Title

Thermodynamically-efficient local computation and the inefficiency of quantum memory compression

Journal

Physical Review Research, 2(2)

ISSN

2643-1564

Authors

Loomis, SP
Crutchfield, JP

Publication Date

2020-04-14

DOI

10.1103/PhysRevResearch.2.023039

Peer reviewed

Thermodynamically-efficient local computation and the inefficiency 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, California 95616, USA

(Received 8 January 2020; accepted 16 March 2020; published 14 April 2020)

Modularity dissipation identifies how locally implemented computation entails costs beyond those required by Landauer's bound on thermodynamic computing. We establish a general theorem for efficient local computation, giving the necessary and sufficient conditions for a local operation to have zero modularity cost. Applied to thermodynamically-generating stochastic processes it confirms a conjecture that classical generators are efficient if and only if they satisfy retrodiction, which places minimum-memory requirements on the generator. This extends immediately to quantum computation: Any quantum simulator that employs quantum memory compression *cannot* be thermodynamically efficient.

DOI: 10.1103/PhysRevResearch.2.023039

I. INTRODUCTION

Recently, Google AI announced a breakthrough in quantum supremacy, using a 54-qubit processor ("Sycamore") to complete a target computation in 200 seconds, claiming the world's fastest supercomputer would take more than 10 000 years to perform a similar computation [1]. Shortly afterward, IBM announced that they had proven the Sycamore circuit could be successfully simulated on the Summit supercomputer, leveraging its 250 PB storage and 200 petaFLOPS speed to complete the target computation in a matter of days [2]. This episode highlights two important aspects of quantum computing: first, the importance of memory and, second, the subtle relationship between computation and simulation.

Feynman [3] broached the notion that quantum computers would be singularly useful for the simulation of quantum processes, without supposing that this would also make them advantageous at simulating classical processes. Here, we explore issues raised by the recent developments in quantum computing, focusing on the problem of simulating classical stochastic processes via stochastic and quantum computers. We show that using quantum computers to simulate classical processes typically requires nonzero thermodynamic cost, while stochastic computers can theoretically achieve zero cost in simulating classical processes. This supports the viewpoint originally put forth by Feynman—that certain types of computers would each be advantageous at simulating certain physical processes—which challenges the current claims of quantum supremacy. Furthermore, we show that in both classical and quantum simulations, thermodynamic efficiency places a lower bound on the required memory of the simulator.

To demonstrate both, we must prove a new theorem on the thermodynamic efficiency of local operations. Correlation is a resource: it has been investigated as such in the formalism of *resource theories* [4] such as that of local operations with classical communication [5], with public communication [6], and many others, as well as the theory of local operations alone, under the umbrella term of *common information* [7–9]. Correlations have long been recognized as a thermal resource [10–13], enabling efficient computation to be performed when taken properly into account. Local operations that act only on part of a larger system are known to never increase the correlation between the part and the whole; most often, they are destructive to correlations and therefore resource-expensive.

Thermodynamic dissipation induced by a local operation—say on system A of a bipartite system AB to make a new joint system CB—is classically proportional to the difference in mutual information [14]:

$$\Delta S_{\text{loc}} = k_{\text{B}} T \left(I[A:B] - I[C:B] \right).$$

This can be asymptotically achieved for quantum systems [15]. By the data processing inequality [16,17], it is always nonnegative: $\Delta S_{\text{loc}} \geqslant 0$. Optimal thermodynamic efficiency is achieved when $\Delta S_{\text{loc}} = 0$.

To identify the conditions, in both classical and quantum computation, when this is so, we draw from prior results on saturated information-theoretic inequalities [18–24]. Specifically, by using a generalized notion of quantum sufficient statistic [24–27], we show that a local operation on part of a system is efficient if and only if it unitarily preserves the minimal sufficient statistic of the part for the whole. Our geometric interpretation of this also draws on recent progress on fixed points of quantum channels [28–31].

Paralleling previous results on ΔS_{loc} [14], our particular interest in locality arises from applying it to thermal transformations that generate and manipulate stochastic processes. This

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

is the study of *information engines* [12,13,32–35]. Rooted in computational mechanics [36–41], which investigates the inherent computational properties of natural processes and the resources they consume, information engines embed stochastic processes and Markovian generators in the physical world, where Landauer's bound for the cost of erasure holds sway [10].

A key result for information engines is the *information-processing second law* (IPSL): the cost of transforming one stochastic process into another by any computation is at least the difference in their Kolmogorov-Sinai entropy rates [33]. However, actual physical generators and transducers of processes, with their own internal memory dynamics, often exceed the cost required by the IPSL [14]. This arises from the temporal locality of a physical generator—only operating from time step to time step, rather than acting on the entire process at once. The additional dissipation ΔS_{loc} induced by this temporal locality gives the true thermodynamic cost of operating an information engine.

Previous work explored optimal conditions for a classical information engine to generate a process. Working from the hidden Markov model (HMM) [42] that determines an engine's memory dynamics, it was conjectured that the HMM must be *retrodictive* to be optimal. For this to hold, the current memory state must be a sufficient statistic of the *future* data for predicting the *past* data [14].

Employing a general result on conditions for reversible local computation, the following confirms this conjecture in the form of an equivalent condition on the HMM's structure. We then extend this, showing that it holds for quantum generators of stochastic processes [15,43–51]. Notably, quantum generators are known to provide potentially unbounded advantage in memory storage when compared with classical generators of the same process [44,45,47–50]. Surprisingly, the advantage is contingent: optimally-efficient generators—those with $\Delta S_{\text{loc}} = 0$—must not benefit from any memory compression. We show this to be true not only for previously published quantum generators but also for a new family of quantum generators derived from time reversal [49,52–54].

While important on its own, this also provides a complementary view of our previous result on quantum generators, which showed that a quantum-compressed generator is never less thermodynamically-efficient than the classical generator it compresses [15]. Combined with our current result, one concludes that a quantum-compressed generator is efficient with respect to the generator it compresses but, to the extent that it is compressed, it cannot be optimally efficient. In short, only classical retrodictive generators achieve the lower bound dictated by the IPSL. Practically, this highlights a pressing need to experimentally explore the thermodynamics of quantum computing.

II. THERMODYNAMICS OF QUANTUM INFORMATION RESERVOIRS

The physical setting of our work is in the realm of *information reservoirs*—systems all of whose states have the same energy level. Landauer's principle for quantum systems says that to change an information reservoir A from state ρ_A to state ρ'_A requires a work cost satisfying the lower bound

$$W \geqslant k_{\text{B}} T \ln 2 \, (\text{H}[\rho_A] - \text{H}[\rho'_A]), \tag{1}$$

where $\text{H}\rho_A$ is the von Neumann entropy [17]. Note that the lower bound $W_{\text{min}} := k_{\text{B}} T \ln 2 \, (H[\rho_A] - H[\rho'_A])$ is simply the change in free energy for an information reservoir. Furthermore, due to an information reservoir's trivial Hamiltonian, all of the work W becomes heat Q. Then the total entropy production—of system and environment—is

$$\Delta S := Q + k_{\text{B}} T \ln 2 \, \Delta \text{H}[A] = W - W_{\text{min}}. \tag{2}$$

Thus, not only does Landauer's principle provide the lower bound but it also reveals that any work exceeding W_{min} represents dissipation.

Reference [14] showed that Landauer's bound may indeed be attained in the quasistatic limit for any channel acting on a *classical* information reservoir. This result generally does not extend to single-shot quantum channels [55]. However, when we consider asymptotically many parallel applications of a quantum channel, we recover the tightness of Landauer's bound [15].

These statements are exceedingly general. To derive useful results, we must place further constraints on the system dynamics to see how Landauer's bound is affected. Reference [14] introduced the following perspective: Consider a bipartite information reservoir AB on which we wish to apply the local channel $\mathcal{E} \otimes I_B$, where $\mathcal{E} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_C)$ maps the states of system A into those of system C, transforming the initial joint state ρ_{AB} to the final state ρ_{CB}. The Landauer bound for $AB \to CB$ is given by $W_{\text{min}} = k_{\text{B}} T \ln 2 \, (\text{H}[\rho_{AB}] - \text{H}[\rho_{CB}])$. However, since we constrained ourselves to use local manipulations, the lowest achievable bound is actually $W_{\text{loc}} := k_{\text{B}} T \ln 2 \, (\text{H}[\rho_A] - \text{H}[\rho_C])$. Thus, we must have dissipation of at least

$$\begin{aligned} \Delta S &\geqslant W_{\text{loc}} - W_{\text{min}} \\ &= k_{\text{B}} T \ln 2 \, (\text{H}[\rho_A] - \text{H}[\rho_{AB}] - \text{H}[\rho_C] + \text{H}[\rho_{CB}]) \\ &= k_{\text{B}} T \ln 2 \, (\text{I}[A : B] - \text{I}[C : B]), \end{aligned} \tag{3}$$

where $\text{I}[A : B] = \text{H}[\rho_A] + \text{H}[\rho_B] - \text{H}[\rho_{AB}]$ is the quantum mutual information. And so we have a minimal *locality dissipation*:

$$\Delta S_{\text{loc}} := k_{\text{B}} T \ln 2 \, (\text{I}[A : B] - \text{I}[C : B]), \tag{4}$$

which arises because we did not use the correlations to facilitate our erasure. See Fig. 1 for a simple example of this phenomenon.

This local form of Landauer's principle is still highly general, but the following shows how to examine it for specific classical and quantum computational architectures The key question we ask is the following: For which architectures can ΔS_{loc} be made to exactly vanish? We first we consider this problem generally and then provide a solution.

III. REVERSIBLE LOCAL COMPUTATION

Suppose we are given a bipartite system AB with state ρ_{AB}. We wish to determine the conditions for a local channel $\mathcal{E}_A \otimes I_B$ that maps A to C:

$$\rho'_{CB} = \mathcal{E}_A \otimes I_B(\rho_{AB}), \tag{5}$$

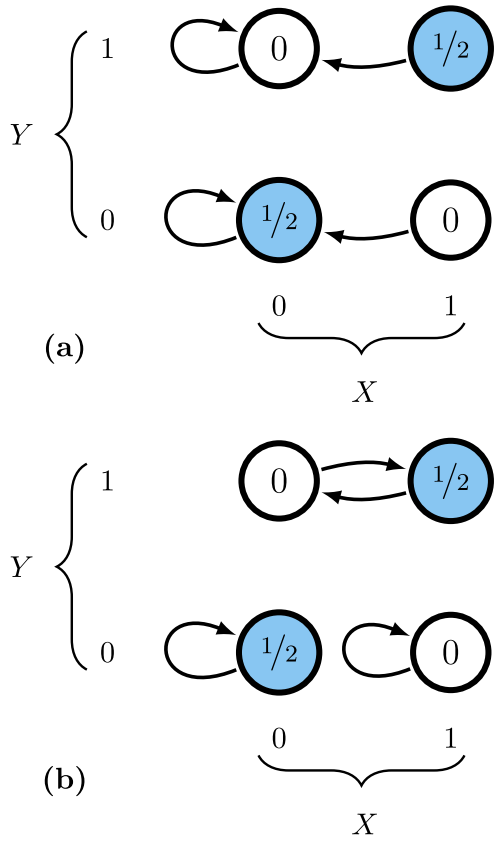

FIG. 1. Thermodynamics of locality: Suppose we have two bits XY in a correlated state where $\frac{1}{2}$ probability is in $XY = 00$ and $\frac{1}{2}$ probability is in $XY = 11$. (a) A thermodynamically irreversible operation can be performed to erase only X (that is, set $X = 0$ without changing Y) if we are not allowed to use knowledge about the state of Y. (b) A reversible operation can be performed to erase X if we are allowed to use knowledge about Y. Both operations have the same outcome given our initial condition, but the nonlocal operation (a) is more thermodynamically costly because it is irreversible. According to Theorem 1, operation (a) is costly since it erases information in X that is correlated with Y.

to preserve the mutual information $I[A : B] = I[C : B]$. Proofs of the following results are provided in the Supplemental Material (SM) [56].

Stating our result requires first defining the quantum notion of a sufficient statistic. Previously, quantum sufficient statistics of A for B were defined when AB is a classical-quantum state [27]. That is, when ρ_{AB} commutes with a local measurement on A. They were also introduced in the setting of sufficient statistics for a family of states [24,25]. This corresponds to the case where AB is quantum-classical—ρ_{AB} commutes with a local measurement on B. Our definition generalizes these cases to fully-quantal correlations between A and B.

We start, as an example, by giving the following definition of a minimum sufficient statistic of a classical joint random variable $XY \sim \Pr(x, y)$ in terms of an equivalence relation. We define the predictive equivalence relation $\sim$ for which $x \sim x'$ if and only if $\Pr(y|x) = \Pr(y|x')$ for all y. The *minimum sufficient statistic* (MSS) $[X]_Y$ is given by the equivalence classes $[x]_Y := \{x' : x \sim x'\}$. Let us denote $\Sigma := [X]_Y$ and let $\Pr(y|\sigma) := \Pr(y|x)$ for any $x \in \sigma$.

This cannot be *directly* generalized to the quantum setting since correlations between A and B cannot always be described in the form of states conditioned on the outcome of a local measurement on A. If the latter were the case, the state would be classical-quantum, but general quantum correlations can be much more complicated than these. However, we can take the most informative local measurement that does not disturb ρ_{AB} and then consider the "atomic" quantum correlations it leaves behind.

Let ρ_{AB} be a bipartite quantum state. A *maximal local commuting measurement* (MLCM) of A for B is any local measurement X with projectors $\{\Pi^{(x)}\}$ on system A such that

$$\rho_{AB} = \bigoplus_x \Pr(X = x)\rho_{AB}^{(x)}, \tag{6}$$

where

$$\Pr(X = x) = \mathrm{Tr}\big[\big(\Pi_X^{(x)} \otimes I_B\big)\rho_{AB}\big], \tag{7}$$

and

$$\Pr(X = x)\rho_{AB}^{(x)} = \big(\Pi_X^{(x)} \otimes I_B\big)\rho_{AB}\big(\Pi_X^{(x)} \otimes I_B\big), \tag{8}$$

and any further local measurement Y on $\rho_{AB}^{(x)}$ disturbs the state:

$$\rho_{AB}^{(x)} \neq \sum_y \big(\Pi_Y^{(y)} \otimes I_B\big)\rho_{AB}^{(x)}\big(\Pi_Y^{(y)} \otimes I_B\big). \tag{9}$$

We call the states $\{\rho_{AB}^{(x)}\}$ *quantum correlation atoms*.

Proposition 1: MLCM uniqueness. Given a state ρ_{AB}, there is a unique MLCM of A for B.

Now, as in the classical setting, we define an equivalence class over the values of the MLCM via the equivalence between their quantum correlation atoms. Classically, these atoms are simply the conditional probability distributions $\Pr(\cdot|x)$; in the classical-quantum setting, they are the conditional quantum states $\rho_B^{(x)}$. Note that each is defined as a distribution on the variable Y or system B. In contrast, the general quantum correlation atoms $\rho_{AB}^{(x)}$ depend on both systems A and B.

The resulting challenge is resolved in the following way: Let ρ_{AB} be a bipartite quantum state and let X be the MLCM of A for B. We define the *correlation equivalence* relation $x \sim x'$ over values of X where $x \sim x'$ if and only if $\rho_{AB}^{(x)} = (U \otimes I_B)\rho_{AB}^{(x')}(U^\dagger \otimes I_B)$ for a local unitary U.

Finally, we define the *minimal local sufficient statistic* (MLSS) $[X]_\sim$ as the equivalence class $[x]_\sim := \{x' : x' \sim x\}$ generated by the relation $\sim$ between correlation atoms. Thus, our notion of sufficiency of A for B is to find the most informative local measurement and then coarse-grain its outcomes by unitary equivalence over their correlation atoms. The correlation atoms and the MLSS $[X]_\sim$ together describe the correlation structure of the system AB.

The machinery is now in place to state our result. The proof depends on previous results regarding the fixed points of stochastic channels [28–31] and saturated information-theoretic inequalities [18–24]. This background and the proof are described in the SM.

Theorem 1: Reversible local operations. Let ρ_{AB} be a bipartite quantum state and let $\mathcal{E}_A \otimes I_B$ be a local operation

with $\mathcal{E}_A : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_C)$. Suppose X is the MLCM of ρ_{AB} and Y, that of $\rho_{CB} = \mathcal{E}_A \otimes I_B(\rho_{AB})$. The decomposition into correlation atoms is

$$\rho_{AB} = \bigoplus_x \Pr_A(x)\rho_{AB}^{(x)}, \tag{10}$$

$$\rho_{CB} = \bigoplus_y \Pr_C(y)\rho_{CB}^{(y)}. \tag{11}$$

Then, $I[A:B] = I[C:B]$ if and only if $\mathcal{E}_A$ can be expressed by Kraus operators of the form:

$$K^{(\alpha)} = \bigoplus_{x,y} e^{i\phi_{xy\alpha}} \sqrt{\Pr(y,\alpha|x)}\, U^{(y|x)}, \tag{12}$$

where $\phi_{xy\alpha}$ is any arbitrary phase and $\Pr(y,\alpha|x)$ is a stochastic channel that is nonzero only when $\rho_{AB}^{(x)}$ and $\rho_{CB}^{(y)}$ are equivalent up to a local unitary operation $U^{(y|x)}$ that maps $\mathcal{H}_A^{(x)}$ to $\mathcal{H}_C^{(y)}$.

The theorem's classical form follows as a corollary:

Corollary 1: Reversible local operations, classical. Let XY be a joint random variable and let $\Pr(Z=z|X=x)$ be a channel from $\mathcal{X}$ to some set $\mathcal{Z}$, resulting in the joint random variable ZY. Then $I[X:Y] = I[Z:Y]$ if and only if $\Pr(Z=z|X=x) > 0$ only when $\Pr(Y=y|Z=z) = \Pr(Y=y|X=x)$ for all y.

In light of the previous section, there is a simple thermodynamic interpretation of Theorem 1 and Corollary 1: local channels that circumvent dissipation due to their locality (i.e., those which have $\Delta S_{\text{loc}} = 0$) are precisely those channels that preserve the sufficiency structure of the joint state. They may create and destroy any information that is not stored in the sufficient statistic and the correlation atoms. However, the sufficient statistic itself must be conserved and the correlation atoms must be only unitarily transformed.

We now turn to apply this perspective to classical and quantum *generators*—systems that use thermodynamic mechanisms to produce stochastic processes. We compute the necessary and sufficient conditions for these generators to have zero locality dissipation: $\Delta S_{\text{loc}} = 0$. And so, in this way we determine precise criteria for when they are thermodynamically efficient.

IV. THERMODYNAMICS OF CLASSICAL GENERATORS

A classical generator is the physical implementation of a hidden Markov model (HMM) [42] $\mathfrak{G} = (\mathcal{S}, \mathcal{X}, \{T^{(x)}_{s's}\})$, where (here) $\mathcal{S}$ is countable, $\mathcal{X}$ is finite, and for each $x \in \mathcal{X}$, $\mathbf{T}^{(x)}$ is a matrix with values given by a stochastic channel from $\mathcal{S}$ to $\mathcal{S} \times \mathcal{Y}$, $T^{(x)}_{s's} := \Pr_{\mathfrak{G}}(s', x|s)$. We define generators to use *recurrent* HMMs, which means the total transition matrix $T_{s's} := \sum_x T^{(x)}_{s's}$ is irreducible. In this case, there is a unique stationary distribution $\pi_{\mathfrak{G}}(s)$ over states $\mathcal{S}$ satisfying $\pi_{\mathfrak{G}}(s) > 0$, $\sum_s \pi_{\mathfrak{G}}(s) = 1$, and $\sum_s T_{s's}\pi_{\mathfrak{G}}(s) = \pi_{\mathfrak{G}}(s')$.

During its operation, a generator's function is to produce a *stochastic process*—for each ℓ, a probability distribution $\Pr_{\mathfrak{G}}(x_1 \ldots x_\ell)$ over words $x_1 \ldots x_\ell \in \mathcal{X}^\ell$. The probabilities for words of length ℓ generated by $\mathfrak{G}$ are defined by

$$\Pr_{\mathfrak{G}}(x_1 \ldots x_\ell) := \sum_{s_0,\ldots,s_\ell \in \mathcal{S}^{\ell+1}} T^{(x_\ell)}_{s_\ell s_{\ell-1}} \cdots T^{(x_1)}_{s_1 s_0} \pi(s_0). \tag{13}$$

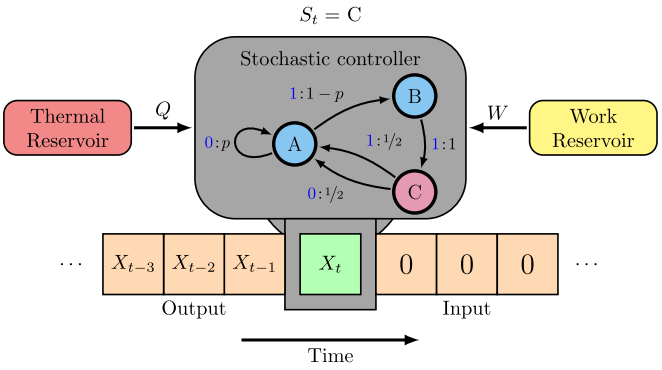

FIG. 2. Information ratchet sequentially generating a symbol string on an empty tape: At time step t, S_t is the random variable for the ratchet state. The generated symbols in the generated (output) process are denoted by X_{t-1}, X_{t-2}, X_{t-3}, ... The most recently generated symbol X_t (green) is determined by the internal dynamics of the ratchet's memory by using heat Q from the thermal reservoir as well as work W from the work reservoir. (*Inside ratchet*) Ratchet memory dynamics and symbol emission 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. Graphically, this is represented by a hidden Markov model, depicted here as a state-transition diagram in which nodes are states s and edges represent transitions $s \to s'$ labeled by the generated symbol and associated probability: $x : \Pr(s', x|s)$. (Reproduced with permission from Ref. [15].)

Typically, we view a generator as operating over discrete time, writing out a sequence of symbols from $x \in \mathcal{X}$ on a tape, while internally transforming its memory state; see Fig. 2. Starting with an initial state $S_0 \sim \pi(s)$ and empty tape at time $t = 0$, the entire system at time t is described by the joint random variable $X_1 \ldots X_t S_t$, with distribution

$$\Pr_{\mathfrak{G}}(x_1 \ldots x_t, s_t) := \sum_{s_0,\ldots,s_{t-1} \in \mathcal{S}^t} T^{(x_t)}_{s_t s_{t-1}} \cdots T^{(x_1)}_{s_1 s_0} \pi_{\mathfrak{G}}(s_0). \tag{14}$$

Continuing this technique, one can compute the joint random variable $X_1 \ldots X_t S_t X_{t+1} S_{t+1}$.

This picture of a generator as operating on a tape while continually erasing and rewriting its internal memory allows us to define the possible thermodynamics, also shown in Fig. 2. Erasure generally requires work, drawn from the work reservoir, while the creation of noise often allows the extraction of work, which is represented in our sign convention by drawing negative work from the reservoir. Producing a process $X_1 \ldots X_t \sim \Pr(x_1 \ldots x_t)$ of length t has an associated work cost $W \geqslant -k_B T \ln 2 H[X_1 \ldots X_t]$. The negative sign, as discussed, indicates work $k_B T \ln 2 H[X_1 \ldots X_t]$ may be transferred from the thermal reservoir to the work reservoir. For large t, this can be asymptotically expressed by the work rate $W/t \geqslant -k_B T \ln 2\, h_\mu$, where

$$h_\mu := \lim_{t\to\infty} \frac{1}{t} \mathrm{H}[X_1 \ldots X_t] \tag{15}$$

is the process's *Kolmogorov-Sinai entropy rate* [33]. This is a reasonable description of the average entropy rate of a process that is *stationary*—that is, $\Pr(X_t \ldots X_{t+\ell} = x_1 \ldots x_{\ell-1})$ is independent of t—and *ergodic*. Said differently, for large

t a typical realization $x_1 \ldots x_t$ contains the word $\widehat{x}_1 \ldots \widehat{x}_\ell$ approximately $t \times \Pr(\widehat{x}_1 \ldots \widehat{x}_\ell)$ times. Recurrent generators produce exactly these sorts of processes.

Now, a given generator cannot necessarily be implemented as efficiently as the minimal work rate $W_{\min} := -k_B T \ln 2\, h_\mu$ indicates. This is because a generator acts temporally locally, only being able to use its current memory state to generate the next memory state and symbol. The true cost at time t must be bounded below by $W_{\rm loc} := W_{\min} + \Delta S_{\rm loc}$, where in this case the locality dissipation is [14]

$$\Delta S_{\rm loc} = k_B T \ln 2 \left(I[S_t : X_1 \ldots X_t] - I[S_{t+1} X_{t+1} : X_1 \ldots X_t]\right). \tag{16}$$

In this case, the dissipation does not represent work lost to heat but rather the increase in tape entropy that did not facilitate converting heat into work. To understand this in some detail, this section identifies the necessary and sufficient conditions for efficient generators—those with $\Delta S_{\rm loc} = 0$.

To state our result for classical generators, we must introduce two further notions regarding generators. As before, proofs of results are given in the SM. Consider a partition of $\mathcal{S}$: $\mathcal{P} = \{\mathcal{P}_\theta\}$, $\mathcal{P}_\theta \cap \mathcal{P}_{\theta'}$, $\bigcup_\theta \mathcal{P}_\theta = \mathcal{S}$, labeled by index θ. Let

$$\Pr_{\mathfrak{G}^{\mathcal{P}}}(\theta', x|\theta) := \sum_{\substack{s' \in \mathcal{P}_{\theta'} \\ s \in \mathcal{P}_\theta}} \Pr_{\mathfrak{G}}(s', x|s)\pi(s|\theta), \tag{17}$$

with $\pi_{\mathfrak{G}}(s|\theta) = \pi_{\mathfrak{G}}(s)/\pi_{\mathfrak{G}^{\mathcal{P}}}(\theta)$ and $\pi_{\mathfrak{G}^{\mathcal{P}}}(\theta) = \sum_{s \in \mathcal{P}_\theta} \pi_{\mathfrak{G}}(s)$. We say a partition $\{\mathcal{P}_\theta\}$ is *mergeable* with respect to the generator $\mathfrak{G} = (\mathcal{S}, \mathcal{X}, \{T^{(x)}_{s's}\})$ if the merged generator $\mathfrak{G}^{\mathcal{P}} = (\mathcal{P}, \mathcal{X}, \{\widetilde{T}^{(x)}_{\theta'\theta}\})$, with transitions $\widetilde{T}^{(x)}_{\theta'\theta} := \Pr(\theta', x|\theta)$, generates the same process as the original.

Pertinent to our goals here is the notion of *retrodictive equivalence*. Let $\Pr_{\mathfrak{G}}(x_1 \ldots x_t|s_t) := \Pr_{\mathfrak{G}}(x_1 \ldots x_t, s_t)/\pi_{\mathfrak{G}}(s_t)$. Given two states $s, s' \in \mathcal{S}$ of a generator $(\mathcal{S}, \mathcal{X}, \{T^{(x)}_{s's}\})$, we say that $s \sim s'$ if $\Pr_{\mathfrak{G}}(x_1 \ldots x_t|s) = \Pr_{\mathfrak{G}}(x_1 \ldots x_t|s')$ for all words $x_1 \ldots x_t$. The equivalence class $[S_t]_\sim$ is the sufficient statistic of S_t for predicting the past symbols $X_1 \ldots X_t$. The set $\mathcal{P}_\sim := \{[s]_\sim : s \in \mathcal{S}\}$ of equivalence classes is a partition on $\mathcal{S}$ that we index by σ.

Proposition 2. Given a generator $(\mathcal{S}, \mathcal{X}, \{T^{(x)}_{s's}\})$, the partition $\mathcal{P} := \mathcal{P}_\sim$ induced by retrodictive equivalence is mergeable.

We now state our theorem for efficient classical generators:

Theorem 2. A generator $\mathfrak{G} = (\mathcal{S}, \mathcal{X}, \{T^{(x)}_{s's}\})$ satisfies $I[S_t : X_1 \ldots X_t] = I[S_{t+1} X_{t+1} : X_1 \ldots X_t]$ for all t if and only if the retrodictively state-merged generator $\mathfrak{G}^{\mathcal{P}} = (\mathcal{P}_\sim, \mathcal{X}, \{\widetilde{T}^{(x)}_{\sigma'\sigma}\})$ satisfies $\widetilde{T}^{(x)}_{\sigma'\sigma} \propto \delta_{\sigma, f(\sigma', x)}$ for some function $f : \mathcal{S} \times \mathcal{X} \to \mathcal{S}$.

We say that a generator $\mathfrak{G} = (\mathcal{S}, \mathcal{X}, \{T^{(x)}_{s's}\})$ satisfying $T^{(x)}_{s's} \propto \delta_{s, f(s', x)}$ for some f is *co-unifilar*. The dual property $T^{(x)}_{s's} \propto \delta_{s', f(s, x)}$ for some f is called *unifilar* [57]. For every process, there is a unique generator, called the *reverse ϵ-machine,* constructed by retrodictively state-merging any co-unifilar generator [50]. Similarly, by using a different partition called *predictive* equivalence on states, any unifilar generator for a process can be state-merged into a unique generator called the *forward ϵ-machine* of that process [50].

The reverse ϵ-machine has the following property: Let $\overrightarrow{X}_t := X_{t+1} X_{t+1} \ldots$ represent all future generated symbols, the reverse ϵ-machine state Σ_t at time t is the minimum sufficient statistic of $\overrightarrow{X}_t$ for predicting $X_1 \ldots X_t$. Any generator whose state S_t is a sufficient statistic of $\overrightarrow{X}_t$ for $X_1 \ldots X_t$ is called a *retrodictor*. The reverse ϵ-machine can then be considered the minimal retrodictor.

Reference [14] conjectured that the necessary and sufficient condition for $\Delta S_{\rm loc} = 0$ is that the generator in question is a retrodictor. In the SM we confirm this by establishing that the conditions of Theorem 2 imply that the generator is a retrodictor.

A similar result, for classical generators, was presented in Ref. [34] where a lower bound on $\Delta S_{\rm loc}$ was derived for predictive generators [Eq. (A23) in Ref. [34]]. A consequence of this bound is that $\Delta S_{\rm loc} = 0$ only when the predictor is also a retrodictor. However, this bound does not extend to nonpredictive generators. In contrast, Theorem 2 applies to all generators.

Our result is complemented by another recent result [35] which demonstrated how, from a predictive generator, one can construct a sequence of generators that asymptotically approach a retrodictor and whose dissipation $\Delta S_{\rm loc}$ asymptotically approaches zero. Helpfully, this result points to possible perturbative extensions of Theorem 2.

These results bear on the trade-off between dissipation and *memory* for classical generators. The reverse (forward) ϵ-machine, being a state-merging of any co-unifilar (unifilar) generator, is minimal with respect to the co-unifilar (unifilar) generators via all quantifications of the memory, such as the number of memory states $|\mathcal{S}|$ and the entropy $H[S]$ [50].

As a consequence, we now see that the above showed that any *thermodynamically efficient* generator can be state-merged into a co-unifilar generator. This means it can be further state-merged into the reverse ϵ-machine of the process it generates. In short, thermodynamic efficiency comes with a memory constraint. And, when the memory falls below this constraint, *dissipation must be present*.

V. THERMODYNAMICS OF QUANTUM MACHINES

A process's forward ϵ-machine, a key player in the previous section, may be concretely defined as the unique generator $\mathfrak{G} = (\mathcal{S}, \mathcal{X}, \{T^{(x)}_{s's}\})$ for a given process satisfying [38]

(1) *Recurrence*: $T_{s's} := \sum_x T^{(x)}_{s's}$ is an irreducible matrix;

(2) *Unifilarity*: $T^{(x)}_{s's} > 0$ only when $s' = f(s, x)$ for some function $f : \mathcal{S} \times \mathcal{X} \to \mathcal{S}$;

(3) *Predictively Distinct States*: $\Pr(x_t x_{t+1} \ldots x_\ell|s_t) = \Pr(x_t x_{t+1} \ldots x_\ell|s'_t)$ for all ℓ and $x_t x_{t+1} \ldots x_\ell$ implies $s_t = s_{t'}$.

ϵ-machines are a process's minimal unifilar generators, in the sense that they are smallest with respect to the number of memory states $|\mathcal{S}|$, the entropy $H[S]$, and all other ways of measuring memory, such as the Rényi entropies $H_\alpha[S] := \frac{1}{1-\alpha} \log_2[\sum_s \pi_{\mathfrak{G}}(s)^\alpha]$. In this, they are unique.

However, one can implement ϵ-machines with even lower memory costs, by encoding them in a quantum system and generating symbols by means of a noisy measurement. This encoding is called a *q-machine*. In terms of qubits, as a unit of size, these implementations can generate the same

process at a much lower memory cost than the ϵ-machine's bit-based memory cost. It has also been shown that these quantum implementations have a lower locality cost W_{loc} than their corresponding ϵ-machine, and so they are more thermodynamically efficient [15]. This section identifies the constraints for quantum generators to have zero dissipation; that is, $\Delta S_{\text{loc}} = 0$. We show that this results in a peculiar pair of constraints. First, the forward ϵ-machine memory must not be smaller than the memory of the reverse ϵ-machine. (This mirrors the results of Theorem 2 in the SM.) Second, the quantum generator achieves no compression. That is, the memory of the quantum generator in qubits is precisely the memory of the forward ϵ-machine in bits. Thus, compression of memory and perfect thermodynamic efficiency are exclusive outcomes.

To state this precisely, we review q-machines and introduce several new definitions to capture their properties (see the SM for the proofs).

Given a forward ϵ-machine $\mathfrak{G} = (\mathcal{S}, \mathcal{X}, \{T^{(x)}_{s's}\})$, for any set of phases $\{\phi_{xs} : x \in \mathcal{X}, s \in \mathcal{S}\}$ there is an encoding $\{|\psi_s\rangle : s \in \mathcal{S}\}$ of the memory states $\mathcal{S}$ into a Hilbert space $\mathcal{H}_S$ and a set of Kraus operators $\{K^{(x)} : x \in \mathcal{X}\}$ on said Hilbert space such that

$$K^{(x)} |\psi_s\rangle = e^{i\phi_{xs}} \sqrt{T^{(x)}_{f(s,x),s}} \, |\psi_{f(s,x)}\rangle \, . \tag{18}$$

This expression implicitly defines the Kraus operators given the encoding $\{|\psi_s\rangle\}$. The encoding, in turn, is determined up to a unitary transformation by the following constraint on their overlaps:

$$\langle\psi_r|\psi_s\rangle = \sum_{x\in\mathcal{X}} e^{i(\phi_{xs}-\phi_{xr})} \sqrt{T^{(x)}_{r',r} T^{(x)}_{s',s}} \, \langle\psi_{r'}|\psi_{s'}\rangle \, , \tag{19}$$

where $r' = f(r, x)$ and $s' = f(s, x)$. This equation has a unique solution for every choice of phases $\{\phi_{xs}\}$ [51].

We note that if $\pi_{\mathfrak{G}}(s)$ is the ϵ-machine's stationary distribution, then the stationary state of this quantum generator is given by the ensemble

$$\rho_\pi = \sum_s \pi_{\mathfrak{G}}(s) \, |\psi_s\rangle \, \langle\psi_s| \tag{20}$$

and satisfies

$$\rho_\pi = \sum_x K^{(x)} \rho_\pi K^{(x)\dagger} . \tag{21}$$

When we say that a quantum generator *uses less memory than its classical counterpart*, we mean that $\dim \mathcal{H}_S \leqslant |\mathcal{S}|$, $\mathrm{H}[\rho_\pi] \leqslant \mathrm{H}[S]$, and further that $\mathrm{H}_\alpha[\rho_\pi] \leqslant \mathrm{H}_\alpha[S]$, where $\mathrm{H}_\alpha[\rho_\pi] := \frac{1}{1-\alpha} \log_2 \mathrm{Tr}[\rho_\pi^\alpha]$ are the Rényi-von Neumann entropies [44,49,50].

To see this quantum generator as a physical system, as in Fig. 2, requires us to interpret the tape as being written on as a series of copies of a single Hilbert space $\mathcal{H}_A$ that represents one cell on the tape. On $\mathcal{H}_A$ we define the computational basis $\{|x\rangle : x \in \mathcal{X}\}$ in which outputs will be written. The system at time t can be described by using the joint Hilbert space $\mathcal{H}_{A_1} \otimes \mathcal{H}_{A_t} \otimes \mathcal{H}_S$, where each $\mathcal{H}_{A_k}$ is unitarily equivalent to $\mathcal{H}_A$, and the state is

$$\rho_{\mathfrak{G}}(t) := \sum_{x_1\ldots x_t} |x_1 \ldots x_t\rangle \, \langle x_1 \ldots x_t| \otimes K^{(x_t\ldots x_1)} \rho_\pi K^{(x_t\ldots x_1)\dagger} , \tag{22}$$

where $K^{(x_t\ldots x_1)} = K^{(x_t)} \cdots K^{(x_1)}$ and $|x_1 \ldots x_t\rangle = |x_1\rangle_{A_1} \otimes \cdots \otimes |x_t\rangle_{A_t}$. From this we get the process generated by the ϵ-machine and quantum generator in terms of the Kraus operators as

$$\Pr_{\mathfrak{G}}(x_1 \ldots x_t) := \mathrm{Tr}[K^{(x_t\ldots x_1)} \rho_\pi K^{(x_t\ldots x_1)\dagger}]. \tag{23}$$

Let us now briefly discuss the thermodynamic properties of quantum generators, homing in on our main result about conditions for their efficiency. The previous section discussed how a process, to be generated, requires the minimal work rate $W_{\text{min}} = -k_{\text{B}} T \ln 2 \, h_\mu$. However, this is not typically achievable for classical generators. The same principle holds for quantum generators: Since they act temporally locally, the true cost at time t is bounded below by $W_{\text{loc}} = W_{\text{min}} + \Delta S_{\text{loc}}$ and the locality dissipation ΔS_{loc} has the same form:

$$\Delta S_{\text{loc}} = k_{\text{B}} T \ln 2 \, (I[S_t : A_1 \ldots A_t] - I[S_{t+1} A_{t+1} : A_1 \ldots A_t]). \tag{24}$$

There are two crucial differences, though. First, the mutual information I above is the *quantum* mutual information derived from the von Neumann entropy. Second, even the work rate W_{loc} is not necessarily achievable in the single-shot case [55]. However, it may be attained for asymptotically parallel generation [15]. We will not concern ourselves with this second problem here. Our intent is to focus, as in the previous section, on the necessary and sufficient conditions for $\Delta S_{\text{loc}} = 0$.

The preceding material was, in fact, review. We now introduce a simple partition that may be constructed on the memory states of the ϵ-machine for a given quantum implementation. Specifically, we define the *maximal commuting partition* (MCP) on $\mathcal{S}$ to be the most refined partition $\{\mathcal{B}_\theta\}$ such that the overlap matrix $\langle\psi_r|\psi_s\rangle$ is block diagonal. That is, $\{\mathcal{B}_\theta\}$ is such that $\langle\psi_r|\psi_s\rangle = 0$ if $r \in \mathcal{B}_\theta$ and $s \in \mathcal{B}_{\theta'}$ for $\theta \neq \theta'$.

Our result on thermodynamically-efficient quantum generators is as follows:

Theorem 3: Maximally-efficient quantum generator. Let $\mathfrak{G} = (\mathcal{S}, \mathcal{X}, \{T^{(x)}_{s's}\})$ be a given process's ϵ-machine. Suppose we build from it a quantum generator with encoding $\{\psi_s\}$ and Kraus operators $\{K^{(x)}\}$. Let $\mathcal{B} := \{\mathcal{B}_\theta\}$ be the MCP of $\mathcal{S}$. Then the quantum generator has $\Delta S_{\text{loc}} = 0$ if and only if the partition $\mathcal{B}$ is trivially maximal—in that $|\mathcal{B}_\theta| = 1$ for each θ—and the retrodictively state-merged generator $\mathfrak{G}^{\mathcal{B}}$ of $\mathfrak{G}$ is co-unifilar.

We previously found that, in the limit of asymptotically parallel generation, a quantum generator is always more thermodynamically efficient than its corresponding ϵ-machine, in that it has a lower dissipation [15]. Yet this does not imply that dissipation can be made to vanish for quantum generators of a process. In fact, only for processes whose forward ϵ-machine is also a retrodictor can dissipation be made to vanish. In these cases, the memory states will be orthogonally encoded, and so no memory compression is achieved, which is seen

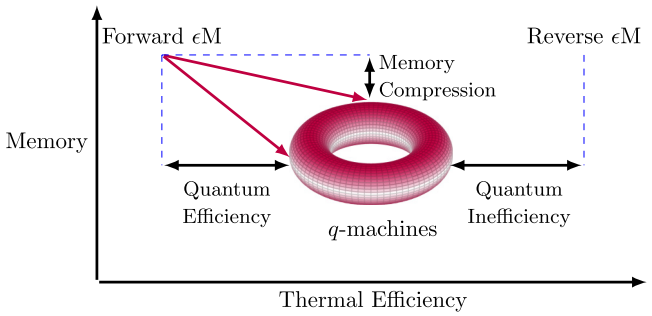

FIG. 3. Performance trade-offs for q-machines, whose variety and dependence on phases $\{\phi_{xs}\}$ is depicted by a torus: Under all ways of quantifying memory, the q-machines constructed from a predictor achieve nonnegative memory compression [50], and they also have a smaller dissipation ΔS_{loc}, rendering them more thermodynamically efficient [15]. However, to achieve positive compression, they must also have a nonzero ΔS_{loc}, rendering them less efficient than a classical retrodictor.

by the trivial maximality of $\mathcal{B}$. The situation is heuristically represented in Fig. 3.

VI. THERMODYNAMICS OF REVERSE q-MACHINES

We showed that forward ϵ-machines compressed via the q-machine cannot achieve the efficiency of a classical retrodictor. However, one may wonder what happens to a retrodictor's optimal efficiency if it is directly compressed. We now demonstrate a method for such compression, derived from the time-reversal of the q-machine, and prove that even here any nonzero compression of memory precludes optimal efficiency.

A process's reverse ϵ-machine may be defined similarly to the forward ϵ-machine as the unique generator $\mathfrak{G} = (\mathcal{S}, \mathcal{X}, \{T^{(x)}_{s's}\})$ for a given process satisfying:

(1) *Recurrence*: $T_{s's} := \sum_x T^{(x)}_{s's}$ is an irreducible matrix;

(2) *Co-unifilarity*: $T^{(x)}_{s's} > 0$ only when $s = f(s', x)$ for some function $f : \mathcal{S} \times \mathcal{X} \to \mathcal{S}$;

(3) *Retrodictively Distinct States*: $\Pr(x_1 \ldots x_t|s_t) = \Pr(x_1 \ldots x_t|s'_t)$ for all t and $x_1 \ldots x_t$ implies $s_t = s_{t'}$.

Reverse ϵ-machines are a process's minimal co-unifilar generators, in the sense that they are smallest with respect to the number of memory states $|\mathcal{S}|$, the entropy $\mathrm{H}[S]$, and all other ways of measuring memory, such as the Rényi entropies $\mathrm{H}_\alpha[S] := \frac{1}{1-\alpha}\log_2[\sum_s \pi_{\mathfrak{G}}(s)^\alpha]$.

There is an intricate relationship between forward and reverse ϵ-machines that can only be appreciated in the language of time reversal. The time-reverse of a generator $\mathfrak{G} = (\mathcal{S}, \mathcal{X}, \{T^{(x)}_{s's}\})$ is the generator $\widetilde{\mathfrak{G}} = (\mathcal{S}, \mathcal{X}, \{\widetilde{T}^{(x)}_{s's}\})$ where $\widetilde{T}^{(x)}_{s's} = \pi_s T^{(x)}_{s's}/\pi_{s'}$ [54]. The generator $\widetilde{\mathfrak{G}}$ is associated with the reverse process, $\Pr_{\widetilde{\mathfrak{G}}}(x_1 \ldots x_t) = \Pr_{\mathfrak{G}}(x_t \ldots x_1)$. Note that time reversal preserves both the state space $\mathcal{S}$ and the stationary distribution π_s.

Given a process's forward ϵ-machine $\mathfrak{F}$, its time reverse $\widetilde{\mathfrak{F}}$ is the *reverse* ϵ-machine of the *reverse* process. Conversely, given a process's reverse ϵ-machine $\mathfrak{G}$, its time reverse $\widetilde{\mathfrak{G}}$ is the *forward* ϵ-machine of the *reverse* process. Since the stationary distribution and state space are preserved under time reversal, $\mathfrak{F}$ and $\widetilde{\mathfrak{F}}$ have the same memory costs, as do $\mathfrak{G}$ and $\widetilde{\mathfrak{G}}$. However, somewhat surprisingly, this does not mean that $\mathfrak{F}$ and $\mathfrak{G}$ have the same memory costs [53].

Previous work compared the results of compressing the forward ϵ-machine $\mathfrak{F}$ of a process *and* the forward ϵ-machine $\widetilde{\mathfrak{G}}$ of the reverse process using the q-machine formalism. The result, for compressing $\widetilde{\mathfrak{G}}$, is a q-machine that generates the reverse process—remarkably, with identical cost to the q-machine constructed from $\mathfrak{F}$ [49].

The q-machine constructed from $\widetilde{\mathfrak{G}}$ generates a quantum process and as such can itself undergo quantum time reversal [52], resulting in a new process that is generated by what we call the *reverse q-machine*. Just as the q-machine compresses $\widetilde{\mathfrak{G}}$, the reverse q-machine is a compression of $\mathfrak{G}$. Although the reverse q-machine is derived from the q-machine via time reversal, there is genuinely new physics present, because the dissipation ΔS_{loc} [Eq. (24)] is not invariant under time reversal. Thus, they must be approached as a separate case from the traditional q-machine when examining their thermodynamic efficiency.

The details of the time reversal are handled in the SM. Here, we present the resulting technique for compressing the reverse ϵ-machine. Given a reverse ϵ-machine $\mathfrak{G} = (\mathcal{S}, \mathcal{X}, \{T^{(x)}_{s's}\})$, for any set of phases $\{\phi_{xs} : x \in \mathcal{X}, s \in \mathcal{S}\}$ there is an encoding $\{|\psi_s\rangle : s \in \mathcal{S}\}$ of *orthogonal* states into a Hilbert space $\mathcal{H}_S$ and a set of Kraus operators $\{K^{(x)} : x \in \mathcal{X}\}$ on said Hilbert space such that

$$K^{(x)}|\psi_s\rangle = \sum_{s' \in \mathcal{S}} e^{i\phi_{xs'}} \sqrt{T^{(x)}_{s's}}\,|\psi_{s'}\rangle. \tag{25}$$

The orthogonality of $\{|\psi_s\rangle\}$ allows us to turn this into an explicit definition of the Kraus operators:

$$K^{(x)} = \sum_{s' \in \mathcal{S}} e^{i\phi_{xs'}} \sqrt{T^{(x)}_{s'f(s',x)}}\,|\psi_{s'}\rangle\langle\psi_{f(s',x)}|. \tag{26}$$

The stationary state ρ_π of this machine is, unlike the q-machine, generically not expressible as an ensemble of the encoding states $\{|\psi_s\rangle\}$. If this were so, the orthogonality of $\{|\psi_s\rangle\}$ would make them a diagonalizing basis for ρ_π, and we would achieve no memory compression. Rather, compression is achieved for the reverse q-machine precisely because the stationary state ρ_π is generically not diagonal in the encoding states—in contrast with the q-machine, which derived compression from the nonorthogonality of the encoding states.

The reverse q-machine stochastic dynamics Eq. (23) and thermodynamics Eq. (24) are defined precisely as those for q-machines in the previous section. As before, to prove our result we must define a special partition of the generator states. Here, it is important to note a relationship between a process's forward ϵ-machine $\mathfrak{F} = (\mathcal{P}, \mathcal{X}, \{R^{(x)}_{p'p}\})$ and its reverse ϵ-machine $\mathfrak{G} = (\mathcal{S}, \mathcal{X}, \{T^{(x)}_{s's}\})$. Specifically, the state S_t of $\mathfrak{G}$ after seeing the word $x_1 \ldots x_t$ and the state P_t of $\mathfrak{F}$ after the same are related by

$$\Pr_{\widetilde{\mathfrak{G}}}(s_t|x_1 \ldots x_t) = \sum_{p_t} \Pr_{\mathcal{C}}(s_t|p_t) \Pr_{\mathfrak{F}}(p_t|x_1 \ldots x_t) \tag{27}$$

for some channel $\Pr_{\mathcal{C}}(s|p)$. Let λ_p be the stationary distribution of $\mathfrak{F}$'s states and let $\Pr_{\mathcal{E}}(s'|s) = \sum_p \Pr_{\mathcal{C}}(s|p) \Pr_{\mathcal{C}}(s'|p)\lambda_p/\pi_s$. Let $\mathcal{B} = \{\mathcal{B}_\theta\}$ be the ergodic

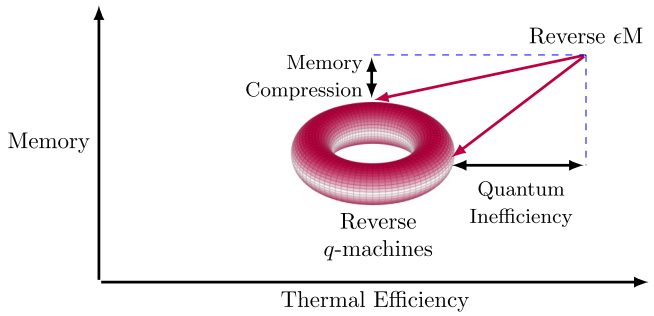

FIG. 4. Performance trade-offs for reverse q-machines, whose variety and dependence on $\{\phi_{xs}\}$ is represented by a torus: Under all quantifications of memory, the reverse q-machines constructed from a retrodictor achieve non-negative memory compression. However, to achieve positive compression, they must also have a nonzero dissipation ΔS_{loc}. The latter renders them less thermodynamically efficient.

partition of $\Pr_{\mathcal{E}}(s'|s)$, such that $\Pr_{\mathcal{E}}(s'|s) > 0$ only when $\theta(s) = \theta(s')$. The SM shows that ρ_π is diagonal in the blocks defined by $\mathcal{B}$.

Our result for reverse q-machines, proven in the SM, can now be stated:

Theorem 4: Maximally-efficient reverse q-machine. Let $\mathfrak{G} = (\mathcal{S}, \mathcal{X}, \{T^{(x)}_{s's}\})$ be a given process' reverse ϵ-machine. Suppose we build from it a reverse q-machine with encoding $\{|\psi_s\rangle\}$ and Kraus operators $\{K^{(x)}\}$. Let $\mathcal{B} := \{\mathcal{B}_\theta\}$ be the MCP of $\mathcal{S}$. Then the reverse q-machine has $\Delta S_{\text{loc}} = 0$ if and only if the partition $\mathcal{B}$ is trivially maximal—in that $|\mathcal{B}_\theta| = 1$ for each θ—and the predictively state-merged generator $\mathfrak{G}^{\mathcal{B}}$ of $\mathfrak{G}$ is unifilar.

Notice that this statement is a similar to that made in the last section and is essentially its time reverse. It implies that the only reverse ϵ-machines which can be quantally compressed are those which are also predictive generators. Also, again the trivial maximality of the ergodic partition $\mathcal{B}$ implies an inability to achieve nonzero compression. A heuristic diagram of the situation is shown in Fig. 4.

In conjunction with the previous section, this is a profound result on the efficiency of quantum memory compression. Distinct from the classical case, where Theorem 2 established that *every* process has certain generators that do achieve zero dissipation, Theorems 3 and 4 imply that only *certain* processes have zero-dissipation quantum generators and, moreover, those particular processes achieve no memory compression. The memory states, being orthogonally encoded, take no advantage of the quantum setting to reduce their memory cost.

VII. CONCLUDING REMARKS

We identified the conditions under which local operations circumvent the thermodynamic dissipation ΔS_{loc} that arises from destroying correlation. We started by showing how a useful theorem can be derived by using recent results on the fixed points of quantum channels. We applied it to the setting of local operations to determine the necessary and sufficient conditions for vanishing ΔS_{loc} in classical and quantum settings, with the aid of a generalized notion of quantum sufficient statistic. We employed this fundamental result to review and extend previous results on the thermodynamic efficiency of generators of stochastic processes. We confirmed a recent conjecture regarding the conditions for vanishing ΔS_{loc} in a classical generator. And, then, we showed that the exact same conditions hold for quantum generators, even to the point of requiring orthogonal encoding of memory states. This implies the profound result that quantum memory compression and perfect efficiency ($\Delta S_{\text{loc}} = 0$) are incompatible.

It is appropriate here to recall the lecture by Feynman in the early days of thinking about quantum computing, in which he observed that quantum systems can only be simulated on classical (even probabilistic) computers with great difficulty, but on a fundamentally-quantum computer they could be more realistically simulated [3]. Here, we considered the task of simulating a classical stochastic process by two means: one by using fundamentally-classical but probabilistic machines and the other by using a fundamentally-quantum machine. Previous results generally indicated quantum machines are advantageous in memory for this task, in comparison with their classical counterparts. Historically, this led to a much stronger notion of "quantum supremacy" than Feynman proposed: quantum computers may be advantageous in *all* tasks [58].

However, the quantum implementation we examined, although advantageous in memory, requires nonzero dissipation in order to cash in on that advantage. Furthermore, not every process necessarily has a quantum generator that achieves zero dissipation. This is in sharp contrast with the classical outcome. And so, this returns us to the spirit of Feynman's vision for simulating physics, in which it may sometimes be the case that the best machine to simulate a classical stochastic process is a classical stochastic computer—at least, thermodynamically speaking.

To further exercise these results, further extensions must be made to quantum generators, beyond the q-machine and its time reverse. We must determine if the exclusive relationship between compression and zero dissipation continues to hold in such extensions. We pursue this question in forthcoming work.

ACKNOWLEDGMENTS

The authors thank Fabio Anza, David Gier, Richard Feynman, Ryan James, Alexandra Jurgens, Gregory Wimsatt, and Ariadna Venegas-Li for helpful discussions and the Telluride Science Research Center for hospitality during visits. As a faculty member, J.P.C. similarly thanks the Santa Fe Institute. This material is based upon work supported by, or in part by, FQXi Grant FQXi-RFP-IPW-1902, the U.S. Army Research Laboratory, and the U.S. Army Research Office under contract W911NF-13-1-0390 and grant W911NF-18-1-0028, and via Intel Corporation support of CSC as an Intel Parallel Computing Center.

[1] F. Arute *et al.*, Quantum supremacy using a programmable superconducting processor, Nature (London) **574**, 505 (2019).

[2] E. Pednault, J. A. Gunnels, G. Nannicini, L. Horesh, and R. Wisnieff, Leveraging secondary storage to simulate deep 54-qubit sycamore circuits, arXiv:1910.09534.

[3] R. Feynman, Simulating physics with computers, Int. J. Theor. Phys. **21**, 467 (1982).

[4] B. Coecke, T. Fritz, and R. W. Spekkens, A mathematical theory of resources, Inf. Comput. **250**, 59 (2016).

[5] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Quantum entanglement, Rev. Mod. Phys. **81**, 865 (2009).

[6] U. M. Maurer, Secret key agreement by public discussion from common information, IEEE Trans. Inf. Theory **39**, 733 (1993).

[7] P. Gács and J. Körner, Common information is far less than mutual information, Probl. Control Inf. Theory **2**, 119 (1973).

[8] A. D. Wyner, The common information of two dependent random variables, IEEE Trans. Inf. Theory **21**, 163 (1975).

[9] G. R. Kumar, C. T. Li, and A. El Gamal, Exact common information, in *Proceedings of the IEEE International Symposium on Information Theory, Honolulu, HI, 2014* (IEEE, Piscataway, NJ, 2014), pp. 161–165.

[10] R. Landauer, Irreversibility and heat generation in the computing process, IBM J. Res. Dev. **5**, 183 (1961).

[11] S. Lloyd, Use of mutual information to decrease entropy: Implications for the second law of thermodynamics, Phys. Rev. A **39**, 5378 (1989).

[12] A. B. Boyd, D. Mandal, and J. P. Crutchfield, Correlation-powered information engines and the thermodynamics of self-correction, Phys. Rev. E **95**, 012152 (2017).

[13] A. B. Boyd, D. Mandal, and J. P. Crutchfield, Leveraging environmental correlations: The thermodynamics of requisite variety, J. Stat. Phys. **167**, 1555 (2017).

[14] A. B. Boyd, D. Mandal, and J. P. Crutchfield, Thermodynamics of Modularity: Structural Costs Beyond the Landauer bound, Phys. Rev. X **8**, 031036 (2018).

[15] S. Loomis and J. P. Crutchfield, Thermal efficiency of quantum memory compression, arXiv:1911.00998 [quant-ph].

[16] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. (Wiley-Interscience, New York, 2006).

[17] M. Nielsen and I. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, New York, 2010).

[18] D. Petz, Sufficient subalgebras and the relative entropy of states of a von Neumann algebra, Commun. Math. Phys. **105**, 123 (1986).

[19] D. Petz, Sufficiency of channels over von Neumann algebras, Q. J. Math. Oxford **39**, 97 (1988).

[20] M. B. Ruskai, Inequalities for quantum entropy: A review with conditions for equality, J. Math. Phys. **43**, 4358 (2002).

[21] P. Hayden, R. Josza, D. Petz, and A. Winter, Structure of states which satisfy strong subadditivity of quantum entropy with equality, Commun. Math. Phys. **246**, 359 (2004).

[22] M. Mosonyi and D. Petz, Structure of sufficient quantum coarse-grainings, Lett. Math. Phys. **68**, 19 (2004).

[23] M. Mosonyi, Ph.D. thesis, KU Leuven, 2005 (unpublished).

[24] D. Petz, *Quantum Information Theory and Quantum Statistics*, (Springer, Berlin, Heidelberg, 2008).

[25] A. Jenčová and D. Petz, Structure of sufficient quantum coarse-grainings, Commun. Math. Phys. **263**, 259 (2006).

[26] A. Łuczak, Quantum sufficiency in the operator algebra framework, Int. J. Theor. Phys. **53**, 3423 (2013).

[27] M. S. Leifer and R. W. Spekkens, A Bayesian approach to compatibility, improvement, and pooling of quantum states, J. Phys. A: Math. Theor. **47**, 275301 (2014).

[28] B. Baumgartner and H. Narnhoffer, The structures of state space concerning quantum dynamical semigroups, Rev. Math. Phys. **24**, 1250001 (2012).

[29] R. Carbone and Y. Pautrat, Irreducible decompositions and stationary states of quantum channels, Rep. Math. Phys. **77**, 293 (2016).

[30] J. Guan, Y. Feng, and M. Ying, The structure of decoherence-free subsystems, arXiv:1802.04904.

[31] V. V. Albert, Asymptotics of quantum channels: Conserved quantities, an adiabatic limit, and matrix product states, Quantum **3**, 151 (2019).

[32] D. Mandal and C. Jarzynski, Work and information processing in a solvable model of Maxwell's demon, Proc. Natl. Acad. Sci. USA **109**, 11641 (2012).

[33] A. B. Boyd, D. Mandal, and J. P. Crutchfield, Identifying functional thermodynamics in autonomous Maxwellian ratchets, New J. Phys. **18**, 023049 (2016).

[34] A. J. P. Garner, J. Thompson, V. Vedral, and M. Gu, Thermodynamics of complexity and pattern manipulation, Phys. Rev. E **95**, 042140 (2017).

[35] A. J. P. Garner, Oracular information and the second law of thermodynamics, arXiv:1912.03217.

[36] J. P. Crutchfield and K. Young, Inferring Statistical Complexity, Phys. Rev. Lett. **63**, 105 (1989).

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

[38] N. F. Travers and J. P. Crutchfield, Equivalence of history and generator ϵ-machines, arXiv:1111.4500.

[39] J. P. Crutchfield, Between order and chaos, Nat. Phys. **8**, 17 (2012).

[40] N. Travers and J. P. Crutchfield, Asymptotic synchronization for finite-state sources, J. Stat. Phys. **145**, 1202 (2011).

[41] N. Travers and J. P. Crutchfield, Exact synchronization for finite-state sources, J. Stat. Phys. **145**, 1181 (2011).

[42] D. R. Upper, Ph.D. thesis, University of California, Berkeley, 1997.

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

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

[45] P. M. Riechers, J. R. Mahoney, C. Aghamohammadi, and J. P. Crutchfield, A closed-form shave from Occam's quantum razor: Exact results for quantum compression, Phys. Lett. A **93**, 052317 (2016).

[46] F. C. Binder, J. Thompson, and M. Gu, A Practical, Unitary Simulator for non-Markovian Complex Processes, Phys. Rev. Lett. **120**, 240502 (2017).

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

[48] C. Aghamohammadi, S. P. Loomis, J. R. Mahoney, and J. P. Crutchfield, Extreme Quantum Memory Advantage for Rare-Event Sampling, Phys. Rev. X **8**, 011025 (2018).

[49] 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).

[50] S. Loomis and J. P. Crutchfield, Strong and weak optimizations in classical and quantum models of stochastic processes, J. Stat. Phys. **176**, 1317 (2019).

[51] 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).

[52] G. Crooks, Quantum operation time reversal, Phys. Rev. A **77**, 034101 (2008).

[53] J. P. Crutchfield, C. J. Ellison, and J. R. Mahoney, Time's Barbed Arrow: Irreversibility, Crypticity, and Stored Information, Phys. Rev. Lett. **103**, 094101 (2009).

[54] C. J. Ellison, J. R. Mahoney, R. G. James, J. P. Crutchfield, and J. Reichardt, Information symmetries in irreversible processes, Chaos **21**, 037107 (2011).

[55] O. C. O. Dahlsten, R. Renner, E. Rieper, and V. Vedral, Inadequacy of von Neumann entropy for characterizing extractable work, New J. Phys. **13**, 053015 (2011).

[56] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevResearch.2.023039 for a notational overview, a review of (i) fixed points of quantum channels, reversible computation, and sufficient statistics, (ii) quantum implementations of classical generators and q-machines and their thermodynamic costs, and (iii) details of example calculations.

[57] R. B. Ash, *Information Theory* (John Wiley and Sons, New York, 1965).

[58] J. Preskill, Quantum computing and the entanglement frontier, arXiv:1203.5813.