

PAPER • OPEN ACCESS

## Provably unbounded memory advantage in stochastic simulation using quantum mechanics

To cite this article: Andrew J P Garner *et al* 2017 *New J. Phys.* **19** 103009

View the [article online](#) for updates and enhancements.

### Related content

- [Measures and applications of quantum correlations](#)  
Gerardo Adesso, Thomas R Bromley and Marco Cianciaruso
- [Multi-partite entanglement can speed up quantum key distribution in networks](#)  
Michael Epping, Hermann Kampermann, Chiara Macchiavello *et al.*
- [Algorithmic independence of initial condition and dynamical law in thermodynamics and causal inference](#)  
Dominik Janzing, Rafael Chaves and Bernhard Schölkopf

### Recent citations

- [Contextuality, memory cost and non-classicality for sequential measurements](#)  
Costantino Budroni
- [Strong and Weak Optimizations in Classical and Quantum Models of Stochastic Processes](#)  
Samuel P. Loomis and James P. Crutchfield
- [Optimal stochastic modeling with unitary quantum dynamics](#)  
Qing Liu *et al*



## PAPER

## Provably unbounded memory advantage in stochastic simulation using quantum mechanics

## OPEN ACCESS

## RECEIVED

31 March 2017

## REVISED

30 June 2017

## ACCEPTED FOR PUBLICATION

28 July 2017

## PUBLISHED

11 October 2017

Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.


 Andrew J P Garner<sup>1,2,7</sup> , Qing Liu<sup>3</sup>, Jayne Thompson<sup>1</sup>, Vlatko Vedral<sup>1,2,4,5</sup> and mile Gu<sup>1,3,6,7</sup>
<sup>1</sup> Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, 117543, Singapore<sup>2</sup> Center for Quantum Information, Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, 100084, People's Republic of China<sup>3</sup> School of Physical and Mathematical Sciences, Nanyang Technological University, 639673, Singapore<sup>4</sup> Atomic and Laser Physics, University of Oxford, Clarendon Laboratory, Parks Road, Oxford, OX1 3PU, United Kingdom<sup>5</sup> Department of Physics, National University of Singapore, 2 Science Drive 3, 117542 Singapore<sup>6</sup> Complexity Institute, Nanyang Technological University, 639673, Singapore<sup>7</sup> Authors to whom any correspondence should be addressed.E-mail: [ajpgarner@nus.edu.sg](mailto:ajpgarner@nus.edu.sg) and [gumile@ntu.edu.sg](mailto:gumile@ntu.edu.sg)
**Keywords:** quantum information, quantum advantage, statistical complexity, theory of simulation, quantum computation, computational mechanics

### Abstract

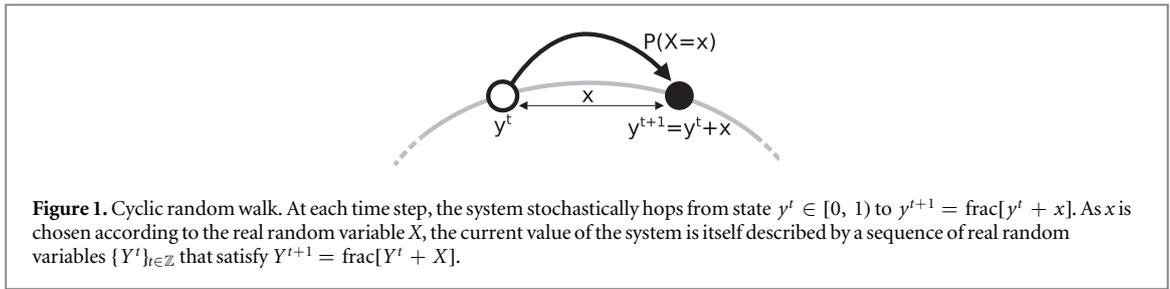
Simulating the stochastic evolution of real quantities on a digital computer requires a trade-off between the precision to which these quantities are approximated, and the memory required to store them. The statistical accuracy of the simulation is thus generally limited by the internal memory available to the simulator. Here, using tools from computational mechanics, we show that quantum processors with a fixed finite memory can simulate stochastic processes of real variables to arbitrarily high precision. This demonstrates a provable, unbounded memory advantage that a quantum simulator can exhibit over its best possible classical counterpart.

Many macroscopic processes we wish to simulate involve the dynamics of real numbers. The dynamical properties we wish to track (e.g. the position of an object) can take on almost any number, seemingly without noticeable quantisation until one goes down to the Planck scale. The simulation of such processes necessitates compromise between the resources allocated and the precision with which we track such properties. Clever implementations to this problem, such as the floating point format [1], form the heart of modern computing technology—but all subscribe to the same trade-off: treating a quantity with higher precision requires the allocation of more memory. To perfectly replicate the future statistics of a continuous variable dynamical system exactly would inevitably require unbounded memory.

The advent of quantum technology, however, opens new possibilities. Not only has this technology shown great potential in solving problems many consider classically intractable [2–6], it has demonstrated the capability to greatly reduce the amount of information one needs to send in certain tasks requiring communication between distributed parties [7–9]. Could the memory required by a quantum machine that simulates dynamical processes likewise scale much more favourably with precision?

Here, we consider the simulation of a class of stochastic systems involving the dynamics of parameters that take on real numbers. Classical simulation of such processes digitally involves ‘coarse-graining’: the parameter at each point in time is approximated to  $n$  bits of precision at some memory cost that scales linearly with  $n$ . We construct quantum simulators that exhibit *unbounded* advantage. The quantum simulator can exactly replicate the statistics of a  $n$  bit classical simulator for arbitrarily large  $n$  using a bounded amount of memory. Thus, quantum simulators can side-step the precision–memory trade-off—finite quantum memory can simulate such processes to arbitrary fixed precision.

This unbounded divergence has practical and foundational consequences. Practically, it suggests that quantum processors may be increasingly advantageous as we wish to simulate ever more memory-intensive systems, such as those arising from big data sets. Foundationally, the minimal memory required to simulate a



process is a well-established measure of structure, known as *statistical complexity* [10–20]. Our work suggests that there are certain processes which grow unboundedly in statistical complexity, but yet remain simple to an observer with quantum capabilities.

## Cyclic random walks

Consider a small bead located on a circular ring of circumference 1 (as per figure 1). Its position can always be described by some real number  $y \in [0, 1)$ . At each discrete time  $t \in \mathbb{Z}$ , the bead's position is stochastically perturbed. This perturbation is described by a real random variable  $X$  that is governed by a continuous probability density function  $P(X)$ , such that

$$Y^{t+1} = \text{frac}[Y^t + X], \quad (1)$$

where  $Y^t$  represents the random variable that governs the location of the bead at time  $t$ , and  $\text{frac}[y] = y - [y] \in [0, 1)$  denotes the fractional part of  $y$ , such that positions differing only by whole rotations around the ring are equivalent. We refer to  $P(X)$  as the *shift function*, and assume the process is stationary, in the sense that  $P(X)$  has no explicit dependence on  $t$ , and rotationally symmetric such that  $X$  has no dependence on the current value of  $Y^t$ . This same formalism describes a diverse range of systems undergoing cyclic random walks, such as the azimuthal motion of gas molecules diffusing in an annular tube, or the position of a single electron travelling through an electric circuit with constant resistance.

We capture the dynamics of  $Y$  formally using the framework for describing stochastic processes. In general, a stochastic process  $\mathcal{P}$  is characterised by a bi-infinite sequence of random variables  $\{Y^t\}_t$ , that governs its value at each discrete time  $t \in \mathbb{Z}$ . For convenience, we often segregate past and future values, such that  $\overleftarrow{Y} = \dots Y^{-1} Y^0$  and  $\overrightarrow{Y} = Y^1 Y^2 \dots$  respectively govern the values in the past and future with respect to time  $t = 0$ . The cyclic random walk above is then entirely captured by the joint probability distribution  $P(\overleftarrow{Y}, \overrightarrow{Y})$  such that for any instance of the process with past values  $\overleftarrow{y}$ , future values  $\overrightarrow{y}$  will be observed with probability  $P(\overrightarrow{Y} = \overrightarrow{y} \mid \overleftarrow{Y} = \overleftarrow{y})$ .

Here, we consider the simulations of the above process to ever increasing precision. We adopt a natural technique of discretizing a continuous process, by introducing a family of stochastic processes  $\{\mathcal{P}_n\}$  that describe discrete approximations of this process, where in each the position of bead is represented to  $n$  bits of precision by a  $n$ -digit binary number. This is done by limiting  $y$  to a discrete set of  $N = 2^n$  equally spaced values,  $y_j = j/N$  (for  $j = 0$  to  $N - 1$ ). At each time-step, the probability that a bead in discrete location  $y_j$  transitions to  $y_k$ , is given by the probability  $p_{jk}$  that a bead initially at  $y_j$  will transition to any value of  $y$  whose  $n$  bit binary representation is  $y_k$ . That is

$$p_{kj} = P(Y^{t+1} = y \in \mathcal{I}_k \mid Y^t = y_j), \quad (2)$$

where  $\mathcal{I}_k = \{y : |y - y_k| < \frac{1}{2N}\}$  represents the interval on the ring that is 'rounded to'  $y_k$ . This results in a Markovian stochastic process that emits a symbol from the finite alphabet  $\{y_k\}$  at each time-step, whose dynamics are governed by the stochastic matrix with elements  $p_{jk}$ . As  $n \rightarrow \infty$ , the statistics of  $\mathcal{P}_n$  approach that of  $\mathcal{P}$ ; at the potential cost of tracking more information<sup>8</sup>.

## Classical simulation costs scale with precision

We can formally describe simulators using the tools of computational mechanics [10–13]. A simulator of a process is a device whose future output behaviour conditioned on any particular past should be statistically indistinguishable to the process itself. Specifically, let the state of the simulator at each time be  $s^t$ , such that in the

<sup>8</sup> An alternative discretization is to calculate the transition probabilities by assuming the initial value of  $y^t$  is uniformly distributed in  $\mathcal{I}_j$ . This yields asymptotically identical statistics as  $N \rightarrow \infty$ , and does not change the results of this article.

subsequent time-step it can output  $y^{t+1}$  and transition to state  $s^{t+1}$ . For this device to be a statistically faithful simulator of a process  $P(\overleftarrow{Y}, \overrightarrow{Y})$ , we require that:

1. For each specific past  $\overleftarrow{y}$  at each time  $t$ , we can deterministically configure the device using a function  $f$  into some state  $s = f(\overleftarrow{y})$ , such that it will produce future outputs  $\overrightarrow{y}$  with probability  $P(\overrightarrow{Y} = \overrightarrow{y} \mid \overleftarrow{Y} = \overleftarrow{y})$ .
2. If a simulator is in state  $s^t = f(\overleftarrow{y}^t)$  at time  $t$ , and outputs  $y^t$  in the subsequent time-step, its internal state must then transition to  $s^{t+1} = f(\overleftarrow{y}^t y^t)$ .

The first condition ensures the simulator can be initialised to simulate desired conditional future statistics; the second that a correctly initialised simulator continues to exhibit statistically correct statistics at every time-step. The memory cost of the simulator corresponds to the storage requirements of this internal state. This cost is bounded from below by the information entropy of the random variable  $S := f(Y)$ . In the asymptotic limit of many independent identically distributed copies of the simulator, this bound is tight as the ensemble of states may be compressed (such as by Shannon's noiseless encoding theorem [21], or Schumacher compression [22, 23]). Physically a simulator can be viewed as a communication channel in time: it represents the exact object Alice must give to Bob at each time-step that captures sufficient past information for Bob to replicate the processes conditional future behaviour.  $f$  is known as the *encoding function*, which describes how the past is encoded within the channel.

This memory cost of the provably optimal classical simulator—known as the *statistical complexity*  $C_\mu$ —is extensively studied in complexity science [10]. This value captures the absolute minimum memory any classical simulator of a process must store, and thus is a prominent quantifier of a process's structure and complexity<sup>9</sup> (e.g. [13–20]). Such an optimal simulator can be explicitly constructed, and corresponds to the simulator that stores in its internal memory the *causal states* of the process [10, 11]: defined by an encoding function  $f$  such that  $f(\overleftarrow{y}) = f(\overleftarrow{y}')$  if and only if  $P(\overrightarrow{Y} \mid \overleftarrow{Y} = \overleftarrow{y}) = P(\overrightarrow{Y} \mid \overleftarrow{Y} = \overleftarrow{y}')$  (i.e. the conditional futures of  $\overleftarrow{y}$  and  $\overleftarrow{y}'$  coincide).

In our cyclic random walks, each  $\mathcal{P}_n$  is a first-order Markov process: the statistics of future outcomes depend only on the most recent value of  $Y^t$ . When this example is discretized, the causal states are thus typically in one-to-one correspondence with the  $2^n$  discrete values that  $Y$  can take<sup>10</sup>. That is,  $\mathcal{P}_n$  has  $2^n$  causal states, labelled  $\{s_j\}_{j=0}^{2^n-1}$ , where  $s_j$  corresponds to the set of pasts ending in  $Y^0 = y_j$ . When the simulator has been running for a sufficiently long time, the probability distribution over the internal memory converges on  $P(S = s_i) = \frac{1}{N}$  for each  $i$ —its steady state, in which all causal states occur with equiprobability. Thus, the classical statistical complexity

$$C_\mu = n, \quad (3)$$

scales linearly with the precision.

## Quantum simulators are memory-efficient

It has recently been shown that quantum processors have the capability to simulate stochastic processes with less memory than is classically possible [25–29]. Here, we construct an explicit quantum simulator for the cyclic random walk. Instead of storing each causal state  $s_i$  directly, our quantum simulator stores a corresponding quantum state

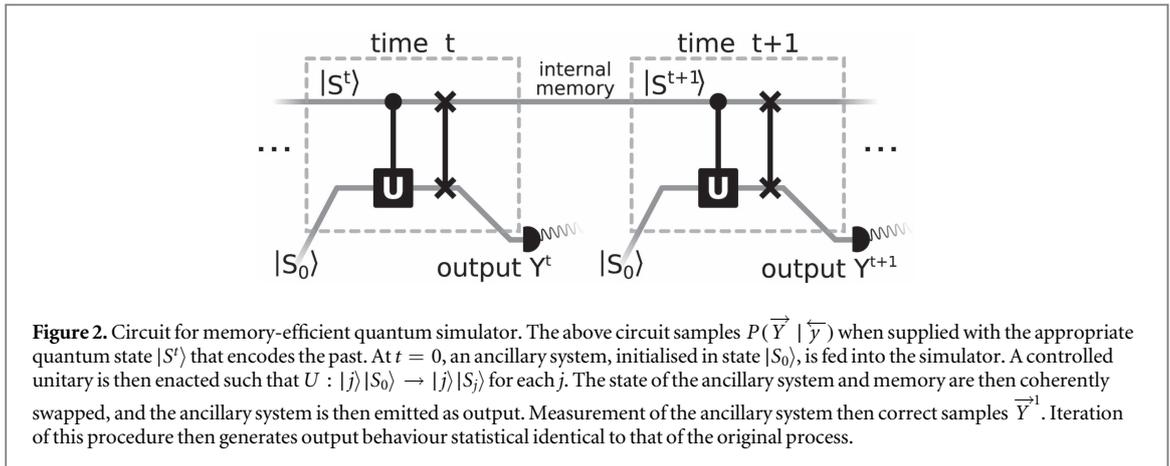
$$|S_j\rangle = \sum_{k=0}^{N-1} \sqrt{p_{kj}} |k\rangle, \quad (4)$$

where  $\{|k\rangle\}$  forms an orthonormal basis.

The stationary state of the quantum simulator is then given by the *quantum ensemble state*  $\rho = \frac{1}{N} \sum_j |S_j\rangle \langle S_j|$  (as all quantum states occur with equiprobability). Thus the memory required to store these states is given by the von Neumann entropy given  $H_Q := -\text{Tr}(\rho \log \rho) = -\sum_k \lambda_k \log \lambda_k$ , where  $\lambda_k$  are the eigenvalues of  $\rho$ . The key improvement here is that  $\{|S_j\rangle\}$  are not in general mutually orthogonal, and thus  $H_Q$  is generally less than  $C_\mu$ .

<sup>9</sup> The statistical complexity is distinct from algorithmic information (Kolmogorov–Chaitin complexity). Statistical complexity is, as the name would imply, intrinsically statistical—concerned with the replication of the statistical behaviour of a process; whereas algorithmic information relates to the compressibility of an exact string [24].

<sup>10</sup> There are exceptions, such as when  $P(x) = 1$  for  $x \in [0, 1)$ , and the system jumps to a completely random point at each time-step; here there is only one causal state for all  $N$ , because the current position no longer affects the future outcomes at all.



Nevertheless a quantum circuit (outlined in figure 2—with details in the [appendix](#)) acting on these quantum states will produce statistically identical outputs to the classical simulator.

The von Neumann entropy of a quantum state is equal to the Shannon entropy of the outcome statistics of a projective measurement on that state, minimised over all choices of projective measurement. This minimisation corresponds to a measurement in the basis in which the state's density matrix is diagonal. A classical probability distribution maps onto a mixed quantum state, diagonal in a fixed basis. As such, the stationary state of the classical simulator can be assigned a quantum state, whose von Neumann entropy is exactly that distribution's Shannon entropy. This allows us to compare the entropic cost of the classical and quantum machines' memories on an equal footing.

## Unbounded advantage of quantum memory

We now come to the main claim of our paper: there are stochastic processes that can be simulated to infinite precision using a finite amount of quantum memory.

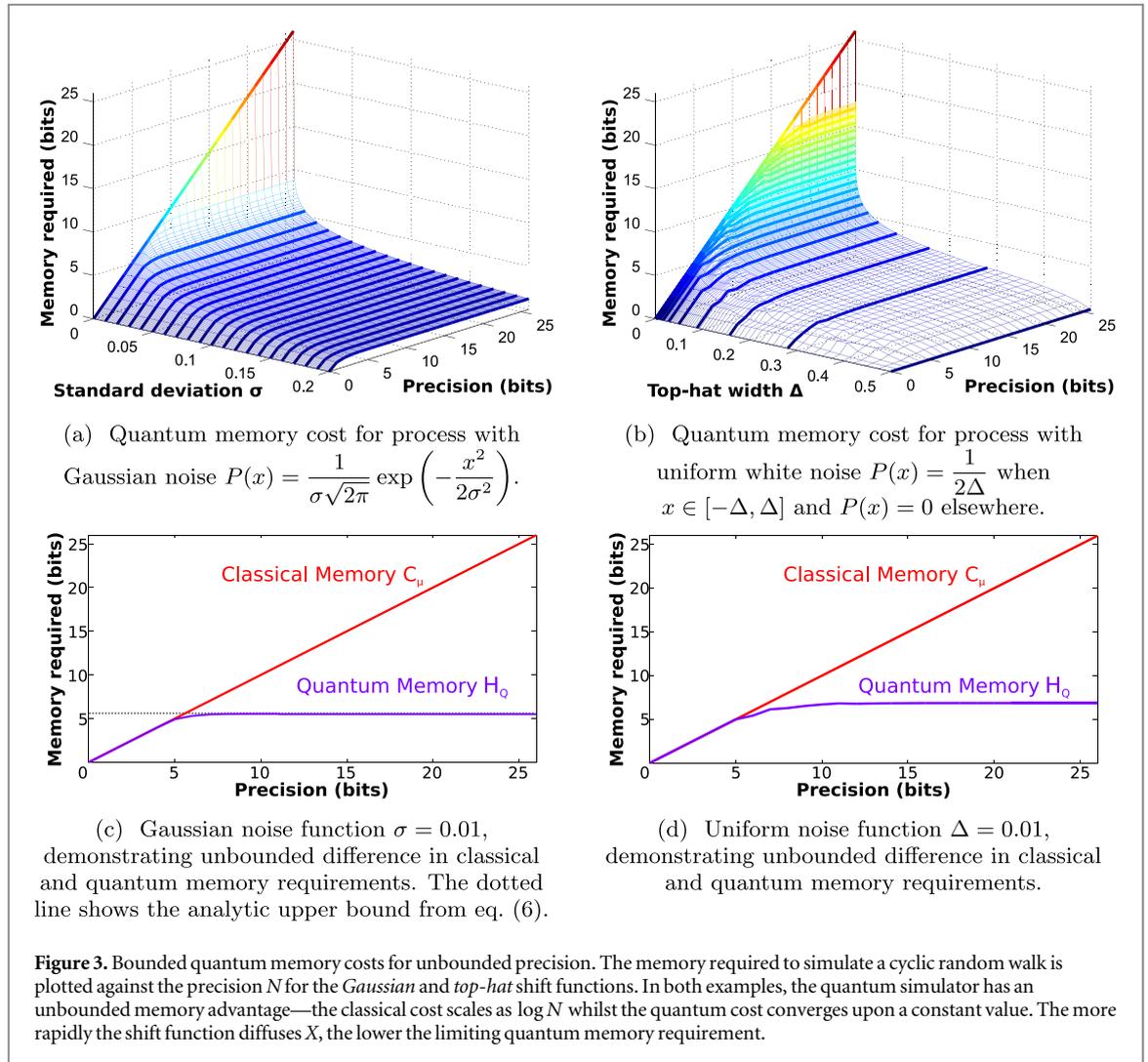
Explicitly, we show that for certain cyclic processes, the quantum ensemble state's eigenvalues  $\{\lambda_k\}_{k=0\dots N-1}$  satisfy  $\lim_{N \rightarrow \infty} \sum_{k=0}^{N-1} -\lambda_k \log \lambda_k = \Omega$  for some finite value  $\Omega$ . Our result relies on first observing that the eigenvalues  $\lambda_k$  can be directly related to transition probabilities  $\{p_{jk}\}$  via the relation

$$\lambda_k = \frac{1}{N} \mathcal{F}[\sqrt{p_{j0}}] \mathcal{F}[\sqrt{p_{(N-j)0}}], \quad (5)$$

where  $\mathcal{F}$  denotes the *discrete Fourier transform* (DFT),  $\mathcal{F}(x_j) = \sum_{j=0}^{N-1} x_j \exp\left(\frac{-2\pi i}{N} jk\right)$ . (The proof relies on invoking the cyclic symmetry of the process—and hence of the transition probabilities—and is explicitly derived in the [appendix](#).) The spread  $p_{j0}$  (as a function of  $j$ ) is an indicator of how quickly a particle diffuses in the random walk. Thus, the Fourier-like relation between  $p_{j0}$  and  $\lambda_k$  indicates an inverse relationship between the amount of diffusion in the cyclic process and the spread of eigenvalues. The greater the variance of  $X$ , the more quickly a particle diffuses, and the smaller the spread of  $\lambda_k$ —resulting in a reduced quantum memory requirement. We now show that for some natural examples, this reduction is sufficiently large that  $H_q$  remains bounded for all  $n$  (as illustrated in figure 3).

**Example 1 (Gaussian noise).** A cyclic process rotating at a constant rate subject to Gaussian noise has a shift function given by a Gaussian distribution  $G_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$  about mean  $\mu$  with standard deviation  $\sigma$ . Here,  $\mu$  characterises the average velocity (in terms of the variable's mean displacement per time-step), and  $\sigma$  the size of the fluctuations. When  $\mu = 0$ , this process corresponds to Gaussian diffusion. For our analysis, we take  $\sigma \ll 1$  and thus ignore fluctuations where the particle travels more than a complete loop around the ring in a single time-step (a value of  $\sigma = 0.1$  ensures that such events are less likely than one part in a million).

As can be seen in figures 3(a) and (c), as the desired precision increases, the memory cost of simulating this process quickly converges onto a constant determined by the fluctuation strength  $\sigma$ ; ultimately, *infinite-precision simulation is possible using only a finite quantum memory*. This behaviour may be understood analytically by seeing that for large  $N$ , the eigenvalues associated with the quantum simulator's internal memory are also given by samples from a Gaussian distribution:  $\lambda_k = G_{0, \frac{1}{4\pi\sigma}}(k)$  for  $k = -\frac{N}{2}, \dots, \frac{N}{2} - 1$ , where for convenience we have cyclicly offset the label of the eigenvalues' indices by  $N$  (proof in [appendix](#)). This demonstrates that increasing  $\sigma$  tightens the spread of eigenvalues, and thus reduces the memory requirement for the quantum simulator.



In the [appendix](#), we prove that as the precision  $n = \log N$  increases, the sum  $\lim_{N \rightarrow \infty} \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} \lambda_k \log \lambda_k$  converges on a finite value, bounded (in bits) by

$$H_Q \leq \frac{1}{2 \ln 2} - (1 + 4\sqrt{2\pi}\sigma) \log_2 2\sqrt{2\pi}\sigma. \tag{6}$$

Thus, for any fixed  $0 < \sigma \ll 1$ , the Gaussian random walk may be simulated to arbitrarily high precision using a quantum simulator of bounded entropy. Moreover, this also implies an unbounded divergence between the classical and the *quantum statistical complexity* [26, 30]  $C_Q$ , which is upper bounded by  $H_Q$ .

**Example 2 (Uniform white noise).** In the second example, we consider a particle that is perturbed by uniformly distributed noise. At each time-step, the particle can move anywhere in the range of  $\mu \pm \Delta$  from its current position with uniform probability, where  $\Delta < \frac{1}{2}$ . Again,  $\mu$  characterises the average velocity, and here  $\Delta$  the size of the fluctuations. The associated shift function is a *top-hat function*, that has a uniform value of  $\frac{1}{2\Delta}$  in the range  $x \in [\mu - \Delta, \mu + \Delta]$  and 0 everywhere else.

The entropy of the quantum simulator,  $H_q$  is plotted for various precision in figures 3(b) and (d). We see that for any fixed  $\Delta > 0$ , the quantum memory required by our simulator converges to a bounded value. As in the Gaussian scenario, the quantum simulator can replicate a classical simulation to any given precision using with finite entropy. In the [appendix](#), we prove this analytically. We show that as  $N \rightarrow \infty$ , the entropy remains finite, and is bounded above by  $H_Q \leq \frac{1.894}{\sqrt{\Delta}} + 3.067$ . In particular, for large  $N$ , the eigenvalues of the relevant ensemble state obey  $\lambda_k = 2\Delta \text{sinc}^2(2k\Delta)$  for  $k = -\frac{N}{2}, \dots, \frac{N}{2} - 1$ , where  $\text{sinc}(x)$  is the *normalised sinc function*,  $\text{sinc}(x) := \frac{1}{\pi x} \sin(\pi x)$ . Larger values  $\Delta$  will result in a smaller spread of eigenvalues, and result is smaller  $H_q$ . For any given  $\Delta > 0$  the entropy is finite in the limit  $N \rightarrow \infty$ . This establishes a second natural example where the

quantum simulator can demonstrate an unbounded memory advantage over its best possible classical counterpart.

## The origin of quantum advantage

The source of classical inefficiency can be understood by considering dynamics on causal states. Consider two instances of  $\mathcal{P}_n$ , one where  $Y^0 = y_j$ , and the other where  $Y^0 = y_{j+1}$ . As their conditional future statistics differ (that is,  $P(\vec{X} | Y^0 = y_j) \neq P(\vec{X} | Y^0 = y_{j+1})$ ), a classical simulator must be configured differently for each instance (corresponding to being initialised in one of two different causal states,  $s_j$  or  $s_{j+1}$ ). Nevertheless, there is finite probability that at the next time-step, both instances of the process emit the same output (up to precision  $n$ ). Should this happen, we would not be able to use the current state of the machine to determine the causal state it was in at the previous time. That is, there is some probability that the distinction between  $s_j$  and  $s_{j+1}$  will never be reflected in the future statistics of the process—a phenomenon known as crypticity [27, 31]. As  $n$  increases, this occurs with greater likelihood (tending to unit probability as  $n \rightarrow \infty$ ), and thus proportionally more information is wasted. Ultimately, in the limit of high precision, a vanishingly small proportion of the information stored in the classical memory is pertinent to the statistical behaviour of the process's future.

Quantum simulators compensate for this waste by mapping these causal states to non-orthogonal quantum states. The quantum state (equation (4)) associated with neighbouring causal states ( $|S_j\rangle$  and  $|S_{j+1}\rangle$ ) also become increasingly similar with increasing  $n$ —resulting in progressively greater savings. Consider the Gaussian scenario, where  $H_q$  is bounded by equation (6). For small  $\sigma$ , the memory cost scales as  $-\log_2 \sigma$ , such that halving the variance of fluctuations at each time-step adds one bit to the memory cost of the quantum simulator. The standard deviation of the shift function has set an effective length scale over which the system must be simulated classically. The statistical behaviour of future outputs from two systems that are initially prepared in points separated by more than one standard deviation are typically distinguishable, and so these points must be stored as nearly orthogonal quantum states at some memory cost. On the other hand, when two points are initially closer than the standard deviation scale, the probability that they could be distinguished by their future behaviour diminishes, and they may be represented by increasingly overlapping quantum states. In this regime, a fixed finite memory can accommodate any desired precision.

We gain further insight into the origins of quantum advantage by considering the cases where it does not appear:  $\sigma = 0$  and  $\Delta = 0$ . In both these cases, the shift function is a Dirac delta distribution. As such, no matter how high the precision, by observing the future outputs, it will *always* be possible to distinguish whether the system came from some site  $s_j$  or its neighbour  $s_{j+1}$ ; the dynamics of the system are wholly reversible. If  $s_j$  always transitions to  $s_k$  and  $s_{j+1}$  always to  $s_{k+1}$ , being able to distinguish between these two sites is crucial to produce the correct statistical behaviour, even as the precision increases. As such, the quantum simulator cannot tolerate overlap between the states  $|s_j\rangle$  and  $|s_{j+1}\rangle$ , and must store them orthogonally (allowing them to be distinguished). In this scenario, the quantum simulator cannot demonstrate any advantage in memory cost over its classical analogue.

## Discussion and outlook

In this article, we presented a task in which quantum mechanics has an unbounded memory advantage over the most memory-efficient classical alternative: the simulation of a classical cyclic stochastic process. We found that the classical simulator has a memory requirement that scales linearly with the precision required, while the quantum simulator's requirement may be bounded by a finite value, even at arbitrarily high fixed precision. This establishes a rare scenario where the scaling advantage of quantum processing can be provably established.

This finding leads to a number of natural open questions—the first being of generality. Certainly, the examples presented are sufficiently simple that such divergences are unlikely to be merely a mathematical oddity. The unbounded quantum advantage relies on  $\{\mathcal{P}_n\}$  having two properties: (a) the number of causal states grows with  $n$ , and (b) the conditional future statistics  $P(\vec{X} | S = s_i)$  between different causal states converges sufficient quickly with  $n$ . If these conditions can be formalised, we may be able to establish similar divergences in much more general scenarios, such as the simulation of non-Markovian or non-cyclic processes. Beyond von Neumann entropy, it would be interesting if similar scaling can be found for other metrics of memory cost, such as the dimension—namely, whether there is an encoding that allows for simulation to arbitrary precision using a Hilbert space of bounded dimension. Meanwhile the inefficiency of classical simulators have shown to directly result in unavoidable increased heat dissipation [32–34]. This hints that quantum processing may allow significant energetic savings for stochastic simulation, especially for systems that become increasingly difficult to simulate as they scale in size.

On a foundational level, the statistical complexity is often regarded as a fundamental measure of a process's intrinsic structure—the rationale being that it quantifies the minimal amount of information about a process's history that must be recorded to allow for predictions about that process's future behaviour. The measure has been applied to understand structure within diverse complex settings: from the dynamics of neurons [14] and the stock market [18], to quantifying self-organisation [15], among other examples [16, 17, 19, 20]. The discovery of more efficient quantum models has led to the idea that the complexity of a system depends on what sort of information we use to observe it [26, 30]. In this context, our results establish a family of processes that can look ever more complex classically, but remain simple quantum-mechanically. It would be fascinating to see if divergences between quantum and classical complexities can be found in existing studies, such as the examples above. Could it be that these systems appear complex classically—but look much simpler when viewed through the lens of quantum theory?

## Acknowledgments

We thank Thomas Elliott, David Garner, Jan-Åke Larsson, and Chengran Yang for helpful comments and discussions. We gratefully acknowledge funding from the John Templeton Foundation Grant 53914 'Occam's Quantum Mechanical Razor: Can Quantum theory admit the Simplest Understanding of Reality?'; the Foundational Questions Institute; the Ministry of Education in Singapore, the Academic Research Fund Tier 3 MOE2012-T3-1-009; and the the National Research Foundation of Singapore (Award Nos. NRF–NRF2016–02 and NRF–CRP14-2014-02).

## Appendix

### Classical costs from computational mechanics

We here present some minimal details from the mathematical framework of computational mechanics [10–13] to substantiate the claim that the classical simulator's minimal memory cost is equal to the precision  $\log N$ .

In computational mechanics, the evolution of a dynamical property (over domain  $\mathcal{Y}$ ) is characterised by a discrete-time stochastic process  $\mathcal{P}$ , written as bi-infinite sequence of random variables  $\{Y^t\}_{t \in \mathbb{Z}}$ , where each random variable  $Y^t$  governs the value  $y^t \in \mathcal{Y}$  of the dynamical property at time  $t$ . The statistical behaviour of a process may be represented in a *causal* manner by writing it as the conditional probability distribution  $P(\vec{Y}_t | \overleftarrow{Y}_t)$ , where  $\vec{Y}^t = Y^{t+1}Y^{t+2}\dots$  is the infinite string of random variables occurring after time  $t$ , and  $\overleftarrow{Y}^t = \dots Y^{t-1}Y^t$  is the infinite string of random variables occurring before (and including) time  $t$ . For *stationary* processes (such as the time-independent cyclic random walks described in this article), this distribution has no explicit time dependence, so we omit the superscript  $t$ .

A faithful *simulator* of process  $\mathcal{P}$  is a machine (or programme) that, having been initialised in accordance with the observation of past  $\overleftarrow{y}^t$ , then generates a series of outputs  $\vec{y}_t$  according to the distribution  $P(\vec{Y}^t = \vec{y}^t | \overleftarrow{Y}^t = \overleftarrow{y}^t)$ . Since storing an infinite string  $\overleftarrow{y}^t$  may require an unbounded amount of memory, one instead configures the internal state of the simulator  $s$  (over configuration space  $\mathcal{S}$ ) according to some function  $s = f(\overleftarrow{y}^t)$ , satisfying  $P(\vec{Y}^t = \vec{y}^t | S = s) = P(\vec{Y}^t = \vec{y}^t | \overleftarrow{Y}^t = \overleftarrow{y}^t)$ , where  $S = f(\overleftarrow{Y}^t)$  is the random variable describing the internal state of the simulator (formed by applying the function  $f$  on each variate of  $\overleftarrow{Y}^t$ ). Moreover, once initiated into state  $s^t$ , when the simulator outputs  $y^t$  in the subsequent time-step, its internal state must then transition to the state  $s^{t+1} = f(\overleftarrow{y}^t y^t)$  (where  $\overleftarrow{y}^t y^t$  indicates the concatenation of  $y^t$  to the end of string  $\overleftarrow{y}^t$ ).

The memory cost of such a simulator is given by the information entropy of  $S$ ,  $H(S) = -\sum_{s_i \in \mathcal{S}} P(S = s_i) \log P(S = s_i)$ . The function  $f$  that minimises this classically corresponds to identifying the *causal state* of a particular past [10, 11], defined by the equivalence relationship:  $\overleftarrow{y} \sim_{\epsilon} \overleftarrow{y}'$  for pasts  $\overleftarrow{y}$  and  $\overleftarrow{y}'$  if and only if  $P(\vec{Y} = \vec{y} | \overleftarrow{X}_t = \overleftarrow{y}) = P(\vec{Y} = \vec{y} | \overleftarrow{X}_t = \overleftarrow{y}')$  for all possible future values  $\vec{y} \in \vec{Y}$ . The causal states are unique for any given process, and so their entropy  $H(S)$  is a property of the process itself known as its *statistical complexity*  $C_{\mu}$ , capturing the intuition that a more complex process requires more memory to simulate.

For Markovian processes, such as discussed in this article, the number of causal states required is equal to the number of unique rows in the stochastic matrix describing the evolution. When these rows are generated by the discretization of a continuous process into  $N$  divisions—such as when they are derived from the cyclic walk's shift function  $P(X)$ —the number of states will be equal to  $N$ , except for very specific (e.g. pathologically fractal) choices of  $P(X)$  and  $N$ . Since by symmetry the probability of the simulator being in any particular state is equal,

the classical memory cost of a simulator hence scales with the number of sites as  $\log N$ , or linearly with the precision  $n = \log_2 N$ .

### Details of the quantum circuit in figure 2

Let us consider figure 2 in more depth (see also [25]). The circuit consists of one persistent internal memory state, and an ‘output tape’—a line of quantum states, which are fed into the system one at a time. Suppose each state on the output tape is initialised into some arbitrary state  $|\phi\rangle$ . For any two quantum states  $|x\rangle$  and  $|y\rangle$  in the same Hilbert space, it is always possible to construct a unitary transformation  $V$  such that  $V|x\rangle = |y\rangle$ . This will be of the form  $|y\rangle\langle x| + \sum_i |y'_i\rangle\langle x'_i|$  where  $|x'_i\rangle$  are states orthogonal to each other and to  $|x\rangle$ , and  $|y'_i\rangle$  are states orthogonal to each other and to  $|y\rangle$ . Thus, in the joint Hilbert space  $\mathcal{H}^N \otimes \mathcal{H}^N$  of two quantum systems of dimension  $N$ , it is possible to build a ‘controlled’ unitary operation  $U$  containing the elements  $|j\rangle\langle j| \otimes |\psi_j\rangle\langle\phi|$  for every  $|\psi_j\rangle$  in an arbitrary (generally non-orthogonal) set of states  $\{|\psi_j\rangle\}_{j=0\dots(N-1)}$ . (Note: the orthogonality of  $\{|j\rangle\}$  allows us to pairwise use the above construction for each  $|\psi_j\rangle$ .)

For a Markovian process discretized such that the stochastic matrix with elements  $p_{jk}$  describes its evolution, the above prescription supplies the unitary operation required for our quantum simulator when we set each  $|\psi_j\rangle = |S_j\rangle = \sum_{k=0}^{N-1} \sqrt{p_{kj}}|k\rangle$ , as per equation (4) (states  $\{|k\rangle\}$  and  $\{|j\rangle\}$  are in the same basis).

We may now evaluate the action of a single time-step (grey dashed box within figure 2). Here, the joint Hilbert space corresponds to that of the internal memory together with the output tape. In the figure, we explicitly wrote the initial state of the output tape as  $|\phi\rangle = |S_0\rangle$ , but this is arbitrary; any  $|\phi\rangle$  could be made into  $|S_0\rangle$  by acting on it first with a unitary gate containing  $|S_0\rangle\langle\phi|$ . At the start of a time step, the internal memory is in state  $|S^t\rangle = |S_j\rangle = \sum_{k=0}^{N-1} \sqrt{p_{kj}}|k\rangle$ . Hence, the joint state of the memory and output tape is initially  $|S_j\rangle \otimes |\phi\rangle$ . After the controlled unitary is applied, the memory and tape will be in the entangled state  $\sum_k \sqrt{p_{kj}}|k\rangle \otimes |S_k\rangle$ . Applying a coherent swap operation (i.e. exchanging the labels of the Hilbert spaces) will take this joint state to  $\sum_k \sqrt{p_{kj}}|S_k\rangle \otimes |k\rangle$ —the state of the system at the end of the grey box.

The tape system is then ejected from the simulator. If one were to measure this state in the  $\{|k\rangle\}$  basis, one projects onto state  $|k\rangle$  with probability  $p_{kj}$ , and hence the output statistics of this measurement match that of the process being simulated. Moreover, after measuring, due to the entanglement, we know that when  $|k\rangle$  is measured, the internal memory must be in state  $|S_k\rangle$ , which is exactly the quantum state that would have been prepared if we had mapping the output statistics onto a classical causal state and then prepared  $|S_k\rangle$  directly. Hence, the quantum circuit in figure 2 can function as a discretized simulator for a Markovian process.

However, it is very important to note that there is no need whatsoever to measure the output tape  $|k\rangle$  for the quantum simulator to continue functioning. If it suits one’s purpose to store the output states in quantum memory (e.g. to perform further quantum information processing on the output data), then the quantum simulator still functions correctly. In this mode of operation, the measurements can be omitted from figure 2, and after  $M$  steps, the simulator would have produced the entangled state

$$|\Phi\rangle = \sum_{i_1} \dots \sum_{i_M} \sqrt{P(Y^t = y_{i_1}, \dots, Y^{t+M} = y_{i_M} | S^t)} |S^{t+M}(S^t, y_{i_1}, \dots, y_{i_M})\rangle \otimes |y_{i_1}\rangle \otimes \dots \otimes |y_{i_M}\rangle, \quad (7)$$

where  $|S^{t+M}(S^t, y_{i_1}, \dots, y_{i_M})\rangle$  is the quantum state that would have been prepared if the system was originally in causal state  $S^t$  then outputted string  $y_{i_1} \dots y_{i_M}$ , and a new causal state directly set according to this output sequence. Measuring the string of output tape subsystems thus still ensures that the internal memory state collapses into the correct causal state  $|S^{t+M}\rangle$ , conditional on the string observed.

In the first mode of operation (as drawn in figure 2), only one ancillary quantum system is required, as it can be reset and re-used between timesteps (the output tape carries away classical information only). In the second mode, the quantum output explicitly fulfils the role of the ancillary system, and a fresh ancillary system (provided by the ‘blank’ output tape set to some fixed choice of pure quantum state) is inserted at each time step. In both modes, the ancillary system does not need to persist between time steps in order for the simulator to continue producing statistically correct outputs. As such, in both cases, it is the von Neumann entropy  $-\text{Tr} \rho \log \rho$  of the first subsystem, which remains within the simulator at all times, that we consider to be the internal memory cost.

### Derivation of discrete eigenspectrum

The quantum machine state corresponding to the system being in classical state  $\alpha$  is given as  $|S_\alpha\rangle = \sum_\beta \sqrt{p_{\beta\alpha}}|\beta\rangle$ . Assuming  $\{p_{\beta\alpha}\}_{\beta\alpha}$  is simply connected, the quantum machine will reach a stationary state  $\rho = \frac{1}{N} \sum_\alpha |S_\alpha\rangle\langle S_\alpha|$ . Rather than directly calculating the entropy of  $\rho$ , we can instead evaluate the entropy of the

associated Gram matrix  $g$ , whose elements  $g_{\alpha\beta}$  are given by the overlaps  $\frac{1}{N}\langle S_\alpha|S_\beta\rangle$ <sup>11</sup>. The circular symmetry of the cyclic random walk ensures that the discretized transition probabilities satisfy  $p_{\alpha\beta} = p_{(\alpha+k)(\beta+k)}$  (that is, the transition probabilities depend only on differences between indices). It hence follows that the Gram matrix associated with  $\rho$  is *circulant* [35]. Since all rows can be derived by cyclic permutation of the top row, we shall drop one index and write the top row as  $g_\alpha = g_{0\alpha}$ . The eigenvalues of the Gram matrix are given by  $\lambda_k = \sum_\alpha g_\alpha \exp\left(-\frac{2\pi i}{N}\alpha k\right)$  for  $k = 0, \dots, N - 1$ , which can immediately be recognised as the DFT of  $\{g_\alpha\}_\alpha$ , which we denote as  $\mathcal{F}(g_{0\alpha})$ .

Moreover, the inner product  $\langle S_0|S_j\rangle = \sum_\alpha \sqrt{p_{\alpha 0}p_{\alpha j}}$ , has the form of a convolution  $\sqrt{p} * \sqrt{q}$ , where we have rewritten  $p_{\alpha j}$  as  $q_{0(\alpha-j)}$  such that  $q$  is the  $N$ -periodic extension of the reflection of  $p$ ;  $q_{0j} = p_{(N-j)0}$  and  $q_{0j} = q_{0(j+N)}$ . We may then apply the circular convolution theorem to find the eigenvalues of  $g$ , and therefore of  $\rho$ :

$$\lambda_k = \frac{1}{N}\mathcal{F}[\sqrt{p_{j0}}]\mathcal{F}[\sqrt{p_{(N-j)0}}]. \tag{8}$$

These eigenvalues can hence be found efficiently by numerical algorithms, such as the fast-Fourier transform.

**Example (Dirac-delta shift function).** Let the shift function be  $P(x) = \delta(x - x_0)$  for some  $x_0 \in [0, 1)$ . It can be seen that all  $p_{j0} = 0$  except for the one at index  $j'$  that incorporates the delta peak where  $p_{j'0} = 1$ . Hence,  $\mathcal{F}(p_{j'0}) = \exp\left(-2\pi i\frac{j'}{N}k\right)$  and  $\mathcal{F}(p_{(N-j)0}) = \exp\left(-2\pi i\frac{(N-j')}{N}k\right)$ , and so  $\lambda_k = \frac{1}{N}$  for all  $k$ . Thus, the von Neumann entropy of the simulator's memory is  $\log N$ .

**Example (Uniform shift function).** Consider the uniform shift function  $P(x) = 1$  for  $x \in [0, 1)$ . Here,  $p_{j0} = \frac{1}{N}$ , and so  $\mathcal{F}[\sqrt{p_{j0}}] = \sqrt{N}$  for  $k = 0$  and 0 for all other  $k$ . As such, we find that the eigenvalue  $\lambda_0 = 1$ , and all other eigenvalues  $\lambda_1 = \dots \lambda_{N-1} = 0$ , and hence the entropy of the Gram matrix is zero, for all values of  $N$ .

**Sampling Fourier transforms**

It will be useful to show an auxiliary relationship between discrete and continuous Fourier transforms. Let  $g(x)$  be a function over the range  $x \in [0, 1]$  that is sampled at  $N$  equally spaced points with values given by  $g_n = g\left(\frac{n}{N}\right)$  for  $n = 0 \dots N - 1$ . We can construct a function  $g_{\text{comb}}(x) = \sum_{n=0}^{N-1} \delta\left(x - \frac{n}{N}\right)g(x)$ , whose Fourier transform is

$$\begin{aligned} \mathcal{F}(g_{\text{comb}}(x)) &= \int_{-\infty}^{\infty} dx \sum_{n=0}^{N-1} \delta\left(x - \frac{n}{N}\right)g(x) \exp(-2\pi i k x) \\ &= \sum_{n=0}^{N-1} g\left(\frac{n}{N}\right) \exp\left(-2\pi i \frac{n}{N}k\right), \end{aligned} \tag{9}$$

which when evaluated at integer  $k$  is exactly the DFT of the samples  $\{g_n\}$ , which we write as  $\{\lambda_k\}$ .

If  $g$  is periodic, it is always possible to offset the position of the sample window of  $g$  by some integer  $c$  without changing the values of  $g$ 's DFT. For the functions we consider in this article, it is more convenient to start at  $-\frac{N}{2}$ , since typically  $g_{-\frac{N}{2}}, g_{\frac{N}{2}} \rightarrow 0$  and  $g_0 = 1$ . Moreover, once the sample window has been set, the values of  $g(x)$  outside this window can not affect  $\lambda_k$ , since they do not feature in the sum. Thus, instead of considering sampling  $g(x)$  across a finite window, we can consider an infinite delta train sampled at the same intervals, but across a function  $g_{\text{once}}(x)$  where  $g_{\text{once}}(x) = g(x)$  inside the range of the sample window (i.e.  $\left[-\frac{1}{2}, \frac{1}{2}\right)$ ) for the window used in this article) and  $g_{\text{once}}(x) = 0$  outside this range. Here

$$\begin{aligned} \lambda_k &= \mathcal{F}(g_{\text{comb}}(x)) = \mathcal{F}\left(g(x) \sum_{n=-\frac{N}{2}}^{\frac{N}{2}} \delta\left(x - \frac{n}{N}\right)\right) \\ &= \mathcal{F}\left(g_{\text{once}}(x) \sum_{n=-\infty}^{\infty} \delta\left(x - \frac{n}{N}\right)\right) \\ &= \mathcal{F}(g_{\text{once}}(x)) * \sum_{m=-\infty}^{\infty} \delta(k - mN), \end{aligned} \tag{10}$$

<sup>11</sup> This works by constructing a fictitious purification of  $\rho$ , given  $|\Psi\rangle = \sum_i \frac{1}{\sqrt{N}}|S_i\rangle \otimes |i\rangle$  (where  $\{|i\rangle\}_i$  is an orthonormal basis) such that  $\text{Tr}_B|\Psi\rangle\langle\Psi| = \rho$  and  $\text{Tr}_A|\Psi\rangle\langle\Psi| = g$ . Since the von Neumann entropy of pure state  $|\Psi\rangle$  is 0, it follows from triangle inequalities that  $H(\rho) = H(g)$ .

where we have used the convolution theorem in the final step. The periodic sampling of  $g(x)$  causes the Fourier transform to be periodic with period  $N$  (a phenomenon known as aliasing), such that  $\lambda_k = \lambda_{k+N}$ ; the convolution with a delta train effectively makes  $\lambda_k$  a *periodic sum* of  $\mathcal{F}(g_{\text{once}}(x))$ . This periodicity allows us the freedom to choose a convenient range of  $k$ . In this article, we will typically use  $-\frac{N}{2}$  to  $\frac{N}{2} - 1$ . If  $\mathcal{F}(g_{\text{once}}(x)) \approx 0$  outside the chosen range, then we can approximate

$$\lambda_k \approx [\mathcal{F}(g_{\text{once}}(x))](k). \tag{11}$$

**Asymptotic limit of eigenvalues**

For large  $N$ , we can derive an expression for  $\lambda_k$  in terms of the probability density function  $P(x)$ . We substitute  $p_{\alpha 0}$  with  $\frac{1}{N}P(\frac{\alpha}{N})$ , which for Riemann-integrable  $P(x)$  is an arbitrarily good approximation in the limit of  $N \rightarrow \infty$ . Similarly, we may substitute  $p_{\alpha j}$  with  $\frac{1}{N}P^\circ(-\frac{j-\alpha}{N})$ , where  $P^\circ(x)$  denotes the 1-periodic extension<sup>12</sup> of  $P(x)$ . Taking the limit of the Riemann sum for a product of two functions, we then see

$$\begin{aligned} \lim_{N \rightarrow \infty} \langle S_0 | S_j \rangle &= \lim_{N \rightarrow \infty} \sum_{\alpha=0}^{N-1} \sqrt{p_{\alpha 0} p_{\alpha j}} \\ &= \lim_{N \rightarrow \infty} \sum_{\alpha=0}^{N-1} \frac{1}{N} \sqrt{P\left(\frac{\alpha}{N}\right) P^\circ\left(-\frac{j-\alpha}{N}\right)} \\ &= \int_0^1 dx \sqrt{P(x) P^\circ(y-x)}, \end{aligned} \tag{12}$$

where  $y = \frac{j}{N}$ . Moreover, since  $P$  only has support in  $[0, 1)$ , we can rewrite the integral limits from  $-\infty$  to  $\infty$ , and conclude that  $\lim_{N \rightarrow \infty} \langle S_0 | S_j \rangle = [\sqrt{P(x)} * \sqrt{P^\circ(-x)}](y)$  sampled at  $y = 0, \frac{1}{N}, \dots, \frac{N-1}{N}$ . Thus by treating  $g_j$  as samples from a function  $g(y = \frac{j}{N})$  at discrete intervals of  $\frac{1}{N}$ , we find that  $g_j \approx \frac{1}{N}g\left(y = \frac{j}{N}\right)$  for large  $N$ , and hence

$$g\left(y = \frac{j}{N}\right) = [\sqrt{P(x)} * \sqrt{P^\circ(-x)}](y). \tag{13}$$

As shown in equation (10), the eigenvalues  $\{\lambda_k\}$  are given by  $\lambda_k = [\mathcal{F}(g_{\text{once}}(x)) * \sum_{m=-\infty}^{\infty} \delta(k - mN)]$  evaluated at integers  $k = 0, 1, \dots, N - 1$ , where  $g_{\text{once}}(y) = g(y)$  over an (arbitrary) single period of  $g(y)$  and takes the value zero elsewhere. Due to the periodic summation, it can be seen also that  $\lambda_k = \lambda_{k+N}$ , and so we are also free to choose the most convenient range for  $k$ , which will typically be from  $-\frac{N}{2}$  to  $\frac{N}{2} - 1$ . If  $[\mathcal{F}(g_{\text{once}})](k) \approx 0$  when  $|k| > \frac{N}{2}$ , then the approximation

$$\lambda_k \approx [\mathcal{F}(g_{\text{once}})](k) \quad \text{for } k = -\frac{N}{2}, \dots, \frac{N}{2} - 1 \tag{14}$$

is reasonable. This assumption amounts taking enough samples of  $g(x)$  to admit a faithful reconstruction of  $g(x)$  under the Nyquist–Shannon theorem [36]. This holds true for the examples we shall now consider, where we will ultimately take large values of  $N$ .

**Example 1 (Gaussian noise).** Suppose the shift function of the particle is given by a Gaussian distribution  $G_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$  about  $\mu$  with standard deviation  $\sigma \ll 1$  such that we can ignore the probability of the particle looping around the ring.

**Derivation of eigenvalues**

We can express  $\sqrt{G_{\mu,\sigma}(x)}$  as a Gaussian:

$$\begin{aligned} \sqrt{G_{\mu,\sigma}(x)} &= \sigma^{-\frac{1}{2}} (2\pi)^{-\frac{1}{4}} \exp\left(-\frac{(x-\mu)^2}{4\sigma^2}\right) \\ &= \sigma^{\frac{1}{2}} (2\pi)^{\frac{1}{4}} \sqrt{2} (\sqrt{2}\sigma)^{-1} (2\pi)^{-\frac{1}{2}} \exp\left(-\frac{(x-\mu)^2}{2(\sqrt{2}\sigma)^2}\right) \\ &= \sigma^{\frac{1}{2}} 2^{\frac{1}{2}} (2\pi)^{\frac{1}{4}} G_{\mu,\sqrt{2}\sigma}(x). \end{aligned} \tag{15}$$

It can be easily verified that  $g_{\mu,\sigma}(-x) = g_{-\mu,\sigma}(x)$ .

<sup>12</sup>Equivalent to wrapping  $x$  to  $[0, 1)$  before evaluating  $P(x)$ .

We also note that  $\mathcal{F}(g_{\mu,\sigma}(x))$  is also Gaussian:

$$\begin{aligned} \mathcal{F}(g_{\mu,\sigma}(x)) &= \exp(2\pi i \mu k) \exp(-2(\pi\sigma)^2 k^2) \\ &= (2\pi)^{\frac{1}{2}} \frac{1}{2\pi\sigma} \exp(-2\pi i \mu k) \cdot (2\pi)^{-\frac{1}{2}} \left(\frac{1}{2\pi\sigma}\right)^{-1} \exp\left(-\frac{k^2}{2\left(\frac{1}{2\pi\sigma}\right)^2}\right) \\ &= (2\pi)^{-\frac{1}{2}} \sigma^{-1} \exp(-2\pi i \mu k) g_{0, \frac{1}{2\pi\sigma}}(k) \end{aligned} \tag{16}$$

Likewise, we can express  $[G_{\mu,\sigma}(x)]^2$  as a Gaussian:

$$\begin{aligned} [G_{\mu,\sigma}(x)]^2 &= \sigma^{-2} (2\pi)^{-1} \exp\left(-\frac{(x-\mu)^2}{\sigma^2}\right) \\ &= \sigma^{-1} (2\pi)^{-\frac{1}{2}} 2^{-\frac{1}{2}} \left(\frac{\sigma}{\sqrt{2}}\right)^{-1} (2\pi)^{-\frac{1}{2}} \exp\left(-\frac{(x-\mu)^2}{2\left(\frac{\sigma}{\sqrt{2}}\right)^2}\right) \\ &= \sigma^{-1} (2\pi)^{-\frac{1}{2}} 2^{-\frac{1}{2}} G_{\mu, \frac{\sigma}{\sqrt{2}}}(x). \end{aligned} \tag{17}$$

Taken together (making sure to substitute in the correctly modified values of  $\mu$  and  $\sigma$ ), this allows us to provide an analytic solution for equation (14) for Gaussian shift functions:

$$\begin{aligned} \lambda_k &= \mathcal{F}(\sqrt{G_{\mu,\sigma}(x)}) \mathcal{F}(\sqrt{G_{\mu,\sigma}(-x)}) \\ &= \mathcal{F}(\sqrt{g_{\mu,\sigma}(x)}) \mathcal{F}(\sqrt{G_{-\mu,\sigma}(x)}) \\ &= 2\sigma (2\pi)^{\frac{1}{2}} \mathcal{F}(G_{\mu, \sqrt{2}\sigma}(x)) \mathcal{F}(G_{-\mu, \sqrt{2}\sigma}(x)) \\ &= 2\sigma (2\pi)^{\frac{1}{2}} (2\pi)^{-1} (\sqrt{2}\sigma)^{-2} \exp(-2\pi i \mu k) \exp(2\pi i \mu k) \cdot \left[G_{0, \frac{1}{2\sqrt{2}\pi\sigma}}(k)\right]^2 \\ &= (2\pi)^{-\frac{1}{2}} \sigma^{-1} \left(\frac{1}{2\sqrt{2}\pi\sigma}\right)^{-1} (2\pi)^{-\frac{1}{2}} 2^{-\frac{1}{2}} G_{0, \frac{1}{4\pi\sigma}}(k) \\ &= G_{0, \frac{1}{4\pi\sigma}}(k). \end{aligned} \tag{18}$$

Hence, we see that choosing Gaussian transfer function with standard deviation  $\sigma \ll 1$  corresponds to a spectrum of eigenvalues with standard deviation  $\frac{1}{4\pi\sigma}$ .

### Upper bound on quantum memory cost

We now demonstrate that the entropy of such a system, given  $H_Q = -\sum_k \lambda_k \log_2 \lambda_k$ , is finite by bounding it from above. For convenience, we write  $\lambda(k) := G_{0, \frac{1}{4\pi\sigma}}(k) = A \exp(-Bk^2)$  where  $A := 2\sqrt{2\pi}\sigma$  and  $B := 8\pi^2\sigma^2$ , and will perform the calculation in units of *nats*. Thus, consider  $c(k) = -\lambda(k) \ln \lambda(k)$ , explicitly

$$c(k) = A \exp(-Bk^2) (Bk^2 - \ln A). \tag{19}$$

By setting  $\frac{dc}{dk} = 2ABk \exp(-Bk^2) (-Bk^2 + \ln A + 1) = 0$ , we find that  $c(k)$  has stationary points at  $k = 0, \pm\infty$  and when

$$k = \pm \sqrt{\frac{\ln A + 1}{B}} = \pm \sqrt{\frac{\ln(2\sqrt{2\pi}\sigma) + 1}{8\pi^2\sigma^2}}. \tag{20}$$

When  $\sigma < \frac{1}{2e\sqrt{2\pi}} \approx 0.073$ , these last two solutions disappear, and since we are in the regime of  $\sigma \ll 1$ , this condition is satisfied. Hence, for small  $\sigma$ ,  $c(k)$  monotonically decreases from its maximum value at  $k = 0$  for both positive and negative  $k$ . This allows us to apply the Maclaurin–Cauchy integral bound (see e.g. [37]),

$$\int_m^\infty c(k) dk \leq \sum_{k=m}^\infty c(k) \leq c(m) + \int_m^\infty c(k) dk, \tag{21}$$

which holds for any monotonically decreasing region  $[m, \infty)$  of a function  $c(k)$  (here,  $m = 0$ ).

Using known results for definite Gaussian integrals,

$$\int_0^\infty e^{-Bx^2} dx = \frac{1}{2} \sqrt{\frac{\pi}{B}} \quad \text{and} \quad \int_0^\infty x^2 e^{-Bx^2} dx = \frac{1}{4} \sqrt{\frac{\pi}{B^3}}, \tag{22}$$

we evaluate

$$\begin{aligned} \int_0^\infty c(k)dk &= AB\frac{1}{4}\sqrt{\frac{\pi}{B^3}} - A \ln A \sqrt{\frac{\pi}{B}} \\ &= \frac{A}{2}\sqrt{\frac{\pi}{B}}\left(\frac{1}{2} - \ln A\right) \\ &= \frac{1}{2}\left(\frac{1}{2} - \ln A\right). \end{aligned} \tag{23}$$

Since  $c(0) = -A \ln A$ , we find from equation (20) that

$$\begin{aligned} \sum_{k=m}^\infty c(k) &\leq \left(\frac{1}{2} - \ln A\right) - A \ln A, \\ &\leq \frac{1}{4} - \left(\frac{1}{2} + A\right) \ln A. \end{aligned} \tag{24}$$

To obtain a bound on  $H_Q$ , we double the above since  $c(k)$  is even, and multiply by  $\frac{1}{\ln 2}$  to convert from *nats* to *bits* (equivalently, change the base  $\ln$  to  $\log_2$  since  $\frac{\ln x}{\ln 2} = \log_2 x$ ):  $H_Q \leq \frac{1}{2 \ln 2} - (1 + 2A) \log_2 A$ . In terms of the shift function's standard deviation  $\sigma$ , this gives our result

$$H_Q \leq \frac{1}{2 \ln 2} - (1 + 4\sqrt{2\pi}\sigma) \log_2 2\sqrt{2\pi}\sigma. \tag{25}$$

In the limit of small  $\sigma$ , the leading term of the entropy thus scales with  $-\log_2 \sigma$ , such that halving the width of the standard deviation adds one bit to the maximum required quantum memory cost.

**Example 2 (Uniform white noise).** The normalised top-hat (rectangular) shift function allowing for jumps of up to  $\pm \Delta$  around a constant displacement  $\mu$  is written

$$S_\Delta(x) = \begin{cases} \frac{1}{2\Delta} & \text{for } \mu - \Delta \leq x \leq \mu + \Delta, \\ 0 & \text{otherwise.} \end{cases} \tag{26}$$

**Derivation of eigenvalues**

Taking the square root of this function alters its normalisation, but not its shape:  $\sqrt{S_\Delta(x)} = \sqrt{2\Delta} S_\Delta(x)$ .

Suppose  $0 < \Delta < \frac{1}{2}$ . In this case,  $S_\Delta(x) * S_\Delta(-x)$  yields the triangle function

$$S_\Delta(x) * S_\Delta(-x) = \begin{cases} 1 - \frac{x}{2\Delta} & \text{for } 0 \leq x \leq 2\Delta, \\ 1 + \frac{x}{2\Delta} & \text{for } -2\Delta \leq x < 0, \\ 0 & \text{otherwise.} \end{cases} \tag{27}$$

This function is independent of the constant displacement  $\mu$ . Indeed, non-zero  $\mu$  only results in perfectly cancelling terms  $e^{2\pi i k \mu}$  and  $e^{-2\pi i k \mu}$  in the Fourier transform.

Basic Fourier analysis tells us that  $S_\Delta(x)$  transforms into a *normalised sinc function* ( $\text{sinc } x = \sin(\pi x)/\pi x$ ), and the triangle function into the square of this:  $\mathcal{F}(S_\Delta(x) * S_\Delta(-x)) = 2\Delta \text{sinc}^2(2k\Delta)$ . As this tends to 0 for large  $k$ , we can approximate the values of  $\lambda_k$  for large  $N$  using equation (14), to find the eigenspectrum

$$\lambda_k = 2\Delta \text{sinc}^2(2k\Delta) \quad \text{for } k = -\frac{N}{2}, \dots, \frac{N}{2} - 1. \tag{28}$$

**Upper bound on quantum memory cost**

Through the careful deployment of mildly intimidating algebra, we can also derive an upper bound on entropy cost of simulating the square shift function. The outline of the proof is as follows. To bound  $\sum_k c(k)$  where  $c(k) = -\lambda_k \ln \lambda_k$ , we first construct a monotonically decreasing function  $d(k)$  that satisfies  $c(k) \leq d(k)$  at every  $k$ , and then show that  $\sum d(k)$  is bounded from above. This sum will hence also upper-bound  $\sum_k c(k)$ . As with the Gaussian example, for algebraic convenience, we will use natural logarithms and only consider the region of positive  $k$ . In the final stage, we will convert from *nats* to *bits*, and use the evenness of  $c(k)$  to arrive at the full bound.

Explicitly, we write

$$c(x) = -2\Delta \frac{\sin^2 x}{x^2} \ln \left( 2\Delta \frac{\sin^2 x}{x^2} \right), \tag{29}$$

where we have made the substitution  $x = 2\pi k \Delta$ .

In the region  $x > 0$ , we can expand

$$c(x) = -2\Delta \frac{\sin^2 x}{x^2} \left[ \ln(\sin^2 x) - 2 \ln \left( \frac{x}{\sqrt{2\Delta}} \right) \right]. \tag{30}$$

The function  $-y \ln y$  has a maximum value of  $\frac{1}{e}$  at  $y = e$ , and so we can upper bound  $c(x)$  by making the substitution of  $-\sin^2 x \ln \sin^2 x$  with  $\frac{1}{e}$ . Since  $\sin^2 x \in [0, 1]$ , in the region  $x > \sqrt{2\Delta}$  where  $4\Delta \ln \left( \frac{x}{\sqrt{2\Delta}} \right) > 0$ , we can likewise upper bound  $c(x)$  by making the substitution of  $\sin^2(x)$  with 1. Thus, for the region  $x > \sqrt{2\Delta}$ , we have a function  $f(x) \geq c(x)$  given

$$f(x) = \frac{2\Delta}{x^2} \left[ \frac{1}{e} + 2 \ln \frac{x}{\sqrt{2\Delta}} \right]. \tag{31}$$

However, as we plan to ultimately apply the Maclaurin–Cauchy integral convergence test, it is only convenient to use this upper bound in the region of  $x$  where  $f(x)$  monotonically decreases. We identify this region by setting  $\frac{df}{dx} = \frac{4\Delta}{x^3} \left( -2 \ln \frac{x}{\sqrt{2\Delta}} + 1 - \frac{1}{e} \right) = 0$ , to find that  $f(x)$  decreases monotonically when  $x \geq \sqrt{2\Delta} \exp \left( \frac{e-1}{2e} \right)$ , descending from its maximum value of  $\exp \left( \frac{1-e}{e} \right)$ .

However, once again consider  $c(x)$ . Since it has the form of  $-y \ln y$ , it follows that in *any* region,  $c(x) \leq \frac{1}{e}$ . Since  $\frac{1-e}{e} > \frac{1}{e}$ , we can then upper bound  $c(x)$  in the region of  $0 \leq x \leq \sqrt{2\Delta}$  to form the monotonically decreasing function  $d(x)$  given

$$d(x) = \begin{cases} \exp \left( \frac{1-e}{e} \right) & 0 \leq x \leq \sqrt{2\Delta} \exp \left( \frac{e-1}{2e} \right) \\ \frac{2\Delta}{x^2} \left[ \frac{1}{e} + 2 \ln \frac{x}{\sqrt{2\Delta}} \right] & x > \sqrt{2\Delta} \exp \left( \frac{e-1}{2e} \right), \end{cases} \tag{32}$$

that is guaranteed to satisfy  $d(x) \geq c(x)$  for all  $x \geq 0$ . At this point, it is convenient to express this again in terms of  $k$ , making the substitution  $k_{\text{split}} = \frac{1}{\pi\sqrt{2\Delta}} \exp \left( \frac{e-1}{2e} \right)$ :

$$d(k) = \begin{cases} \exp \left( \frac{1-e}{e} \right) & 0 \leq k \leq \lceil k_{\text{split}} \rceil \\ \frac{1}{2\Delta\pi^2 k^2} \left[ \frac{1}{e} + 2 \ln(\pi\sqrt{2\Delta} k) \right] & k > \lceil k_{\text{split}} \rceil, \end{cases} \tag{33}$$

where  $\lceil k_{\text{split}} \rceil$  represents the lowest integer above (or including)  $k_{\text{split}}$ . This rounding is necessary since  $k_{\text{split}} = \frac{1}{\pi\sqrt{2\Delta}} \exp \left( \frac{e-1}{2e} \right)$  is in general not an integer. To upper bound  $c(k)$  at all points, we must round up this split between the regions of  $k$ , since  $\exp \left( \frac{1-e}{e} \right)$  upper bounds all  $f(k)$ . (I.e. being slightly too inclusive in the first region will result in a slightly higher value of  $d(k)$  for the first  $k$  satisfying  $k \geq k_{\text{split}}$ .)

Having derived our monotonically decreasing function  $d(k)$ , we are now in a position to show that  $\sum_{k=0}^{\infty} d(k)$  is finite for  $\Delta > 0$ . Writing  $\sum_{k=0}^{\lceil k_{\text{split}} \rceil} d(k) + \sum_{k=\lceil k_{\text{split}} \rceil}^{\infty} d(k)$  (for an upper bound, it is fine if a term is counted twice!), we evaluate the two regions separately. Firstly,

$$\begin{aligned} \sum_{k=0}^{\lceil k_{\text{split}} \rceil} d(k) &\leq \left( 1 + \frac{1}{\pi\sqrt{2\Delta}} \exp \left( \frac{e-1}{2e} \right) \right) \exp \left( \frac{1-e}{e} \right) \\ &= \exp \left( \frac{1-e}{e} \right) + \frac{1}{\pi\sqrt{2\Delta}} \exp \left( \frac{1-e}{e} \right), \end{aligned} \tag{34}$$

where we have used  $\frac{1}{2\pi\Delta} \exp \left( \frac{e-1}{2e} \right) + 1 > \left\lceil \frac{1}{2\pi\Delta} \exp \left( \frac{e-1}{2e} \right) \right\rceil$ . Secondly, using the Maclaurin–Cauchy integral test (see e.g. [37]), we bound

$$\begin{aligned} \sum_{k=\lceil k_{\text{split}} \rceil}^{\infty} d(k) &\leq d(\lceil k_{\text{split}} \rceil) + \int_{\lceil k_{\text{split}} \rceil}^{\infty} d(k) dk \\ &\leq \exp \left( \frac{1-e}{e} \right) + \int_{k_{\text{split}}}^{\infty} d(k) dk, \end{aligned} \tag{35}$$

where the second line follows by substituting  $d(\lceil k_{\text{split}} \rceil)$  with the maximum value of  $d(k)$ , and by failing to round up the lower bound of the integral (thus including an extra contribution equal to  $\int_{k_{\text{split}}}^{\lceil k_{\text{split}} \rceil} d(k) dk \geq 0$ ). This integral may be analytically solved,

$$\begin{aligned}
& \frac{1}{2\pi^2\Delta} \int_{k_{\text{split}}}^{\infty} \frac{1}{k^2} \left[ \frac{1}{e} + 2 \ln(\pi\sqrt{2\Delta}k) \right] \\
&= \left[ \frac{-1}{2\pi^2\Delta} \left( \frac{1}{k} + 2 + 2 \ln(\pi\sqrt{2\Delta}k) \right) \right]_{\frac{1}{\pi\sqrt{2\Delta}} \exp\left(\frac{e-1}{2e}\right)}^{\infty} \\
&= \frac{3}{\pi\sqrt{2\Delta}} \exp\left(\frac{1-e}{2e}\right).
\end{aligned} \tag{36}$$

Combining these two terms, we arrive at:

$$\sum_{k=0}^{\infty} d(k) \leq \frac{4}{\pi\sqrt{2\Delta}} \exp\left(\frac{1-e}{2e}\right) + 2 \exp\left(\frac{1-e}{e}\right). \tag{37}$$

Finally, to bound the entropy  $H_Q = -\sum_{\infty}^{\infty} \lambda_k \log_2 \lambda_k$ , we must double the above ( $c(k)$  is even, and equation (37) bounds only the region  $[0, \infty)$ ), and we convert from nats to bits (by including a factor of  $\frac{1}{\ln 2}$ ):

$$H_Q \leq \frac{8}{\pi \ln 2 \sqrt{2\Delta}} \exp\left(\frac{1-e}{2e}\right) + \frac{4}{\ln 2} \exp\left(\frac{1-e}{e}\right). \tag{38}$$

By evaluating the constant terms, approximately,

$$H_Q \leq \frac{1.894}{\sqrt{\Delta}} + 3.067, \tag{39}$$

yielding our result.

## ORCID iDs

Andrew J P Garner  <https://orcid.org/0000-0002-3747-9997>

## References

- [1] IEEE 2008 754-2008 - *IEEE Standard for Floating-Point Arithmetic* (New York: IEEE) (<https://doi.org/10.1109/IEEESTD.2008.4610935>)
- [2] Deutsch D 1985 Quantum theory, the church-turing principle and the universal quantum computer *Proc. R. Soc. A* **400** 97–117
- [3] Deutsch D and Jozsa R 1992 Rapid solution of problems by quantum computation *Proc. R. Soc. A* **439** 553–8
- [4] Grover L K 1996 A fast quantum mechanical algorithm for database search *Proc. 28th Annual ACM Symp. on Theory of Computing, STOC '96* (New York: ACM) pp 212–9
- [5] Shor P W 1997 Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer *SIAM J. Comput.* **26** 1484–509
- [6] Cleve R, Ekert A K, Macchiavello C and Mosca M 1998 Quantum algorithms revisited *Proc. R. Soc. A* **454** 339–54
- [7] van Dam W 2000 Nonlocality and communication complexity *PhD Thesis* University of Oxford
- [8] de Wolf R M 2001 Quantum computing and communication complexity *PhD Thesis* University of Amsterdam
- [9] Brassard G 2003 Quantum communication complexity *Found. Phys.* **33** 1593–616
- [10] Crutchfield J P and Young K 1989 Inferring statistical complexity *Phys. Rev. Lett.* **63** 105–8
- [11] Shalizi C R and Crutchfield J P 2001 Computational mechanics: pattern and prediction, structure and simplicity *J. Stat. Phys.* **104** 817–79
- [12] Crutchfield J P, Ellison C J and Mahoney J R 2009 Time's barbed arrow: irreversibility, crypticity, and stored information *Phys. Rev. Lett.* **103** 094101
- [13] Crutchfield J P 2011 Between order and chaos *Nat. Phys.* **8** 17–24
- [14] Haslinger R, Klinkner K L and Shalizi C R 2010 The computational structure of spike trains *Neural Comput.* **22** 121–57
- [15] Shalizi C R, Shalizi K L and Haslinger R 2004 Quantifying self-organization with optimal predictors *Phys. Rev. Lett.* **93** 118701
- [16] Marques da Silva J G, Sartorelli J C, Gonçalves W M and Pinto R D 1997 A scale law in a dripping faucet *Phys. Lett. A* **226** 269–74
- [17] Clarke R W, Freeman M P and Watkins N W 2003 Application of computational mechanics to the analysis of natural data: an example in geomagnetism *Phys. Rev. E* **67** 016203
- [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** 179–87
- [19] Li C-B, Yang H and Komatsuzaki T 2008 Multiscale complex network of protein conformational fluctuations in single-molecule time series *Proc. Natl Acad. Sci. USA* **105** 536–41
- [20] Lu C and Brooks R R 2012 P2P hierarchical botnet traffic detection using hidden Markov models *Proc. 2012 Workshop on Learning from Authoritative Security Experiment Results—LASER '12* (New York: ACM) pp 41–6
- [21] Shannon C E 1948 A mathematical theory of communication *Bell Syst. Tech. J.* **27** 379–423, 623–56
- [22] Schumacher B 1995 Quantum coding *Phys. Rev. A* **51** 2738–47
- [23] Winter A 1999 Coding theorems of quantum information theory *PhD Thesis* Universität at Bielefeld
- [24] Ladyman J, Lambert J and Wiesner K 2013 What is a complex system? *Eur. J. Phil. Sci.* **3** 33–67
- [25] Gu M, Wiesner K, Rieper E and Vedral V 2012 Quantum mechanics can reduce the complexity of classical models *Nat. Commun.* **3** 762
- [26] Suen W Y, Thompson J, Garner A J P, Vedral V and Gu M 2017 The classical-quantum divergence of complexity in modelling spin chains *Quantum* **1** 25
- [27] Mahoney J R, Aghamohammadi C and Crutchfield J P 2016 Occams quantum strop: synchronizing and compressing classical cryptic processes via a quantum channel *Sci. Rep.* **6** 20495

- [28] Palsson M S, Gu M, Ho J, Wiseman H M and Pryde G J 2017 Experimentally modeling stochastic processes with less memory by the use of a quantum processor *Sci. Adv.* **3** e1601302
- [29] Riechers P M, Mahoney J R, Aghamohammadi C and Crutchfield J P 2016 Minimized state complexity of quantum-encoded cryptic processes *Phys. Rev. A* **93** 052317
- [30] Aghamohammadi C, Mahoney J R and Crutchfield J P 2017 The ambiguity of simplicity in quantum and classical simulation *Phys. Lett. A* **381** 1223–27
- [31] Mahoney J R, Ellison C J and Crutchfield J P 2009 Information accessibility and cryptic processes *J. Phys. A: Math. Theor.* **42** 362002
- [32] Wiesner K, Gu M, Rieper E and Vedral V 2012 Information-theoretic lower bound on energy cost of stochastic computation *Proc. R. Soc. A* **468** 4058–66
- [33] Still S, Sivak D A, Bell A J and Crooks G E 2012 Thermodynamics of prediction *Phys. Rev. Lett.* **109** 120604
- [34] Garner A J P, Thompson J, Vedral V and Gu M 2017 The thermodynamics of complexity and pattern manipulation? *Phys. Rev. E* **95** 042140
- [35] Gray R M 2005 Toeplitz and circulant matrices: a review *Found. Trends Commun. Inf. Theory* **2** 155–239
- [36] Shannon C E 1949 Communication in the presence of noise *Proc. IRE* **37** 10–21
- [37] Knopp K 1990 *Theory and Applications of Infinite Series* 2nd english edn (New York: Dover)