must similarly be drawn from \( P(\mathcal{X} \mid m = |f(\mathcal{X})\rangle) = P(\mathcal{X} \mid |f(\mathcal{X})\rangle) \).

Current state-of-the-art constructions for quantum causal models \[3\] assign memory states directly from causal states \( s \to |s\rangle \), though the optimal quantum encoding strategy is presently unknown for general processes \[1\]–[22] – we therefore do not designate these quantum metrics as complexity measures. Nevertheless, it has been shown that in general there exists a quantum model with \( C_q \leq C_\mu \) \[9\]. This quantum advantage exploits the possibility to store quantum information in non-orthogonal states \[23\], enabling efficient isolation of predictive features. It has recently been shown that quantum models can also exhibit \( D_q < D_\mu \) \[1\]–[4].

**Dual Poisson processes.** Consider a system that undergoes a series of Poissonian decay events through one of two channels with rates \( \gamma_1 \) and \( \gamma_2 \). After each event, the decay channel for the next emission is chosen randomly, with probability \( p \) or \( \bar{p} = 1 - p \) respectively. The choice of channel is hidden internally in the system, such that an external observer can only see when the decay events occur. Specifically, we consider an observer operating on discrete timesteps \( \Delta t \), recording a 1 when an event occurs, and 0 otherwise. We call the resultant stochastic process a dual Poisson process, and it manifests as a series of 1s separated by strings of 0s. Note that the probabilistic choice of channel occurs only after events (1s), and remains unchanged across non-events (0s). The probability that a contiguous string of 0s (bookended by 1s) is of at least length \( n \) is given by the so-called survival probability \( \Phi(n) \):

\[
\Phi(n) = p \Gamma_1^n + \bar{p} \Gamma_2^n, \tag{1}
\]

where \( \Gamma_j = \exp(-\gamma_j \Delta t) \). We shall now look at the scaling of the memory metrics of causal models for such processes as the temporal precision \( \Delta t \) is refined – making the process increasingly non-Markovian. With arbitrary \( \Phi(n) \) this framework describes general renewal processes \[39\].

**Optimal classical causal model.** Since the observer is unaware of the choice of decay channel, the information they must track reflects their confidence in the chosen rate based on the time since last emission. That is, a causal model of a dual Poisson process must track the number of 0s (\( \bar{N} \)) since the last 1 in order to predict how many more 0s (\( \bar{n} \)) until the next 1 appears; the direction of the arrows signifies that this is information to be obtained either from the past, or in the future. The relevant conditional future distribution is given by

\[
P\left( \bar{N} = \bar{n} \mid \bar{N} \right) = \frac{\Phi(\bar{n} + 1) - \Phi(\bar{n} + \bar{n} + 1)}{\Phi(\bar{n})}. \tag{2}
\]

When \( \gamma_1 \neq \gamma_2 \) and \( p \neq 0,1 \) this conditional distribution is different for every \( \bar{n} \). We can thus treat \( \bar{n} \) as being
each timestep, a constant unitary interaction with a qubit, and using another to probe it [Fig. 3]. At measurement (in the computational basis \(|1\rangle\)) both the memory and probe qubits, after which measurement (in the computational basis \(|0\rangle\) and \(|1\rangle\) ) the conditional probability becomes increasingly small \(\langle 0|\right)\rightarrow \langle 0|\leftarrow \langle 1|\rightarrow \langle 1|\rangle\) \(\langle n|\leftarrow \langle n|\rangle\) corresponding to having observed \(n\) 0s since the last 1. We require

\[
U|\varsigma(n)\rangle|0\rangle = \sqrt{\frac{\Phi(n + 1)}{\Phi(n)}}|\varsigma(n + 1)\rangle|0\rangle + \sqrt{1 - \frac{\Phi(n + 1)}{\Phi(n)}}|\varsigma(0)\rangle|1\rangle, \tag{3}
\]

where the first subspace corresponds to the memory and the second to the probe (reset to \(|0\rangle\) at each timestep). To understand this criterion, consider the required action of \(U\) – for any quantum memory state \(|\varsigma(n)\rangle\) it must take the current memory state and blank probe state (left-hand side) to a state such that: (i) the measurement statistics of the probe in the computational basis are correct according to Eq. (2) (setting \(n\) \(\equiv n\), with \(n\) \(\equiv n\) and the cumulative \(n\) \(\equiv n\) to obtain the probability for 1 and 0 respectively); and (ii) the quantum memory state is updated correctly according to this outcome \(|\varsigma(n + 1)\rangle\) for non-events 0, and reset to \(|\varsigma(0)\rangle\) for events 1. It can be seen that this is satisfied by the right-hand side of the condition, with the weightings corresponding to the probability (amplitudes) of the desired measurement statistics. Note that in principle we have the freedom to add a phase factor to the second term on the right-hand side; we do not include this here as it is not necessary for our construction.

In the Technical Appendix, we show that for any dual Poisson process the condition Eq. (3) is satisfied by the set of quantum memory states \(\{|\varsigma(n)\rangle\}\) corresponding to having observed \(n\) 0s since the last 1. We require

\[
|\varsigma(n)\rangle = \sqrt{p \Gamma_1^n + ig \sqrt{p \Gamma_2^n}}|0\rangle + i \sqrt{1 - g^2} \Gamma_2^n|1\rangle, \tag{4}
\]

where \(g\) is defined in the Technical Appendix, along with an explicit expression for \(U\). Crucially, Eq. (4) evinces that the memory states can be encoded into a single
Cvantage in both ensemble and single-shot settings. the unbounded scaling of the quantum compression ad-

pare the scaling of the quantum memory metrics with

D that our model is (single-shot) minimal and that

system is the smallest possible memory, we can conclude

everywhere except the lines \( P = 0, 1 \) and \( \gamma = 1 \), where \( D_q = 0 \).

qubit, guaranteeing \( C_q \leq 1 \) and \( D_q \leq 1 \). Moreover, since

the process is generically not memoryless, and a binary

system is the smallest possible memory, we can conclude

that our model is (single-shot) minimal and that \( D_q = 1 \)

is the quantum topological complexity. In Fig. 2 we com-

pare the scaling of the quantum memory metrics with

those of the minimal classical model. We thus observe

the unbounded scaling of the quantum compression ad-

vantage in both ensemble and single-shot settings.

Finally, we calculate the quantum statistical memory

\( C_q \) of our models in the continuum limit \((\Delta t \to 0)\) for the

whole dual Poisson process family. Due to the timescale

invariance of \( C_q \) for renewal processes \([13]\), the entire

process family can be characterised by two parameters:

\( \gamma = \gamma_1 / \gamma_2 \) and \( p \). Moreover, due to symmetries of the

processes, \( \{ \gamma, p \} \) yields the same model as \( \{ \gamma^{-1}, \bar{p} \} \). In

Fig. 4 we plot the continuum limit \( C_q \) for a broad param-

er range. Interestingly, we see that for most choices of

parameters \( C_q \ll D_q \), suggesting yet further significant

memory savings can be achieved in the ensemble regime.

We see that \( C_q \) is largest when the two rates are signif-

icantly different (since the choice of decay channel has

larger impact on the future) and the system predomin-

antly picks the faster rate (since the confidence in the

choice of channel is frequently reset).

Relationship to other works. We have shown that

quantum dimensional advantages in causal modelling of

classical stochastic processes can grow without bound.

The highly cross-disciplinary nature of this work nec-

essarily invites comparison with a range of prior and

current research directions, and remarks on these rela-

tionships are in order. Foremost, a number of previous

studies have shown unbounded quantum memory advan-

tages in ensemble settings \([12, 15]\), where the advantage

is contingent on an asymptotically-large set of simula-

tors acting in parallel with a shared memory. A scaling

advantage in terms of dimension has previously been

found for a Markovian process \([11]\), albeit at the cost of

an unboundedly-large alphabet (and hence output regis-

ter). Thus, while theoretically demonstrating the scaling

quantum memory advantages, these advantages are not

presently experimentally feasible due to the need to ei-

erther implement many simulators at once, or assign an an-

cilla of unbounded dimension for the output register. In

contrast, our proposal requires only two qubits to demon-

strate its advantage (and the associated scaling), and so

is eminently more practical to implement; moreover, our

proposal is the smallest possible that could ever demon-

strate such an advantage, in the sense that if either the

memory or output register of a model has fewer than

two states then the process it simulates is trivial and/or

memoryless.

The modelling of quantum dynamics with classical

simulators is well-studied; several works approaching this

problem from a variety of angles show that it typically

requires unbounded classical resources to track the dy-
namics of a finite quantum system, due to the continuous

nature of the Hilbert space it occupies \([5, 8]\). Here, by re-

versing the scenario we show that this problem can turn

into an asset – the very properties of (even simple) quan-

tum systems that make them appear complex to classical

systems can turn complex classical problems into simple

quantum ones.

The use of complex amplitudes to define the mem-

ory state encodings has previously been considered for

models with a small number of states, where it has been

shown to offer reductions in both entropic and dimen-

sional memory requirements relative to quantum models

based on real amplitudes \([9]\). By using such phases to

achieve compression of the memory into a single qubit

we have addressed a conjecture made in this work regard-

ing the potential of complex encodings to offer significant

enhancements to the quantum advantage.

We note that despite a degeneracy in nomenclature,

our framework is distinct from quantum causal mod-
els \([12]\) in the sense of causal inference \([13]\). Yet, the

overlap in terminology is understandable: in such works

the goal is to identify causal relationships between vari-
bles, e.g., to determine one variable causes the value

of another, or if both stem from a common cause; on

the other hand we start from the proposition that the

past causes the future, and seek to identify what the

information in the past observations is that gives rise to

(i.e., causes) the future statistics. Finally, we also note a

resemblance between our discretised models using an an-
cillary system to interrogate a memory qubit and recent

work on models of quantum clocks \([14]\).

Concluding remarks. The single-shot setting of our

advantage is ideal for current and near-future quantum

technologies. Crucially, such dimensional advantages can

be more readily verified than corresponding entropic ad-

vantages; one need only count the dimension of the mem-

ory system, rather than perform full tomography \([4]\).

The small-scale quantum systems required for our pro-

posal are highly amenable to present experimental ca-

FIG. 4: Quantum statistical memory of dual Poisson

processes. For much of the dual Poisson process family, the

information stored by our quantum models is significantly

less than one bit, highlighting that significant further compression

is possible for ensemble simulators. The information stored

is largest when both (i) the emission rates are significantly

different, and (ii) the system is more likely to adopt the faster

rate, as the state of confidence is frequently reset. We see that

\( C_q \) is largest when both (i) the emission rates are signif-

icantly different (since the choice of decay channel has

larger impact on the future) and the system predomi-

nantly picks the faster rate (since the confidence in the

choice of channel is frequently reset).
pabilities; they could for instance be implemented with current two-qubit ion trap experiments, where sequential interaction-measurement-feedback cycles have been realised [43]. Indeed, consideration of resources in the experimentally more straightforward single-shot regime has garnered notable interest in other contexts [46,50].

While we have shown an unbounded dimensional scaling advantage for the dual Poisson process specifically, our findings motivate future work to investigate the typicality of processes that admit such single-shot advantages, particularly with the introduction of complex amplitudes, as well as approximate compression techniques for those that do not. Indeed, the behaviour of a broader range of renewal processes can be captured by generalising Eq. (4) to have different amplitudes and include additional states. A clear direction to extend these techniques for scaling advantages is to other continuous parameters such as spatial co-ordinates [12], or more abstract settings such as continuous belief spaces [20].

Interestingly, while both our quantum model and the optimal classical model provide an approximation of the fully-continuous process for finite timesteps \( \Delta t \), in the quantum case a decrease in timestep size is not accompanied by an increase in memory size. The quantum model memory size \( D_q \) is entirely independent of the timestep, and does not exhibit the classical scaling of memory with precision. This indicates that the limiting factor in the accuracy of quantum models of such processes is not the available memory, but the accuracy with which it can be addressed. Our results already in some sense indicate a robustness of the quantum advantage: errors in the implementation of the quantum model can be accounted for by limiting the precision \( \Delta t \) to not exceed that achievable by the experiment – and the problem of noise exceeding the difference in future statistics for large \( n \) is mitigated by the truncated process. Ultimately, while it would not be possible to witness the scaling difference all the way up to the continuum limit, it can still be shown up to the best achievable precision. We note that while errors present in current quantum technologies would not prevent us from demonstrating that our quantum models can achieve better precision than any classical model at a fixed number of (qu)bits, the possibility to address larger numbers of classical bits with smaller errors than qubits on quantum computers would presently allow classical computers to achieve a higher level of precision. Nevertheless, our results suggest compression tasks as a potential future route for demonstrating absolute superiority of quantum technologies over classical devices, and as a critical application of these incipient devices.

Technical Appendix

Truncated dual Poisson processes. As noted above, \( D_\mu \) remains infinite at any level of discretisation, as there is no maximum \( n \) that the processes cannot exceed, and no merging of different \( n \) into the same causal state. However, keeping this infinite overhead of states provides very little additional predictive power at large \( n \Delta t \), as the probability of the process reaching these states is small, and the difference in conditional probabilities between \( n = n \) and \( n + 1 \) is monotonically decreasing towards zero with increasing \( n \). We therefore introduce a truncated form of the process, where there is a designated ‘terminal’ state at \( n_{\text{term}} \), such that all states at \( n = n \) are merged down into this state. This terminal state must have transition probabilities that are a weighted average of all the merged states; when an event happens the model transitions to \( n = 0 \) as before, but now on non-events the system remains in the \( n = n_{\text{term}} \) state, as opposed to advancing further.

There is a level of subjective choice in how \( n_{\text{term}} \) is selected. Appropriate methods can be for example based on the fidelity of the conditional distributions at larger \( n \), or on the probabilities of reaching such states. For concreteness, we pick a straightforward criterion:

\[
\begin{align*}
\text{\( n_{\text{term}} := \min_{n \in \mathbb{N}} n \Phi(n) \leq \delta (1 - \Phi(1)). \) (5)}
\end{align*}
\]

That is, \( n_{\text{term}} \) is the first state for which the probability of reaching said state is less than a fraction \( \delta \) of the probability of decay in the first timestep. We here use \( \delta = 0.01 \).

Quantum model construction. As described in the main text, a unitary \( U \) and set of quantum memory states \( \{ |s(n)\rangle \} \) that satisfy Eq. (3) form a quantum causal model of the dual Poisson process at a particular set of parameters. Here we show that the memory states given by Eq. (4) form such a set of states, and give the corresponding \( U \). This will prove our main result.

We begin by postulating a unitary operator \( U \) that is stipulated to act in the following manner on two (non-orthogonal) states \( \{ |\phi_1\rangle, |\phi_2\rangle \} \) that we refer to as ‘generator’ states:

\[
U|\phi_j\rangle|0\rangle = \sqrt{\Gamma_j}|\phi_j\rangle|0\rangle + \sqrt{1 - \Gamma_j}|\phi_R\rangle|1\rangle, \quad j \in \{1, 2\}, \tag{6}
\]

with \( j \in \{1, 2\} \), and we refer to \( |\phi_R\rangle := \sqrt{\bar{p}}|\phi_1\rangle + i\sqrt{\bar{p}}|\phi_2\rangle \) as the ‘reset’ state. We define the overlap of the two generator states \( g := \langle \phi_1|\phi_2 \rangle \), noting that it depends on the size of the timesteps \( \Delta t \). Without loss of generality, we can enforce that this quantity be both real and positive. Analogous to current systematic approaches for constructing quantum causal models [32,33] we utilise the relation \( \langle \phi_1|\phi_2 \rangle = \langle \phi_1|0\rangle U^\dagger|\phi_2\rangle|0\rangle \), which arises from the properties of unitary operators. From this, we obtain

\[
g = \frac{\sqrt{(1 - \Gamma_1)(1 - \Gamma_2)}}{1 - \sqrt{\Gamma_1\Gamma_2}}. \tag{7}
\]

Armed with this, we can now express the generator states in the computational basis of a qubit \( \{ |0\rangle, |1\rangle \} \). Without loss of generality, we can assign

\[
\begin{align*}
|\phi_1\rangle &= |0\rangle, \\
|\phi_2\rangle &= g|0\rangle + \sqrt{1 - g^2}|1\rangle. \tag{8}
\end{align*}
\]
We see that after emitting a 1, the memory always transitions to the reset state $|\psi_0\rangle$; it defines the $\bar{n}=0$ state to which the memory returns after the occurrence of an event. This thus corresponds to the memory state $|\psi(0)\rangle$. The remaining $|\xi(n)\rangle$ can be obtained by applying the unitary to the reset state $n$ times and post-selecting on the probe being measured as $|0\rangle$ after each application: $|\xi(n)\rangle \propto (\Pi_0 U)^n |\phi_R\rangle |0\rangle$, where $\Pi_0 = |0\rangle |0\rangle$ is the projector onto the non-event subspace of the probe. Accounting for normalisation, we obtain

$$|\xi(n)\rangle = \sqrt{p_1^{(n)}} |\phi_1\rangle + i \sqrt{p_2^{(n)}} |\phi_2\rangle.$$  

By inserting Eq. (8) into Eq. (9) we recover the memory states as prescribed in Eq. (4). We can similarly express $U$ in the computational basis using Eqs. (8) and (9):

$$U|0\rangle|0\rangle = \sqrt{\Gamma_1} |0\rangle|0\rangle$$
$$+ \sqrt{1-\Gamma_1} (\sqrt{p} + i \sqrt{1-p}) |0\rangle|1\rangle$$
$$+ i \sqrt{1-\Gamma_1} \sqrt{1-g^2} |1\rangle|1\rangle$$

and

$$U|1\rangle|0\rangle = \sqrt{\Gamma_2} |0\rangle|0\rangle$$
$$+ \sqrt{1-\Gamma_2} |1\rangle|0\rangle$$
$$+ (\sqrt{1-\Gamma_2} - \sqrt{1-\Gamma_1} g) \sqrt{p} + i \sqrt{1-p} g |0\rangle|1\rangle$$
$$+ i (\sqrt{1-\Gamma_2} - \sqrt{1-\Gamma_1} g) \sqrt{1-g^2} |1\rangle|1\rangle.$$ (10)

The remaining two columns describing the action $U|0\rangle|1\rangle$ and $U|1\rangle|1\rangle$ are not uniquely defined by the model, and remain a free choice provided the unitarity of $U$ is upheld. We can also construct a pair of Kraus operators $\{E_0, E_1\}$ that describe the effective evolution of the memory conditional on the observed output and may be used to update the memory when tracking an external system that behaves according to the process. These operators are defined according to $E_j = \langle j | U | 0 \rangle$, where the states in this expression belong to the probe subspace [38]; they can be readily obtained from Eqs. (10) and (11).

Finally, we verify that this model produces the correct survival probability for the process Eq. (1). This is found from the probability of recovering a contiguous string of $n$ 0s after starting from the reset state:

$$\Phi(n) = \langle \phi_R | (U^\dagger \Pi_0)^n (\Pi_0 U)^n |\phi_R\rangle |0\rangle$$
$$= p_1^{(n)} + p_2^{(n)}.$$ (12)

Thus, our construction faithfully replicates the process, and may be used to track the dynamics of any dual Poisson process, at any level of discretisation, thus proving our main result. Notably, our quantum models are free from the need to introduce a truncation at long times as was done in the classical case.

As would be expected, the quantum memory states Eq. (9) and $U$ Eqs. (10) and (11) depend on the particular parameters defining the specific dual Poisson process to be modelled. Nevertheless, once initialised in a particular memory state $|\xi(n)\rangle$ corresponding to our observed past (which, being a qubit state, can always be prepared with at most three rotations from $\{R_y, R_z\}$ — corresponding to rotations of a qubit around the $y$ and $z$ axis of the Bloch sphere [54]), the model operates by repeated applications of the same unitary $U$, each followed by measurement and reset of the ancilla to simulate the future statistics. Being a two-qubit unitary, $U$ can always be synthesised by at most fifteen rotations from $\{R_y, R_z\}$ and three CNOT gates [57, 58], irrespective of the parameters.

Considering current state-of-the-art ion trap experiments with gate fidelities in excess of 99.9% for two qubit gates, and 99.99% for single qubit gates [59, 60], we can naively estimate their fidelity in implementing $U$ as being at least 99.5% when multiplying the fidelity in implementing the elementary gates (i.e., 0.9993 $\times$ 0.999915). This estimate is likely too pessimistic, as it neglects that some single qubit gates are applied in parallel to different qubits, and that the errors can act in opposing directions to partially cancel each other out. Nevertheless, we can use this as a back-of-the-envelope benchmark for implementability. Using the parameters of Fig. 2 ($\gamma_1 = 1, \gamma_2 = 1, p = 0.9$) with $\log_2(1/\Delta t) = 4$, from Eqs. (10) and (11) we obtain

$$U = \left(\begin{array}{ccc} \Phi_0 & -0.1788 & \# \\ 0 & 0.9692 & \# \\ 0.6891 + 0.1230i & -0.1605 - 0.0287i & \# \\ 0.1940i & -0.0452i & \# \end{array}\right),$$ (13)

where $\#$ indicates a ‘free’ value to be chosen subject to the constraint that the columns form mutually orthogonal vectors. By inspection, we can see that any implementation of this with at least 99.5% fidelity must capture the salient features to a high degree of accuracy. At this value of $\Delta t$ the quantum scaling advantage is already clearly visible, and so is practical to verify with current ion trap experiments.

Calculating statistical complexity and quantum statistical memory. Prior work on the computational mechanics of renewal processes [40] has established that the steady-state probabilities of the causal states are proportional to their survival probabilities. That is, $P(\bar{n}) = \mu \Phi(\bar{n})$, where $\mu^{-1} := \sum_{n=0}^{\infty} \Phi(n)$. For dual Poisson processes,

$$\mu = \frac{(1 - \Gamma_1)(1 - \Gamma_2)}{p(1 - \Gamma_2) + p(1 - \Gamma_1)}.$$ (14)
Thus, we have that \( C_\mu = -\sum_{n=0}^{\infty} P(n) \log_2[P(n)] \) and
\[
\tilde{C}_\mu = -\sum_{n=0}^{n_{\text{term}}-1} P(n) \log_2[P(n)] - P(n_{\text{term}}) \log_2[P(n_{\text{term}})],
\]
where \( P(n_{\text{term}}) = \sum_{n=n_{\text{term}}}^{\infty} P(n) \). Further, using Eq. (1) with \( \rho = \sum_{n=0}^{\infty} \langle n \rangle |\langle n \rangle| \) we obtain
\[
\rho = \mu \left( \frac{\rho_{11}}{1 - \Gamma_1^2} + \frac{\rho_{22}}{1 - \Gamma_2^2} + \frac{g \sqrt{(1 - g^2)} \rho_{12} - i \sqrt{(1 - g^2)} \rho_{21}}{1 - \Gamma_1 1^2} \right),
\]
which may be straightforwardly diagonalised to find the two eigenvalues \( \{\lambda_1, \lambda_2\} \), and hence calculate \( C_q = -\lambda_1 \log_2(\lambda_1) - \lambda_2 \log_2(\lambda_2) \).

Acknowledgements

This work was funded by the Lee Kuan Yew Endowment Fund (Postdoctoral Fellowship), grant 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 RG190/17, National Research Foundation Fellowship NRF-NRF2016-02, and National Research Foundation and Agence Nationale de la Recherche joint Project No. NRF2017-NRFANR004 VanQuTe. T.J.E., C.Y., and F.C.B. thank the Centre for Quantum Technologies for their hospitality.

[1] J. Thompson, A. J. P. Garner, J. R. Mahoney, J. P. Crutchfield, V. Vedral, and M. Gu, Physical Review X 8, 031013 (2018).
[2] S. Loomis and J. P. Crutchfield, arXiv:1808.06369 (2018).
[3] Q. Liu, T. J. Elliott, F. C. Binder, C. Di Franco, and M. Gu, Physical Review A 99, 062110 (2019).
[4] 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, arXiv:1812.04251 (2018).
[5] A. Monras and A. Winter, Journal of Mathematical Physics 57, 015219 (2016).
[6] A. Cabello, M. Gu, O. Gühne, J.-A. Larsson, and K. Wiesner, Physical Review A 94, 052127 (2016).
[7] A. Cabello, M. Gu, O. Gühne, and Z.-P. Xu, Physical Review Letters 120, 130401 (2018).
[8] P. Warszawski and H. M. Wiseman, arXiv:1905.10935 (2019).
[9] M. Gu, K. Wiesner, E. Rieper, and V. Vedral, Nature Communications 3, 762 (2012).
[10] J. R. Mahoney, C. Aghamohammadi, and J. P. Crutchfield, Scientific Reports 6, 20495 (2016).
[11] M. S. Palsson, M. Gu, J. Ho, H. M. Wiseman, and G. J. Pryde, Science Advances 3, e1601302 (2017).
[12] A. J. P. Garner, Q. Liu, J. Thompson, V. Vedral, et al., New Journal of Physics 19, 103009 (2017).
[13] C. Aghamohammadi, J. R. Mahoney, and J. P. Crutchfield, Scientific Reports 7 (2017).
[14] T. J. Elliott and M. Gu, npj Quantum Information 4, 18 (2018).
[15] T. J. Elliott, A. J. P. Garner, and M. Gu, New Journal of Physics 21, 013021 (2019).
[16] A. Khintchine, Mathematische Annalen 109, 604 (1934).
[17] J. P. Crutchfield and K. Young, Physical Review Letters 63, 105 (1989).
[18] C. R. Shalizi and J. P. Crutchfield, Journal of Statistical Physics 104, 817 (2001).
[19] J. P. Crutchfield, Nature Physics 8, 17 (2012).
[20] A. R. Cassandra, L. P. Kaebbling, and M. L. Littman, in proceedings of the Twelfth AAAI National Conference on Artificial Intelligence (AAAI Press, 1994), pp. 1023–1028.
[21] L. P. Kaebbling, M. L. Littman, and A. W. Moore, Journal of artificial intelligence research 4, 237 (1996).
[22] R. Haslinger, K. L. Klinkner, and C. R. Shalizi, Neural Computation 22, 121 (2010).
[23] S. E. Marzen, M. R. DeWeese, and J. P. Crutchfield, Frontiers in Computational Neuroscience 9, 105 (2015).
[24] C.-B. Li, H. Yang, and T. Komatsu, Proceedings of the National Academy of Sciences 105, 536 (2008).
[25] D. Kelly, M. Dillingham, A. Hudson, and K. Wiesner, PloS one 7, e29703 (2012).
[26] J. B. Park, J. W. Lee, J.-S. Yang, H.-H. Jo, and H.-T. Moon, Physica A: Statistical Mechanics and its Applications 379, 179 (2007).
[27] R. W. Clarke, M. P. Freeman, and N. W. Watkins, Physical Review E 67, 016203 (2003).
[28] A. J. Palmer, C. W. Fairall, and W. A. Brewer, IEEE Transactions on Geoscience and Remote Sensing 38, 2056 (2000).
[29] D. P. Varn, G. S. Canright, and J. P. Crutchfield, Physical Review B 66, 174110 (2002).
[30] F. Ghafari Jouneghani, M. Gu, J. Ho, J. Thompson, W. Y. Suen, H. M. Wiseman, and G. J. Pryde, arXiv:1711.03661 (2017).
[31] F. C. Binder, J. Thompson, and M. Gu, Physical Review Letters 120, 240502 (2018).
[32] C. Aghamohammadi, J. R. Mahoney, and J. P. Crutchfield, Physics Letters A 381, 1223 (2016).
[33] P. M. Riechers, J. R. Mahoney, C. Aghamohammadi, and J. P. Crutchfield, Physical Review A 93, 052317 (2016).
[34] W. Y. Suen, J. Thompson, A. J. P. Garner, V. Vedral, and M. Gu, Quantum 1, 25 (2017).
[35] J. Thompson, A. J. P. Garner, V. Vedral, and M. Gu, npj Quantum Information 3, 6 (2017).
[36] C. Aghamohammadi, S. P. Loomis, J. R. Mahoney, and J. P. Crutchfield, Physical Review X 8, 011025 (2018).
[37] C. Yang, F. C. Binder, V. Narasimhachar, and M. Gu, Physical Review Letters 121, 260602 (2018).
[38] M. A. Nielsen and I. Chuang, Quantum Computation and Quantum Information (2000).
[39] W. L. Smith, Journal of the Royal Statistical Society. Series B (Methodological) pp. 243–302 (1958).
[40] S. E. Marzen and J. P. Crutchfield, Entropy 17, 4891 (2015).
[41] S. Marzen and J. P. Crutchfield, Journal of Statistical Physics 168, 109 (2017).
[42] J.-M. A. Allen, J. Barrett, D. Horsman, C. M. Lee, and R. W. Spekkens, Physical Review X 7, 031021 (2017).
[43] J. Pearl, Causality: models, reasoning and inference, vol. 29 (Springer, 2000).
[44] M. P. Woods, R. Silva, G. Pütz, S. Stupar, and R. Renner, arXiv:1806.00491 (2018).
[45] V. Negnevitsky, M. Marinelli, K. Mehta, H.-Y. Lo, C. Flühmann, and J. P. Home, Nature 563, 527 (2018).
[46] R. Renner, International Journal of Quantum Information 6, 1 (2008).
[47] L. Del Rio, J. Åberg, R. Renner, O. Dahlsten, and V. Vedral, Nature 474, 61 (2011).
[48] F. G. S. L. Brandao and N. Datta, IEEE Transactions on Information Theory 57, 1754 (2011).
[49] L. Wang and R. Renner, Physical Review Letters 108, 200501 (2012).
[50] M. Tomamichel and M. Hayashi, IEEE Transactions on Information Theory 59, 7693 (2013).
[51] J. Åberg, Nature Communications 4, 1925 (2013).
[52] M. Horodecki and J. Oppenheim, Nature Communications 4, 2059 (2013).
[53] N. Y. Halpern, A. J. P. Garner, O. C. O. Dahlsten, and V. Vedral, New Journal of Physics 17, 095003 (2015).
[54] M. Tomamichel, Quantum Information Processing with Finite Resources: Mathematical Foundations, vol. 5 (Springer, 2015).
[55] K. Korzekwa, M. Lostaglio, J. Oppenheim, and D. Jennings, New Journal of Physics 18, 023045 (2016).
[56] B. Regula, K. Fang, X. Wang, and G. Adesso, Physical Review Letters 121, 010401 (2018).
[57] F. Vatan and C. Williams, Physical Review A 69, 032315 (2004).
[58] B. Kraus and J. I. Cirac, Physical Review A 63, 062309 (2001).
[59] T. P. Harty, D. T. C. Allcock, C. J. Ballance, L. Guidoni, H. A. Janacek, N. M. Linke, D. N. Stacey, and D. M. Lucas, Physical Review Letters 113, 220501 (2014).
[60] C. J. Ballance, T. P. Harty, N. M. Linke, M. A. Sepiol, and D. M. Lucas, Physical Review Letters 117, 060504 (2016).