UC Davis
UC Davis Previously Published Works

Title
Thermal Efficiency of Quantum Memory Compression

Permalink

Journal

ISSN

Authors
Loomis, Samuel P
Crutchfield, James P

Publication Date

DOI

Thermal Efficiency of Quantum Memory Compression

Samuel P. Loomis[*] and James P. Crutchfield[†]
Complexity Sciences Center and Physics Department,
University of California at Davis, One Shields Avenue, Davis, CA 95616

(Dated: March 24, 2020)

Quantum coherence allows for reduced-memory simulators of classical processes. Using recent results in single-shot quantum thermodynamics, we derive a minimal work cost rate for quantum simulators that is quasistatically attainable in the limit of asymptotically-infinite parallel simulation. Comparing this cost with the classical regime reveals that quantizing classical simulators not only results in memory compression but also in reduced dissipation. We explore this advantage across a suite of representative examples.

PACS numbers: 05.45.-a 89.75.Kd 89.70.+c 05.45.Tp

LIST OF CORRECTIONS

This Letter demonstrates the potential for a quantum machine to perform a particular task—namely, simulating hidden Markov models—using not only less memory storage, but also reduced thermodynamic dissipation. To do so it synthesizes recent developments in quantum memory compression [1–10] and quantum thermodynamics [11–22] into a fruitful new crossover framework for studying the thermodynamics of quantum simulators for stochastic processes.

We begin by reviewing *computational mechanics*, which seeks to analyze how natural systems manipulate information to produce and transform stochastic processes [23–27]. More recently, computational mechanics examined the thermodynamics of computation, generalizing Landauer's principle for memory erasure [28] to derive the *information processing Second Law* [29], which gives the minimal cost of transforming one stochastic process in to another, and the *thermodynamics of modularity* [30], which determines the implementation costs for transformations. An important result in this classical regime is that the Shannon-entropy Landauer bound on average work cost for a computation can be achieved for a single implementation of that computation [30]. That is, the (quasistatic) single-shot cost of a computation is the same as the cost for asymptotically-infinite parallel implementations.

We next consider the relationship between memory and thermodynamics in the quantum regime. Quantum computational mechanics recently explored how to simulate and transform *classical* stochastic processes using *quantum* systems [1, 2, 5, 6], even constructing experimental implementations [3, 4]. Generally, quantum simulators of complex processes require less memory (measured by the quantum-state von Neumann entropy) than classical (measured by the statistical complexity—classical-state Shannon entropy) [7, 8].

Quantum thermodynamics [12], though recently advancing via thermal resource theories [11, 13–15, 17] and single-shot thermodynamics [16, 19–21], has not yet been applied to examine quantum simulators. However, it is known that Landauer's lower bound, as given in the form of Shannon and von Neumann entropies, is not generally attainable. A more nuanced view is necessary [16, 19]. Due to this, single-shot and asymptotic analyses must be performed separately when transitioning from classical to quantum regimes.

Using these recent results in quantum thermodynamics, we calculate achievable and lower bounds on the work cost rate for the quasistatic implementation of quantum simulators, in both the single-shot and asymptotic regimes. These bounds reveal a direct relationship between memory compression achieved by a quantum implementation and the increase in extractable work via the same. We then elucidate the nature of this trade-off across a suite of examples.

Ratchets, generators, and processes. By *computation* we mean transducing an input sequence $\ldots x_0 x_1 x_2 \ldots$ into a new output sequence $\ldots y_0 y_1 y_2 \ldots$. When a computation is done *online* the implementation acts on a single input symbol at a time and immediately determines the output symbol based on *stored memory* of the past inputs and outputs. The graphical representation of such a transducer, see Fig. 1, suggests calling them *ratchets*, as done previously [29].

The following restricts itself to *finite generators* that take only a trivial (constant) input, produce as output a stochastic process taking values in a finite alphabet $\mathcal{X}$, and employ a finite set $\mathcal{S}$ of memory states. This is sufficient to explore the central relationships between memory and thermodynamics.

A finite generator's operation is described by the probabilities $\Pr(s', x | s)$ of emitting symbol x and ending in a
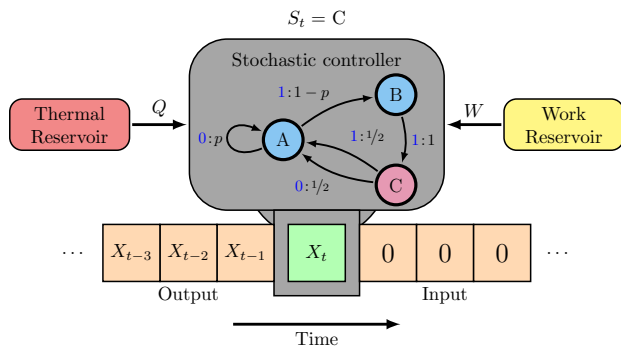
FIG. 1. Information ratchet sequentially generates a symbol string on an empty tape: At time step t, S_t is the random variable for the ratchet state. The generated symbols in the output process are denoted by $X_{t-1}, X_{t-2}, X_{t-3}, \ldots$. The most recently generated symbol X_t (green) is determined by the internal dynamics of the ratchet's memory, using heat Q from the thermal reservoir as well as work W from the work reservoir. *(Ratchet interior.)* The memory dynamics and symbol production are governed by the conditional probabilities $\Pr(s_{t+1}, x_t|s_t)$, where s_t is the current state at time t, x_t is the generated symbol and s_{t+1} is the new state. Diagrammatically, this is a hidden Markov model—a labeled, directed graph in which nodes are states s and edges represent transitions $s \to s'$ labeled by the emitted symbol and associated probability $x : \Pr(s', x|s)$.

final memory state s', if starting in memory state s. This is depicted as the *hidden Markov model* (HMM) [31] in the ratchet interior (Fig. 1).

An HMM is *irreducible* if the matrix $\Pr(s'|s) := \sum_x \Pr(s', x|s)$ is irreducible. Functionally, this means that every state s' can be reached from any other state s in the state transition diagram.

An irreducible HMM has a unique *stationary distribution* $\Pr_0(s)$ over its memory states such that $\sum_s \Pr(s'|s) \Pr_0(s) = \Pr_0(s')$. We let $S'XS \sim \Pr(s', x|s) \Pr_0(s)$ represent the joint random variable of the generated symbol and the memory state before and after generation. Over many time steps, we accrue the joint random variable $S_{t+1}X_t \ldots S_2X_1S_1 \sim \Pr(s_{t+1}, x_t|s_t) \ldots \Pr(s_2, x_1|s_1) \Pr_0(s_1)$. The sequence of random variables $X_t \ldots X_1$, typically not independent of each other, describes t samples of the *stochastic process* simulated by the generator.

Most results on generators and their quantum counterparts encompass those with at least the two following properties:

1. *Predictivity*: $s' = f(x, s)$. The next state is always determined by the previous state and the generated symbol: $\Pr(s', x|s) \propto \delta_{s', f(x,s)}$ for some f. (Elsewhere known as *unifilarity* [32].)

2. *Minimality*: For any two states s and s', if $\Pr(x_t \ldots x_1|s) = \Pr(x_t \ldots x_1|s')$ for all t, then $s = s'$. This ensures that no two states predict the same future distributions. A nonminimal HMM is minimized by merging predictively equivalent states.

For any process, there is a unique generator satisfying these two properties, called the ϵ-*machine* [26, 27].

Given an ϵ-machine, from its stationary distribution we can calculate its *statistical complexity* (*i.e.*, memory) $C_\mu := \mathrm{H}[S]$ [23] and *non-Markovity* $N_\mu := \mathrm{H}[S'|X]$, using the Shannon entropy and conditional entropy, respectively [33].

Physical implementations. We define a generator's *implementation* as the sextuplet $(\mathcal{H}_S, \mathcal{H}_X, \mathcal{H}_A, \mathcal{H}_B, U, \mathcal{E})$ consisting of four Hilbert spaces (*memory, output, auxiliary, bath* systems); a unitary U on all four; and an ensemble $\{|\psi_s\rangle : s \in \mathcal{S}\}$ embedding the classical memory states into the memory system $\mathcal{H}_S$, respectively. The auxiliary system starts in a given pure state $|0\rangle_A$, while the bath is taken to start in a thermal state. Following convention for information reservoirs, we consider the memory, output, and auxiliary systems to be energyless, though the bath system may have some nontrivial Hamiltonian H_B. Under these conditions, we require $[U, H_B] = 0$, following the rules for microscopic energy conservation [11]. Furthermore, as long as we begin and end in an information reservoir, we may assume our operations are performed via Hamiltonian control, with minimal work costs defined by the state-averaged changes in energy level over a quasistatic erasure of system A [12, 18, 20].

When the auxiliary and the bath are traced out, the implementation must take the form of the positive map:

$$\mathcal{T}(|\psi_s\rangle \langle \psi_s| \otimes |0\rangle \langle 0|) = \sum_{x,s'} \Pr(s', x|s) |\psi_{s'}\rangle \langle \psi_{s'}| \otimes |x\rangle \langle x|.$$

Resetting the thermal bath has no associated cost—it may simply be brought into contact with a larger bath. However, if an auxiliary system is used, its reset (erasure) cost must be taken into account.

Classical implementation. Two concrete types of implementation have been considered previously. The first addresses efficiently implementing the generator via classical thermodynamics. Using Hamiltonian control, Ref. [30] showed that *any* stochastic channel can be implemented in a way that achieves the Landauer bound. In particular, applying a channel $\Pr(y|x)$ to a random variable X, resulting in Y, can be performed with the

work cost $W = k_\mathrm{B} T \left(\mathrm{H}\left[X\right] - \mathrm{H}\left[Y\right] \right)$. For generators this means we can achieve the work cost per time step of:

$$W_\mu = k_\mathrm{B} T \ln 2 \left(\mathrm{H}\left[S\right] - \mathrm{H}\left[S'X\right] \right) \ . \qquad (1)$$

Written differently:

$$W_\mu = k_\mathrm{B} T \ln 2 \left(C_\mu - N_\mu - \mathrm{H}\left[X\right] \right) \ . \qquad (2)$$

This is, in a sense, maximally efficient: The work can never be made lower than Landauer's bound, so this is the best we can possibly do over all classical implementations. Noting that $\mathrm{H}\left[S'\right] = \mathrm{H}\left[S\right]$, we find $W_\mu = -k_\mathrm{B} T \ln 2 \ \mathrm{H}\left[X|S'\right] \leq 0$, so predictive generators may extract positive work.

Much of the cost W_μ is due to the local nature of the generator: it does not have access to previously generated symbols to choose its operations. The information processing Second Law (IPSL) [29], when applied to generators, states that the work cost per symbol for generating a process *with* access to the previous symbols is bounded by:

$$W \geq -k_\mathrm{B} T \ln 2 \ h_\mu \ ,$$

where $h_\mu = \lim_{t\to\infty} \frac{1}{t} \mathrm{H}\left[X_1 \ldots X_t\right]$ is the *entropy rate* of the process being generated. Generally, $W_\mu \geq -k_\mathrm{B} T \ln 2 \ h_\mu$ [30].

Classical generators require memory, quantified in the case of the ϵ-machine by the statistical complexity C_μ. Memory cost can be reduced by embedding the memory states into a nonorthogonal quantum ensemble. This motivates the use of quantum implementations of generators.

Quantum implementations. In a quantum implementation, we apply a unitary operator to SXA alone. This unitary is divided into two parts: $U_{SXA} = \left(1_S \otimes U_{XA}\right)\left(U_{SX} \otimes 1_A\right)$. The first operation U_{SX} evolves the memory system and the output system to achieve the necessary correlation. While the second U_{XA} entangles the output and the auxiliary to represent the effect of a measurement device. The first takes the form:

$$U_{SX} \left|\psi_s\right\rangle \left|0\right\rangle = \sum_{x,s'} e^{i\phi_{xs}} \sqrt{\mathrm{Pr}(x|s)} \left|\psi_{f(s,x)}\right\rangle \left|x\right\rangle \ ,$$

where $\left|x\right\rangle$ form an orthogonal computational basis representing the generated symbols. When the generator is predictive, a unitary performing this transformation exists for any choice of the arbitrary phases ϕ_{xs} [10].

For a quantum implementation of an ϵ-machine, we can measure its memory cost by the quantum complexity

$C_q := \mathrm{H}_q\left[S\right]$ and its quantum non-Markovity as $N_q := \mathrm{H}_q\left[S'|X\right]$, where $H_q(S)$ is the von Neumann entropy of the stationary state ρ_S on system S and $\mathrm{H}_q\left[S'|X\right]$ is the conditional entropy of the state ρ'_{SX} after implementing U_{SX}. These states have the form:

$$\rho_S = \sum_s \mathrm{Pr}_0(s) \left|\psi_s\right\rangle \left\langle\psi_s\right| \ \text{and}$$

$$\rho'_{SX} = \sum_{s',x,s} \mathrm{Pr}(s',x|s)\, \mathrm{Pr}_0(s) \left|\psi_s\right\rangle \left\langle\psi_s\right| \otimes \left|x\right\rangle \left\langle x\right| \ .$$

For any ϵ-machine quantum implementation, we have $C_q \leq C_\mu$, with strict equality only when the ϵ-machine is retrodictive [2, 9].

In general, the single-shot case—that implements a single copy of a generator—cannot achieve Landauer's bound. Synthesizing several results in quantum erasure and information processing [19–21], the Supplementary Material [34] (SM) derives our first main result: a single generator can be implemented with a work cost of no more than:

$$\begin{aligned}
\frac{W_q^\epsilon}{k_\mathrm{B} T \ln 2} \leq{} & \mathrm{H}_{\max}^{\epsilon^2/4}\left[S\right] - \mathrm{H}_{\min}^{\epsilon^2/64}\left[S'|X\right] \\
& - \mathrm{H}_{\min}^{\epsilon^2/64}\left[X\right] + O\left(\log \frac{1}{\epsilon}\right) ,
\end{aligned} \qquad (3)$$

with a probability of failure less than ϵ. Rather than use the Shannon entropies of the classical work, this is expressed in the smooth conditional entropies of quantum information theory [35–39] as applied to the states ρ_S, ρ'_{SX}, and $\rho'_X = \mathrm{Tr}_S\left(\rho'_{SX}\right)$.

Suppose, instead of implementing a single copy of a generator, we implement N generators in parallel, each producing an independent realization of the desired process. The *asymptotic equipartition property* of smooth entropies [37] then shows that the work rate $W_q := \lim_{\epsilon\to 0} \lim_{N\to\infty} W_q^\epsilon/N$ is given by:

$$\begin{aligned}
\frac{W_q}{k_\mathrm{B} T \ln 2} &= \mathrm{H}_q\left[S\right] - \mathrm{H}_q\left[S'|X\right] - \mathrm{H}\left[X\right] \\
&= C_q - N_q - \mathrm{H}\left[X\right]
\end{aligned} \qquad (4)$$

where $\mathrm{H}_q\left[\cdot\right]$ is the von Neumann entropy [40]. C_q, N_q, and W_q are functions of the quantum implementation chosen; in particular, $W_q = W_q(\phi_{xs})$ is a function of the phases. The SM shows that this is always at least as small as the classical cost: $W_q(\phi_{xs}) \leq W_\mu$. Combining this with the IPSL and W_μ's negativity gives:

$$-k_\mathrm{B} T \ln 2 \ h_\mu \leq W_q(\phi_{xs}) \leq W_\mu \leq 0 \ , \qquad (5)$$

for all $\{\phi_{xs}\}$. Thus, the quantum implementation of a

predictive generator offers improvement over the classical implementation in the work that can be thermodynamically extracted.

Equations (3) to (5) are our three primary results. In the remainder, we explore in a suite of example generators the relationship between the *memory compression* $\Delta_q C := C_\mu - C_q$ and the *work advantage* $\Delta_q W := W_\mu - W_q$. (The suite covers the qualitatively distinct behaviors observed in our numerical exploration.) We find that the *efficiency of compression* $e_q = \Delta_q W / (k_B T \Delta_q C \ln 2)$—the improvement in work cost for each bit of compression achieved—is a key quantity for monitoring the behavior of quantum implementations.

Markov Generators. A Markov chain $X_1 \ldots X_t$ is a chain of random variables X_t, where each variable is conditionally independent of the past given its predecessor: $\Pr(x_t | x_{t-1} \ldots x_0) = \Pr(x_t | x_{t-1})$ for all t. In a sense, a Markov chain is its own generator—one in which memory states are also the produced symbols: $\mathcal{S} = \mathcal{X}$. For Markov generators, which type has historically dominated physical modeling, knowing the produced symbol X_t automatically determines the next state S_{t+1}, as they are identical. Their non-Markovity $N_\mu = 0$ (hence the name of that quantity) and $C_\mu = \mathrm{H}[X]$.

As a consequence, the relationship between memory compression and work advantage is particularly direct. Classical work extraction is simply $W_\mu = C_\mu - \mathrm{H}[X] = 0$, indicating that Markov chain generation is thermodynamically neutral at best. However, $W_q = C_q - \mathrm{H}[X] \leq 0$, such that quantally compressed Markov chain generators are indeed capable of work extraction. The memory and work advantages also take on a simple relationship: $\Delta_q W = k_B T \Delta_q C$ and so they are maximally efficient: $e_q = 1$.

R, k-Golden Mean Hidden Generators. However, measurements are typically not a process' internal state; thus, we must address hidden Markov generators. In addition to non-Markovity N, another means of quantifying how distant a process' generator is from being Markov is the *Markov order*: the smallest integer R such that $\mathrm{H}[S_R | X_{R-1} \ldots X_1] = 0$. In other words, R is the largest number of symbols we must see before the generator's state is known. It is infinite for most processes [25]; for Markov-generated processes $R = 1$.

A dual notion to the Markov order is the *cryptic order*. This is the smallest integer k such that $\lim_{t \to \infty} \mathrm{H}[S_k | X_1 \ldots X_t] = 0$. This is a more general condition than that for Markov order: consequently, $k \leq R$ for all processes.
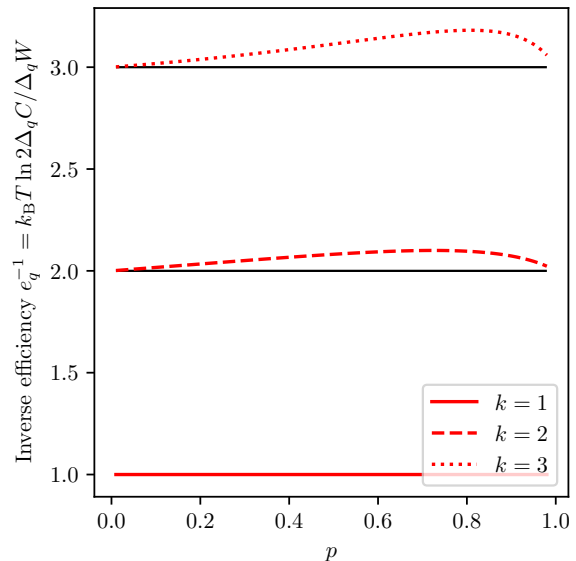

FIG. 2. R, k-Golden Mean Generator thermal efficiency: Inverse compression efficiency e_q^{-1} depends only on the crypticity k and transition parameter p. Black lines added at integer k for comparison.

There is a family of generators—R, k-*Golden Mean Generators*—that for each integer pair, R and k, contains a family that generates processes with Markov order R and cryptic order k, parametrized by a transition probability p. (This family is defined in the SM.) Additionally, for each R, k-Golden Mean Generator the SM shows that (i) the quantum generators are degenerate and each $\{\phi_{x,k}\}$ gives the same quantum generator and (ii) the compression efficiency $e_q(R, k)$ of the quantum generator depends only on the cryptic order k. Numerical calculations for $k = 1, 2, 3$ are shown in Fig. 2. Note, there, the apparent crypticity bound:

$$e_q(k) \leq \frac{1}{k} \ . \tag{6}$$

Nemo Generator. Most processes have infinite cryptic and Markov orders [24]: $R = \infty$ and $k = \infty$. We explored an example of this, the Nemo generator, whose state-transition diagram is displayed in the SM. Its behavior differs from R, k-Golden Mean Generators in two key respects. First, whereas each R, k-Golden Mean Generator has only one geometrically distinct quantum implementation, the Nemo generator's space of work and quantum compression trade-offs is one-dimensional, parametrized by the phase $\Phi = 2\phi_{0A} + 2\phi_{0C} + \phi_{1C} - \phi_{1A} - \phi_{1B}$. Second, the efficiency bound Eq. (6) clearly does not hold. (See Fig. S1(e).) If it did, then $e_q = 0$. Instead, numerical exploration shows e_q is bounded away from zero and,
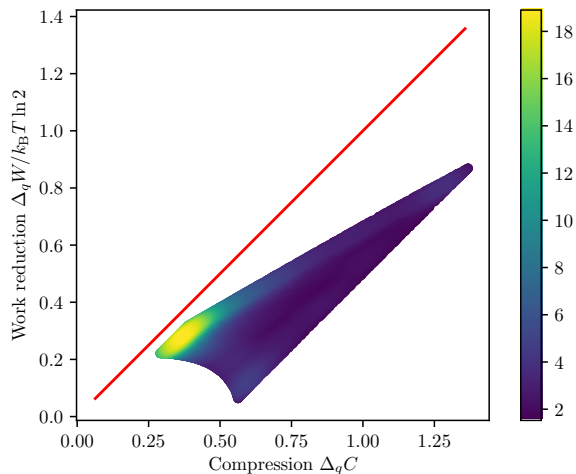
FIG. 3. Two-Step-Erase Generator: A complex relationship between $\Delta_q W$ and $\Delta_q C$ appears that is not captured by a single efficiency ϵ_q. The density in the plot assumes uniform distribution over phases $\{\phi_{xs}\}$, with blue indicating low density and yellow indicating high density.

incidentally, only varies within a small range, such that $e_q \approx 0.3885 \pm 0.025$.

Two-Step Erase Generator. The previous two generators had relatively simple quantum implementations with either complete degeneracy or only filling out a one-dimensional curve in their *work-compression* (W/C) charts. This is not the generic behavior of quantum generators. To illustrate this, we now also examine a generator, termed the *Two-Step Erase* Generator, whose compression thermodynamics is more qualitatively typical of the generators we explored. Its state-transition diagram is also given in the SM.

Figure 3 presents a W/C-chart that plots out every achievable $(\Delta_q C, \Delta_q W/k_B T)$ pair over the range of possible phases $\{\phi_{xs}\}$ that determine the quantum implementation, colored by density. Density is determined by assuming uniform distribution over the phases $\{\phi_{xs}\}$. We note that the "high-advantage" regions, where both $\Delta_q C$ and $\Delta_q W$ are large, are actually spanned by only a small volume of generators, while regions with lower advantages are spanned by a large volume, indicating that high advantage may not be robust.

Closing Remarks. We derived single-shot and asymptotic work costs for quantum generator implementations that are quasistatically attainable. The first of these results opens the pathway for single-shot comparisons between classical and quantum resources in process generation, while the second allows direct comparison in terms of asymptotic quantities. We demonstrated that, when it comes to quantizing predictive generators, one can

"have their cake and eat it too" with regards to thermodynamics and memory compression: advantage in both ($\Delta_q C \geq 0$ and $\Delta_q W \geq 0$) can be simultaneously attained.

We analyzed asymptotic thermal efficiencies in four generator classes, demonstrating a diversity of trade-offs between work and memory advantages. For every predictive generator of a process, there is a well-defined family of quantum implementations. However, the scope of their variety ranges from simple (R, k-Golden Mean) to highly complex (Two-Step Erase) generators. Even when the family of generators is simple, there exist fascinating and complex relationships between the work advantage, memory advantage, and the generator's computational properties.

Forthcoming work employs the work bounds here to compare quantum ϵ-machines to classical and quantum non-predictive generators, seeking the conditions for optimal work cost over all generators of a given stochastic process [41]. It also remains to relax certain assumptions underlying our work, such as the free availability of empty tapes (which must be prepared), and the quasistatic assumption which requires infinite time. We believe that the work completed here provides a foundation for these and other further extensions.

* sloomis@ucdavis.edu
† chaos@ucdavis.edu

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

[2] J. R. Mahoney, C. Aghamohammadi, and J. P. Crutchfield. Occam's quantum strop: Synchronizing and compressing classical cryptic processes via a quantum channel. *Scientific Reports*, 6(20495), 2016. 1, 3, 9

[3] M. S. Palsson, M. Gu, J. Ho, H. M. Wiseman, and G. J. Pryde. Experimentally modeling stochastic processes with less memory by the use of a quantum processor. *Science Advances*, 3(2), 2017. 1

[4] F. G. Jouneghani, M. Gu, J. Ho, J. Thompson, W. Y. Suen, H. M. Wiseman, and G. J. Pryde. Observing the

ambiguity of simplicity via quantum simulations of an Ising spin chain. 2014. arXiv:1711.03661 [quant-ph]. 1

[5] F. C. Binder, J. Thompson, and M. Gu. A practical, unitary simulator for non-markovian complex processes. *Phys. Rev. Lett.*, 120:240502, 2017. 1, 2

[6] J. Thompson, A. J. P. Garner, V. Vedral, and M. Gu. Using quantum theory to simplify inputâĂŞoutput processes. *npj Quantum Information*, 3:6, 2017. 1

[7] P. M. Riechers, J. R. Mahoney, C. Aghamohammadi, and J. P. Crutchfield. Minimized state-complexity of quantum-encoded cryptic processes. *Phys. Rev. A*, 93(5):052317, 2016. 1, 8

[8] J. Thompson, A. J. P. Garner, J. R. Mahoney, J. P. Crutchfield, V. Vedral, and M. Gu. Causal asymmetry in a quantum world. *Phys. Rev. X*, 8:031013, 2018. 1

[9] S. Loomis and J. P. Crutchfield. Strong and weak optimizations in classical and quantum models of stochastic processes. *J. Stat. Phys*, 176:1317âĂŞ1342, 2019. 3, 5

[10] Q. Liu, T. J. Elliot, F. C. Binder, C. Di Franco, and M. Gu. Optimal stochastic modeling with unitary quantum dynamics. *Phys. Rev. A*, 99:062110, 2019. 1, 3, 2

[11] D. Janzing, P. Wocjan, R. Zeier, R. Geiss, and Th. Beth. Thermodynamic cost of reliability and low temperatures: Tightening landauer's principle and the second law. *Int. J. Theor. Physics.*, 39(12), 2000. 1, 2

[12] S. Vinjanampathy and J. Anders. Quantum thermodynamics. *Contemporary Physics*, 57:4:545–579, 2016. 1, 2

[13] M. Lostaglio. Thermodynamic laws for populations and quantum coherence: A self-contained introduction to the resource theory approach to thermodynamics. 1, 2

[14] M. Horodecki and J. Oppenheim. Fundamental limitations for quantum and nanoscale thermodynamics. *Nature Comm.*, 4:2059, 2013. 1

[15] F. G. S. L. Brandão, M. Horodecki, J. Oppenheim, J. M. Renes, and R. W. Spekkens. Resource theory of quantum states out of thermal equilibrium. *Phys. Rev. Lett.*, 111:250404, 2013. 1

[16] F. G. S. L. Brandão, M. Horodecki, N. Ng, J. Oppenheim, and S. Wehner. The second laws of quantum thermodynamics. *PNAS*, 112 (11):3275–3279, 2015. 1

[17] M. Lostaglio, D. Jennings, and T. Rudolph. Description of quantum coherence in thermodynamic processes requires constraints beyond free energy. *Nature Comm.*, 6:6383, 2015. 1, 2

[18] A. M. Alhambra, L. Masanes, J. Oppenheim, and C. Perry. Fluctuating work: From quantum thermodynamical identities to a second law equality. *Phys. Rev. X*, 6, 2016. 2

[19] O. C. O. Dahlsten, R. Renner, E. Rieper, and V. Vedral. Inadequacy of von Neumann entropy for characterizing extractable work. *New J. Phys.*, 13, 2011. 1, 3

[20] L. del Rio, J. Åberg, R. Renner, O. Dahlsten, and V. Vedral. The thermodynamic meaning of negative entropy. *Nature*, 474:61–63, 2011. 2, 3, 4

[21] P. Faist, F. Dupuis, J. Oppenheim, and R. Renner. The minimal work cost of information processing. *Nature Comm.*, 6:7669, 2015. 1, 3, 4

[22] Y. Guryanova, N. Friis, and M. Huber. Ideal projective measurements have infinite resource costs. *Quantum*, 4, 2020. 1, 2, 4

[23] J. P. Crutchfield and K. Young. Inferring statistical complexity. *Phys. Rev. Let.*, 63:105–108, 1989. 1, 2

[24] J. P. Crutchfield, C. J. Ellison, and J. R. Mahoney. Time's barbed arrow: Irreversibility, crypticity, and stored information. *Phys. Rev. Lett.*, 103(9):094101, 2009. 4

[25] R. G. James, J. R. Mahoney, C. J. Ellison, and J. P. Crutchfield. Many roads to synchrony: Natural time scales and their algorithms. *Phys. Rev. E*, 89:042135, 2014. 4

[26] N. F. Travers and J. P. Crutchfield. Equivalence of history and generator ϵ-machines. arxiv.org:1111.4500 [math.PR]. 2

[27] J. P. Crutchfield. Between order and chaos. *Nature Physics*, 8:17–24, 2012. 1, 2

[28] R. Landauer. Irreversibility and heat generation in the computing process. *IBM J. Res. Develop.*, 5(3):183–191, 1961. 1

[29] A. B. Boyd, D. Mandal, and J. P. Crutchfield. Identifying functional thermodynamics in autonomous Maxwellian ratchets. *New J. Physics*, 18:023049, 2016. 1, 3

[30] A. B. Boyd, D. Mandal, and J. P. Crutchfield. Thermodynamics of modularity: Structural costs beyond the Landauer bound. *Physical Review X*, 8(3):031036, 2018. 1, 2, 3

[31] D. R. Upper. *Theory and Algorithms for Hidden Markov Models and Generalized Hidden Markov Models.* PhD thesis, University of California, Berkeley, 1997. Published by University Microfilms Intl, Ann Arbor, Michigan. 2

[32] R. B. Ash. *Information Theory.* John Wiley and Sons, New York, 1965. 2

[33] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* Wiley-Interscience, New York, second edition, 2006. 2

[34] Supplemental Material at [URL inserted by publisher] reviews energy flow conventions, quantum implementations of classical generators and costs, and examples. 3

[35] R. Renner and S. Wolf. Smooth renyi entropy and applications. *International Symposium on Information Theory, 2004. ISIT 2004. Proceedings.*, 2011. 3, 2

[36] R. Renner. *Security of Quantum Key Distribution.* PhD thesis, ETH Zurich, 2005. arXiv:quant-ph/0512258. 2

[37] M. Tomamichel. *A Framework for Non-Asymptotic Quantum Information Theory.* PhD thesis, ETH Zurich, 2012. arXiv:1203.2142 [quant-ph]. 3

[38] A. Vitanov, F. Dupuis, M. Tomamichel, and R. Renner. Chain rules for smooth min- and max-entropies. *IEEE Trans. Info. Th.,*, 59(5):2603–2612, 2013. 3

[39] M. Tomamichel. *Quantum Information Processing with Finite Resources.* Springer, Cham, Switzerland, 2016. 3

[40] M. Nielsen and I. Chuang. *Quantum Computation and Quantum Information.* Cambridge University Press, New York, 2010. 3, 2

[41] S. Loomis and J. P. Crutchfield. Thermodynamically-efficient local computation and the inefficiency of quan-

tum memory compression. *Phys. Rev. R*, in press, 2020. arXiv:2001.02258. 5

Supplementary Materials

Thermal Efficiency of Quantum Memory Compression

Samuel P. Loomis and James P. Crutchfield

The Supplementary Materials calls out energy flow directionality, reviews quantum implementations of classical generators and the thermodynamic cost of these implementations, and provides details on the example calculations.

ENERGY FLOW CONVENTION

The main text appeals to a particular direction of energy flow. This is particularly at issue in applying the information processing Second Law (IPSL) from Ref. [29]. There, the IPSL is stated in the form:

$$W \leq k_{\mathrm{B}} T(h'_\mu - h_\mu) \ ,$$

where the former entropy rate is that of the output tape and the latter, that of the input tape. In short, this is in the case in which the "work done" W is interpreted as the "work the ratchet does on a work reservoir"—the work extracted.

Here, work is defined as the "work done on the tape, taken from the work reservoir", so is opposite in sign. Then one has:

$$W \geq k_{\mathrm{B}} T(h_\mu - h'_\mu) \ .$$

And so, the rules of energy flow are consistent, appearing here just with opposite direction. With simulators, as here, $h_\mu = 0$ and therefore work is extractable while generating a process.

QUANTUM IMPLEMENTATIONS OF CLASSICAL GENERATORS

We define a classical generator as a triplet $(\mathcal{S}, \mathcal{X}, \{\mathbf{T}^{(x)} : x \in \mathcal{X}\})$ where $T^{(x)}_{s's} = \Pr(s', x|s)$. $\mathcal{S}$ is the finite set of memory states, $\mathcal{X}$ is the finite alphabet of produced symbols, and $\Pr(s', x|s)$ determines the transition-and-production dynamic of the generator.

To analyze the thermodynamics of physical implementations of generators, we must establish rules that circumscribe what we consider physically allowed and the correspondence to thermodynamic quantities such as work and heat.

Here, we consider the *resource theory of thermal operations* [14, 15]. Generally, on a quantum system S we allow operations of the form:

$$\mathcal{E}(\rho_S) := \mathrm{Tr}_B \left(U \rho_S \otimes \frac{e^{-\beta H_B}}{Z_B} U^\dagger \right) \ , \tag{S1}$$

where S and B are auxiliary systems with Hamiltonians H_S and H_B, B a thermal bath, and U acts on the joint Hilbert space of $\mathcal{H}_S$ and $\mathcal{H}_B$. The unitary operator U satisfies the rule of microscopic energy conservation, where we constrain $[U, H_S + H_B] = 0$.

Recall from the main body that an implementation $(\mathcal{H}_S, \mathcal{H}_X, \mathcal{H}_A, \mathcal{H}_B, U, \mathcal{E})$ of a generator involves the memory space $\mathcal{H}_S$, symbol space $\mathcal{H}_X$, auxiliary space $\mathcal{H}_A$, and bath space $\mathcal{H}_B$; the ensemble $\mathcal{E} = \{|\psi_s\rangle : s \in \mathcal{S}\}$; and a unitary acting on $\mathcal{H}_S \otimes \mathcal{H}_X \otimes \mathcal{H}_A \otimes \mathcal{H}_B$, such that the channel:

$$\mathcal{T}_{SX}(\rho_{SX}) := \mathrm{Tr}_{AB} \left(U \rho_{SX} \otimes |0\rangle \langle 0|_A \otimes \rho_B U^\dagger \right)$$

satisfies:

$$\mathcal{T}_{SX}\left(|\psi_s\rangle\langle\psi_s|_S \otimes |0\rangle\langle0|_X\right) = \sum_{s',x} \Pr\left(s',x|s\right)|\psi_{s'}\rangle\langle\psi_{s'}| \otimes |x\rangle\langle x|_X \ . \tag{S2}$$

Suppose that there are Hamiltonians H_S, H_X, H_A, and H_B for each system such that $\rho_{\text{sys}} = Z_{\text{sys}}^{-1}\exp\left(-\beta H_{\text{sys}}\right)$ is the Gibbs distribution of its Hamiltonian. Then the resource theory of thermal operations requires $[U, H_S + H_X + H_A + H_B] = 0$ [11].

In quantum mechanics, the rule of microscopic energy conservation $[U, H_S + H_X + H_A + H_B] = 0$ brings coherence with respect to the Hamiltonian into play as a resource [13, 17]. The type of systems we consider here are what are often, in the literature of information engines, called *information reservoirs*: systems whose Hamiltonian is trivially degenerate, so that energetics does not play a direct role in their dynamics. On such systems, tracking coherence is no longer at issue, as all operators commute with a degenerate Hamiltonian.

In this case, we can restrict ourselves to Hamiltonian control protocols for the erasure of the auxiliary system, as in [20], where the Hamiltonian is degenerate at the beginning and end of the procedure, and stays in one basis during the procedure. In this case the work cost is computed by adding changes in the energy levels, weighted by the relative probabilities of being in those levels [12].

Among quantum implementations, the only form that has been studied for generators is the *unitary* implementation, which itself is only valid for predictive generators [5, 10]. In this implementation, the bath system B is not used and the unitary operator is split into two steps, $U = U_2 U_1$, where $U_1 = U_{SX} \otimes 1_A$ and $U_2 = 1_S \otimes U_{XA}$.

In the first step, the *evolution step*, we act only on the memory and the output SX with the unitary U_{SX} defined by the action:

$$U_{SX}|\psi_s\rangle|0\rangle = \sum_x e^{i\phi_{xs}}\sqrt{\Pr(x|s)}|\psi_{f(x,s)}\rangle|x\rangle \ . \tag{S3}$$

The unitarity evolution of U_{SX} in fact defines the overlap matrix $\Omega_{rs} := \langle\psi_r|\psi_s\rangle$ via the recursive formula [10]:

$$\Omega_{rs} = \sum_x \sqrt{\Pr(x|r)\Pr(x|s)}e^{i(\phi_{xs}-\phi_{xr})}\Omega_{f(r,s)f(x,s)} \ . \tag{S4}$$

For any choice of phases, an overlap matrix Ω_{rs} and unitary U_{SX} exist.

In the second step—the *measurement step*—the symbol is observed, sending the pure state $U_{SX}|\psi_s\rangle\langle\psi_s|\otimes|0\rangle\langle0|U_{SX}^\dagger$ to the mixed state in Eq. (S2). This is done by coupling the system X to the auxiliary system A and applying a unitary so that:

$$U_{XA}|x\rangle_X|0\rangle_A \propto |x\rangle_X|x\rangle_A \ .$$

When the auxiliary is discarded (or, more realistically, reset) we are left with the state on SX, as desired.

The perfect preparation, or reset, of the auxiliary in state $|0\rangle_A$ cannot be performed with finite resources [22]. In the following section we consider the work costs achievable in the single-shot setting where a probability of failure ϵ is allowed for the reset step. The underlying protocol is quasistatic (see Ref. [20]), requiring infinite time to complete. Here, we focus on fundamental work limits in the single-shot and asymptotically parallel regimes, but not in the finite time regime, which is an important direction of future extension.

INFORMATION-THEORETIC DEFINITIONS

This section employs the Shannon [33], von Neumann [40], and smooth conditional entropies [35, 36]. The Shannon entropy for a random variable $X \sim \Pr(x)$ and von Neumann entropy for a system A (no relation to the auxiliary A

in the previous section) with density matrix ρ_A are, respectively:

$$\mathrm{H}\left[X\right] \equiv -\sum_x \mathrm{Pr}(x) \log_2 \mathrm{Pr}(x) \text{ and}$$

$$\mathrm{H}_q\left[S\right] \equiv -\mathrm{Tr}\left(\rho_S \log_2 \rho_S\right) \ .$$

For bipartite variables XY and bipartite quantum systems AB, these quantities beget the conditional entropies and mutual informations:

$$\mathrm{H}\left[X|Y\right] \equiv \mathrm{H}\left[XY\right] - \mathrm{H}\left[Y\right] \ ,$$
$$\mathrm{H}_q\left[A|B\right] \equiv \mathrm{H}_q\left[AB\right] - \mathrm{H}_q\left[B\right] \ ,$$
$$\mathrm{I}[X:Y] \equiv \mathrm{H}\left[X\right] + \mathrm{H}\left[Y\right] - \mathrm{H}\left[XY\right] \ , \text{ and}$$
$$\mathrm{I}_q[A:B] \equiv \mathrm{H}_q\left[A\right] + \mathrm{H}_q\left[B\right] - \mathrm{H}_q\left[AB\right] \ .$$

For two systems A and B with joint state ρ_{AB}, the min- and max-entropies are given by [35–37, 39]:

$$\mathrm{H}_{\min}\left[A|B\right]_\rho \equiv \min_{\sigma_B} \sup\{\lambda : \rho_{AB} \leq 2^{-\lambda} 1_A \otimes \sigma_B\} \text{ and}$$

$$\mathrm{H}_{\max}\left[A|B\right]_\rho \equiv \max_{\sigma_B} 2 \log_2 F\left(\rho_{AB}, 1_A \otimes \sigma_B\right) \ ,$$

where $F(\rho, \sigma) = \mathrm{Tr}\left(\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\right)$ is the *fidelity*. The smooth conditional entropies are optimizations of these quantities over all $\tilde{\rho}_{AB}$ within the ϵ-ball $B_\epsilon(\rho_{AB})$; that is, all states such that $\sqrt{1 - F\left(\tilde{\rho}_{AB}, \rho_{AB}\right)} < \epsilon$:

$$\mathrm{H}_{\min}^\epsilon\left[A|B\right] \equiv \max_{\tilde{\rho}_{AB}} \mathrm{H}_{\min}\left[A|B\right]_{\tilde{\rho}} \text{ and}$$

$$\mathrm{H}_{\max}^\epsilon\left[A|B\right] \equiv \min_{\tilde{\rho}_{AB}} \mathrm{H}_{\max}\left[A|B\right]_{\tilde{\rho}} \ .$$

When B is uncorrelated with A, $\rho_{AB} = \rho_A \otimes \rho_B$, the resulting quantities are independent of B and so we have the marginal smooth entropies $\mathrm{H}_{\max}^\epsilon\left[A\right]$ and $\mathrm{H}_{\min}^\epsilon\left[A\right]$. We will utilize a result on smooth conditional entropies that generalizes the chain rule on von Neumann entropy [38]. We state two somewhat streamlined versions of the theorem here. For any $\delta > 0$ and systems AB:[1]

$$\mathrm{H}_{\max}^\delta\left[B|A\right] \leq \mathrm{H}_{\max}^{4\delta}\left[AB\right] - \mathrm{H}_{\min}^\delta\left[A\right] + O\left(\log\frac{1}{\delta}\right) \text{ and} \tag{S5}$$

$$\mathrm{H}_{\min}^\delta\left[B|A\right] \leq \mathrm{H}_{\min}^{4\delta}\left[AB\right] - \mathrm{H}_{\min}^\delta\left[A\right] + O\left(\log\frac{1}{\delta}\right) \tag{S6}$$

GENERATOR IMPLEMENTATION COSTS

We import the following result from Ref. [20]: Given a system S correlated with an auxiliary A, and any $\epsilon > 0$, there is a procedure for erasing A while preserving S, with probability of failure ϵ, that has a work cost of no more than:[2]

$$\frac{W}{k_\mathrm{B}T\ln 2} \leq \mathrm{H}_{\max}^{\epsilon^2/16}\left[A|S\right] + O\left(\log\frac{1}{\epsilon}\right) \ . \tag{S7}$$

The smooth conditional entropy also provides a lower bound on the attainable work cost[3] [21]: $W/k_\mathrm{B}T\ln 2 \geq \mathrm{H}_{\max}^{\sqrt{2\epsilon}}\left[A|S\right]$. While this bound is finite in the limit $\epsilon \to 0$, the upper bound Eq. (S7) diverges and so does not

---

[1] Comparing our statements with Ref. [38], note that ignore the third system C and swap A and B. Rather than use the four parameters ϵ, ϵ', ϵ'', and f, we use the single parameter δ such that $\epsilon = 4\delta$, $\epsilon' = \epsilon'' = \delta$, and $f = O\left(\log\frac{1}{\delta}\right)$. Then Eq. (S5) and Eq. (S6) correspond to the sixth and first equations on page 2 of Ref. [38].

[2] Again, comparing with Ref. [20], our Eq. (S7) is drawn from Thm. 1. System S from Ref. [20] is system A here, system O from Ref. [20] is system S here, and δ from Ref. [20] is ϵ here. In Ref. [20], take $\epsilon = \delta^2/13$ and $\Delta = 2\log\left(1/\epsilon\right) = 2\log 13 + 4\log\left(1/\delta\right)$. The term $2\log 13$ is irrelevant in the $\delta \to 0$ limit so we include it in the big-O term.

[3] In this case, comparing to Ref. [21], note that their system E is our system A, their system X' is our system S, and finally we have directly replaced $\bar{\epsilon}$ with $\sqrt{2\epsilon}$.

guarantee any particular attainable work cost. Additionally, it must be noted that the protocol used by Ref. [20] is quasistatic, requiring infinite time to complete. This is consistent with previous results on, say, the unattainability of perfect measurements with finite resources [22].

We can use Eq. (S7) to prove a generalization of the *detailed* Landauer cost—that is, to show the conditions under which Landauer's bound is quasistatically attainable. Suppose we have a quantum channel $\mathcal{E}$ we wish to implement and we do so on a system S with mixed state ρ_S. The target state is $\rho'_S = \mathcal{E}(\rho_S)$. We perform the map in the following way. Using the Stinespring dilation of $\mathcal{E}$, we couple S to an auxiliary system A in state $|0\rangle\langle0|_A$ and perform a unitary operation on both systems:

$$\rho'_{SA} = U_{SA}\rho_S \otimes |0\rangle\langle0|_A U^{\dagger}_{SA} ,$$

such that $\mathcal{E}(\rho_S) = \text{Tr}_A(\rho'_{SA})$. At the end of the procedure we must erase A. This can be done with cost Eq. (S7). This form of the cost for implementing a channel is given in Ref. [21], where an argument similar to the following was used to derive Landauer's lower bound in the macroscopic limit. Here, we apply the same logic to show that Landauer's bound is also *attainable* in the macroscopic (simulating an infinite number of parallel channels) limit.

Applying Eq. (S5) to Eq. (S7), we have:

$$\frac{W}{k_{\text{B}}T\ln 2} \leq \text{H}^{\epsilon^2/4}_{\text{max}}[S'A'] - \text{H}^{\epsilon^2/16}_{\text{min}}[S'] + O\left(\log\frac{1}{\epsilon}\right) .$$

However, $\text{H}^{\epsilon^2/4}_{\text{max}}[S'A'] = \text{H}^{\epsilon^2/4}_{\text{max}}[S]$ by unitary equivalence, so we have the erasure cost:

$$\frac{W}{k_{\text{B}}T\ln 2} \leq \text{H}^{\epsilon^2/4}_{\text{max}}[S] - \text{H}^{\epsilon^2/16}_{\text{min}}[S'] + O\left(\log\frac{1}{\epsilon}\right) . \tag{S8}$$

Since we can perform the initial unitary with no work, this is the only work cost involved in implementing the channel. To summarize: The channel $\mathcal{E}$ can be performed on the system S with a work cost not exceeding Eq. (S8).

Now, suppose we choose instead to implement parallel generation of our process. That is, we have N independent systems on which we want to implement N independent copies of the channel $\mathcal{E}$ with probability of error less than $\epsilon > 0$. Naturally, the work cost becomes:

$$\frac{W}{k_{\text{B}}T\ln 2} \leq \text{H}^{\epsilon^2/4}_{\text{max}}\left[S^{\otimes N}\right] - \text{H}^{\epsilon^2/16}_{\text{min}}\left[S'^{\otimes N}\right] + O\left(\log\frac{1}{\epsilon}\right) .$$

Significantly, the error term does not depend on N. When we further account for the Asymptotic Equipartition Theorem of smooth conditional entropies, we have the remarkable result for the work rate:

$$\frac{W}{Nk_{\text{B}}T\ln 2} \leq \text{H}[S] - \text{H}[S'] + O\left(\sqrt{\frac{1}{N}\log\frac{1}{\epsilon}}\right) . \tag{S9}$$

With Landauer's bound sandwiching the work from below, we find a tight result on the achievable work cost. By scaling error with N, for instance $\epsilon \sim 2^{-\sqrt{N}}$, Landauer's bound can, in the limit $N \to \infty$, be achieved for quantum channels. In the single-shot regime, the bound of Eq. (S8) gives us a somewhat less certain range of achievability.

This can be applied directly to the implementation of *generators* discussed in the previous section. In the single-shot setting, we have:

$$\frac{W}{k_{\text{B}}T\ln 2} \leq \text{H}^{\epsilon^2/4}_{\text{max}}[S] - \text{H}^{\epsilon^2/16}_{\text{min}}[S'X] + O\left(\log\frac{1}{\epsilon}\right) .$$

Applying Eq. (S6) this becomes:

$$\frac{W}{k_{\text{B}}T\ln 2} \leq \text{H}^{\epsilon^2/4}_{\text{max}}[S] - \text{H}^{\epsilon^2/64}_{\text{min}}[S'|X] - \text{H}^{\epsilon^2/64}_{\text{min}}[X] + O\left(\log\frac{1}{\epsilon}\right) . \tag{S10}$$

Finally, consider the asymptotic limit of N parallel generators producing N independent copies of a stochastic process. The Asymptotic Equipartition Theorem again gives the result:

$$\frac{W_q}{k_{\mathrm{B}}T\ln 2} \leq \mathrm{H}_q[S] - \mathrm{H}_q[S'|X] - \mathrm{H}[X] + O\left(\sqrt{\frac{1}{N}}\log\frac{1}{\epsilon}\right) . \tag{S11}$$

Our last result is the inequality Eq. (5). We note that:

$$\begin{aligned} W_q &= \mathrm{H}_q[S] - \mathrm{H}_q[S'|X] - \mathrm{H}[X] \\ &= \mathrm{H}_q[S'] - \mathrm{H}_q[S'|X] - \mathrm{H}[X] \\ &= \mathrm{I}_q[S':X] - \mathrm{H}[X] . \end{aligned}$$

The same form can be given for W_μ in terms of the Shannon entropies:

$$\begin{aligned} W_\mu &= \mathrm{H}[S] - \mathrm{H}[S'|X] - \mathrm{H}[X] \\ &= \mathrm{I}[S':X] - \mathrm{H}[X] . \end{aligned}$$

Now, in the quantum model $\mathrm{I}_q(S':X)$ is the mutual information of the state:

$$\rho'_{SX} = \sum_{x,s,s'} \mathrm{Pr}_0(s)\mathrm{Pr}(s',x|s)\,|\psi_{s'}\rangle\langle\psi_{s'}|\otimes|x\rangle\langle x| ,$$

which can be derived from the classical variables $S'X$ by the local mappings $s' \mapsto |\psi_{s'}\rangle$ and $x \mapsto |x\rangle$. Then by the data processing inequality: $\mathrm{I}_q(S':X) \leq \mathrm{I}(S':X)$. This proves that $W_q \leq W_\mu$.

EXAMPLE GENERATORS

Understanding the behavior of our example generators (see Fig. S1) requires discussing gauge properties of quantum implementations.

The physical properties of each quantum generator are entirely determined by its overlap matrix $\Omega_{rs} = \langle\psi_r|\psi_s\rangle$. However, this in itself contains nonphysical degrees of freedom [9]. None of the invariant geometry of our generators is modified under the transformation $|\psi_s\rangle \mapsto e^{i\Psi_s}|\psi_s\rangle$ on the signal states. Thus, these represent a *gauge transformation*. In terms of the overlap matrix, this means that our generators are invariant under the transformations $\Omega_{rs} \mapsto e^{i(\Psi_s - \Psi_r)}\Omega_{rs}$.

It is helpful (especially for the Nemo process) to consider these gauge properties in terms of how they act on the phases $\{\phi_{xs}\}$ that determine the quantum generator. Applying the gauge transformation to the consistency formula gives:

$$\Omega_{rs} = \sum_x \sqrt{\mathrm{Pr}(x|r)\mathrm{Pr}(x|s)}e^{i(\tilde{\phi}_{xs} - \tilde{\phi}_{xr})}\Omega_{f(x,r)f(x,s)} ,$$

where:

$$\tilde{\phi}_{xs} = \phi_{xs} - \Psi_s + \Psi_{f(x,s)} \tag{S12}$$

is the induced transformation on the generator's phases. Equation (S12) can be taken as a fundamental description of the gauge transformation.

Using Eq. (S12) allows us to determine the *gauge invariants*—that is, combinations of the phases $\{\phi_{xs}\}$ that do not change under a gauge transformation. In this case, the gauge invariants are best understood graphically, in terms of the hidden Markov models from before. Each phase $\{\phi_{xs}\}$ can be understood as being assigned to an edge, while each phase in the gauge transformation $\{\Psi_s\}$ can be seen as being assigned to a state.
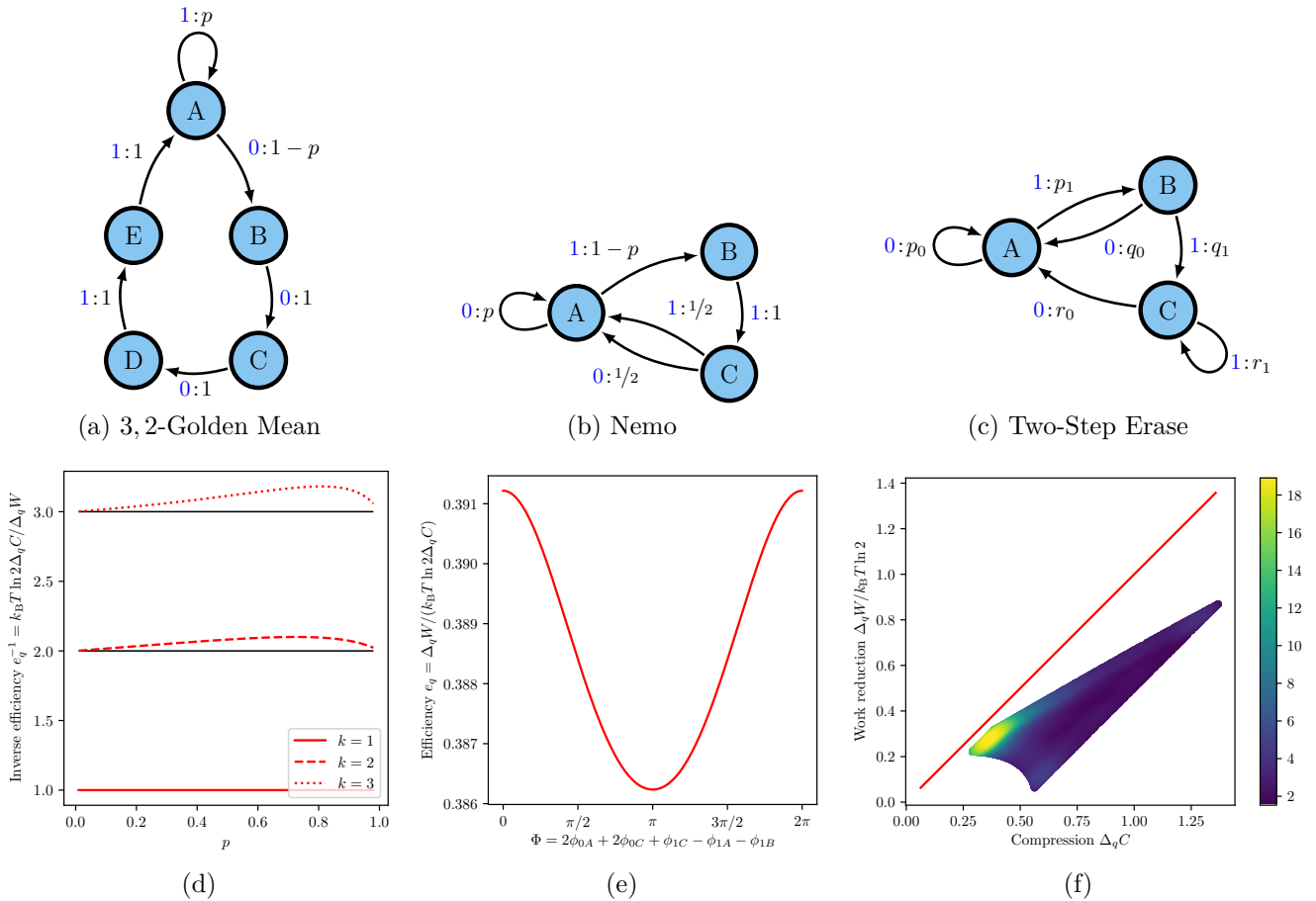
(a) $3, 2$-Golden Mean (b) Nemo (c) Two-Step Erase

(d) (e) (f)

FIG. S1. *(a)* The R, k-Golden Mean Generator for $R = 3$ and $k = 2$. See main text for general construction. *(b)* The Nemo generator, which has infinite Markov and cryptic orders. *(c)* The Two-Step Erase generator always erases its memory to A upon generating a 0, but erases its memory to C upon generating two consecutive 1's. *(d)* The inverse compression efficiency e_q^{-1} of the R, k-Golden Mean generator depends only on the crypticity k and transition parameter p. Black bars added at integer values for comparison. *(e)* For the Nemo generator, the overlap matrix Ω, which determines all the quantum-information-theoretic properties of the implementation, depends only on the combined phase $\Phi = 2\phi_{0A} + 2\phi_{0C} + \phi_{1C} - \phi_{1A} - \phi_{1B}$. Consequently, the efficiency e_q depends on this quantity and the parameter p. Numerical exploration shows that the variation of e_q due to Φ is quite small in amplitude and varies sinusoidally. We plot this variation for $p = 0.5$. *(f)* For general p, q, r of the Two-Step Erase process (showing $p = 1/2$, $q = 1/5$, and $r = 2/5$ here), we find a complex relationship between $\Delta_q W$ and $\Delta_q C$ appears. This is not captured by a single efficiency e_q. The density in the plot assumes uniform distribution over phases $\{\phi_{xs}\}$, with blue indicating low density and yellow indicating high density.

For each loop of edges, we can take a linear combination of the constituent edges' phases ϕ_{xs}, adding positive and negative signs based on the direction of the edges. These loop sums are the gauge invariants. For instance, the Nemo process has $\Phi_0 = \phi_{0A}$, $\Phi_1 = \phi_{1C} - \phi_{0C}$, and $\Phi_2 = \phi_{1A} + \phi_{1B} + \phi_{1C}$ as gauge invariants.

(R, k)-**Golden Mean Generators**

An R, k-Golden Mean Generator is one with $R + k$ memory states. These states can be considered to belong to two groups: the A state, which is the only nondeterministic state and the B-states $\mathcal{B} \equiv \{B_1, \ldots, B_{R+k-1}\}$. The B-states are further broken down into a Markov part $\mathcal{R} \equiv \{B_1, \ldots, B_{R-1}\}$ and a cryptic part $\mathcal{K} \equiv \{B_R, \ldots, B_{R+k-1}\}$. The

dynamic on the generator is given by:

$$\Pr\left(s', 0|s\right) = \begin{cases} 1 - p & s = A, \ s' = B_1 \\ 1 & s = B_r, \ s' = B_{r+1}, \ 0 \le r < R \\ 0 & \text{otherwise} \end{cases}$$

and

$$\Pr\left(s', 1|s\right) = \begin{cases} p & s', s = A \\ 1 & s = B_r, \ s' = B_{r+1}, \ R \le r \le R + k - 2 \\ 1 & s = B_{R+k-1}, \ s' = A \\ 0 & \text{otherwise} \end{cases}.$$

We can check that:

$$\Pr{}_0(s) = \begin{cases} \frac{1}{1+(R+k-1)(1-p)} & s = A \\ \frac{1-p}{1+(R+k-1)(1-p)} & s = A \end{cases}.$$

is the stationary distribution. Letting $Z = 1 + (R + k - 1)(1 - p)$, we have:

$$\Pr\left(s', 0, s\right) = \begin{cases} \frac{1-p}{Z} & s = A, \ s' = B_1 \\ \frac{1-p}{Z} & s = B_r, \ s' = B_{r+1}, \ 0 \le r < R \\ 0 & \text{otherwise} \end{cases}$$

and

$$\Pr\left(s', 1, s\right) = \begin{cases} \frac{p}{Z} & s', s = A \\ \frac{1-p}{Z} & s = B_r, \ s' = B_{r+1}, \ R \le r \le R + k - 2 \\ \frac{1-p}{Z} & s = B_{R+k-1}, \ s' = A \\ 0 & \text{otherwise} \end{cases}.$$

It is helpful to also have:

$$\Pr(X = 0) = \frac{R(1 - p)}{Z} \ ,$$

$$\Pr(X = 1) = \frac{(k - 1)(1 - p) - 1}{Z} \ ,$$

$$\Pr\left(s'|0\right) = \begin{cases} \frac{1}{R} & s = A, \ s' = B_1 \\ \frac{1}{R} & s = B_r, \ s' = B_{r+1}, \ 0 \le r < R \\ 0 & \text{otherwise}, \end{cases}$$

and

$$\Pr\left(s'|1\right) = \begin{cases} \frac{p}{(k-1)(1-p)-1} & s', s = A \\ \frac{1-p}{(k-1)(1-p)-1} & s = B_r, \ s' = B_{r+1}, \ R \le r \le R + k - 2 \\ \frac{1-p}{(k-1)(1-p)-1} & s = B_{R+k-1}, \ s' = A \\ 0 & \text{otherwise} \end{cases}.$$

First, we wish to show that regardless of the chosen phases $\{\phi_{xs}\}$ we get the equivalent quantum model. Recall that

the formula defining the overlaps is given by:

$$\Omega_{rs} = \sum_x \sqrt{\Pr(x|r)\Pr(x|s)} e^{i(\phi_{xs}-\phi_{xr})} \Omega_{f(r,s)f(x,s)} \ .$$

In this case, we have:

$$\Omega_{AB_{R+k-1}} = \sqrt{p} e^{i(\phi_{1B_{R+k-1}} - \phi_{1A})} \ ,$$
$$\Omega_{B_r B_s} = e^{i(\phi_{1B_r} - \phi_{1B_s})} \Omega_{B_{r+1} B_{s+1}} \ \text{and}$$
$$\Omega_{AB_r} = \sqrt{p} e^{i(\phi_{1B_r} - \phi_{1A})} \Omega_{AB_{r+1}} \ ,$$

which has the solution:

$$\frac{\Omega_{AB_{R+m}}}{\sqrt{p^{k-m}}} = e^{i\left(\sum_{j=m}^{k-1} \phi_{1B_{R+j}} - (k-m)\phi_{1A}\right)} \ \text{and}$$
$$\frac{\Omega_{B_{R+m} B_{R+n}}}{\sqrt{p^{m-n}}} = e^{i\left(\sum_{j=n}^{k-1} \phi_{1B_{R+j}} - \sum_{j=m}^{k-1} \phi_{1B_{R+j}} - (m-n)\phi_{1A}\right)} \ .$$

Note that under the gauge transformation $\Psi_A = k\phi_{1A}$ and $\Psi_{B_m} = \sum_{j=m}^{k-1} \phi_{1B_{R+j}} + m\phi_{1A}$, we can eliminate phases and end up simply with:

$$\begin{aligned}\Omega_{AB_{R+m}} &= \sqrt{p^{k-m}} \ \text{and}\\ \Omega_{B_{R+m} B_{R+n}} &= \sqrt{p^{m-n}}\end{aligned} \ . \tag{S13}$$

We note that this matrix only explicitly depends upon k and not R. This extends a result from Ref. [7] to all R and k, as well as to all choices of phase $\{\phi_{xs}\}$.

We can also apply these probabilities to compute the efficiency. The conditional entropies are:

$$\begin{aligned}\mathrm{H}\left[S'|X=0\right] &= \log R \ \text{and}\\ \mathrm{H}\left[S'|X=1\right] &= \log\left(k(1-p)+p\right) - \frac{(k-1)(1-p)}{k(1-p)+p}\log(1-p) \ .\end{aligned}$$

Under compression, the $X=0$ term does not change: $\mathrm{H}_q\left[S'|X=0\right] = \log R$. We will not compute the compressed term for $X=1$ since we need only note that it is a function of k and p and not of R.

The classical and quantum memories can be evaluated as:

$$\begin{aligned}C_\mu &= \log Z + \frac{1}{Z}C_\mu^{(\mathcal{K})} - \frac{1}{Z}(R-1)(1-p)\log(1-p) \ \text{and}\\ C_q &= \log Z + \frac{1}{Z}C_q^{(\mathcal{K})} - \frac{1}{Z}(R-1)(1-p)\log(1-p) \ ,\end{aligned}$$

where:

$$\begin{aligned}C_\mu^{(\mathcal{K})} &= (k-1)\log(1-p) \ \text{and}\\ C_q^{(\mathcal{K})} &= \mathrm{Tr}\left(\Omega Z P^{(\mathcal{K})} \log\left(\Omega Z P^{(\mathcal{K})}\right)\right) \ .\end{aligned}$$

are the contributions to complexity from only the states in $\mathcal{K}$. These contributions are only functions of k and p.

Then we see that the efficiency has the numerator and denominator:

$$\Delta_q C = \frac{1}{Z}\left(C_\mu^{(\mathcal{K})} - C_q^{(\mathcal{K})}\right) \ \text{and} \tag{S14}$$

$$\frac{\Delta_q W}{k_\mathrm{B} T \ln 2} = \frac{1}{Z}\left(C_\mu^{(\mathcal{K})} - C_q^{(\mathcal{K})}\right) + \frac{(k-1)(1-p)-1}{Z}\mathrm{H}_q\left[S'|X=1\right] \ . \tag{S15}$$

In this final form, we see that Z cancels in the ratio $e_q = k_{\mathrm{B}}T \ln 2\Delta_q C/\Delta_q W$, and all that remains are functions that depend only on k and p.

Nemo Generator

For the Nemo Generator, we know that not all phases $\{\phi_{xs}\}$ give equivalent implementations. To analyze the situation in more detail, we make use of the gauge invariants.

The gauge invariants of the Nemo implementations are:

$$\begin{aligned}
\Phi_0 &= \phi_{0A} \ , \\
\Phi_1 &= \phi_{1C} - \phi_{0C} \ , \text{ and} \\
\Phi_2 &= \phi_{1A} + \phi_{1B} + \phi_{1C} \ .
\end{aligned} \tag{S16}$$

We work to express the overlap matrix in terms of these invariants.

Recall that the formula defining the overlaps. For the Nemo process, this gives the system of equations:

$$\begin{aligned}
\Omega_{AB} &= \sqrt{1-p}\, e^{i(\phi_{1C}-\phi_{1A})} \ , \\
\Omega_{BC} &= \frac{1}{\sqrt{2}} e^{i(\phi_{1C}-\phi_{1B})} \ , \text{ and} \\
\Omega_{CA} &= \sqrt{\frac{p}{2}}\, e^{i(\phi_{0A}-\phi_{0C})} + \sqrt{\frac{1-p}{2}}\, e^{i(\phi_{1A}-\phi_{1C})}\Omega_{AB} \ ,
\end{aligned}$$

which has the solution:

$$\begin{aligned}
\Omega_{AB} &= \frac{\sqrt{p(1-p)}}{1+p} e^{i(\phi_{1C}-\phi_{1A}+\phi_{0A}-\phi_{0C})} \ , \\
\Omega_{BC} &= \frac{\sqrt{p}}{1+p} e^{i(\phi_{1C}-\phi_{1A}+\phi_{0A}-\phi_{0C})} \ , \text{ and} \\
\Omega_{CA} &= \frac{\sqrt{2p}}{1+p} e^{i(\phi_{0A}-\phi_{0C})} \ .
\end{aligned}$$

Now, we gauge fix ϕ_{1A} and ϕ_{1B} so that Ω_{AB} and Ω_{BC} are phaseless. The result is:

$$\begin{aligned}
\Omega_{AB} &= \frac{\sqrt{p(1-p)}}{1+p} \ , \\
\Omega_{BC} &= \frac{\sqrt{p}}{1+p} \ , \text{ and} \\
\Omega_{CA} &= \frac{\sqrt{2p}}{1+p} e^{i(2\Phi_0+2\Phi_1-\Phi_2)} \ .
\end{aligned} \tag{S17}$$

We see that the overlap matrix then only depends on the gauge invariants in the single phase $\Phi = 2\Phi_0 + 2\Phi_1 - \Phi_2$. This generalizes a result from Ref. [2] to all input phases $\{\phi_{xs}\}$.