The classical–quantum divergence of complexity in modelling spin chains

Whei Yeap Suen[1], Jayne Thompson[1], Andrew J. P. Garner[1], Vlatko Vedral[2 1 3], and Mile Gu[4 5 1]

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

[2]Atomic and Laser Physics, University of Oxford, Clarendon Laboratory, Parks Road, Oxford, OX1 3PU, United Kingdom.

[3]Department of Physics, National University of Singapore, 3 Science Drive 2, Singapore 117543

[4]School of Mathematical and Physical Scieces, Nanyang Technological University, Singapore

[5]Complexity Institute, Nanyang Technological University, Singapore

August 1, 2017

arXiv:1511.05738v4 [quant-ph] 31 Jul 2017

The minimal memory required to model a given stochastic process—known as the statistical complexity—is a widely adopted quantifier of structure in complexity science. Here, we ask if quantum mechanics can fundamentally change the qualitative behaviour of this measure. We study this question in the context of the classical Ising spin chain. In this system, the statistical complexity is known to grow monotonically with temperature. We evaluate the spin chain's quantum mechanical statistical complexity by explicitly constructing its provably simplest quantum model, and demonstrate that this measure exhibits drastically different behaviour: it rises to a maximum at some finite temperature then tends back towards zero for higher temperatures. This demonstrates how complexity, as captured by the amount of memory required to model a process, can exhibit radically different behaviour when quantum processing is allowed.

Statistical complexity emerges from the scientific ideal that we understand reality through cause and effect – the more precisely we can isolate the causes of natural things, the greater our comprehension. In this context, the more causes one must postulate to fully replicate the behaviour of a process, the more complicated that process appears. This motivates the statistical complexity as a measure of the process's intrinsic structure and complexity, since it represents the minimal amount of causal information one must record about past observations of a phenomena to model the statistics of observations made at future times [1].

Following these motivations, there has been significant work on developing a complete framework for statistical complexity. There are methods for evaluating the statistical complexity of general stationary stochastic processes [2, 3], and for estimating its value directly from observational statistics [4, 5, 6]. Meanwhile, the corresponding optimal models can be systematically constructed. If a process has a statistical complexity of C, we can systematically replicate the process's statistical behaviour using a model that records only C bits of information about the past. This analytic tractability, combined with a clear operational motivation, has propelled the use of statistical complexity within complexity science as a key measure of structure. Its field of study – computational mechanics – has been applied to analyse structure in diverse settings [7, 8, 9, 10], an early example being the Ising spin chain [11].

Conventional studies only consider building classical models. Yet, recent advances show that even when modelling the same classical process, a quantum model may require less input information [12, 13, 14, 15, 16]. This motivates an important question: could statistical complexity exhibit very different qualitative behaviour in the quantum regime? If true, this would imply that many existing studies in statistical complexity could draw very different conclusions when taking quantum information processing into account. Progress in this direction, however, has faced significant hurdles. While there are constructions for quantum models that use less memory than the classical limit, it is not known whether these quantum models are optimal. Thus the *quantum statistical complexity* has never been explicitly evaluated for any process where there is a quantum advantage, and existing quantum models provide only upper bounds [13].

Here, we present the first provably optimal quantum model for a process with a quantum advantage – a model of sequential measurement outcomes along a classical Ising spin chain. We determine the Ising chain's exact quantum statistical complexity. We compare this to existing studies of the spin chain's classical statistical complexity [11]. Whereas the classical statistical complexity monotonically increases with temperature, the quantum measure attains a

Whei Yeap Suen: wheiyeap@u.nus.edu
Mile Gu: mgu@quantumcomplexity.org

maximal value at some finite temperature, then decreases monotonically to zero. This demonstrates that the quantum and classical statistical complexities can diverge significantly in their qualitative behaviour. A system can become increasingly more difficult to model classically as certain physical parameters are varied, and yet simpler to model quantum mechanically.

Framework. Consider a stochastic system that at each discrete time $t \in \mathbb{Z}$ emits an output x_t from some configuration space $\mathcal{X}$. The statistical behaviour at each time t is characterised by a random variable X_t. At each time t, the resulting string of outputs may be partitioned into a *past* $\overleftarrow{X}_t := \dots X_{t-1}X_t$ comprised of outputs already emitted, and a *future* $\vec{X}_t := X_{t+1}X_{t+2}\dots$ consisting of outputs yet to come. The system's output behaviour can be captured by a stochastic process – a probability distribution over bi-infinite strings $\mathrm{P}(\overleftarrow{X}_t, \vec{X}_t)$ that describes how past observations correlate with future behaviour. Here we consider *stationary* processes, where $P(\vec{X}_t) = P(\vec{X}_0)$ for any $t \in \mathbb{Z}^+$, and so hereafter drop the subscript t from infinite strings.

Each instance of the process will have some specific past $\overleftarrow{x}$, with a corresponding conditional future distribution $\mathrm{P}(\vec{X} \,|\, \overleftarrow{X} = \overleftarrow{x})$. For any non-trivial process, $\mathrm{P}(\vec{X} \,|\, \overleftarrow{X} = \overleftarrow{x})$ depends on $\overleftarrow{x}$ – this reflects the ideal that prediction is meaningful only when the past is correlated with the future. A predictive model, then, is an algorithmic abstraction of $\mathrm{P}(\vec{X} \,|\, \overleftarrow{X})$ that exhibits statistically identical conditional future behaviour. Each model specifies an encoding function ϵ that maps each possible past $\overleftarrow{x}$ onto an internal state $s \in \mathcal{S}$ of some physical system, such that systematic actions on this system at future time-steps will generate a string of outputs obeying $\mathrm{P}(\vec{X} \,|\, \overleftarrow{X} = \overleftarrow{x})$.

The stochastic process can then be modeled by Markovian dynamics on $\mathcal{S}$. At each time-step t, a model in state s_t generates x according to $P(X_{t+1} = x|s_t) = P(X_{t+1} = x|\overleftarrow{x}_t)$, and updates its internal state to $s_{t+1} = \epsilon(\overleftarrow{x}')$, where $\overleftarrow{x}' = \dots x_{t-1}x_t x$ is the updated past. This implies that if we have two black boxes, one containing an instance of the original process with specific past $\overleftarrow{x}$, and another containing a machine initialized in state $s = \epsilon(\overleftarrow{x})$, then there will be no way to discriminate between them based on their observed output behaviour. In particular, $\mathrm{P}(\vec{X} \,|\, \overleftarrow{X} = \overleftarrow{x}) = \mathrm{P}(\vec{X} \,|\, S = s)$, where S is the random variable governing $s = \epsilon(\overleftarrow{x})$. The resulting models satisfy *unifilarity*, such that there is no uncertainty in S_{t+1} when given X_{t+1} and S_t – an important mathematical property that ensures a model's internal state contains no information about future outputs outside of what is available from the past [25]. The amount of information such a model tracks is then given by the Shannon entropy, $H(S) = -\sum_{s\in\mathcal{S}} p_s \log p_s$, where p_s is the probability $\epsilon(\overleftarrow{x}) = s$.

Clearly many different predictive models exist for each stochastic process. The brute force approach would be to take ϵ as the identity map, yielding a model that demands the entire past as input. While such a model will no doubt work, it is clearly inefficient. Take for example, the modelling of a completely random process. Here, as all pasts are equally likely, such a model would demand an unbounded amount of past information. On the other hand, if the process is completely random, the future does not depend on the past, and thus no past information should be require to correctly sample the process's conditional future.

Simplest Classical Models. Computational mechanics provides a framework for constructing the provably simplest predictive models of a process - the ones for which $H(S)$ is minimized [1, 3]. This optimality can be achieved by introducing the equivalence relation

$$\overleftarrow{x} \sim_\epsilon \overleftarrow{x}' \text{ iff } \mathrm{P}\left(\vec{X} \,|\, \overleftarrow{X} = \overleftarrow{x}\right) = \mathrm{P}\left(\vec{X} \,|\, \overleftarrow{X} = \overleftarrow{x}'\right). \quad (1)$$

Let $\mathcal{S} = \{s\}$ be the resulting set of equivalence classes on the set of all pasts, and the encoding function be defined such that $\epsilon(\overleftarrow{x}) = s$ iff $\overleftarrow{x} \in s$. The resulting map represents a model that stores not which past was observed, but rather only which equivalence class that past belongs in. Given $s = \epsilon(\overleftarrow{x})$, it is possible to systematically sample $\mathrm{P}(\vec{X} \,|\, \overleftarrow{X} = \overleftarrow{x})$. Specifically, consider a Markov finite state machine with state space $\mathcal{S}$. At each time-step, the machine's behaviour is dictated by transition elements $T^r_{ij} = \mathrm{P}(S_{t+1} = s_j, X_{t+1} = r \,|\, S_t = s_i)$: the probability that a machine initially in state s_i will transition to s_j whilst outputting $r \in \mathcal{X}$ (e.g. figure 1). This ensures that when initialized in the appropriate causal state $\epsilon(\overleftarrow{x})$, the machine can sequentially generate an arbitrarily long string $x_{t+1}x_{t+2}\dots$ with probability $\mathrm{P}(X_{t+1}X_{t+2}\dots \,|\, \overleftarrow{x})$.

In literature, such machines are referred to as ϵ-machines. Meanwhile, the states $s \in \mathcal{S}$ are named *causal states*, emphasizing that they capture all the causal information contained in $P(\overleftarrow{X}, \vec{X})$ [1, 3]. The amount of information required to store these causal states

$$C_\mu = -\sum_{s\in\mathcal{S}} p_s \log p_s \quad (2)$$

thus defines the *statistical complexity* of a stochastic process – representing the minimal amount of past information one needs about the process to replicate its conditional future statistics.

Spin Chains. We consider the 1D Ising system with nearest-neighbour interactions (as in [11]) of length $2N + 1$ with periodic boundary conditions. This system is described by the Hamiltonian [18]

$$\mathcal{H}_N = \sum_{k=-N\dots N} -Jx_k x_{k+1} - Bx_k, \quad (3)$$

where J is the coupling parameter, B is an external magnetic field, and x_k is the state of the spin at site

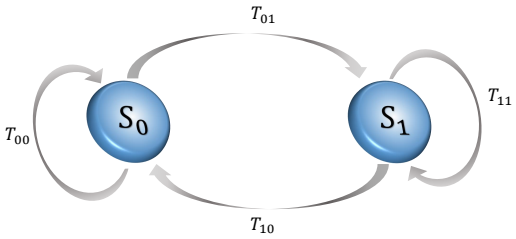


Figure 1: The ϵ-machine for the Ising chain consists of two causal states $s_0 = \{\overleftarrow{x}, x_0 = 1\}$, and $s_1 = \{\overleftarrow{x}, x_0 = -1\}$. For this particular process, the ϵ-machine has the special property such that $T_{ij}^{(r)} = 0$ for all $r \neq j$, and thus its dynamics can be entirely captured by transition probabilities $T_{ij} = T_{ij}^{j}$, the probability a machine in state s_i will transition to state s_j and output j, and coincides with the probability $P(X_{t+1} = j|X_t = i)$. The exact values for T_{ij} depend on B, J and T, and can be found in [17, 18].

k, which takes values $+1$ for spin up, -1 for spin down. At temperature T, the spin configurations at thermal equilibrium obey the Boltzmann distribution, such that

$$\mathrm{P}(X_{-N:N}) \propto \mathrm{e}^{-\mathcal{H}_N/k_B T}. \tag{4}$$

where $X_{-N:N} = X_{-N}, \ldots X_{N-1}, X_N$ and X_k is the random variable that governs x_k.

The statistical complexity of such Ising chains has been studied in the thermodynamic limit $N \to \infty$ [11]. In this framework we consider a device that scans the chain from left to right. Suppose at time t the device measures spin x_t. Let $\overleftarrow{X}_t$ govern the statistics of all spins measured so far, to the left of the device, and $\vec{X}_t$ all spins to the right of the device yet to be measured. The probability distribution $\mathrm{P}(\overleftarrow{X}_t, \vec{X}_t)$ then represents the total statistics of what the device observes. The resulting *statistical complexity* then captures the minimum memory requirement of the device which, given access to the state of spins $\overleftarrow{x}_t$ left of its current site, can accurately sample from the conditional distribution $\mathrm{P}(\vec{X}_t \,|\, \overleftarrow{X}_t = \overleftarrow{x}_t)$ to simulate spin statistics it has yet to see. As the spins in the chain are correlated, we expect $\mathrm{P}(\vec{X}_t \,|\, \overleftarrow{X}_t) \neq \mathrm{P}(\vec{X}_t)$, and thus the memory requirement to be strictly non-zero. In the thermodynamic limit, $\mathrm{P}(\overleftarrow{X}_t, \vec{X}_t)$ becomes a stationary stochastic process that is invariant with respect to t. Thus we can drop the index t and adopt the standard tools of computational mechanics.

The associated ϵ-machine can be systematically constructed. Here we summarize prior work [11]. A key observation is that $\mathrm{P}(\overleftarrow{X}, \vec{X})$ is Markovian. Furthermore, $\mathrm{P}(\vec{X} \,|\, x_0 = +1) \neq \mathrm{P}(\vec{X} \,|\, x_0 = -1)$ for any finite T. Thus the Ising system has exactly two causal states, s_0 and s_1, coinciding respectively with the set of pasts where the device last observed $+1$ or -1. Once initialized in the correct causal state, the resulting ϵ-machine can replicate the correct conditional future statistics according to figure 1.

Quantum Statistical Complexity. A number of approaches to reducing the entropy of classical models using quantum processing have been previously proposed [12, 14], but the question of whether such models are optimal has not yet been formally studied. To evaluate the quantum statistical complexity, we first need to define quantum models in general, and then define optimal quantum models.

Definition 1 (Quantum models)**.** *A quantum model is an ordered triple $\mathcal{Q} = (f, \Omega, \mathcal{M})$, where Ω is a set of quantum states, f is a encoding function that that maps each possible past $\overleftarrow{x}$ to a quantum state $|s_{\overleftarrow{x}}\rangle \in \Omega$, and $\mathcal{M}$ is a quantum procedure (that is, a completely-positive trace-preserving map) such that application of $\mathcal{M}$ on a physical system Ξ in state $f(\overleftarrow{x})$ should (i) generate an output x with probability $P(X = x|\overleftarrow{X} = \overleftarrow{x})$ and (ii) transition Ξ into state $f(\overleftarrow{x}')$, where $\overleftarrow{x}' = \overleftarrow{x}x = \ldots x_{-1}x_0x$ is the resulting past after x is observed.*

Condition (i) ensures that the quantum model generates statistically identical future predictions to its classical ε-machine counterpart. Meanwhile (ii) ensures that the model's internal state updates accordingly at each time-step such that it remains 'synchronized,' so that the internal state of machine at any point in time always correctly encodes the past. In this case a series of L repeated application of $\mathcal{M}$ acting on a physical system Ξ initially in state $f(\overleftarrow{x})$, will generate a random variable $X_{1:L}|\overleftarrow{x} := X_1 \ldots X_L|\overleftarrow{x}$ with probability distribution $P(X_{1:L}|\overleftarrow{X} = \overleftarrow{x})$, for any desired $L \in \mathbb{Z}^+$. In the limit $L \to \infty$ the quantum model will therefore replicate the process's future behaviour by generating a random variable $\vec{X}$ with probability distribution $P(\vec{X}|\overleftarrow{X} = \overleftarrow{x})$, whenever it is supplied with a system Ξ in state $f(\overleftarrow{x})$. This definition assumes encoded states to be pure to maintain unifilarity, in line with classical models. However, the results we shall present here continue to hold even if this assumption is dropped (see Lemma 3 in appendix).

The entropy $H(\rho)$ of the stationary mixture of encoded states ρ represents the amount of past information the model must store to correctly simulate the future. This motivates

$$C(\mathcal{Q}) = H(\rho) = -\mathrm{Tr}(\rho \log \rho), \tag{5}$$

as the *complexity* of a quantum model, where $\rho = \sum_i p_i \, |s_i\rangle \langle s_i|$ when $p_i = \mathrm{P}(f(\overleftarrow{x}) = |s_i\rangle)$ is the probability of the model being in the quantum state $|s_i\rangle$. An optimal quantum model is then the quantum model with the least complexity:

Definition 2 (Optimality)**.** *A quantum model $\mathcal{Q}$ of a process $\mathrm{P}(\overleftarrow{X}, \vec{X})$ is optimal for $\mathrm{P}(\overleftarrow{X}, \vec{X})$ if and only if for any other quantum model $\mathcal{Q}'$ of $\mathrm{P}(\overleftarrow{X}, \vec{X})$, $C(\mathcal{Q}) \leq C(\mathcal{Q}')$.*

Unlike the classical ϵ-machine, there may be more than one optimal quantum model. At the very least,

for any optimal quantum model $\mathcal{Q}_0$ with a set of states Ω_0, one can form another quantum model $\mathcal{Q}_0'$ with states Ω_0' trivially related to those in Ω_0 by a unitary transformation. Since the von Neumann entropy is invariant under unitary transformations, $\mathcal{Q}_0'$ will also be an optimal model.

Definition 3 (Quantum Causal States)**.** *If* $\mathcal{Q}_0 = (f_0, \Omega_0, \mathcal{M}_0)$ *is an optimal quantum model for a process* $\mathrm{P}(\overleftarrow{X}, \vec{X})$*, then* Ω_0 *are a set of* quantum causal states *for* $\mathrm{P}(\overleftarrow{X}, \vec{X})$.

This furnishes the necessary background for a definition of quantum statistical complexity.

Definition 4 (Quantum Statistical Complexity)**.** *Consider a stochastic process* $\mathrm{P}(\overleftarrow{X}, \vec{X})$ *with an optimal quantum model* $\mathcal{Q}_0$*. Let* $H(\rho)$ *be the amount of information stored by* $\mathcal{Q}_0$*. The quantum statistical complexity of* $\mathrm{P}(\overleftarrow{X}, \vec{X})$ *is defined as* $C_q = H(\rho)$.

Thus the quantum statistical complexity represents the minimal past information an optimal quantum model must record in order to accurately sample from $\mathrm{P}(\vec{X} \,|\, \overleftarrow{X} = \overleftarrow{x})$. Evaluating the quantum statistical complexity is generally non-trivial. While general methods of constructing better-than-classical quantum models are known [12, 14, 15], it remains an open question if and when such models are optimal. Thus to date, C_q has never been computed. We do not know in general exactly how much simpler a process can be when modelled quantum mechanically, nor if the quantum statistical complexity exhibits qualitatively different behaviour.

Here we aim to construct a provably optimal quantum model for the Ising spin chain. To do this, we first introduce two useful lemmas.

Lemma 1 (Causal state correspondence)**.** *For any stochastic process* $\mathrm{P}(\overleftarrow{X}, \vec{X})$*, there exists an optimal quantum model* $\mathcal{Q}_0 = \{f_0, \Omega_0, \mathcal{M}_0\}$ *of* $\mathrm{P}(\overleftarrow{X}, \vec{X})$ *such that for any two pasts* $\overleftarrow{x}$*,* $\overleftarrow{x}'$*,* $f_0(\overleftarrow{x}) = f_0(\overleftarrow{x}')$ *if and only if* $\epsilon(\overleftarrow{x}) = \epsilon(\overleftarrow{x}')$.

That is, there exists an optimal quantum model that has a one-to-one correspondence between the causal states and quantum causal states. This lemma implies there is never any benefit in differentiating two different pasts with coinciding conditional futures. The formal proof is based on the concavity of the von Neumann entropy, and is supplied in the technical appendix.

The second lemma places a constraint on how similar two quantum causal states can be, if we require the model to still be capable of generating correct future output statistics.

Lemma 2 (Maximum fidelity constraint)**.** *Consider a stochastic process* $\mathrm{P}(\overleftarrow{X}, \vec{X})$ *with a valid a quantum model* $\mathcal{Q} = \{f, \Omega, \mathcal{M}\}$*, where* $\Omega = \{|s_i\rangle\}_i$ *consists of quantum states in one-to-one correspondence with classical causal states* $\{s_i\}_i$*. Let* $\sigma_i = |s_i\rangle\langle s_i|$*. Then the following statement is true:* $F(\sigma_i, \sigma_j) \leq F^{ij}_{\max}$ *for all* i *and* j*, where* F *is the Fidelity and* $F^{ij}_{\max} = F\Big(\mathrm{P}(\vec{X} \,|\, s_i), \mathrm{P}(\vec{X} \,|\, s_j)\Big)$[1].

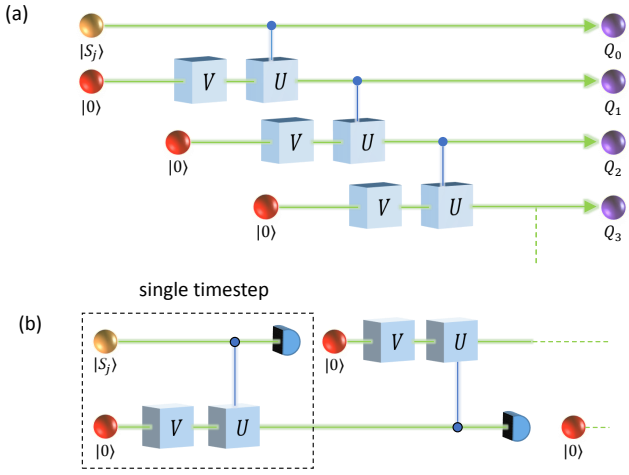


Figure 2: (a) Quantum circuit of how the quantum ϵ-machine can sample from the desired probability distribution $\mathrm{P}(\vec{X} \,|\, \overleftarrow{X} = \overleftarrow{x})$ when given a quantum causal state $|s_j\rangle = \epsilon_q(\overleftarrow{x})$. We introduce two unitaries: V that maps $|0\rangle$ to $|s_0\rangle$, and U such that $U|s_0\rangle = |s_1\rangle$. The machine is first supplied with an ancillary qubit that is systematically initialized in state $|s_0\rangle$ by application of V. Subsequently, it applies U on the ancillary qubit, controlled on the memory (system in state $|s_i\rangle$), where the control unitary transformation maps $|k\rangle|\phi\rangle \to |k\rangle U^k|\phi\rangle$ and $k \in \{0, 1\}$ indexes the computational basis. The original causal state is then emitted from the machine (denoted by Q_0). Repetition of this procedure yields a large entangled chain of qubits $Q_1, Q_2, \ldots$ that contains a superposition of all possible futures. Measurement of Q_k with respect to observable $\mathcal{Z} = |0\rangle\langle 0| - |1\rangle\langle 1|$ then yields the correct statistics for x_k. (b) Alternatively, we can view the above circuit as an iterative procedure. We can measure each as soon as it is emitted, which will collapse the corresponding ancillary qubit to the desired quantum causal state for the subsequent timestep.

The formal proof follows from the monotonicity of the fidelities under valid quantum operations, and is supplied in the technical appendix.

An optimal quantum model of the spin chain. We now apply these general results to the 1D Ising spin chain. To discuss the quantum statistical complexity, we must identify an optimal quantum model of the process. We shall proceed by presenting a model $\mathcal{Q}_0 = (\epsilon_q, \Omega_0, \mathcal{M}_0)$, based on the construction in [12]. We will later prove that this model is optimal for the Ising spin chain, and hence refer to $\mathcal{Q}_0$ as the *quantum ϵ-machine* for this process. This model uses

[1] We remind readers that the fidelity between two classical probability distributions is a special case of the fidelity between two quantum states, hence we use the same symbol for both cases.

two quantum states $\Omega_0 = \{|s_0\rangle, |s_1\rangle\}$ where

$$|s_0\rangle = \sqrt{T_{00}}\,|0\rangle + \sqrt{T_{01}}\,|1\rangle\,,$$
$$|s_1\rangle = \sqrt{T_{10}}\,|0\rangle + \sqrt{T_{11}}\,|1\rangle \qquad (6)$$

and T_{ij} is the transition probability for the classical ϵ-machine of the process in state s_i to output $(-1)^j$ and transition to s_j. The associated encoding map is $\epsilon_q : \bar{x} \to |\epsilon(\bar{x})\rangle$, where ϵ is the corresponding map from $\bar{x}$ to classical causal states. As shown explicitly in figure 2, there exists a quantum procedure $\mathcal{M}_0$ that allows this model to systematically sample from $\mathrm{P}(X_1 \ldots X_L \,|\, \overleftarrow{X} = \bar{x})$ for any L, by repeated applications of $\mathcal{M}_0$ acting on a physical system Ξ initialized in state $\epsilon_q(\bar{x})$. Thus the resulting triple $\mathcal{Q}_0 = (\epsilon_q, \Omega_0, \mathcal{M}_0)$ is indeed an accurate quantum model for the Ising spin chain.

This model has an internal entropy

$$H(\rho) = -\mathrm{Tr}\,(\rho \log \rho)\,, \qquad (7)$$

where

$$\rho = p_0\,|s_0\rangle\langle s_0| + p_1\,|s_1\rangle\langle s_1| \qquad (8)$$

such that p_0 (p_1) is the probability of any spin being up (down). Since $|s_0\rangle$ and $|s_1\rangle$ are generally not mutually orthogonal, we see immediately that this machine stores less information than any possible classical model. That is $C_q \leq H(\rho) < C_\mu$.

To establish that $H(\rho)$ in indeed the quantum statistically complexity (i.e. that $H(\rho) = C_q$), we need to establish that $\mathcal{Q}_0$ is an optimal quantum model for the Ising system.

Central Result. *The quantum ϵ-machine $\mathcal{Q}_0 = (\epsilon_q, \Omega_0, \mathcal{M}_0)$ is an optimal quantum model for the 1D Ising spin chain, and hence its internal entropy $H(\rho)$ corresponds to the quantum statistical complexity C_q of this system.*

The proof is rather involved, and we present the full details in the technical appendix. This proof relies on the Markovian nature of the 1D Ising system, and that it only has two casual states. The basic approach is to first note that it is sufficient to examine quantum models with two states (Lemma 1), and then directly evaluate the fidelity $F[\mathrm{P}(\vec{X} \,|\, s_0), \mathrm{P}(\vec{X} \,|\, s_1)]$ for the spin chain and show that it coincides with $F(|s_0\rangle\langle s_0|, |s_1\rangle\langle s_1|)$. Lemma 2 is then invoked together with the monotonic relationship between the fidelity of two[2] quantum states and the entropy of their statistical mixture to establish optimality of $\mathcal{Q}_0$.

This result formally establishes equation (7) as an analytical expression for C_q, allowing us to compare its qualitative behaviour with C_μ.

[2] This monotonic relationship does not hold for mixtures of more than two states [19]. This presents a significant challenge for generalizations of this proof to processes with more than two causal states.

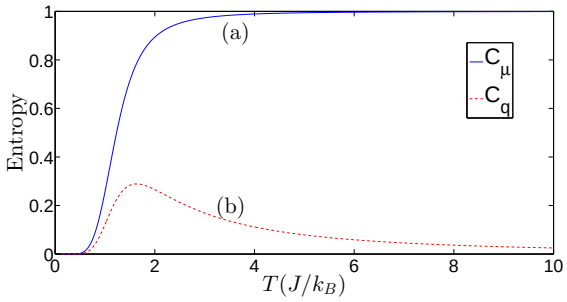


Figure 3: Plot of classical and quantum statistical complexity for various temperature of the Ising spin chain, with coupling parameter $J = 1$ and the magnetic field strength $B = 0.3$. Whereas the classical statistical complexity (a) monotonically increases with temperature, its quantum counterpart (b) attains maximal value for finite T.

Qualitative Divergences. The quantum statistical complexity and its classical counterpart exhibit very different qualitative behaviour (figure 3). The classical statistical complexity C_μ increases monotonically with temperature for all finite temperatures T. This reflects the fact that at hotter temperatures, all spin states become more equally likely. Thus the two classical causal states become progressively more equiprobable, resulting in a higher associated memory cost. Note that at infinite temperature, C_μ drops discontinuously to zero. This can be understood through the observation that at this limit, the spins observed on the left become completely uncorrelated with those on the right, and thus no memory is necessary to simulate the spin chain.

In contrast, the quantum statistical complexity C_q peaks at some finite temperature $T_{\max}$. Progressive *increases* in temperature past this point result in a decrease in the amount of quantum memory required – C_q decays smoothly to 0 as $T \to \infty$. We conclude that for $T > T_{\max}$, the quantum and classical analogues of statistical complexity diverge in qualitative behaviour. The Ising system takes progressively more memory to simulate classically with increasing temperature, this relation is reversed in quantum models for high temperatures. Indeed, as $T \to \infty$, the efficiency ratio C_μ/C_q becomes unbounded.

At this stage, we note an important caveat. The statistical complexity is an entropic quantity, and thus takes operational meaning in the independent identically-distributed limit. Specifically, suppose a processes has respective classical and quantum statistical complexity of C_μ and C_q. This implies that if we wish to replicate the futures of an ensemble of N such processes in the limit of large N, then we would either need NC_μ bits or NC_q qubits. However, if we are simulating only a single instance of a stochastic process, $C_q < C_\mu$ does not necessarily imply that the process can be simulated with less qubits than bits. Indeed, modelling a single instance of the spin chain would require either 1 bit or 1 qubit. It would therefore be

interesting to see if these qualitative divergences can exist using single shot measures, such as a max entropy of the input (known in complexity literature as the topological state complexity [1]).

Discussion. The statistical complexity – the amount of historical information we must retain about a system's past in order to simulate its future behaviour – is a popular operational quantifier of structure and complexity. Here we study how this measure can change when quantum processing is allowed. Using a 1D Ising system, we showed that the classical and quantum statistical complexity can diverge drastically in qualitative behaviour. Whereas classical models identify spin chains at very high temperatures as the most memory intensive to simulate and thus the most complex, quantum models identify these as being very simple. Our techniques also establish two necessary properties that any optimal quantum model must satisfy, which contribute towards computing quantum statistical complexity in more general settings.

More foundationally, these results suggest that complexity is an observer-dependent concept. Different observers may construe a system to be highly complex or extremely simple depending on whether they could reason quantum mechanically. This motivates a number of questions: how pervasive are such divergences and how do they impact our current understanding of what is complex? Certainly, the Ising model is used to model diverse phenomena in the physical and biological sciences [20, 21, 22, 23]. Meanwhile generalisations of statistical complexity have been used to study the dynamics of self-organization [24] and the structure of input-output processes [26], and it would be interesting to see how such results may change when viewed through the lens of quantum theory.

Note added. During the process of referee review, further work has highlighted the qualitative difference between C_q and C_μ leads to ambiguity about the simplest parameter regime for a particular system, and discussed generality of this phenomenon [28]. The unbounded efficiency ratio between C_q and C_μ has also since been shown in the context of progressively longer N-nearest neighbour Dyson-Ising systems [29] (in the limit of $C_q \to 0$, as in this article), as well as in simulating discretizations of continuous systems [30] and stochastic processes in continuous time [31] (where C_q is bounded, but C_μ grows unboundedly). Meanwhile, the differences in C_μ and C_μ have since being extended to cover input-output processes that react differently to different inputs [32].

Acknowledgements

We thank Felix Binder, Michele Dall'Arno, Alex Monras, Blake Pollard, Karoline Wiesner, and Benjamin Yadin for enlightening discussion. We are grateful for funding from the National Research Foundation of Singapore and in particular NRF Award No. NRF–NRFF2016–02 and NRF CRP Award No. NRF-CRP14-2014- 02; 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; the National Natural Science Foundation of China Grants 11450110058, 61033001 and 61361136003; the 1000 Talents Program of China; and the Oxford Martin School.

References

[1] J. P. Crutchfield and K. Young, Physical Review Letters **63**, 105 (1989).
[2] J. P. Crutchfield, Physica D: Nonlinear Phenomena **75**, 11 (1994).
[3] C. R. Shalizi and J. P. Crutchfield, Journal of statistical physics **104**, 817 (2001).
[4] C. R. Shalizi et al., Ph.D. thesis, University of Wisconsin–Madison (2001).
[5] C. R. Shalizi and K. L. Shalizi, in *Proceedings of the 20th conference on Uncertainty in artificial intelligence* (AUAI Press, 2004), pp. 504–511.
[6] D. Kelly, M. Dillingham, A. Hudson, and K. Wiesner, PloS one **7**, e29703 (2012).
[7] N. Perry and P.-M. Binder, Physical Review E **60**, 459 (1999).
[8] J. E. Hanson and J. P. Crutchfield, Physica D: Nonlinear Phenomena **103**, 169 (1997).
[9] W. Gonçalves, R. Pinto, J. Sartorelli, and M. De Oliveira, Physica A: Statistical Mechanics and its Applications **257**, 385 (1998).
[10] P. Tino and M. Koteles, IEEE Transactions on Neural Networks **10**, 284 (1999).
[11] J. Crutchfield and D. Feldman, Physical Review E **55**, 1239R (1997).
[12] M. Gu, K. Wiesner, E. Rieper, and V. Vedral, Nature communications **3**, 762 (2012).
[13] R. Tan, D. R. Terno, J. Thompson, V. Vedral, and M. Gu, The European Physical Journal Plus **129**, 1 (2014).
[14] J. Mahoney, C. Aghamohammadi, and J. Crutchfield, Scientific Reports **6** (2016).
[15] P. M. Riechers, J. R. Mahoney, C. Aghamohammadi, and J. P. Crutchfield, Phys. Rev. A **93** (2015).
[16] M. S. Palsson, M. Gu, J. Ho, H. M. Wiseman, and G. J. Pryde, Science Advances, **3.2**, e1601302 (2017).
[17] D. P. Feldman, Ph.D. thesis, University of California, Davis (1998).
[18] J. M. Yeomans, *Statistical mechanics of phase transitions* (Clarendon Press, 1992).

[19] R. Jozsa and J. Schlienz, Physical Review A **62**, 012301 (2000).
[20] E. Schneidman, M. J. Berry, R. Segev, and W. Bialek, Nature **440**, 1007 (2006).
[21] T. L. Hill, The Journal of Chemical Physics **30**, 383 (1959).
[22] B. Simon and R. B. Griffiths, Communications in Mathematical Physics **33**, 145 (1973).
[23] S. Torquato, Physical biology **8**, 015017 (2011).
[24] C. R. Shalizi, K. L. Shalizi, and R. Haslinger, Physical review letters **93**, 118701 (2004).
[25] J. P. Crutchfield, C. J. Ellison, R. G. James, and J. R. Mahoney, Chaos **30**, 3 (2010).
[26] N. Barnett and J. P. Crutchfield, Journal of Statistical Physics **161**, 404 (2015).
[27] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information* (Cambridge university press, 2010).
[28] C. Aghamohammadi, J. R. Mahoney and J. P. Crutchfield, Physics Letters A **381**, 1223–1227 (2017).
[29] C. Aghamohammadi, J. R. Mahoney and J. P. Crutchfield, arXiv:1609.03650v2
[30] A. J. P. Garner, Q. Liu, J. Thompson, V. Vedral and M. Gu, arXiv:1609.04408
[31] T.J. Elliott, M. Gu, arXiv:1704.04231
[32] J. Thompson, A. J. P. Garner, V. Vedral and M. Gu, npj Quantum Information **3**, 6 (2017).

TECHNICAL APPENDIX

The ϵ-machine for Ising spin chains. Let $\mathcal{P} = \mathrm{P}\left(\overleftarrow{X}, \vec{X}\right)$ denote the probability distribution of an Ising spin chain. Crutchfield and Feldman [11] show that

$$\mathrm{P}\left(\vec{X} \,|\, \overleftarrow{X} = \overleftarrow{x}\right) = \mathrm{P}\left(\vec{X} \,|\, X_0 = x_0\right), \tag{9}$$

that is, the probability of observing the right half spin configurations given the knowledge of the left half spin configuration, requires only the knowledge of x_0[17]. This implies that any two pasts $\overleftarrow{x} = \ldots, x_{-1}, x_0$, $\overleftarrow{x}' = \ldots, x'_{-1}, x'_0$ belong to the same causal state provided $x_0 = x'_0$. Thus, the ϵ-machine for this process has two causal states, $s_j = \{\overleftarrow{x} : x_0 = -1^j\}$, with $j = 0, 1$. The probability of finding the system in either causal state is thus given by the probability of finding any spin in the system to be up or down, denoted by p_0 and p_1 respectively.

As the causal states of the system are in one to one correspondence with x_0, $T^{(r)}_{ij}$ is non-zero only when $r = (-1)^j$. Thus, the notation for the ϵ-machine's dynamics can be simplified by writing $T^{(+1)}_{i0} = T_{i0}$ and $T^{(-1)}_{i1} = T_{i1}$. Furthermore the exact transition probabilities between causal states

$$T_{ij} = \mathrm{P}\big(X_{k+1} = (-1)^j \,|\, X_k = (-1)^i\big) \tag{10}$$

align with the probabilities governing the spin at location $k+1$ given the state of the spin at location k. The exact values of these transition probabilities vary as a function of temperature, and are well known [18].

Lemma 1 (Causal state correspondence)**.** *For any stochastic process* $\mathrm{P}(\overleftarrow{X}, \vec{X})$*, there exists an optimal quantum model* $\mathcal{Q}_0 = \{f_0, \Omega_0, \mathcal{M}_0\}$ *of* $\mathrm{P}(\overleftarrow{X}, \vec{X})$ *such that for any two pasts* $\overleftarrow{x}$*,* $\overleftarrow{x}'$*,* $f_0(\overleftarrow{x}) = f_0(\overleftarrow{x}')$ *if and only if* $\epsilon(\overleftarrow{x}) = \epsilon(\overleftarrow{x}')$*.*

Proof. We first prove the "only if" by contradiction. Consider two pasts $\overleftarrow{x}$ and $\overleftarrow{x}'$ where $\epsilon(\overleftarrow{x}) \neq \epsilon(\overleftarrow{x}')$. This implies $\mathrm{P}(\vec{X} \,|\, \overleftarrow{x}) \neq \mathrm{P}(\vec{X} \,|\, \overleftarrow{x}')$. For some quantum model $\mathcal{Q} = \{f, \Omega, \mathcal{M}\}$, let $|\phi_{\overleftarrow{x}}\rangle = f(\overleftarrow{x})$ and $|\phi_{\overleftarrow{x}'}\rangle = f(\overleftarrow{x}')$. Set density matrices $\sigma_{\overleftarrow{x}} = |\phi_{\overleftarrow{x}}\rangle\langle\phi_{\overleftarrow{x}}|$ and $\sigma_{\overleftarrow{x}'} = |\phi_{\overleftarrow{x}'}\rangle\langle\phi_{\overleftarrow{x}'}|$. Now if $\sigma_{\overleftarrow{x}} = \sigma_{\overleftarrow{x}'}$, then there cannot exist a systematic procedure $\mathcal{M}$ that can output different statistics on input of either $|\phi_{\overleftarrow{x}}\rangle$ or $|\phi_{\overleftarrow{x}'}\rangle$. Thus no quantum model can generate the correct conditional statistics for both $\overleftarrow{x}$ and $\overleftarrow{x}'$. Therefore for any quantum model that simulates $\mathcal{P}$, $\sigma_{\overleftarrow{x}} \neq \sigma_{\overleftarrow{x}'}$ whenever $\epsilon(\overleftarrow{x}) \neq \epsilon(\overleftarrow{x}')$.

The "if" direction will be demonstrated using concavity of entropy. We show that for an arbitrary quantum model $\mathcal{Q}'$, one can always find a quantum model $\mathcal{Q}$ with $C(\mathcal{Q}) \leq C(\mathcal{Q}')$ which satisfies $\sigma_{\overleftarrow{x}} = \sigma_{\overleftarrow{x}'}$ if $\epsilon(\overleftarrow{x}) = \epsilon(\overleftarrow{x}')$.

Consider the set of pasts in the same classical causal state s_i. Let $P_i = \mathrm{P}(\overleftarrow{x} \in s_i) > 0$ be the total probability in the stationary state that the past is in causal

state s_i. Let ρ be the stationary state of quantum model $\mathcal{Q}'$, which we divide into contributions from pasts in s_i and pasts not in s_i: $\rho = \sum_{\overleftarrow{x} \notin s_i} \mathrm{P}(\overleftarrow{x}) \sigma_{\overleftarrow{x}} + \sum_{\overleftarrow{x} \in s_i} \mathrm{P}(\overleftarrow{x}) \sigma_{\overleftarrow{x}}$, where $\mathrm{P}(\overleftarrow{x}) = \mathrm{P}(\overleftarrow{X} = \overleftarrow{x})$. By writing $\rho_{\neg i} = \frac{1}{1-P_i} \sum_{\overleftarrow{x} \notin s_i} p(\overleftarrow{x}) \, \sigma_{\overleftarrow{x}}$, and $q(\overleftarrow{x}) = p(\overleftarrow{x}) / P_i$, we write $\rho = (1-P_i) \, \rho_{\neg i} + P_i \sum_{\overleftarrow{x} \in s_i} q(\overleftarrow{x}) \, \sigma_{\overleftarrow{x}}$, and can then factor out the summation such that $\rho = \sum_{\overleftarrow{x} \in s_i} q(\overleftarrow{x}) \, [(1 - P_i) \, \rho_{\neg i} + P_i \sigma_{\overleftarrow{x}}]$.

From the concavity of entropy,

$$\begin{aligned} H(\rho) &\geq \sum_{\overleftarrow{x} \in s_i} q(\overleftarrow{x}) \, H((1 - P_i) \, \rho_{\neg i} + P_i \sigma_{\overleftarrow{x}}) \\ &\geq \min_{\overleftarrow{x} \in s_i} \{ H((1 - P_i) \, \rho_{\neg i} + P_i \sigma_{\overleftarrow{x}}) \}. \end{aligned} \tag{11}$$

Let $\overleftarrow{x}_m \in s_i$ be the past that minimizes this bound, with associated density operator $\tilde{\sigma} = \sigma_{\overleftarrow{x}_m}$. For every quantum model where $\overleftarrow{x} \in s_i$ are assigned to arbitrary $\sigma_{\overleftarrow{x}}$, there is also a valid quantum model that assigns the state $\tilde{\sigma}$ for all $\overleftarrow{x} \in s_i$, since for all pasts in s_i, repeatedly applying the measurement process $\mathcal{M}$ to $\tilde{\sigma}$ will give the desired future statistics. The stationary state of this model is $\tilde{\rho} = (1 - P_i) \, \rho_{\neg i} + P_i \tilde{\sigma}$, with an entropy of $H(\tilde{\rho}) = H((1 - P_i) \, \rho_{\neg i} + P_i \tilde{\sigma}) \leq H(\rho)$. It follows by considering every causal state s_i in turn, one can construct a quantum model with lower or equal entropy that maps all $\overleftarrow{x} \in s_i$ to the same quantum state σ_i.

Hence, for any arbitrary quantum model, there is always another valid quantum model with the same or lower entropy that satisfies $\sigma_{\overleftarrow{x}} = \sigma_{\overleftarrow{x}'}$ if and only if $\epsilon(\overleftarrow{x}) = \epsilon(\overleftarrow{x}')$. As such, we can restrict our search for an optimal model to those that satisfy the criterion of this lemma.

Finally, we must establish if there is a valid quantum model at all. This is guaranteed by the existence of a classical ϵ-machine, which is a special case of quantum models (with the causal states encoded onto orthogonal quantum states).

□

Lemma 2 (Maximum fidelity constraint)**.** *Consider a stochastic process* $\mathrm{P}(\overleftarrow{X}, \vec{X})$ *with a valid a quantum model* $\mathcal{Q} = \{f, \Omega, \mathcal{M}\}$*, where* $\Omega = \{|s_i\rangle\}_i$ *consists of quantum states in one-to-one correspondence with classical causal states* $\{s_i\}_i$*. Let* $\sigma_i = |s_i\rangle \langle s_i|$*. Then the following statement is true:* $F(\sigma_i, \sigma_j) \leq F_{\max}^{ij}$ *for all* i *and* j*, where* F *is the Fidelity and* $F_{\max}^{ij} = F\Big(\mathrm{P}(\vec{X} \,|\, s_i), \mathrm{P}(\vec{X} \,|\, s_j)\Big)$.

Proof. The proof of follows from the monotonicity of F under trace preserving quantum operation (see, e.g. chapter 9 of [27]). A measurement $\mathcal{M}$ that extracts statistics must induce a quantum operation $\sigma_i \mapsto \mathcal{R}(\sigma_i)$. Let $|\psi\rangle$ and $|\phi\rangle$ be purifications of σ_i and σ_j into a joint system AB such that $\sigma_i = \mathrm{tr}_B(|\psi\rangle \langle\psi|)$, $\sigma_j = \mathrm{tr}_B(|\phi\rangle \langle\phi|)$, and $F(\sigma_i, \sigma_j) = |\langle\psi|\phi\rangle|$. Next, we introduce an environment E for the quantum operation $\mathcal{R}$, such that $\mathcal{R}(\sigma_i) = \mathrm{tr}_E \left(U \left(\sigma_i \otimes |0\rangle \langle 0| \right) U^\dagger \right)$. We then note that $\mathcal{R}(\sigma_i) = \mathrm{tr}_{BE} \left(U \left(|\psi\rangle \langle\psi| \otimes |0\rangle \langle 0| \right) U^\dagger \right)$. This also applies to $\mathcal{R}(\sigma_j)$. By Uhlmann's theorem [27],

$$\begin{aligned} F\left(\mathcal{R}(\sigma_i), \mathcal{R}(\sigma_j)\right) &\geq |\langle\psi| \langle 0| U^\dagger U |\phi\rangle |0\rangle| && (12) \\ &= |\langle\psi|\phi\rangle| && (13) \\ &= F(\sigma_i, \sigma_j). && (14) \end{aligned}$$

Hence, there is a maximum fidelity $F_{\max}$ between the quantum states σ_i and σ_j set by $F_{\max} = F\Big(\mathrm{P}(\vec{X} \,|\, s_i), \mathrm{P}(\vec{X} \,|\, s_j)\Big)$. If this bound is violated, $\mathcal{Q}$ is not a valid quantum model. □

Optimality of the Quantum ϵ-Machine. We prove the following theorem:

Central Result. *Let* $\mathcal{P} = \mathrm{P}(\vec{X}, \overleftarrow{X})$ *denote the stochastic process that governs that Ising system, and* s_0 *and* s_1 *be the causal states of this process, with transition probabilities* T_{ij} *in the corresponding* ϵ*-machine. Then there is an optimal quantum model* $\mathcal{Q}_0 = \{\epsilon_q, \Omega_0, \mathcal{M}_0\}$ *where* Ω_0 *consists of the pure quantum states* $|s_0\rangle$ *and* $|s_0\rangle$*, given by*

$$\begin{aligned} |s_0\rangle &= \sqrt{T_{00}} \, |0\rangle + \sqrt{T_{01}} \, |1\rangle \,, \\ |s_1\rangle &= \sqrt{T_{10}} \, |0\rangle + \sqrt{T_{11}} \, |1\rangle \,. \end{aligned} \tag{15}$$

Proof. From Lemma 1, we know that there is an optimal quantum model with two quantum causal states. Let them be $|\psi_0\rangle$ and $|\psi_1\rangle$. The complexity of this model monotonically decreases with increasing fidelity $F(|\psi_0\rangle\langle\psi_0|, |\psi_1\rangle\langle\psi_1|)$ between the two causal states. From Lemma 2, we know that $F(|\psi_0\rangle\langle\psi_0|, |\psi_1\rangle\langle\psi_1|) \leq F_{\max} = F\Big(\mathrm{P}(\vec{X} \,|\, s_0), \mathrm{P}(\vec{X} \,|\, s_1)\Big)$ for any valid quantum model. Therefore, if there is valid quantum model such that the fidelity $F(|\psi_0\rangle\langle\psi_0|, |\psi_1\rangle\langle\psi_1|) = F_{\max}$, then this model is optimal.

The fidelity between the two distributions $\mathrm{P}(\vec{X} \,|\, s_0)$ and $\mathrm{P}(\vec{X} \,|\, s_1)$ is given by

$$\begin{aligned} &F(\mathrm{P}(\vec{x} \,|\, s_0) \,, \mathrm{P}(\vec{x} \,|\, s_1)) \\ &\quad = \sum_{\vec{x}_i} \sqrt{\mathrm{P}\Big(\vec{X} = \vec{x}_i \,|\, s_0\Big) \, \mathrm{P}\Big(\vec{X} = \vec{x}_i \,|\, s_1\Big)} \\ &\quad = \sum_{x_0, x_1, \ldots} \sqrt{\mathrm{P}(x_0, x_1, \ldots \,|\, s_0)} \sqrt{\mathrm{P}(x_0, x_1, \ldots \,|\, s_1)}, \end{aligned} \tag{16}$$

where x_k represents the processes' output at time $t = k$. The Ising system is a Markovian process where $T_{ij}^{(r)} = T_{ij}^{(r)} \delta_{j \frac{1-r}{2}}$, allowing us to express the transition probabilities as T_{ij}. Under these circumstances we can invoke the chain rule of probability and re-express:

$$F(\mathrm{P}(\vec{x} \,|\, s_0) \,, \mathrm{P}(\vec{x} \,|\, s_1))$$

$$
\begin{aligned}
&= \sum_{x_1,x_2,\ldots} \sqrt{\mathrm{P}(x_1 \,|\, s_0)\,\mathrm{P}(x_2 \,|\, x_1)\ldots} \\
&\qquad\qquad \times \sqrt{\mathrm{P}(x_1 \,|\, s_1)\,\mathrm{P}(x_2 \,|\, x_1)\ldots} \\
&= \sum_{x_1} \sqrt{\mathrm{P}(x_1 \,|\, s_0)\,\mathrm{P}(x_1 \,|\, s_1)} \sum_{x_2,x_3,\ldots} \mathrm{P}(x_2, x_3 \ldots \,|\, x_1) \\
&= \sum_{x_1} \sqrt{\mathrm{P}(x_1 \,|\, s_0)\,\mathrm{P}(x_1 \,|\, s_1)} \\
&= \sqrt{T_{00}T_{10}} + \sqrt{T_{01}T_{11}}. \qquad (17)
\end{aligned}
$$

It can be easily verified that for the quantum states $|s_0\rangle$ and $|s_1\rangle$ presented in equation (15), $F(|s_0\rangle\langle s_0|, |s_1\rangle\langle s_1|) = \sqrt{T_{00}T_{10}} + \sqrt{T_{01}T_{11}}$, and hence these states saturate the maximal possible fidelity for $\mathcal{P}$.

As such, the only remaining step is to demonstrate the existence of an $\mathcal{M}$ that extracts the necessary statistics. This is presented constructively by the quantum circuit in figure 2 of the text. Hence, $\mathcal{Q}_0$ is a valid quantum model, and moreover, since it consists of two pure states that saturate the maximum fidelity, it is an optimal quantum model. □

Since we have found an optimal quantum model $\mathcal{Q}_0$ for the Ising spin chain, this demonstrates that the spin chain's quantum statistical complexity $C_q = C(\mathcal{Q}_0)$. We remark that there is in fact a family of optimal models. Fixing the first quantum causal state $|\sigma_0\rangle$ as $|0\rangle$ without loss of generality (by symmetry), we can write the second quantum causal state $|\sigma_1\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle + e^{i\phi}\sin\left(\frac{\theta}{2}\right)|1\rangle$, where $\theta \in (0, \pi]$ is determined by the maximum fidelity, but we are free to choose any $\phi \in [0, 2\pi)$ and any $|1\rangle$ such that $\langle 0|1\rangle = 0$. All such assignments of $|\sigma_0\rangle$ and $|\sigma_1\rangle$ are all related by unitary transformations, so if there is a measurement $\mathcal{M}$ that extracts the statistics from one such model, a measurement can be found for all such models. Moreover, since the von Neumann entropy is basis-independent, all such choices of model have the same complexity. It thus follows that every assignment of pure states separated by the minimum distance are optimal.

We also establish that the Main Theorem continues to hold when we consider encoding $\overleftarrow{x}$ into mixed quantum states, by the following lemma:

Lemma 3. *Consider a more general quantum model $\tilde{\mathcal{Q}} = \{\tilde{f}, \tilde{\Omega}, \tilde{\mathcal{M}}\}$, where the encoding map $\tilde{f}$ takes each $\overleftarrow{x}$ to some possibly mixed state $\sigma_{\overleftarrow{x}} \in \tilde{\Omega}$. (Clearly, the set of all $\mathcal{Q}$ forms a subset of the set of $\tilde{\mathcal{Q}}$.) Let $P(\overleftarrow{X}, \vec{X})$ denote a Markovian stochastic process with two causal states s_0 and s_1, and two possible outputs. If there exists a quantum model $\mathcal{Q}_0 = \{f_0, \Omega_0, \mathcal{M}_0\}$ where Ω_0 consists of two pure quantum states σ_0 and σ_1 (in one-to-one correspondence with classical causal states), saturating the maximum fidelity $F_{\max} = F\Big(\mathrm{P}(\vec{X} \,|\, s_0), \mathrm{P}(\vec{X} \,|\, s_1)\Big)$, then this $\mathcal{Q}_0$ is optimal.*

Proof. It has already been established by the Main Result that such a $\mathcal{Q}_0$ is optimal amongst pure state quantum models. We shall show that there are no alternative models with a lower entropy than $\mathcal{Q}_0$, even when we allow for encoding onto mixed quantum states.

First, note that the argument of Lemma 1 trivially generalizes to mixed states, we may restrict our search within models with two internal states.

Let $\mathcal{Q}'$ be a quantum model with $\Omega' = \{\tilde{\sigma_0}, \tilde{\sigma_1}\}$, such that at least one state (say, $\tilde{\sigma_1}$) is mixed. As $\mathcal{Q}'$ is a valid quantum model for $P(\overleftarrow{X}, \vec{X})$, there is a measurement process that yields outcome $x_1 \in \{-1, 1\}$ with probability $P(X_1 = x_1 | S_0 = s_i)$ when acting on $\tilde{\sigma_i}$. Let this measurement have elements $\{M, \mathbb{I} - M\}$, such that $\mathrm{Tr}\,(M\tilde{\sigma_0}) = T^{(1)}_{0j}$, $\mathrm{Tr}\,(M\tilde{\sigma_1}) = T^{(1)}_{1j}$, while $\mathrm{Tr}\,((\mathbb{I} - M)\tilde{\sigma_0}) = T^{(-1)}_{0j}$, $\mathrm{Tr}\,((\mathbb{I} - M)\tilde{\sigma_1}) = T^{(-1)}_{1j}$.

The convexity of quantum states allows us to express $\tilde{\sigma_1}$ as a convex combination of pure states $\tilde{\sigma_1} = \sum_i \lambda_i |\psi_i\rangle\langle\psi_i|$. Moreover, the subset of quantum states ρ satisfying the linear constraint $\mathrm{Tr}\,(M\rho) = k$ is also a convex set. Thus, we can decompose σ_1 into some $\{\lambda_i, |\psi_i\rangle\}_i$ such that $\mathrm{Tr}\,(M|\psi_i\rangle\langle\psi_i|) = \mathrm{Tr}\,(M\tilde{\sigma_1})$ for each i individually.

Letting p_0 and p_1 be the probabilities of classical causal states s_0 and s_1 respectively, we write the stationary state as $\tilde{\rho} = p_0\tilde{\sigma}_0 + p_1\tilde{\sigma}_1$, and explicitly factorise:

$$
\begin{aligned}
\tilde{\rho} &= p_0\tilde{\sigma}_0 + p_1 \sum_i \lambda_i |\psi_i\rangle\langle\psi_i| \\
&= \sum_i \lambda_i \left(p_0\tilde{\sigma}_0 + p_1 |\psi_i\rangle\langle\psi_i|\right). \qquad (18)
\end{aligned}
$$

Hence, from concavity of entropy,

$$
\begin{aligned}
H(\tilde{\rho}) &\geq \sum_i \lambda_i H(p_0\tilde{\sigma}_0 + p_1|\psi_i\rangle\langle\psi_i|) \\
&\geq \min_i \{H(p_0\tilde{\sigma}_0 + p_1|\psi_i\rangle\langle\psi_i|)\}. \qquad (19)
\end{aligned}
$$

Let $|\psi_{\min}\rangle$ be the choice of state that minimizes the above expression. Any model that encodes onto the pair of states $\tilde{\sigma}_0$ and $|\psi_{\min}\rangle$ is guaranteed to have complexity equal or lower than $C(\mathcal{Q}')$. Moreover since $\{M, \mathbb{I} - M\}$ acting on $\tilde{\sigma}_0$ and $|\psi_{\min}\rangle$ produces output statistics with fidelity $F_{\max} = F(P(\vec{X}|s_0), P(\vec{X}|s_1))$, by monotonicity of fidelity under contractive maps we have $F(\tilde{\sigma}_0, |\psi_{\min}\rangle\langle\psi_{\min}|) \leq F_{max}$.

Exactly the same argument can be made to find some substitution $|\phi_{\min}\rangle$ for $\tilde{\sigma}_0$, generating a new encoding which is guaranteed to satisfy the maximum fidelity, and to result in equal or lower complexity. This construction shows that no valid mixed state encoding can exceed all pure states encodings. Hence, if we can find a pure state model such that the two internal states saturate the fidelity bound, then it will have lower entropy than any mixed state model. □