Intrinsic and Measured Information in Separable Quantum Processes

David Gier* and James P. Crutchfield†
Complexity Sciences Center and Physics and Astronomy Department, University of California at Davis, One Shields Avenue, Davis, CA 95616

(Dated: March 1, 2023)

Stationary quantum information sources emit sequences of correlated qudits—that is, structured quantum stochastic processes. If an observer performs identical measurements on a qudit sequence, the outcomes are a realization of a classical stochastic process. We introduce quantum-information-theoretic properties for separable qudit sequences that serve as bounds on the classical information properties of subsequent measured processes. For sources driven by hidden Markov dynamics we describe how an observer can temporarily or permanently synchronize to the source's internal state using specific positive operator-valued measures or adaptive measurement protocols. We introduce a method for approximating an information source with an independent and identically-distributed, Markov, or larger memory model through tomographic reconstruction. We identify broad classes of separable processes based on their quantum information properties and the complexity of measurements required to synchronize to and accurately reconstruct them.

CONTENTS

* dgier@ucdavis.edu
† chaos@ucdavis.edu

I. INTRODUCTION

Determining a quantum system's state requires grappling with multiple sources of uncertainty, including several that do not arise in classical physics. Irreducible limits on measurement, in particular, have been a hallmark of quantum physics since Heisenberg introduced the position-momentum uncertainty principle in 1927 [1]. Similar incompatible measurements exist for generic pure quantum states [2].

For a 2-level quantum system, or *qubit*, it is impossible to simultaneously measure the value of a spin in the x-, y-, and z-directions. (Stated mathematically, the Pauli matrices σ_x, σ_y, and σ_z do not commute.) Additionally, a single measurement in each basis is insufficient. One must measure many copies in each basis to specify the distribution of outcomes. As a result, determining an unknown qubit state through quantum state tomography requires measuring a large ensemble of identical copies with a set of mutually unbiased bases [3] or a single informationally-complete *positive operator-valued measure* (POVM) [4].

These sources of uncertainty are familiar in quantum physics. Contrast them with when an observer receives a sequence of correlated qubits. Measuring them one by one, what will they see? And, what then can they infer about the resources necessary to generate these qubit strings? The following answers these questions by teasing apart the sources of apparent randomness and correlation in measured quantum processes.

A. Quantum and Classical Randomness

Also in 1927, von Neumann formulated quantum mechanics in terms of statistical ensembles and quantified the entropy of these *mixed quantum states*. In doing so, he extended Gibbs' work on statistical ensembles and classical thermodynamic entropies to the quantum domain [5]. A mixed quantum state ρ has an entropy $S(\rho) = -tr(\rho \log \rho)$, now known as the *von Neumann entropy*. ($tr(\cdot)$ is the trace operator.) $S(\rho) = 0$ if and only if ρ is a pure (nonmixed) quantum state. On the one hand, the von Neumann entropy is key to understanding quantum systems, particularly those with *entangled* subsystems that exhibit nonclassical correlations. On the other, the uncertainty $S(\rho)$ quantifies a generic feature of statistical ensembles. It does not correspond to any particular quantum mechanical effect.

These two forms of uncertainty—due to ensembles and to quantum indeterminacy—are combined within the framework of quantum information theory, which generalizes classical information theory to quantum observables [6]. One notable example is noiseless coding. Shannon quantified the information content produced by a noiseless classical independent and identically-distributed (i.i.d.) information source—one that emits a state drawn from the same distribution at each timestep [7]. Schumacher's quantum noiseless coding theorem generalized this to quantum information sources. This gave a new physical interpretation of the von Neumann entropy: For an i.i.d. quantum source emitting state ρ, $S(\rho)$ is the number of qubits required for a reliable compression scheme [8].

B. Sources with Memory

Non-i.i.d. stationary information sources inject additional forms of uncertainty. For example, a source may have an internal memory that induces correlations between sequential qubits and therefore between measurement outcomes. Such correlations may be purely classical or uniquely quantal in nature. As we will show, an experimenter who assumes (incorrectly) that such a source is i.i.d. and

then applies existing tomographic methods will not detect these correlations and so will overestimate the source's randomness and underestimate its compressibility.

Classical memoryful sources are described within the framework of computational mechanics, in which stationary dynamical systems serve as information sources with their own internal states and dynamic [9]. Sequential finite-precision measurements of a dynamical system form a discrete-time stochastic process. The resulting process' statistics allow one to construct a model of the source and calculate its asymptotic entropy rate, internal memory requirements, and other physically-relevant properties [10, 11]. Importantly, the uncertainty associated with sequential measurements of a classical information source can be reduced, sometimes substantially, by an observer capable of synchronizing to the process' internal states [12, 13].

Subjecting an open quantum system to sequential qudit probes presents a similar but more general challenge, as the amount of information an observer can glean from an individual qudit through measurement is limited. Recent results established that applying particular measuring instruments to qubits induces complex behavior in measurement sequences [14]. Here, we extend these results by studying properties of the quantum states themselves in addition to particular sequences of measurement outcomes.

The following introduces novel quantum-information-theoretic properties for sequences of separable—i.e., nonentangled—qudits. We build on previous results that focused on entropy rates, compression limits, and optimal coding strategies for stationary quantum information sources [15–17], as well as on results for specific experimentally-motivated deviations from the i.i.d. assumption [18]. The approach is distinct from but complements recent efforts on quantum stochastic processes in which an observer measures a quantum system directly. This is complicated due to the latter's interaction with an inaccessible environment that induces memory effects in sequential measurement outcomes [19–23].

Section II introduces classical processes, separable qudit processes, and methods of transforming from one to the other via classical-quantum channels and measurement channels. Then Section III, in concert with App. A, defines the entropies associated with quantum and classical processes, respectively. Adapting Ref. [11]'s entropy hierarchy, we employ discrete-time derivatives and integrals to obtain a family of distinct quantitative measures of quantum process randomness and correlation. We prove that, for projective or informationally-complete measurements, the sequences of measurement outcomes form classical processes whose information properties are bounded by those of the quantum process being measured.

Section IV then surveys examples of increasingly-structured separable qubit and qutrit processes. Section V discusses how an observer can synchronize to a memoryful source—i.e., determine its internal state—through sequential measurement. Section VI uses the resulting catalog of possible process behaviors to answer practical questions for an observer of a quantum process attempting to perform tomography. Finally, Section VII draws out lessons and proposes future directions and applications, most notably extending the results to the experimentally-realizable generation of arbitrary entangled qudit states [24, 25] and using correlations as a resource to perform thermodynamic quantum information processing [26, 27].

II. STOCHASTIC PROCESSES

We consider the output of an information source to be a discrete-time, stationary stochastic process. If the source output is a classical random variable—X_t for each timestep t—we can directly apply the methods of computational mechanics [11]. Our goal is to extend these methods to describe separable sequences of qudits, each represented by a pure state in d-dimensional Hilbert space: $|\psi_t\rangle \in \mathcal{H}^d$ at each timestep t. Given such a qudit sequence, one can perform repeated, identical measurements such that the outcomes form a classical stochastic process. Since one can choose to measure qudit states in many different bases, the properties of the classical measured process are determined by both the state of the correlated qudits and the measurement choice. Thus, the relationship between a quantum process and classical measured processes is one-to-many. Figure 1 illustrates this setup.

A. Classical Processes

A *classical stochastic process* is defined by a probability measure $\Pr(\overleftrightarrow{X})$ over a chain of random variables:

$$\overleftrightarrow{X} \equiv ...X_{-2}X_{-1}X_0X_1X_2\ldots \ ,$$

with each X_t taking on values drawn from a finite alphabet $\mathcal{X}$. A block of ℓ consecutive random variables is denoted $X_{0:\ell} = X_0X_1\cdots X_{\ell-1}$. The indexing is left-inclusive and right-exclusive. A particular bi-infinite process *realization* is denoted $\overleftrightarrow{x} \equiv ...x_{-2}x_{-1}x_0x_1x_2...$ with events x_t taking values in a discrete set $\mathcal{X}$. Realizations of a block of length ℓ are known as *words* and denoted $x_{0:\ell}$. The set of all words of length ℓ is $\mathcal{X}^\ell$.

We consider processes that are *stationary*, meaning that

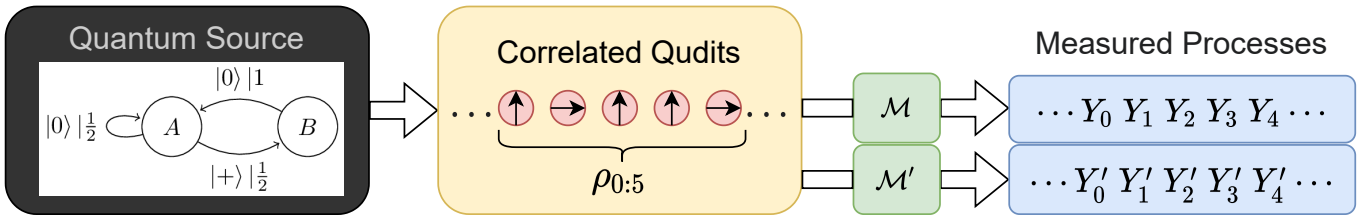


FIG. 1. A stationary quantum information source emits qudits that are correlated due to the source's internal memory. An experimenter measures these qudits in different ways ($\mathcal{M}$ or $\mathcal{M}'$) resulting in a family of classical stochastic processes.

word probabilities $\Pr(X_{0:\ell})$ are time-independent:

$$\Pr(X_{0:\ell}) = \Pr(X_{t:t+\ell}) \ ,$$

for all $t \in \mathbb{Z}$ and $\ell \in \mathbb{Z}^+$. A stationary process's statistics are fully described by the set of length-ℓ word distributions $\Pr(X_{0:\ell})$. A block of length ℓ has at most $|\mathcal{X}|^\ell$ possible realizations (words).

One important subclass of processes are *independently and identically distributed* (i.i.d.) processes. The joint block probabilities of an i.i.d. process take the form:

$$\Pr(X_{0:\ell}) \equiv \Pr(X_0)\Pr(X_1)\cdots\Pr(X_{\ell-1}) \ , \qquad (1)$$

for all $\ell \in \mathbb{Z}^+$. This factoring of the block probabilities results in no statistical correlations between any random variables. Due to stationarity $\Pr(X_t) = \Pr(X_0)$ for all t for an i.i.d. process.

Another commonly-studied subclass consists of the *Markov processes* for which the distribution for each X_t depends only on the immediately preceding random variable X_{t-1}. For Markov processes the joint probabilities for finite-length blocks factor as:

$$\Pr(X_{0:\ell}) \equiv \Pr(X_0)\Pr(X_1|X_0)\cdots\Pr(X_{\ell-1}|X_{\ell-2}) \ , \quad (2)$$

where $\Pr(X|Y)$ is the probability distribution of random variable X conditioned on random variable Y.

Finally, there is the markedly-larger subclass of *hidden Markov processes* that have an internal Markov dynamic that is not directly observable. Though the joint probabilities do not factor as in Eq. (2), the internal Markov dynamic restricts the process statistics as we describe next.

B. Presentations

A *presentation* of a process is a model consisting of a set of internal states and a transition dynamic between those states that together reproduce the process' statistics exactly. A given process may have many presentations. We focus on those depicted with state transition diagrams (directed graphs) that generate stationary, discrete-time stochastic processes in a natural way.

A *Markov chain* is a process presentation defined by the pair $(\mathcal{X}, T)$:

- A finite alphabet $\mathcal{X}$ of m symbols $x \in \mathcal{X}$, and
- A $m \times m$ transition matrix T. That is, if the source emits symbol x_i, with probability $T_{ij} = \Pr(x_j|x_i)$ it emits symbol x_j next.

The stationary distribution for a Markov chain is denoted π, and it is a distribution over causal states in $\mathcal{S}$ that satisfies $\pi = \pi T$. For a Markov chain the set of internal states is exactly the set of emitted symbols since the probability distribution for the next symbol is completely determined by the previous symbol. We represent each state as a node in a graph and each transition as a directed edge between nodes labeled by the associated probability.

Markov chains are sufficient to represent Markov processes, but we can describe the more general class of hidden Markov processes by allowing for internal states not directly observable. These processes are generated by *hidden Markov chains* (HMCs), defined by the triple, $(\mathcal{S}, \mathcal{X}, \mathcal{T})$:

- A finite set $\mathcal{S} = \{\sigma_1, \ldots, \sigma_n\}$ of internal states,
- A finite alphabet $\mathcal{X}$ of m symbols $x \in \mathcal{X}$, and
- A set $\mathcal{T} = \{T^x : x \in \mathcal{X}\}$ of m $n \times n$ symbol-labeled transition matrices. That is, if the source is in state σ_i, with probability $T^x_{ij} = \Pr(x, \sigma_j|\sigma_i)$ it emits symbol x while transitioning to state σ_j.

We represent each possible transition between states as an edge between their nodes labeled with the emitted symbol and the transition probability. An HMC's stationary distribution π uniquely satisfies $\pi = \pi \sum_x T^x$.

Any HMC that exactly reproduces a process' statistical features is a *generative* HMC. This is an important distinction, since only some of those also belong to the more

restrictive class of *predictive* HMCs. An HMC is predictive if its state at time $t+1$ is completely determined by the state at time t and the emitted symbol. This property is known as *unifilarity*.

At this point we must emphasize the difference between a process and a particular presentation of that process. This distinction is critical when designating processes and models to be 'classical' or 'quantum'. A discrete-time classical stochastic processes is classical because it consists of a chain of *classical* random variables. Markov chains and HMCs are classical models because their internal states are dynamics are both *classical*. One may instead construct a presentation of a classical process with a set of *quantum* states that the model transitions between via some *quantum* dynamic. An observer can recover the classical process' statistics by taking sequential measurements on either the system or on ancilla qudits that interact with the system at each timestep. The simulation of classical stochastic processes with quantum resources is the objective of *quantum computational mechanics*. There, a class of quantum models (q-simulators) shows advantage in terms of memory requirements over provably-minimal classical predictive models (ϵ-machines) [28–32]. Likewise different presentations of a quantum processes may have an underlying dynamic that is either classical or quantum. We turn now to quantum processes and their presentations.

C. Quantum Processes

Discrete-time classical stochastic processes consist of one classical random variable for each timestep. Likewise discrete-time quantum stochastic processes consist of one quantum state $|\psi_t\rangle \in \mathcal{H}^d$ at each timestep. We first describe an i.i.d. quantum information source and then generalize to sources with memory.

1. Memoryless

A discrete-time *quantum information source* emits a d-level quantum system or *qudit* at each timestep. The statistical mixture of the infinite qudit sequences emitted by a source is a *quantum process*. As in the classical setting, different classes of quantum processes are distinguished by their temporal correlations. Now, however, for quantum sources we must use quantum information theory to account for both classical and quantal correlations.

First, consider the output of an i.i.d. (memoryless) quantum information source. Let $\mathcal{H}$ be a d-dimensional Hilbert space with pure states $|\psi_x\rangle \in \mathcal{H}$. A d-level i.i.d. quantum information source consists of a set $\mathcal{Q}$ of pure qudit states and a probability distribution over those states such that $\Pr(|\psi_x\rangle) > 0$ for all $|\psi_x\rangle \in \mathcal{Q}$. We refer to $\mathcal{Q}$ as a quantum alphabet and consider only quantum alphabets with a finite number of pure states.

At each discrete timestep t, the source emits state $|\psi_x\rangle$ with probability $\Pr(|\psi_x\rangle)$. The resulting ensemble is described by the $d \times d$ density matrix:

$$\rho_{iid} = \sum_{|\psi_x\rangle \in \mathcal{Q}} \Pr(|\psi_x\rangle) \, |\psi_x\rangle \langle\psi_x| \ . \tag{3}$$

This particular pure-state decomposition of ρ_{iid} is not unique. Moreover, an observer cannot determine $\mathcal{Q}$ through observations—unless $\mathcal{Q}$ consists of only one state—since many pure-state ensembles correspond to the same density matrix.

If an i.i.d. source emits ρ_{iid} at each timestep then the quantum process generated by the source is simply the infinite tensor product state:

$$|\Psi_{iid}\rangle = \cdots \otimes \rho_{iid} \otimes \rho_{iid} \otimes \rho_{iid} \otimes \cdots \ . \tag{4}$$

2. Memoryful

We cannot describe non-i.i.d. sources using a single probability distribution over $\mathcal{Q}$ but must introduce a probability distribution over *sequences* of states drawn from $\mathcal{Q}$. We do this by associating each element of $\mathcal{Q}$ with an element in the symbol alphabet $\mathcal{X}$ of an underlying classical stochastic process $\overleftrightarrow{X}$. Infinite qudit sequences then inherit probabilities from $\overleftrightarrow{X}$. This construction results in qudit sequences that are *separable*—i.e., not entangled.

We express the relationship between symbols and pure quantum states via a memoryless classical-quantum channel $\mathcal{E} : \mathcal{X} \to \mathcal{H}$, taking $x \to |\psi_x\rangle$. This is also known as a *preparation channel* (or encoder); see App. B for more.

Preparation channels are dual to measurement channels, described later, that map quantum states to classical probability distributions and, via sampling, to particular symbols.

For the classical process $\overleftrightarrow{X}$ whose realizations consist of symbols $x \in \mathcal{X}$, the associated quantum alphabet $\mathcal{Q}$ is constructed by passing each element of $\mathcal{X}$ through $\mathcal{E}$ such that $\mathcal{Q} = \{\mathcal{E}(x)\}$. Thus, $\mathcal{Q}$ is completely determined by $\mathcal{X}$ and $\mathcal{E}$. For example, in the i.i.d. case Eq. (3) can also be written $\rho_{iid} = \mathcal{E}(X)$, and each possible pure-state decomposition of ρ_{iid} can now be interpreted as a different combination of classical random variable X and preparation channel $\mathcal{E}$.

In a slight abuse of notation we write $|\Psi_{-\infty:\infty}\rangle = \mathcal{E}(\overleftrightarrow{X})$ to indicate that quantum process $|\Psi_{-\infty:\infty}\rangle$ is formed by

passing each random variable of $\overleftrightarrow{X}$ through the classical-quantum channel $\mathcal{E}$.

Note that an infinite qudit sequence (separable or entangled) can be viewed as a one-dimensional lattice of qudits indexed by $t \in \mathbb{Z}$. These possibly-entangled states can be described in full generality using an operator-algebraic approach. We ground our formal definition of quantum processes in this mathematical setting. (Reference [17] provides a more detailed treatment of observable algebras for entangled qudit sequences over $\mathbb{Z}$.)

Let $\mathcal{B}_t$ be the d-dimensional matrix algebra describing all possible observables on lattice site t. (For $d = 2$, the space of observables is spanned by the identity and the 2×2 complex Pauli (Hermitian, unitary) matrices.) The state of the qudit at site t can be described by the density matrix ρ_t acting on $\mathcal{H}_t$ of dimension d. For a block of ℓ consecutive qudits, all observables can be described by the joint algebra over ℓ sites of the lattice: $\mathcal{B}_{0:\ell} = \bigotimes_{t=0}^{\ell-1} \mathcal{B}_t$, and the state of this block is $\rho_{0:\ell}$ acting on $\mathcal{H}_{0:\ell} = \bigotimes_{t=0}^{\ell-1} \mathcal{H}_t$, a Hilbert space of dimension d^ℓ. Combining all local algebras allows one to define an algebra $\mathcal{B}$ over the infinite lattice. A quantum process is a particular state over the infinite lattice and can be written as $|\Psi_{-\infty:\infty}\rangle$.

As a necessary first step and to more readily adapt information-theoretic tools from classical processes, we return to the more restricted case: separable sequences of qudits drawn from a finite alphabet $\mathcal{Q}$ of pure qudit states. Given a classical word $w = x_0 x_1 \ldots x_{\ell-1} \in \mathcal{X}^\ell$ and a preparation channel $\mathcal{E} : \mathcal{X} \to \mathcal{Q}$, a qudit sequence takes the form:

$$\begin{aligned} |\psi_w\rangle &= \mathcal{E}(w) \\ &= |\psi_{x_0}\rangle \otimes |\psi_{x_1}\rangle \otimes \cdots \otimes |\psi_{x_{\ell-1}}\rangle \ , \end{aligned} \tag{5}$$

where ℓ is the length of the sequence and $|\psi_w\rangle \in \mathcal{H}_{0:\ell}$. Note that $\dim(\mathcal{H}_{0:\ell}) = d^\ell$, $\mathcal{Q} = \{\mathcal{E}(x)\}$, and the number of possible qudit sequences of length ℓ is $|\mathcal{Q}|^\ell$ or—assuming all $|\psi_x\rangle$ are distinguishable—$|\mathcal{X}|^\ell$.

A *separable quantum process* is then defined by $|\Psi_{-\infty:\infty}\rangle = \mathcal{E}(\overleftrightarrow{X})$. Different preparations—i.e., different combinations of $\overleftrightarrow{X}$ and $\mathcal{E}$—may produce the same quantum process.

The set of length-ℓ-block density matrices for a quantum process is given by:

$$\begin{aligned} \rho_{0:\ell} &= \mathcal{E}(X_{0:\ell}) \\ &= \sum_{w \in \mathcal{X}^\ell} \Pr(|\psi_w\rangle) \, |\psi_w\rangle \langle\psi_w| \ , \end{aligned} \tag{6}$$

where $|\psi_w\rangle$ are the separable vectors given in Eq. (5). Conveniently, their probabilities are determined by those of the underlying classical stochastic process $\Pr(\overleftrightarrow{X})$: $\Pr(|\psi_w\rangle) = \Pr(w)$. Each $\rho_{0:\ell}$ is a finite subsystem of the pure quantum state $|\Psi_{-\infty:\infty}\rangle$ over the infinite lattice. We use left-inclusive/right-exclusive indexing for density matrices as well.

For a given $\rho_{0:\ell}$, one can also obtain a purification in a finite-dimensional Hilbert space [33]. It is important to note that, since $\rho_{0:\ell}$ does not have a unique pure-state decomposition, one cannot generally reconstruct the probabilities $\Pr(|\psi_w\rangle)$ from it. Rather, $\rho_{0:\ell}$ contains only information accessible to an observer. And, if $\mathcal{Q}$ contains nonorthogonal qudit states ($\langle\psi_x|\psi_{x'}\rangle \neq 0$ for some $|\psi_x\rangle$,$|\psi_{x'}\rangle \in \mathcal{Q}$) then an observer cannot unambiguously distinguish them.

In addition to separability, we also focus on *stationary* quantum processes, meaning:

$$\rho_{0:\ell} = \rho_{t:\ell+t} \ , \tag{7}$$

for all $\ell \in \mathbb{Z}^+$ and $t \in \mathbb{Z}$. If $\overleftrightarrow{X}$ is stationary, then $|\Psi_{-\infty:\infty}\rangle = \mathcal{E}(\overleftrightarrow{X})$ will be stationary by construction.

For an i.i.d. quantum process the joint probabilities of $\overleftrightarrow{X}$ factor as in Eq. (1), giving the quantum process the form of Eq. (4). The length-ℓ-block density matrix is represented by a product state:

$$\rho_{0:\ell} = \bigotimes_{t=0}^{\ell-1} \rho_{iid} \ , \tag{8}$$

with ρ_{iid} taking the form in Eq. (3).

For an underlying classical process $\overleftrightarrow{X}$ that is Markov, there are additional subtleties. Joint probabilities of $\overleftrightarrow{X}$ factor as in Eq. (2), so the joint probabilities of $|\Psi_{-\infty:\infty}\rangle = \mathcal{E}(\overleftrightarrow{X})$ also factor so that:

$$\begin{aligned} \Pr(|\psi_w\rangle) = \Pr(|\psi_{x_0}\rangle) \Pr(|\psi_{x_1}\rangle \,|\, |\psi_{x_0}\rangle) \\ \cdots \Pr(|\psi_{x_{\ell-1}}\rangle \,|\, |\psi_{x_{\ell-2}}\rangle) \ . \end{aligned} \tag{9}$$

However, an observer cannot reliably distinguish between different states $|\psi_x\rangle$ when measuring a quantum process, and the underlying Markov dynamic is hidden from observation. Thus, the general setting for memoryful quantum processes is that of hidden Markov processes. These are best introduced using concrete models that directly represent a process' structure.

D. Presentations of Quantum Processes

A *presentation* for a quantum process is a model with internal states and a transition dynamic between them that emits pure quantum states, rather than classical

symbols. As for presentations of classical processes, we depict them with state transition diagrams. When $\overleftrightarrow{X}$ is a Markov or hidden Markov process, $|\Psi_{-\infty:\infty}\rangle = \mathcal{E}(\overleftrightarrow{X})$ can be represented with an extension of Ref. [14]'s classically-controlled qubit sources (cCQS), as follows.

A *hidden Markov chain quantum source* (HMCQS) is a triple $(\mathcal{S}, \mathcal{Q}, \mathcal{T})$ consisting of:

- A finite set $\mathcal{S} = \{\sigma_1, \ldots, \sigma_n\}$ of internal states,
- A finite alphabet $\mathcal{Q} = \{|\psi_0\rangle, \ldots, |\psi_{m-1}\rangle\}$ of pure qudit states, with each $|\psi_x\rangle \in \mathcal{H}^d$, and
- A set $\mathcal{T} = \{T^x : |\psi_x\rangle \in \mathcal{Q}\}$ of m $n \times n$ transition matrices. That is, if the source is in state σ_i, with probability $T^x_{ij} = \Pr(|\psi_x\rangle, \sigma_j | \sigma_i)$ it emits qudit $|\psi_x\rangle$ while transitioning to internal state σ_j.

As with HMCs, the stationary distribution for an HMCQS satisfies $\pi = \pi \sum_x T^x$.

Any HMCQS that exactly reproduces a quantum process is a *generative* HMCQS, or generator of the process. Though quantum models cannot be predictive in the same sense as classical models, we can define an analog to classical unifilarity. A HMCQS is *quantum unifilar* if, for every state $\sigma \in \mathcal{S}$, there exists a possible measurement such that the internal state at time $t+1$ is completely determined if the state at time t is σ. We discuss several implications of quantum unifilarity later.

We call a HMCQS a *classical* controller of a quantum process since there is nothing quantal about its internal states or transition dynamic. This is in contrast to related classes of quantum models that evolve a finite quantum system according to a quantum operation (defined via a set of Kraus operators) at each timestep. These include *Quantum Markov Chains* (QMCs) [34] and *Hidden Quantum Markov Models* (HQMMs) [35–37]. While HMCQS emit separable quantum states, QMC and HQMM generate sequences of measurement outcomes (each corresponding to a particular Kraus operator) that form classical stochastic processes.

Anticipating future effort, we consider it worthwhile to draw out several observations on entanglement between successive qudits at this point. Entanglement means that finite-length qudit sequences are not separable and so are not described by Eq. (5). Moreover, their sequence probabilities cannot be straightforwardly defined with reference to an underlying classical stochastic process.

That said, there are systematic ways of defining stationary $|\Psi_{-\infty:\infty}\rangle$ such that the set $\rho_{0:\ell}$ of marginals describe all measurements over blocks of ℓ qudits. For example, if the source's internal structure consists of a D-dimensional quantum system interacting unitarily with one qudit per timestep, it generates a matrix product state (MPS) with a maximum bond dimension of D [38]. If the source operates stochastically (rather than unitarily), then many different MPSs can be emitted with varying probabilities. The collection is then described by *matrix-product density operators* (MPDO) [39]. We refer to these as *entangled qudit processes*. Their dynamical and informational analyses are left for elsewhere. The present goal is to layout the basics for those efforts.

E. Measured Processes

An agent observing a quantum process has many ways to measure it. Let M represent a measurement applied to the qudit in state ρ. In general, M is a *positive operator-valued measure* (POVM) described by a set of positive semi-definite Hermitian operators $\{E_y\}$ on the Hilbert space $\mathcal{H}$ of dimension d. Each E_y corresponds to a possible measurement outcome y, and POVM elements must sum to the identity:

$$\sum_y E_y = \mathbb{I} \ .$$

Projection-valued measures (PVMs) are an important subclass of POVMs with an additional constraint: operators E_y must be orthogonal projectors. PVMs have at most d elements. A PVM consisting only of rank-one projectors on $\mathcal{H}^d$ is a *von Neumann measurement* and has exactly d elements [5].

A set of measurements applied to a block of ℓ qudits are described by some block POVM $\mathcal{M}_{0:\ell}$ with elements $\{E_{y_{0:\ell}}\}$ on the Hilbert space $\mathcal{H}_{0:\ell}$ of dimension d^ℓ. $\mathcal{M}_{0:\ell}$ may include measurements in the joint basis of multiple qudits—measurements essential for fully characterizing entangled processes.

For separable processes we focus on "local" measurements—operators on a single qudit. The measurement operator for a block of ℓ qudits then takes a tensor product structure:

$$\mathcal{M}_{0:\ell} = \bigotimes_{t=0}^{\ell-1} M_t \ , \tag{10}$$

where each M_t is a POVM on $\mathcal{H}_t$.

If we apply the same local POVM M to each qudit, then

$$\mathcal{M}_{0:\ell} = \bigotimes_{t=0}^{\ell-1} M \ . \tag{11}$$

We refer to this as a *repeated POVM measurement*.

An observer can also have multiple POVMs at their disposal and apply different measurements at different time

steps according to some *measurement protocol*. We describe measurement protocols in more detail shortly.

For simplicity, the following ignores $\rho_{0:\ell}$'s post-measurement state and considers only the measurement outcomes $y_{0:\ell}$. Thus, we take $\mathcal{M}_{0:\ell}$ to be a stochastic map $\rho_{0:\ell} \to Y_{0:\ell}$, the random variables representing measurement outcomes.

When applying a measurement of the form of Eq. (11) to a finite block of ℓ qudits, the outcomes factor into a block of ℓ classical random variables:

$$\begin{aligned} Y_{0:\ell} &= Y_0 Y_1 \cdots Y_{\ell-1} \\ &= \mathcal{M}_{0:\ell}(\rho_{0:\ell}) \ , \end{aligned}$$

where the possible values of each Y_t are the POVM measurement outcomes $y \in \mathcal{Y}$. There are $|\mathcal{Y}|^\ell$ possible realizations of $Y_{0:\ell}$. We write a realization (word) of length-ℓ as $y_{0:\ell}$.

The probability of any particular measurement outcome for a block of ℓ qudits in state $\rho_{0:\ell}$ is:

$$\Pr(y_{0:\ell}) = tr(E_{y_{0:\ell}} \rho_{0:\ell}) \ .$$

For $\rho_{0:\ell}$ with the separable form of Eq. (6) and identical POVM measurements on each qudit as in Eq. (11), we can decompose $E_{y_{0:\ell}}$ into ℓ local operators E_{y_t}:

$$\begin{aligned} \Pr(y_{0:\ell}) &= tr\left(E_{y_{0:\ell}} \sum_{w \in \mathcal{X}^\ell} \Pr(w) \left|\psi_w\right\rangle \left\langle\psi_w\right| \right) \\ &= tr\left(\sum_{w \in \mathcal{X}^\ell} \Pr(w) \prod_{t=0}^{\ell-1} E_{y_t} \left|\psi_{x_t}\right\rangle \left\langle\psi_{x_t}\right| \right) . \end{aligned} \quad (12)$$

For a separable qudit process, a sequence $y_{0:\ell}$ of local measurement outcomes can also be interpreted as the result of sending random variables $X_{0:\ell}$ from $\overleftrightarrow{X}$ over the same memoryless noisy channel $\mathcal{C} : \mathcal{X} \to \mathcal{Y}$. $\mathcal{C}$ decomposes into the deterministic preparation $\mathcal{E}$ and our stochastic measurement $\mathcal{M}$:

$$\mathcal{C} = \mathcal{M} \circ \mathcal{E} \ .$$

(Appendix B presents a more thorough description of the classical-quantum channels $\mathcal{E}$ and $\mathcal{M}$.)

This construction makes it clear that measurement outcomes correspond to classical random variables Y_t that take values $y \in \mathcal{Y}$ and form a classical process $\overleftrightarrow{Y}$ with probabilities defined by Eq. (12). To express the relationship between a quantum process and a measured classical process we write:

$$\overleftrightarrow{Y} = \mathcal{M}(|\Psi_{-\infty:\infty}\rangle) \ ,$$

where $\mathcal{M}$ is a repeated, local POVM. If the qudit process is separable, we can also write:

$$\overleftrightarrow{Y} = \mathcal{C}(\overleftrightarrow{X}) \ .$$

F. Adaptive Measurement Protocols

An observer does not need to repeat the same measurement on every qudit but may apply different POVMs at different time steps according to some algorithm. If the agent uses past measurement outcomes to inform their choice of POVM we say they are using an *adaptive measurement protocol*.

The following limits discussion to measurement protocols that have a deterministic finite automata (DFA) as their underlying controller. Similar constructions combining quantum measurement and DFAs have appeared in the context of quantum grammars [40–42].

A *deterministic quantum measurement protocol* (DQMP) is defined by the quintuple $(\mathcal{S}, S_0, M, \mathcal{Y}, \delta)$:

- A finite set $\mathcal{S} = \{\sigma_1, \ldots, \sigma_n\}$ of internal states,
- A unique start state $S_0 \in \mathcal{S}$,
- A set of POVMs $M = \{M_s\}_{\sigma_s \in \mathcal{S}}$, one for each internal state,
- An alphabet $\mathcal{Y}$ of m symbols corresponding to different measurement outcomes, and
- A deterministic transition map $\delta : \mathcal{S} \times \mathcal{Y} \to \mathcal{S}$.

If $\mathcal{S}$ consists of only S_0, then the DQMP is a repeated POVM measurement for POVM M_{S_0}. When $\mathcal{S}$ has more than one internal state, the POVMs corresponding to different states may have the same or a different number of elements. Likewise the symbol sets corresponding to their measurement outcomes may be disjoint or symbols may be repeated.

We can place the following bounds on the size of the set $\mathcal{Y}$: $m \leq \sum_s |\{E_{s,y}\}|$, where $\{E_{s,y}\}$ is the set of operators corresponding to POVM M_s and $m \geq \max_s |\{E_{s,y}\}|$, the size of the POVM with the most elements.

For DQMP $\mathcal{M}$ and qudit process $|\Psi_{-\infty:\infty}\rangle$, obtaining a measured process $\overleftrightarrow{Y} = \mathcal{M}(|\Psi_{-\infty:\infty}\rangle)$ is generically more difficult than for the case of repeated POVM measurements. When an observer begins using protocol $\mathcal{M}$ at $t = 0$ they experience two distinct operating regimes: first the *transient* dynamic, then the *recurrent* dynamic. We briefly outline this process and return to the subject when we describe synchronization—a task deeply related to the transient dynamic—in Section V.

$\mathcal{M}$ begins in state S_0 at $t = 0$. For a given (stationary, ergodic) input $|\Psi_{-\infty:\infty}\rangle$, as $t \to \infty$ the DQMP approaches a stationary distribution over a subset of its internal states $\pi = \{\Pr(\sigma_i) > 0, \text{ for all } \sigma_i \in \mathcal{S}_r\}$, where $\mathcal{S}_r \subseteq \mathcal{S}$ is the set of recurrent states. This distribution (and even which states are in $\mathcal{S}_r$) depends on $|\Psi_{-\infty:\infty}\rangle$. The recurrent dynamic is determined by this stationary distribution and the transition probabilities between states in $\mathcal{S}_r$. Any state not in $\mathcal{S}_r$ is in the transient state set $\mathcal{S}_t$.

The transient dynamic describes how $\mathcal{M}$ goes from S_0 at $t = 0$ to its recurrent dynamic over $\mathcal{S}_r$, which may occur at a finite time $t = t_{sync}$ or only asymptotically as $t \to \infty$. In general two dynamics produce two distinct measured processes $\overleftrightarrow{Y}_r$, which is stationary and ergodic by construction, and $Y_{0:t_{sync}}$ which is not. The final measured process has two components—i.e. $\overleftrightarrow{Y} = \{\overleftrightarrow{Y}_r, Y_{0:t_{sync}}\}$.

G. Discussion

These nested layers of complication suggest working through a concrete example and restating the overall goals.

Imagine an observer measures a single qubit from a quantum source that emitted state ρ_t, using a projective measurement in the computational basis $M_{01} = \{|0\rangle\langle 0|, |1\rangle\langle 1|\}$. The possible measurement outcomes $y_0 = 0$ and $y_0 = 1$ occur with probabilities:

$$\begin{aligned}\Pr(y_0 = 0) &= tr(|0\rangle\langle 0|\,\rho_t) \\ \Pr(y_0 = 1) &= tr(|1\rangle\langle 1|\,\rho_t)\ ,\end{aligned}$$

respectively. These two values determine the distribution for the random variable Y_0 and, by applying the same projective measurement to $\rho_{0:\ell}$, we completely determine the statistics $\Pr(Y_{0:\ell})$ of the measured block. Continuing this procedure for $\ell \to \infty$ defines the measured process $\overleftrightarrow{Y}$.

Naturally, the observer can also choose to apply measurements in another basis; e.g., $M_\pm = \{|+\rangle\langle +|, |-\rangle\langle -|\}$ where $|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$. This typically results in a measured process $\overleftrightarrow{Y}'$ with radically-different statistical features.

Finally, an observer could use an adaptive measurement protocol $\mathcal{M}$. They start in state $s = S_0$ and measure with M_{01}. If $y_0 = 0$ they stay in S_0 and continue using M_{01}. If $y_0 = 1$, they transition to a new internal state and use $M_\pm$ on the next qubit. Regardless of the outcome of $M_\pm$ they return to S_0 and measure the next qubit with M_{01}. The measured process $\overleftrightarrow{Y}''$ will be distinct from both $\overleftrightarrow{Y}$ and $\overleftrightarrow{Y}'$ and may consist of both a transient and recurrent component.

With this setting laid out, we can now more precisely state the questions the following development answers:

1. Given the density matrices $\rho_{0:\ell}$ describing sequences of ℓ separable qudits, what are the general properties of sequences $Y_{0:\ell}$ of measurement outcomes? This is Section III's focus. There, $\rho_{0:\ell}$'s quantum information properties bound the classical information properties of measurement sequences $Y_{0:\ell}$ for certain classes of measurements.
2. Given a hidden Markov chain quantum source, when is an observer with knowledge of the source able to determine the internal state (synchronize)? Can the observer remain synchronized at later times? Section V addresses this.
3. If an observer encounters an unknown qudit source, how accurately can the observer estimate the informational properties of the emitted process through tomography with limited-resources? How can they build approximate models of the source if they reconstruct $\rho_{0:\ell}$ for some finite ℓ? This is Section VI's subject.

Additionally, Section IV illustrates these general results and the required analysis methods using specific examples of qudit processes.

III. INFORMATION IN QUANTUM PROCESSES

We wish to develop an information-theoretic analysis of quantum processes for which the observed sequences depend on the observer's choice of measurement. (Much of this parallels the classical information measures reviewed in App. A.) This requires a more general approach using density matrices $\rho_{0:\ell}$ that contain all the information necessary to describe the outcome of any measurement performed on ℓ-qudit blocks. We use quantum information theory to study properties of the set of $\rho_{0:\ell}$'s and then relate them to classical properties of measurement sequences described in App. A. We begin by briefly reviewing several basic quantities in quantum information theory. References [6] and [33] give a more complete picture of the subject.

A. von Neumann Entropy

In quantum information theory the von Neumann entropy plays a role similar to that of the Shannon entropy in classical information theory. Given a quantum mixed

quantum state ρ, the *von Neumann entropy* is:

$$\begin{aligned} S(\rho) &= -tr(\rho \log_2 \rho) \\ &= -\sum_i \lambda_i \log_2 \lambda_i \ , \end{aligned} \tag{13}$$

where λ_i's are the eigenvalues of the density matrix ρ. $S(\rho) = 0$ if and only if ρ is a pure state. We use $\log_2(\cdot)$, therefore the units of the von Neumann entropy will be *bits*.

From Eq. (13), the von Neumann entropy is the Shannon entropy of the eigenvalue distribution of density matrix ρ. Therefore:

$$S(\rho) = \min_M \mathrm{H}\left[M(\rho)\right] \ , \tag{14}$$

where $\mathrm{H}[\cdot]$ is the Shannon entropy and the minimum is taken over the set of all rank-one POVMs. The minimum will always be a PVM with projectors that compose ρ's eigenbasis [6]. We use brackets to indicate that $M(\rho)$ is a classical probability distribution over measurement outcomes.

To monitor correlations between two quantum systems we use the quantum relative entropy:

$$S(\rho\|\sigma) \equiv tr(\rho \log_2 \rho) - tr(\rho \log_2 \sigma) \ , \tag{15}$$

where ρ and σ are the density operators of the two systems. The quantum relative entropy is nonnegative:

$$S(\rho\|\sigma) \geq 0 \ , \tag{16}$$

with equality if and only if $\rho = \sigma$, a result known as *Klein's Inequality* [33].

The *joint quantum entropy* for a state ρ^{AB} of a bipartite system AB is:

$$\begin{aligned} S(A,B) &= S(\rho^{AB}) \\ &= -tr(\rho^{AB} \log_2 \rho^{AB}) \ , \end{aligned}$$

We can further define a *conditional quantum entropy* of system A conditioned on system B as:

$$S(A|B) = S(A,B) - S(B) \ , \tag{17}$$

where $S(B) = S(\rho^B)$. Note that $S(A|B) \neq S(B|A)$.

In contrast to the classical case, the conditional quantum entropy may be negative—a phenomenon leveraged in super-dense coding protocols [6]. Equivalently, the conditional quantum entropy can be written using the quantum relative entropy as:

$$\begin{aligned} S(A|B) &= -S(\rho^{AB}\|\mathbb{I}^A \otimes \rho^B) \\ &= \log_2(d_A) - S\left(\rho^{AB}\Big\|\frac{\mathbb{I}^A}{d_A} \otimes \rho^B\right) \ , \end{aligned} \tag{18}$$

where $\mathbb{I}^A$ is the identity operator on Hilbert space $\mathcal{H}^A$ with dimension d_A.

The *quantum mutual information* between quantum subsystems A and B is given by:

$$\begin{aligned} S(A{:}B) &= S(A) - S(A|B) \\ &= S(A) + S(B) - S(A,B) \ . \end{aligned} \tag{19}$$

The quantum mutual information is symmetric and nonnegative. If the joint system AB is in a pure state, then $S(A,B)$ will be zero, and $S(A) = S(B)$. It can also be expressed as a quantum relative entropy as:

$$S(A{:}B) = S\left(\rho^{AB}\|\rho^A \otimes \rho^B\right) \ . \tag{20}$$

A few additional well-known properties of the von Neumann entropy facilitate later results. First, each $\rho_{0:\ell}$ for separable qudit sequences is a finite mixture of states formed from length-ℓ words of an underlying classical process, so the following will be useful:

Lemma 1. *Consider a random variable X that takes values $x \in \{0, 1, \ldots, n\}$ with corresponding probabilities $\{p_0, p_1, \ldots p_n\}$. Given a set of density matrices $\{\rho_0, \rho_1, \ldots, \rho_n\}$, the following inequality holds [33]:*

$$S\left(\sum_{x=0}^{n} p_x \rho_x\right) \leq \mathrm{H}[X] + \sum_{x=0}^{n} p_x S(\rho_x) \ ,$$

with equality if and only if the all ρ_x have support on orthogonal subspaces.

Second, since we use quantum channels to both prepare and measure a qudit process, we make use of the fact that the quantum relative entropy is monotonic [6]:

$$S(\rho\|\sigma) \geq S(\mathcal{E}(\rho)\|\mathcal{E}(\sigma)) \ , \tag{21}$$

where $\mathcal{E}$ is any quantum channel. This inequality becomes an equality if and only if there exists a recovery map $\mathcal{R}$ such that $\mathcal{R}(\mathcal{E}(\rho)) = \rho$ and $\mathcal{R}(\mathcal{E}(\sigma)) = \sigma$ [43].

B. Quantum Block Entropy

Since stationary qudit processes are correlated across time, we explore how the von Neumann entropy for qudit blocks scales with block size. The following gives bounds

on the possible measurement sequences one can observe from a quantum information source. As the von Neumann entropy generalizes Shannon entropy, the results here (and many of the proofs) are natural generalizations of those in App. A. We also note that exactly determining $\rho_{0:\ell}$ becomes practically infeasible for large ℓ. And so, Section VI addresses how to approximate properties and models for qudit processes when restricted to measurements of finite length blocks.

For a qudit process we define the quantum block entropy as the von Neumann entropy of a block of ℓ consecutive qudits:

$$S(\ell) \equiv -tr(\rho_{0:\ell} \log_2 \rho_{0:\ell}) \ .$$

If $\rho_{0:\ell}$ is a pure state, $S(\ell) = 0$. By the same logic as the classical case, $S(0) \equiv 0$.

Many properties of the classical block entropy hold for $S(\ell)$.

Proposition 1. *For a stationary qudit process $S(\ell)$ is a nondecreasing function of ℓ.*

Proof. *As a consequence of the strong subadditivity of the von Neumann entropy [33]:*

$$S(\rho^A) + S(\rho^C) \leq S(\rho^{AB}) + S(\rho^{BC}) \ . \tag{22}$$

Let $\rho^{ABC} = \rho_{0:2\ell+1}$, where $\rho^A = \rho_{0:\ell}$, $\rho^B = \rho_\ell$, and $\rho^C = \rho_{\ell+1:2\ell+1}$.

Incorporating qudit process stationarity, we rewrite Eq. (22):

$$\begin{aligned} S(\rho_{0:\ell}) + S(\rho_{\ell+1:2\ell+1}) &\leq S(\rho_{0:\ell+1}) + S(\rho_{\ell:2\ell+1}) \\ S(\ell) + S(\ell) &\leq S(\ell+1) + S(\ell+1) \\ S(\ell) &\leq S(\ell+1) \ , \end{aligned} \tag{23}$$

where Eq. (23) follows from stationarity.

Thus, $S(\ell) \leq S(\ell+1)$, for all $\ell \geq 0$, and $S(\ell)$ is a nondecreasing function of ℓ.

Proposition 2. *For a stationary qudit process $S(\ell)$ is concave.*

Proof. *The von Neumann entropy is strongly subadditive [33], meaning that:*

$$S(\rho^{ABC}) + S(\rho^B) \leq S(\rho^{AB}) + S(\rho^{BC}) \ . \tag{24}$$

For $\ell \geq 3$, let $\rho^{ABC} = \rho_{0:\ell}$ where $\rho^A = \rho_0$, $\rho^B = \rho_{1:\ell-1}$, and $\rho^C = \rho_{\ell-1}$

We can rewrite Eq. (24) by incorporating the stationarity of qudit processes:

$$\begin{aligned} S(\rho_{0:\ell}) + S(\rho_{1:\ell-1}) &\leq S(\rho_{0:\ell-1}) + S(\rho_{1:\ell}) \\ S(\ell) + S(\ell-2) &\leq S(\ell-1) + S(\ell-1) \\ S(\ell) - 2S(\ell-1) + S(\ell-2) &\leq 0 \ , \end{aligned} \tag{25}$$

where Eq. (25) follows from stationarity.

Thus, $S(\ell)$ is concave.

For a separable qudit process formed by passing a classical process through a classical-quantum channel, their block entropies are related in the following way:

Proposition 3. *Let $|\Psi_{\overleftarrow{X}:\overrightarrow{X}}\rangle = \mathcal{E}(\overleftrightarrow{X})$. The block entropies of $|\Psi_{-\infty:\infty}\rangle$ and $\overleftrightarrow{X}$ obey:*

$$S(\rho_{0:\ell}) \leq \mathrm{H}\left[X_{0:\ell}\right] \ ,$$

for all ℓ, with equality if and only if $\mathcal{Q}$ consists of $|\mathcal{X}|$ orthogonal pure states in $\mathcal{H}$ of dimension $d \geq |\mathcal{X}|$.

Proof. *Recall from Eq. (6) that for separable qudit processes:*

$$\rho_{0:\ell} = \sum_{w \in \mathcal{X}^\ell} \Pr(w) \left|\psi_w\right\rangle \left\langle\psi_w\right| \ ,$$

with each $|\psi_w\rangle$ taking the separable form of Eq. (5) and $w = x_{0:\ell}$.

We first note that, for all symbols in $\mathcal{X}$ to be associated with orthogonal qudit states, the minimum dimension of the Hilbert space is $d_{min} = |\mathcal{X}|$.

With $\rho_{0:\ell}$ written as a mixture of separable qudit words, we apply Lemma 1 to obtain:

$$\begin{aligned} S(\rho_{0:\ell}) &\leq \mathrm{H}\left[X_{0:\ell}\right] + \sum_{w \in \mathcal{X}^\ell} \Pr(w) S(|\psi_w\rangle) \\ &\leq \mathrm{H}\left[X_{0:\ell}\right] \ , \end{aligned}$$

where $w = x_{0:\ell}$. The second term evaluates to zero in the final line since each $|\psi_w\rangle$ is a pure state. That is, $S(|\psi_w\rangle) = 0$ for all $w \in \mathcal{X}^\ell$.

Equality occurs if and only if the states $|\psi_w\rangle$ have support on orthogonal subspaces, which requires that $d \geq |\mathcal{X}|$ and all elements of $\mathcal{Q}$ are orthogonal.

We cannot use $S(\rho_{0:\ell})$ to bound the block entropy of a measured process $\overleftrightarrow{Y}$ for *general* POVM measurements. For the case where the measurement $\mathcal{M}_{0:\ell}$ consists only of rank-one POVMs (including all PVMs), however, the following holds:

$$S(\rho_{0:\ell}) \leq \mathrm{H}\left[\mathcal{M}_{0:\ell}(\rho_{0:\ell})\right] \ , \tag{26}$$

with equality if and only if the measurement is performed in the minimum-entropy (eigen)basis of $\rho_{0:\ell}$. This follows directly from Eq. (14).

Proposition 4. *Let $\overleftrightarrow{Y} = \mathcal{M}(|\Psi_{-\infty:\infty}\rangle)$, where $\mathcal{M}$ is a repeated rank-one POVM measurement. The block entropies of $|\Psi_{-\infty:\infty}\rangle$ and $\overleftrightarrow{Y}$ then obey:*

$$S(\rho_{0:\ell}) \leq \mathrm{H}\left[Y_{0:\ell}\right] \ ,$$

for all ℓ, with equality if and only if $|\Psi_{-\infty:\infty}\rangle$ is a separable process with an orthogonal alphabet $\mathcal{Q}$, and $\mathcal{M}$ uses a POVM whose operators include one projector for each element in $\mathcal{Q}$.

Proof. *The bound follows directly from Eq. (26) because repeated rank-one POVM measurements are a subclass of the more general measurement sequence $\mathcal{M}_{0:\ell}$.*

The condition for equality also follows from Eq. (26) but requires more justification. First, we consider measuring the single-qubit marginal ρ_0 with POVM M. For equality each element of the eigenbasis of ρ_0 ($\{|e_i\rangle\}$) must have a corresponding operator in M that is a projector on that eigenspace ($E_i = |e_i\rangle\langle e_i|$).

We can write $\rho_0 = \sum_i p_i |e_i\rangle\langle e_i|$. Since we apply the same POVM to each qudit, all blocks of length-ℓ must have eigenstates of the form $\bigotimes_{t=0}^{\ell-1} |e_t\rangle$—i.e., they must take the separable form of Eq. (5)—making $|\Psi_{-\infty:\infty}\rangle$ a separable process with a quantum alphabet $\mathcal{Q}$ of orthogonal states. The M consisting of projectors onto the states in $\mathcal{Q}$ is then the minimum-entropy measurement over blocks $\rho_{0:\ell}$. Note that there may be other elements of the POVM that are not projectors if the probability of those measurement outcomes is 0 when applied to $\rho_{0:\ell}$. (If the process does not make use of that part of Hilbert space, it does not matter how it is measured.)

To summarize, in the case of separable qudit processes, $S(\ell)$ is upper-bounded by the underlying classical process' block entropy $\mathrm{H}\left[X_{0:\ell}\right]$. For repeated measurement with rank-one POVMs, $S(\ell)$ serves as a lower bound on the block entropy of all classical measured processes $\mathrm{H}\left[Y_{0:\ell}\right]$. There is no direct relationship between $\mathrm{H}\left[X_{0:\ell}\right]$ and $\mathrm{H}\left[Y_{0:\ell}\right]$. Rather, it depends on the specifics of $\mathcal{E}$ and $\mathcal{M}$.

C. von Neumann Entropy Rate

The von Neumann entropy rate of a qudit process is:

$$s = \lim_{\ell\to\infty} \frac{S(\ell)}{\ell} \ . \tag{27}$$

The units of s are *bits per timestep*. This quantity is equivalent to the *mean entropy*, first introduced in the context of quantum statistical mechanics [44]. The limit exists for all stationary processes [45]. Operationally, s is the optimal coding rate for a stationary quantum source [17].

Proposition 5. *For a stationary qudit process we can equivalently write the von Neumann entropy rate as:*

$$s = \lim_{\ell\to\infty} S(\rho_0|\rho_{-\ell:0}) \ . \tag{28}$$

Proof. *This proof closely follows the proof for classical entropy rate in Ref. [46]. We begin by showing that $\lim_{\ell\to\infty} S(\rho_0|\rho_{-\ell:0})$ exists and then that it is equivalent to the limit in Eq. (27). The limit exists if $S(\rho_0|\rho_{-\ell:0})$ is a decreasing, nonnegative function of ℓ:*

$$\begin{aligned} S(\rho_0|\rho_{-\ell:0}) &= S(\rho_{-\ell:1}) - S(\rho_{-\ell:0}) \\ &= S(\ell+1) - S(\ell) \qquad (29) \\ &\geq 0 \ , \qquad (30) \end{aligned}$$

where Eq. (29) follows from stationarity and Eq. (30) follows from the nondecreasing nature of $S(\ell)$. This, combined with the fact that $S(\ell)$ is concave, means $\lim_{\ell\to\infty} S(\rho_0|\rho_{-\ell:0})$ exists.

Now, we establish that $\lim_{\ell\to\infty} S(\rho_0|\rho_{-\ell:0}) = \lim_{\ell\to\infty} S(\ell)/\ell$. Through repeated application of Eq. (17) to a block of length-ℓ we obtain the following chain rule for the von Neumann entropy:

$$S(\rho_{0:\ell}) = \sum_{j=0}^{\ell-1} S(\rho_j|\rho_{j-1}...\rho_0) \ . \tag{31}$$

We can modify indices (due to stationarity) and divide both sides by ℓ to obtain:

$$\frac{S(\rho_{0:\ell})}{\ell} = \frac{1}{\ell}\sum_{i=1}^{\ell} S(\rho_i|\rho_{i-1}...\rho_1) \ . \tag{32}$$

The final steps require the following result, known as the Cesáro mean [46]:

Lemma 2. *If $a_n \to a$ and $b_n = \frac{1}{n}\sum_{i=1}^{n} a_i$, then $b_n \to a$.*

Taking the limit of both sides of Eq. (32) and applying Lemma 2, we find:

$$\lim_{\ell\to\infty} \frac{S(\rho_{0:\ell})}{\ell} = \lim_{\ell\to\infty} S(\rho_\ell|\rho_{\ell-1}...\rho_1) \ . \tag{33}$$

Using stationarity, $\lim_{\ell\to\infty} S(\rho_\ell|\rho_{\ell-1}\ldots\rho_1) = \lim_{\ell\to\infty} S(\rho_0|\rho_{-\ell:0})$. When combined with Eq. (33) this proves that our two definitions of s are equivalent.

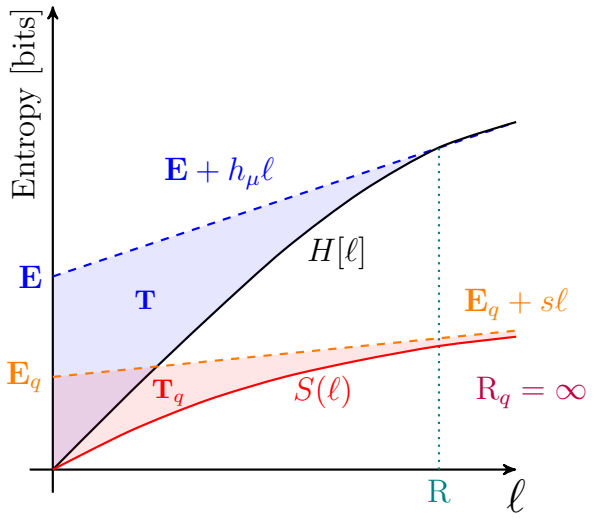


FIG. 2. Convergence of the block entropies to their linear asymptotes. $\mathrm{H}[\ell]$ is the block entropy for a finitary classical process with Markov Order R (see App. A), and $S(\ell)$ is the quantum block entropy for a finitary quantum process with infinite Quantum Markov order R_q. For the classical process $\mathbf{E}$ is the excess entropy, and h_μ is its Shannon entropy rate. Similarly, for a quantum process $\mathbf{E}_q$ is the quantum excess entropy, and s is its von Neumann entropy rate. The area of the blue shaded region is the classical transient information $\mathbf{T}$ and the area of the red shaded region is the quantum transient information $\mathbf{T}_q$.

To motivate a number of the following results it is important to appreciate that simply because a process has a von Neumann entropy rate given by Eq. (27) does not imply that an observer is able to perform a measurement on any qudit such that the uncertainty in that individual measurement is s. Rather, obtaining s corresponds to the measurement basis over the entire chain of qudits for which the distribution of outcomes has the minimal Shannon entropy. This basis is highly nonlocal or otherwise experimentally infeasible for many nontrivial examples. As in the classical case, s appears graphically as the slope of $S(\ell)$ as $\ell \to \infty$, as shown in Fig. 2.

For separable qudit processes, we can also relate s to the classical entropy rate of the underlying process $\overleftrightarrow{X}$, as follows:

Proposition 6. *Let $|\Psi_{-\infty:\infty}\rangle = \mathcal{E}(\overleftrightarrow{X})$. The von Neumann entropy rate s of $|\Psi_{-\infty:\infty}\rangle$ then obeys the bound:*

$$s \leq h_\mu^X \ ,$$

where h_μ^X is the Shannon entropy rate of $\overleftrightarrow{X}$. Equality occurs if and only if $\mathcal{Q}$ consists of $|\mathcal{X}|$ orthogonal pure states in $\mathcal{H}^d$ of dimension $d \geq |\mathcal{X}|$.

Proof. *Divide both sides of Prop. 3 by ℓ and take the limit of both sides as $\ell \to \infty$ to obtain:*

$$\lim_{\ell\to\infty} \frac{S(\rho_{0:\ell})}{\ell} \leq \lim_{\ell\to\infty} \frac{\mathrm{H}[X_{0:\ell}]}{\ell} \ .$$

From Eq. (27), the left side is s, and from Eq. (A1) the right side h_μ^X. The condition for equality is inherited from Prop. 3, concluding the proof.

Restricting once again to repeated measurements with rank-one POVMs, we can prove the following bound for measured processes:

Proposition 7. *Let $\overleftrightarrow{Y} = \mathcal{M}(|\Psi_{-\infty:\infty}\rangle)$ and let $\mathcal{M}$ be a repeated rank-one POVM. The measured entropy rate h_μ^Y then obeys:*

$$s \leq h_\mu^Y \ ,$$

where s is the von Neumann entropy rate of $|\Psi_{-\infty:\infty}\rangle$, with equality if and only if $|\Psi_{-\infty:\infty}\rangle$ is a separable process with an orthogonal alphabet $\mathcal{Q}$, and $\mathcal{M}$ uses a POVM whose operators include a projector for each element in $\mathcal{Q}$.

Proof. *Divide both sides of Prop. 4 by ℓ and take the limit of both sides as $\ell \to \infty$ to obtain:*

$$\lim_{\ell\to\infty} \frac{S(\rho_{0:\ell})}{\ell} \leq \lim_{\ell\to\infty} \frac{\mathrm{H}[Y_{0:\ell}]}{\ell} \ .$$

The left side is s and the right side h_μ^Y. The conditions for equality are inherited from Prop. 4, concluding the proof.

D. Quantum Redundancy

Unlike a classical process, the maximum entropy rate for a qudit process depends on the size of the Hilbert space rather than on the size of the alphabet $\mathcal{Q}$. For Hilbert space of dimension d, the largest possible value of s is $\log_2(d)$, corresponding to an i.i.d. sequence of qudits, each in a maximally-mixed state $\rho_{iid} = \mathbb{I}/d$.

A qudit process can be compressed down to its von Neumann entropy rate s. The amount that it can be compressed is the *quantum redundancy*:

$$\mathbf{R}_q \equiv \log_2(d) - s \ . \tag{34}$$

Statistical biases in individual qudits and temporal correlations between them offer opportunities for compression. $\mathbf{R}_q$ includes the effects of both.

For separable qudit processes we can bound the quantum redundancy using properties of the underlying classical process:

Proposition 8. *Let $\overleftrightarrow{X}$ be a classical process with redundancy $\mathbf{R}^X$, symbol alphabet $\mathcal{X}$, and entropy rate h_μ^X, and let $|\Psi_{-\infty:\infty}\rangle = \mathcal{E}(\overleftrightarrow{X})$ be a qudit process with redundancy $\mathbf{R}_q$, Hilbert space of dimension d, and entropy rate s.*

For $d \geq |\mathcal{X}|$:

$$\mathbf{R}_q \geq \mathbf{R}^X \ ,$$

with equality if and only if $d = |\mathcal{X}|$ and $\mathcal{Q}$ consists of $|\mathcal{X}|$ orthogonal pure states.

For $d < |\mathcal{X}|$,

$$\mathbf{R}_q < \mathbf{R}^X + (h_\mu^X - s) \ ,$$

where the term $(h_\mu^X - s)$ is always positive from Prop. (6).

Proof. *First, consider $d = |\mathcal{X}|$:*

$$\begin{aligned} \mathbf{R}_q &= \log_2(|\mathcal{X}|) - s \\ &= \mathbf{R}^X + h_\mu^X - s \\ &\geq \mathbf{R}^X \ , \end{aligned}$$

The final line comes from Prop. 6, as does the condition for equality.

For $d > |\mathcal{X}|$:

$$\begin{aligned} \mathbf{R}_q &= \log_2(d) - s \\ &> \log_2(|\mathcal{X}|) - s \\ &> \mathbf{R}^X \ . \end{aligned}$$

There is no opportunity for equality. In this case, $\mathcal{Q}$ will not span $\mathcal{H}$, naturally leading to more redundancy.

Finally, for $d < |\mathcal{X}|$:

$$\begin{aligned} \mathbf{R}_q &< \log_2(d) - s \\ &< \log_2(|\mathcal{X}|) - s \\ &< \mathbf{R}^X + (h_\mu^X - s) \ , \end{aligned}$$

concluding the proof.

We can also compare the classical redundancy of a measured process (obtained through repeated use of a rank-one POVM) to the quantum redundancy of the qudit process being measured.

Proposition 9. *Let $\overleftrightarrow{Y}$ be a measured process such that $\overleftrightarrow{Y} = \mathcal{M}(|\Psi_{-\infty:\infty}\rangle)$, and $\mathcal{M}$ a repeated rank-one POVM. Let $\overleftrightarrow{Y}$ have redundancy $\mathbf{R}^Y$, and let $|\Psi_{-\infty:\infty}\rangle$ have quantum redundancy $\mathbf{R}_q$. Then:*

$$\mathbf{R}_q \leq \mathbf{R}^Y \ , \tag{35}$$

with equality if and only if $d = |\mathcal{Y}|$, $|\Psi_{-\infty:\infty}\rangle$ is a separable process with an orthogonal alphabet $\mathcal{Q}$, and $\mathcal{M}$ uses a POVM whose operators include a projector for each element in $\mathcal{Q}$.

Proof. *A rank-one POVM on $\mathcal{H}$ must have at least d elements, therefore $|\mathcal{Y}| \geq d$:*

$$\begin{aligned} \mathbf{R}_q &= \log_2(d) - s \\ &\leq \log_2(|\mathcal{Y}|) - s \\ &\leq \mathbf{R}^Y + h_\mu^Y - s \\ &\leq \mathbf{R}^Y \ . \end{aligned}$$

Going from the first line to the second provides a condition for equality: $d = |\mathcal{Y}|$. Proposition 7 is used in the final line and provides the other conditions for equality.

E. Quantum Entropy Gain

We can take discrete-time derivatives of $S(\ell)$, as was done for $\mathrm{H}[\ell]$ in [11]. This process is summarized in App. A. We call the first derivative of $S(\ell)$ the *quantum entropy gain*:

$$\Delta S(\ell) \equiv S(\ell) - S(\ell - 1) \ , \tag{36}$$

for $\ell > 0$. The units for the quantum entropy gain are *bits per timestep*, and we set the boundary condition $\Delta S(0) = \log_2(d)$. Since $S(\ell)$ is monotone-increasing, $\Delta S(\ell) \geq 0$.

The quantum entropy gain is the amount of additional uncertainty introduced by including the ℓ^{th} qudit in a block, where that uncertainty is quantified by the von Neumann entropy.

By combining Eq. (36) and Eq. (17), we can write $\Delta S(\ell)$ as:

$$\Delta S(\ell) = S(\rho_0 | \rho_{-\ell:0}) \ .$$

This allows relating the quantum entropy gain and the von Neumann entropy rate as follows:

$$s = \lim_{\ell \to \infty} \Delta S(\ell) \ . \tag{37}$$

Thus, paralleling the classical case, the quantum entropy gain serves as a finite-ℓ approximation of the von Neumann

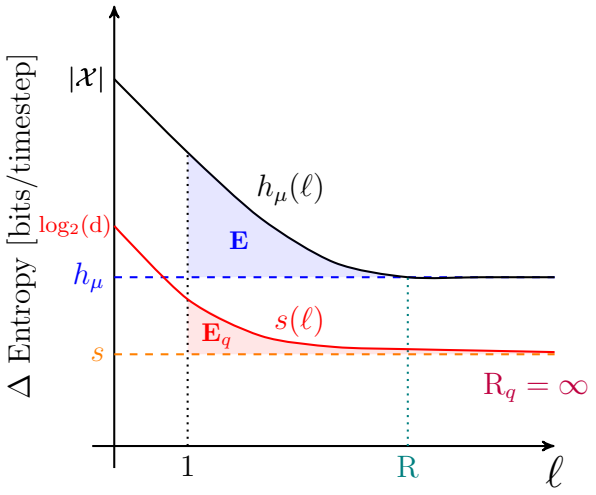


FIG. 3. Convergence of $\Delta H(\ell)$ (for a finitary classical process with Markov order R) and $\Delta S(\ell)$ (for a finitary quantum process with infinite quantum Markov order R_q) to the processes' entropy rates, h_μ and s. The shaded areas are the classical (blue) and quantum (red) excess entropies.

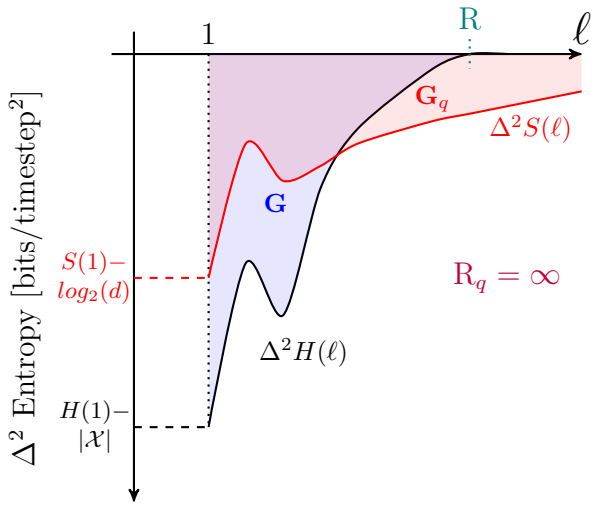


FIG. 4. Convergence of the predictability ($\Delta^2 H(\ell)$ and $\Delta^2 S(\ell)$) to 0 for a finitary classical process with Markov order R and a finitary quantum process with infinite Quantum Markov order R_q. Note that the predictability is not monotonic (unlike the block entropy and its first derivative). The overlapping shaded areas represent the magnitude of the classical (blue) and quantum (red) predictabilities, $\mathbf{G}$ and $\mathbf{G}_q$, which are negative by convention.

entropy rate:

$$s(\ell) \equiv \Delta S(\ell) \ . \tag{38}$$

$s(\ell)$ serves as the best estimate for the entropy rate of a qudit process that can be made by an observer who only has access to measurement statistics for length-ℓ blocks of qudits.

The way in which the entropy rate estimate converges and its relationship to other information properties of a qudit process are summarized in Fig. 3.

F. Quantum Predictability Gain

We call the second derivative of $S(\ell)$ the *quantum predictability gain*, given by:

$$\begin{aligned}\Delta^2 S(\ell) &\equiv \Delta s(\ell) \\ &= s(\ell) - s(\ell - 1) \ ,\end{aligned}$$

for $\ell > 0$. The units of $\Delta^2 S(\ell)$ are *bits per timestep*^{2}. Since $S(\ell)$ is concave, then $\Delta^2 S(\ell) \leq 0$. $|\Delta^2 S(\ell)|$ quantifies how much an observer's estimate of the von Neumann entropy rate s improves if they are enlarge their observations from blocks of $\ell - 1$ to blocks of ℓ qudits. The generic convergence behavior of $\Delta^2 S(\ell)$ is shown in Fig. 4.

For all higher-order discrete derivatives of $S(\ell)$ (as with the classical block entropy):

$$\lim_{\ell \to \infty} \Delta^n S(\ell) = 0, n \geq 2 \ .$$

This follows directly from the existence of the limit in Eq. (27) for stationary quantum states.

G. Total Quantum Predictability

Up to this point, introducing new information-theoretic characteristics of separable quantum processes proceeded by taking discrete-time derivatives of the von Neumann block entropy. We can likewise integrate the functions $\Delta^n S(\ell)$, as is done for the classical case with Eq. (A3). While this starts off straightforwardly, a number of interesting new informational quantities emerge.

These properties of qudit processes take the general form:

$$\mathcal{I}_n^q \equiv \sum_{\ell=\ell_0}^{\infty} [\Delta^n S(\ell) - \lim_{\ell \to \infty} \Delta^n S(\ell)] \ ,$$

where ℓ_0 is the first value of ℓ for which $\Delta^n S(\ell)$ is defined.

$\mathcal{I}_2^q$, the first of these, monitors the convergence of the quantum predictability gain $\Delta^2 S(\ell)$ to its limit of 0 for $\ell \to \infty$. We use $\ell_0 = 1$ to get the *total quantum pre-*

dictability $\mathbf{G}_q$:

$$\begin{aligned}\mathbf{G}_q &\equiv \mathcal{I}_2^q \\ &= \sum_{\ell=1}^{\infty} \Delta^2 S(\ell) \ . \end{aligned} \tag{39}$$

The units of $\mathbf{G}_q$ are *bits per timestep*. Note that $\mathbf{G}_q \leq 0$ because $\Delta^2 S(\ell) \leq 0$ for all ℓ.

$\mathbf{G}_q$ can be interpreted by relating it to a previously-established property of qudit processes: the quantum redundancy.

Proposition 10. *For a stationary qudit process:*

$$\mathbf{G}_q = -\mathbf{R}_q \ .$$

Proof. *Applying Eq. (A3) to Eq. (39) we find that:*

$$\begin{aligned}\mathbf{G}_q &= \lim_{\ell\to\infty} \Delta S(\ell) - \Delta S(0) \\ &= s - \log_2(d) \\ &= -\mathbf{R}_q \ . \end{aligned}$$

Here, the second line follows from Eq. (37) and the third from Eq. (34).

Thus, $|\mathbf{G}_q|$ is the total amount of predictable information per timestep for a qudit process.

Rather immediately, one sees that the amount of information in an individual qudit decomposes into:

$$\log_2(d) = |\mathbf{G}_q| + s \ .$$

$|\mathbf{G}_q|$ is the amount of quantum information within a qudit that is predictable. Whereas, s is the amount of information that is irreducibly random.

The relation $|\mathbf{G}_q| = \mathbf{R}_q$ can be combined with Props. 8 and 9 to prove two corollaries.

Corollary 1. *Let* $\overleftrightarrow{X}$ *be a classical process with total predictability* $\mathbf{G}^X$ *and symbol alphabet* $\mathcal{X}$*, and entropy rate* h_μ^X*, and let* $|\Psi_{-\infty:\infty}\rangle = \mathcal{E}(\overleftrightarrow{X})$ *be a qudit process with total quantum predictability* $\mathbf{G}_q$*, Hilbert space of dimension* d *and entropy rate* s*.*
For $d \geq |\mathcal{X}|$*:*

$$|\mathbf{G}_q| \geq |\mathbf{G}^X| \ ,$$

with equality if and only if $d = |\mathcal{X}|$ *and* $\mathcal{Q}$ *consists of* $|\mathcal{X}|$ *orthogonal pure states.*
For $d < |\mathcal{X}|$*,*

$$|\mathbf{G}_q| < |\mathbf{G}^X| + (h_\mu^X - s) \ ,$$

where the term $(h_\mu^X - s)$ *is always positive from Prop. (6).*

Proof. *This follows immediately from combining Props. 8 and 10.*

Corollary 2. *Let* $\overleftrightarrow{Y}$ *be a measured process such that* $\overleftrightarrow{Y} = \mathcal{M}(|\Psi_{-\infty:\infty}\rangle)$*, and* $\mathcal{M}$ *is a repeated rank-one POVM. Let* $\overleftrightarrow{Y}$ *have redundancy* $\mathbf{G}^Y$*, and let* $|\Psi_{-\infty:\infty}\rangle$ *have quantum redundancy* $\mathbf{G}_q$*. Then:*

$$|\mathbf{G}_q| \leq |\mathbf{G}^Y| \ ,$$

with equality if and only if $d = |\mathcal{Y}|$*,* $|\Psi_{-\infty:\infty}\rangle$ *is a separable process with an orthogonal alphabet* $\mathcal{Q}$*, and* $\mathcal{M}$ *uses a POVM whose operators include a projector for each element in* $\mathcal{Q}$*.*

Proof. *This follows immediately from combining Props.* 9 *and* 10*.*

Graphically, the total quantum predictability is the area between the predictability gain curve and its linear asymptote of 0, as seen in Fig. 4. The von Neumann entropy rate and total predictability lend insight into compression limits for a stationary sources. They do not say, however, whether that compression is achievable due to bias within individual qudit states or correlations between qudits. For that we must continue our way back up the entropy hierarchy.

H. Quantum Excess Entropy

The convergence of $s(\ell)$ to the true von Neumann entropy rate s is quantified with the *quantum excess entropy*:

$$\begin{aligned}\mathbf{E}_q \equiv \mathcal{I}_1^q &= \sum_{\ell=1}^{\infty} \left[\Delta S(\ell) - \lim_{\ell\to\infty} \Delta S(\ell)\right] \\ &= \sum_{\ell=1}^{\infty} \left[s(\ell) - s\right] \ . \end{aligned} \tag{40}$$

The units for $\mathbf{E}_q$ are *bits*. Paralleling the classical case, we refer to any qudit process with finite $\mathbf{E}_q$ as *finitary* and those with infinite $\mathbf{E}_q$ as *infinitary*.

We can further express $\mathbf{E}_q$ in terms of the asymptotic behavior of $S(\ell)$.

Proposition 11. *The quantum excess entropy can be written as:*

$$\mathbf{E}_q = \lim_{\ell\to\infty} \left[S(\ell) - s\ell\right] \ . \tag{41}$$

Proof. *We evaluate the discrete integral in Eq. (40) with Eq. (A3) using partial sums:*

$$\begin{aligned}\mathbf{E}_q &= \lim_{\ell\to\infty}\sum_{m=1}^{\ell}\left[\Delta s(m) - s\right] \\ &= \lim_{\ell\to\infty}\left[S(\ell) - S(0) - s\ell\right] \\ &= \lim_{\ell\to\infty}\left[S(\ell) - s\ell\right] \ ,\end{aligned}$$

since $S(0) = 0$ by definition.

For finitary quantum processes, $\mathbf{E}_q$ is the area between the entropy gain curve and its asymptote s as seen in Fig. 3. It also appears in Fig. 2 as the vertical offset of the linear asymptote to the $S(\ell)$ curve.

This leads to a natural scaling of the quantum block entropy as:

$$S(\ell) \sim \mathbf{E}_q + s\ell \ ,$$

as $\ell \to \infty$.

A clearer interpretation of $\mathbf{E}_q$ as a quantum mutual information is provided by the following proposition:

Proposition 12. *The quantum excess entropy can be written as:*

$$\mathbf{E}_q = \lim_{\ell\to\infty} S(\rho_{-\ell:0}{:}\rho_{0:\ell}) \ ,$$

where $\rho_{-\ell:0}$ and $\rho_{0:\ell}$ are two blocks of ℓ consecutive qudits with a shared boundary.

Proof. *The quantum mutual information, from Eq. (19), between two neighboring blocks of ℓ qudits can be expressed as:*

$$\begin{aligned}S(\rho_{0:\ell}{:}\rho_{-\ell:0}) &= S(\rho_{0:\ell}) - S(\rho_{0:\ell}|\rho_{-\ell:0}) \\ &= S(\ell) - \sum_{t=0}^{\ell-1} S(\rho_t|\rho_{-\ell:t-1}) \ ,\end{aligned}$$

where the final line is obtained through repeated application of Eq. (17).

Taking $\ell \to \infty$:

$$\begin{aligned}\lim_{\ell\to\infty} S(\rho_{0:\ell}{:}\rho_{-\ell:0}) &= \lim_{\ell\to\infty}\left[S(\ell) - \sum_{t=0}^{\ell-1} S(\rho_t|\rho_{-\ell:t-1})\right] \\ &= \lim_{\ell\to\infty}\left[S(\ell) - s\ell\right] \ ,\end{aligned}$$

where the final line follows from Eq. (32) and stationarity.

This final expression is equivalent to the form of $\mathbf{E}_q$ derived in Prop. 11, concluding the proof.

$\mathbf{E}_q$ is therefore a measure of all the correlations between two halves of the infinite sequence of qudits. $\mathbf{E}_q = 0$, if and only if a source is i.i.d. (with $S(\ell) = s\ell$).

We can relate $\mathbf{E}_q$ for a separable qudit process to $\mathbf{E}^X$ of the underlying classical process:

Proposition 13. *Let $\overleftrightarrow{X}$ be a classical process with alphabet $\mathcal{X}$ and excess entropy $\mathbf{E}^X$, and let $|\Psi_{-\infty:\infty}\rangle = \mathcal{E}(\overleftrightarrow{X})$ be a qudit process with alphabet $\mathcal{Q}$ and quantum excess entropy $\mathbf{E}_q$.*

Then:

$$\mathbf{E}_q \leq \mathbf{E}^X \ ,$$

with equality if and only if $\mathcal{Q}$ consists of $|\mathcal{X}|$ orthogonal states.

Proof. *Consider $X_{-\ell:\ell}$, a block of length 2ℓ of the classical process $\overleftrightarrow{X}$. We can write realizations of $X_{-\ell:\ell}$ into a classical register to form the state:*

$$\rho^C_{-\ell:\ell} = \sum_{x_{-\ell:\ell}} \Pr(x_{-\ell:\ell}) \left|x_{-\ell:\ell}\right\rangle \left\langle x_{-\ell:\ell}\right| \ , \tag{42}$$

where all $|x_{-\ell:\ell}\rangle$ are orthogonal. Then we pass each symbol through the preparation channel $\mathcal{E}$ to obtain blocks of our qudit process $\rho_{-\ell:\ell} = \mathcal{E}^{2\ell}(\rho^C_{-\ell:\ell})$, where $\mathcal{E}^{2\ell} = \bigotimes_{i=0}^{2\ell}\mathcal{E}$.

We can express the quantum mutual information as a quantum relative entropy, Eq. (20), giving the following relation:

$$\begin{aligned}\mathrm{I}\left[X_{-\ell:0}{:}X_{0:\ell}\right] &= S(\rho^C_{-\ell:0}{:}\rho^C_{0:\ell}) \\ &\geq S(\mathcal{E}^\ell(\rho^C_{-\ell:0}){:}\mathcal{E}^\ell(\rho^C_{0:\ell})) \\ &\geq S(\rho_{-\ell:0}{:}\rho_{0:\ell}) \ ,\end{aligned} \tag{43}$$

where Eq. (43) comes from the monotonicity of the quantum relative entropy in Eq. (21). The condition for equality comes from Eq. (21) as well. The set of states for which the recovery map must exist is $\mathcal{Q}$, and this is only possible if all $|\psi_x\rangle$ are distinguishable—i.e., orthogonal.

Using Eq. (A9), to write the excess entropy of $\overleftrightarrow{X}$ as a limit we see that:

$$\begin{aligned}\lim_{\ell\to\infty}\mathrm{I}\left[X_{-\ell:0}{:}X_{0:\ell}\right] &\geq \lim_{\ell\to\infty} S(\rho_{-\ell:0}{:}\rho_{0:\ell}) \\ \mathbf{E}^X &\geq \mathbf{E}_q \ .\end{aligned}$$

A similar bound appears when we apply a repeated POVM measurement to the quantum process to obtain a classical process.

Proposition 14. *Let $\overleftrightarrow{Y}$ be a measured process such that $\overleftrightarrow{Y} = \mathcal{M}(|\Psi_{-\infty:\infty}\rangle)$, and $\mathcal{M}$ is a repeated rank-one POVM. Let $\overleftrightarrow{Y}$ have excess entropy $\mathbf{E}^Y$, and let $|\Psi_{-\infty:\infty}\rangle$*

have quantum excess entropy $\mathbf{E}_q$*. Then:*

$$\mathbf{E}_q \geq \mathbf{E}^Y \ ,$$

with equality if and only if $|\Psi_{-\infty:\infty}\rangle$ *is a separable process with an orthogonal alphabet* $\mathcal{Q}$ *and* $\mathcal{M}$ *uses a POVM whose operators include a projector for each element in* $\mathcal{Q}$*.*

Proof. *Consider* $\rho_{-\ell:\ell}$*—a block of length* 2ℓ *of the quantum process* $|\Psi_{-\infty:\infty}\rangle$*—and let* $\mathcal{M}$ *be a repeated measurement of rank-one POVM* M *with elements* $\{E_y\}$ *so that* $M(\rho_t) = \sum_y \Pr(y) |y\rangle \langle y|$*,* $\Pr(y) = tr(E_y \rho_t)$*, and all* $|y\rangle$ *are orthogonal. The repeated measurement applied over a block of length* ℓ *is then* $\mathcal{M}_{0:\ell} = \bigotimes_{i=0}^{\ell-1} M$*.*

We express the quantum mutual information as a quantum relative entropy (using Eq. (20)), and apply Eq. (21) to obtain:

$$\begin{aligned} S(\rho_{-\ell:0}{:}\rho_{0:\ell}) &\geq S\left(\mathcal{M}_{0:\ell}(\rho_{-\ell:0}){:}\mathcal{M}_{0:\ell}(\rho_{0:\ell})\right) \\ &\geq S\Bigg(\sum_{y_{-\ell:0}} \Pr(y_{-\ell:0}) \bigotimes_{t=-\ell}^{0} |y_t\rangle \langle y_t| : \\ &\qquad \sum_{y_{0:\ell}} \Pr(y_{0:\ell}) \bigotimes_{t=0}^{\ell-1} |y_t\rangle \langle y_t| \Bigg) \\ &\geq \mathrm{I}\left[Y_{-\ell:0} : Y_{0:\ell}\right] \ . \end{aligned} \tag{44}$$

The condition for equality comes from Eq. (21) as well. The set of states for which the recovery map must exist is $\{|y\rangle\}$*, and this is only possible if all* $|\psi_x\rangle \in \mathcal{Q}$ *are orthogonal, and* M *contains a projector for each* $|\psi_x\rangle$*. By the same argument as Prop. 4,* $|\Psi_{-\infty:\infty}\rangle$ *must be a separable process, and* $\mathcal{Q}$ *must consist of orthogonal states.*

Taking the limit $\ell \to \infty$ *we see that:*

$$\lim_{\ell\to\infty} S(\rho_{-\ell:0}{:}\rho_{0:\ell}) \geq \lim_{\ell\to\infty} \mathrm{I}\left[Y_{-\ell:0} : Y_{0:\ell}\right]$$
$$\mathbf{E}_q \geq \mathbf{E}^Y .$$

Combining the above proofs we obtain the following corollary relating the excess entropies of the underlying classical process $\overleftrightarrow{X}$ and the measured process $\overleftrightarrow{Y}$:

Corollary 3. *Let* $\overleftrightarrow{X}$ *be a classical process with excess entropy* $\mathbf{E}^X$ *and alphabet* $\mathcal{X}$*, let* $|\Psi_{-\infty:\infty}\rangle = \mathcal{E}(\overleftrightarrow{X})$ *be a separable qudit process with alphabet* $\mathcal{Q}$*, and let* $\overleftrightarrow{Y}$ *be a measured process with excess entropy* $\mathbf{E}^Y$ *such that* $\overleftrightarrow{Y} = \mathcal{M}(|\Psi_{-\infty:\infty}\rangle)$ *where* $\mathcal{M}$ *is a repeated rank-one POVM.*

Then:

$$\mathbf{E}^X \geq \mathbf{E}^Y \ ,$$

with equality if and only if $\mathcal{Q}$ *consists of* $|\mathcal{X}|$ *orthogonal states and* $\mathcal{M}$ *uses a POVM whose operators include a projector for each element in* $\mathcal{Q}$*.*

Proof. *This follows immediately from combining Props. 13 and 14.*

An exact value of $\mathbf{E}_q$ typically requires characterizing infinite-length sequences of qudits. However, we can write a finite-ℓ estimate of $\mathbf{E}_q$ using Eq. (41):

$$\mathbf{E}_q(\ell) = S(\ell) - \ell s(\ell) \ , \tag{45}$$

that generally underestimates $\mathbf{E}_q$'s true value.

I. Quantum Transient Information

We now turn to look at how the quantum block entropy curve converges to its linear asymptote $\mathbf{E}_q + s\ell$. We define the quantum transient information as:

$$\mathbf{T}_q \equiv -\mathcal{I}_0^q = \sum_{\ell=1}^{\infty} [\mathbf{E}_q + s\ell - S(\ell)] \ . \tag{46}$$

The units of $\mathbf{T}_q$ are *bits*×*time steps*.

$\mathbf{T}_q$ is represented graphically as the area between the $S(\ell)$ curve and its linear asymptote for $\ell \to \infty$, as seen in Fig. 2. We will see that $\mathbf{T}_q$ distinguishes between periodic qudit processes that cannot be distinguished with previous information quantities such as $\mathbf{E}_q$ and s.

Proposition 15. *The transient quantum information* $\mathbf{T}_q$ *can be written:*

$$\mathbf{T}_q = \sum_{\ell=1}^{\infty} \ell \left[s(\ell) - s\right] \ .$$

Proof. *The proof reduces to the straightforward proof for transient information of a classical stochastic process in Ref. [11]. It depends only upon Eq. (46) and Eq. (A3), that have the same form in the quantum case as the classical case.*

This expression allows us to estimate $\mathbf{T}_q$ for a given quantum process as:

$$\mathbf{T}_q(\ell) = \sum_{m=1}^{\ell-1} m \left[s(m) - s(\ell)\right] \ , \tag{47}$$

that generally underestimates the true value of $\mathbf{T}_q$.

$\mathbf{T}_q$ is related to the minimal amount of information necessary for an observer to synchronize to a HMCQS. We say an observer is *synchronized* when they are able to determine a source's internal state. If $S(\ell)$ converges to

its linear asymptote at finite ℓ, then there exists an optimal POVM on $\rho_{0:\ell}$ (in the eigenbasis of $\rho_{0:\ell}$) that exactly determines the HMCQS's internal state. Note that this is not guaranteed to be a repeated POVM or even consist of local POVMs. The information within that measurement that is useful for synchronization is quantified by $\mathbf{T}_q$. In contrast, if $S(\ell)$ does not converge for finite-ℓ, then no such POVM over any finite block of qudits exists, and an observer can (at best) only converge asymptotically to the source's internal state. We will see in Section V that even this is not possible for most sources when we restrict ourselves to local measurements.

J. Quantum Markov Order

The quantum Markov order corresponds to the number of previous qudits on which the next qudit is conditionally dependent. A process has quantum Markov order R_q if R_q is the smallest value for which the following property holds:

$$S(\rho_0|\rho_{-R_q:0}) = S(\rho_0|\rho_{-\infty:0}).$$

A graphical interpretation is that when the block size reaches R_q, $S(\ell)$ levels off and has a constant slope—which is s, as seen in Fig. 2. As a consequence R_q is the value of ℓ for which $\Delta S(\ell)$ and $\Delta^2 S(\ell)$ converge to their asymptotic values, as seen in Figs. 3 and 4 respectively.

Note that a classical process is referred to as 'Markov' if $R = 1$. If a separable qudit process has an underlying classical process that is Markov, then it obeys the property in Eq. (9) but does not generally have $R_q = 1$.

Consider a separable quantum process $|\Psi_{-\infty:\infty}\rangle = \mathcal{E}(\overleftrightarrow{X})$ where $\overleftrightarrow{X}$ has Markov order R^X and $|\Psi_{-\infty:\infty}\rangle$ has quantum Markov order R_q. R_q can be equal to, less than, or greater than R^X. We will give a simple example of each case:

- $R_q = R^X$ trivially if $\mathcal{Q}$ consists of orthogonal states, in which case they have identical block entropy curves, via Prop. 3.
- $R_q < R^X$ if $R^X > 0$ and all symbols in $\mathcal{X}$ are mapped to the same pure state $|\psi_x\rangle$. In this case $R_q = 0$, as the resulting process is i.i.d.
- $R_q > R^X$ when $R^X > 0$, $|\mathcal{X}| = |\mathcal{Q}|$ and $\mathcal{Q}$ consists of nonorthogonal states. Frequently, $R_q = \infty$ since arbitrary sequences of nonorthogonal states cannot reliably be distinguished with a finite POVM.

Similar rules apply when comparing R_q to the Markov order R^Y of measured process $\overleftrightarrow{Y} = \mathcal{M}(|\Psi_{-\infty:\infty}\rangle)$:

- $R_q = R^Y$ if $|\Psi_{-\infty:\infty}\rangle$ is a separable process with an orthogonal alphabet $\mathcal{Q}$, and $\mathcal{M}$ uses a POVM whose operators include a projector for each element in $\mathcal{Q}$, via Prop. 4.
- $R_q < R_Y$ if $\mathcal{Q}$ consists of orthogonal states, $R_q = 1$ and $\mathcal{M}$ is a repeated rank-one POVM that does not include projectors onto the states in $\mathcal{Q}$.
- $R_q > R_Y$ if $R_q > 0$ and $\mathcal{M}$ is a repeated POVM measurement with the one-element POVM, $\mathbb{I}$. Note that this is not a rank-one POVM.

Now, with a toolbox of quantum information properties in hand, the following section calculates (or estimates) their values for paradigmatic examples of qudit processes.

IV. QUDIT PROCESSES

We now present a variety of separable qudit processes, organized roughly by increasing structural complexity and demonstrate the extent to which their behavior can be quantified with Section III's informational measures. Their properties are summarized in Table I near the end when we have their informational measures in hand.

A. I.I.D. Processes

Recall that an i.i.d. (independent and identically-distributed) qudit process has no classical or quantum correlation between any of the qudits, and its length-ℓ density matrices are in a product state such that:

$$\rho_{0:\ell} = \bigotimes_{i=0}^{\ell-1} \rho_{iid} \ ,$$

where $\rho_{iid} \in \mathcal{H}$ is the density matrix for a single timestep. The quantum block entropy takes the form $S(\ell) = \ell S(\rho_{iid})$, thus the von Neumann entropy rate is $s = S(\rho_{iid})$. This implies that a repeated projective measurement exists for which the measured process $\overleftrightarrow{Y}$ has a classical entropy rate of $h_\mu = S(\rho_{iid})$. The measurement that realizes this bound consists of orthogonal projectors P_y, each of which is constructed from an eigenvector of ρ_{iid}. In the special case where ρ_{iid} is the maximally-mixed state, $s = \log_2(d)$ and any set of orthogonal projectors on a block $\rho_{0:\ell}$ gives a uniform distribution over all measurement outcomes $y_{0:\ell}$. Since there are no correlations between qudits $\mathbf{E}_q = 0$, $\mathbf{T}_q = 0$, and $R_q = 0$ trivially. The output of a single-qudit source with uncorrelated noise can be considered an i.i.d. qudit process.

B. Quantum Presentations of Classical Processes

Any classical process with alphabet $\mathcal{X}$ can be represented by a qudit process with orthogonal alphabet $\mathcal{Q}$ where each symbol $x \in \mathcal{X}$ corresponds to a pure state $|\psi_x\rangle \in \mathcal{Q}$. In this case the encoding $\mathcal{E}$ is trivial and one can recover the underlying process $\overleftrightarrow{X}$ via repeated measurement with orthogonal projectors $\{P_x = |\psi_x\rangle\langle\psi_x|\}$. This requires that $d \geq |\mathcal{X}|$.

The information measures for the underlying process, the qudit process, and measurement outcomes obey:

$$\begin{aligned} \mathrm{H}\left[X_{0:\ell}\right] &= S(\rho_{0:\ell}) \\ &\leq \mathrm{H}\left[Y_{0:\ell}\right] \ , \end{aligned}$$

with equality for repeated measurement with a rank-one POVM whose elements include the projector set $\{P_x\}$.

From this relation we can see that many quantum information quantities such as s, $\mathbf{E}_q$, $\mathbf{T}_q$, and R_q are equal to the classical properties of $\overleftrightarrow{X}$. Exceptions include quantities that depend on the relationship between d and $|\mathcal{X}|$, such as the quantum redundancy.

Since the quantum-classical channel $\mathcal{E}$ is trivial, the output process $\overleftrightarrow{Y}$ is the result of passing $\overleftrightarrow{X}$ through a noisy classical channel, where the level of noise depends on the particular measurement scheme $\mathcal{M}$.

For a repeated rank-one POVM $\mathrm{H}\left[X_{0:\ell}\right] \leq \mathrm{H}\left[Y_{0:\ell}\right]$ for all ℓ from Props. (3) and (4) . Similar inequalities explicitly relate other classical process properties such as the entropy rates ($h_\mu^X \leq h_\mu^Y$), the predictabilities ($|\mathbf{G}^X| \geq |\mathbf{G}^Y|$), and the excess entropies (see Cor. 3).

C. Periodic Processes

A classical stochastic process $\overleftrightarrow{X}$ is periodic with period p if it consists of repetitions of a template sequence—a length-p block of symbols. A periodic separable qudit process $|\Psi_{-\infty:\infty}\rangle = \mathcal{E}(\overleftrightarrow{X})$ is one for which the underlying classical process $\overleftrightarrow{X}$ is periodic.

For classical periodic processes the block entropy curve reaches a maximal value at its Markov order $R = p$ and thereafter remains constant with increasing ℓ ($h_\mu = 0$). That maximal value is the excess entropy, which is entirely determined by the period according to the formula $\mathbf{E} = \log_2(p)$. Finally, an observer attempting to synchronize to different length-p templates may encounter more or less uncertainty in the process depending on the template itself, a feature captured by the transient information $\mathbf{T}$ [11].

Periodic qudit processes share many of these properties (as we show via Section III's results), but also exhibit richer behavior since they can consist of nonorthogonal qudit states. Fig. 5 shows the quantum block entropies for the period-3 process consisting of the repeated quantum word $|\psi_{00\phi}\rangle = |0\rangle \otimes |0\rangle \otimes |\psi(\phi)\rangle$ where $|\psi(\phi)\rangle = \cos\frac{\phi}{2}|0\rangle + \sin\frac{\phi}{2}|1\rangle$. For $\phi = \pi$ we recover the block entropy for the classical period-3 word '001'. As ϕ decreases $|\psi(\phi)\rangle$ becomes less distinguishable from $|0\rangle$.

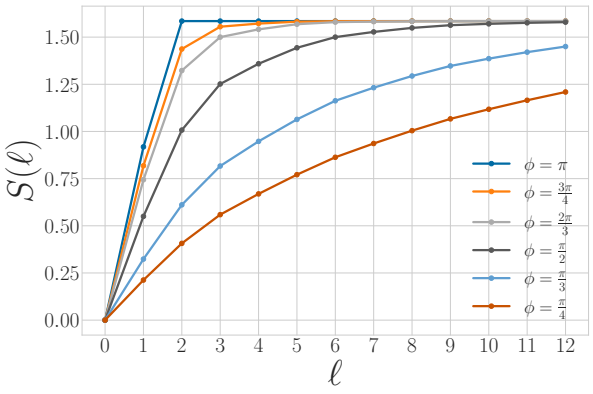


FIG. 5. Quantum block entropies $S(\ell)$ versus length ℓ for the periodic process emitting the state $|\psi_{00\phi}\rangle$ with different values of ϕ. Each curve approaches a maximum value of $\mathbf{E}_q = \log_2 3$. Larger values of ϕ correspond to more distinguishable alphabet states and lower values of $\mathbf{T}_q$.

From Props. 6 and 13 it follows that periodic qudit processes with period p have $s = 0$ and $\mathbf{E}_q \leq \log_2(p)$. (And, $\mathbf{E}_q = \log_2(p)$ unless two classical symbols in $\mathcal{X}$ are sent to the same pure state in $\mathcal{Q}$ in a way that reduces the effective period of the qudit process to less than p.) However, the quantum block entropy curve does not necessarily reach its maximal value of $\mathbf{E}_q$ for $\ell = p$ if $\mathcal{Q}$ contains nonorthogonal states. In this case, $R_q \to \infty$ since an observer cannot unambiguously distinguish where one length-p block begins and another one ends with any finite measurement.

Though all period-p qudit processes have the same von Neumann entropy rate and quantum excess entropy, they may be distinguished by their values for the quantum transient information in two different ways.

First, different quantum alphabets $\mathcal{Q}$ give different values of $\mathbf{T}_q$. Fig. 5 demonstrates that, for a 2-state qubit alphabet, as the states become more less distinguishable, the quantum transient information increases. For $\phi = \pi$ (orthogonal alphabet), $\mathbf{T}_q \approx 2.33$ bits×time steps, whereas for $\phi = \frac{\pi}{2}$, $\mathbf{T}_q \approx 4.22$ bits×time steps. These values of $\mathbf{T}_q$ (and others in this section) are numerically approximated using Eq. (47) with $\ell = 12$.

Second, $\mathbf{T}_q$ can distinguish between different length-p words. Reference [11] shows that $\mathbf{T}$ can distinguish between different period-5 classical words ('00001', '00011',

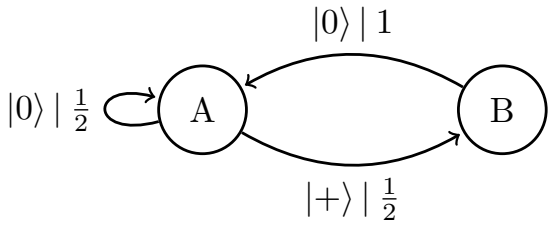


FIG. 6. $|0\rangle$-$|+\rangle$ Quantum Golden Mean Process generator.

and '00101'), and $\mathbf{T}_q$ generalizes this behavior. Whereas all period-3 words are equivalent to '001' under global bit swap and translations, the same is not true for period-5 words. Section V discusses synchronizing to period-5 qudit sources in more detail and relates that task to the value of $\mathbf{T}_q$.

D. Quantum Golden Mean Processes

The classical Golden Mean process consists of all binary strings with no consecutive '1's. It is a Markov process ($R = 1$) since the joint probabilities $\Pr(X_{0:\ell})$ for blocks factor as in Eq. (2) with $\Pr(0|0) = \frac{1}{2}$, $\Pr(1|0) = \frac{1}{2}$, $\Pr(0|1) = 1$, and $\Pr(1|1) = 0$. For the classical Golden Mean $h_\mu = \frac{2}{3}$ bits per symbol and $\mathbf{E} \approx 0.2516$ bits.

Replacing the classical symbol alphabet with the quantum alphabet $\mathcal{Q} = \{|0\rangle, |+\rangle\}$, where $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, gives the $|0\rangle$-$|+\rangle$ Quantum Golden Mean Process, introduced in Ref. [14]. Figure 6 shows its generator.

We can further generalize this process to the $|0\rangle$-$|\psi(\phi)\rangle$ Quantum Golden Mean process with quantum alphabet $\{|0\rangle, |\psi(\phi)\rangle = \cos\frac{\phi}{2}|0\rangle + \sin\frac{\phi}{2}|1\rangle\}$. This process' quantum entropy rate is shown in Fig. 7 for different values of ϕ. For $\phi = \pi$, $|\psi(\phi)\rangle = |1\rangle$ and we recover the classical Golden Mean process. As ϕ decreases to 0, the states in $\mathcal{Q}$ become less distinguishable and s decreases, as expected from Prop. 6.

Also in Fig. 7 we see the entropy rate h_μ^Y of the measured processes obtained by applying a repeated PVM to the $|0\rangle$-$|\psi(\phi)\rangle$ Quantum Golden Mean process. This PVM consists of projectors parametrized by the angle θ, where $M_\theta = \{|\psi(\theta)\rangle\langle\psi(\theta)|, |\psi(\theta+\pi)\rangle\langle\psi(\theta+\pi)|\}$. As mandated by Prop. 7, $h_\mu^Y \geq s$ for all ϕ and θ.

Figure 8 demonstrates how another of Section III's quantum information properties, the quantum excess entropy $\mathbf{E}_q$, relates to the excess entropy $\mathbf{E}^Y$ of the classical measured process. For all ϕ and θ the bound from Prop. 14

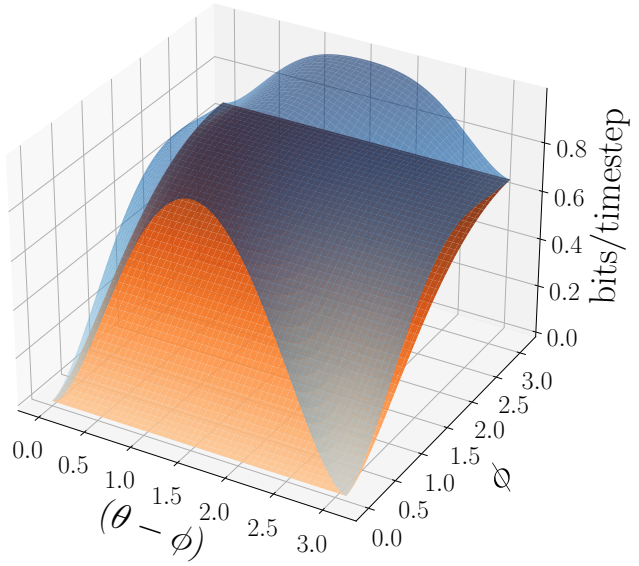


FIG. 7. von Neumann entropy rate s (lower, orange surface) and measured entropy rate h_μ^Y (higher, blue surface) for the $|0\rangle$-$|\phi\rangle$ Quantum Golden Mean process measured with the repeated PVM M_θ. s increases as ϕ does and the alphabet becomes more distinguishable. For $\phi = \pi$ and $\theta = 0, \pi$ we recover the classical Golden Mean. For $(\theta - \phi) = \frac{\pi}{2}$, M_θ applied to $|\phi\rangle$ is a maximum entropy PVM (distribution of measurement outcomes is 50-50). Maxima of h_μ^Y lie in this region.

($\mathbf{E}_q \geq \mathbf{E}^Y$) holds, and $\mathbf{E}^Y$ has maxima at $\phi = \pi$, and $\theta = 0, \pi$, when $\mathcal{Q} = \{|0\rangle, |1\rangle\}$ and $M = M_{01}$. These quantities were estimated using $\ell = 10$.

Underlying the $|0\rangle$-$|\psi(\phi)\rangle$ Quantum Golden Mean is the classical Golden Mean process, which is Markov. Thus, it obeys the quantum Markov property of Eq. (9) despite the fact that it has infinite quantum Markov order for most values of ϕ. This has implications for an observer's ability to synchronize to a Quantum Golden Mean source, which Section V explores.

E. 3-Symbol Quantum Golden Mean

Figure 9 shows the generator of the 3-Symbol Quantum Golden Mean Process with alphabet $\mathcal{Q} = \{|0\rangle, |1\rangle, |+\rangle\}$. Though its generator shares the same internal states and transition probabilities as the $|0\rangle$-$|+\rangle$ Quantum Golden Mean, the 3-Symbol Quantum Golden Mean does not have a one-to-one correspondence between the quantum alphabet $\mathcal{Q}$ and the generator states $\mathcal{S}$ ($|0\rangle \to A$ and $|+\rangle \to B$). Instead $|\mathcal{Q}| = 3$, and these three states cannot all be orthogonal with $d = 2$.

However, unlike the $|0\rangle$-$|+\rangle$ Quantum Golden Mean, we can calculate the quantum entropy rate directly from the

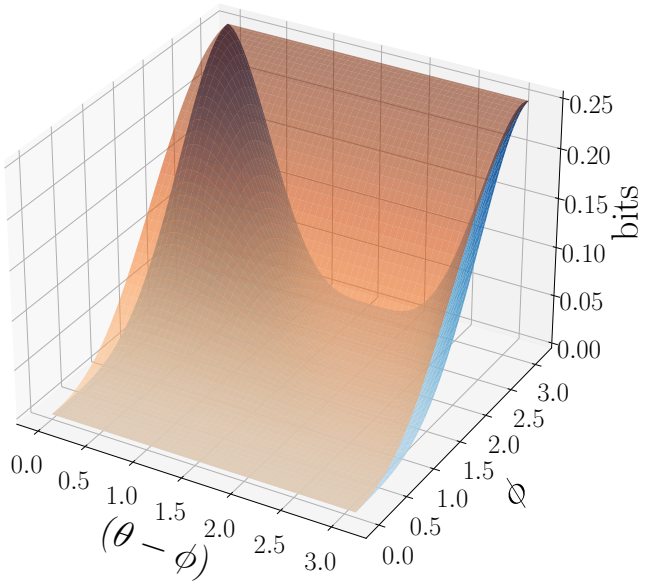


FIG. 8. Quantum excess entropy $\mathbf{E}_q$ (higher, orange surface) and measured excess entropy $\mathbf{E}^Y$ (lower, blue surface) for the $|0\rangle$-$|\phi\rangle$ Quantum Golden Mean process measured with repeated PVM M_θ. $\mathbf{E}_q$ increases with ϕ since $|0\rangle$ and $|\phi\rangle$ become more distinguishable. $\mathbf{E}_Y$ is maximized for $(\theta - \phi) = 0, \pi$, since M_θ can best determine if $|\phi\rangle$ was emitted.

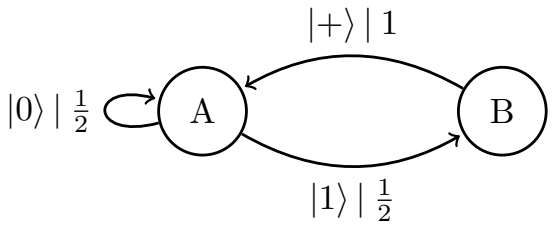


FIG. 9. 3-Symbol Quantum Golden Mean Process Generator.

generator because (1) it has the property of *quantum unifilarity* and (2) it is possible to synchronize to it. Both are discussed at length in Section V. For now, a HMCQS is *quantum unifilar* if and only if, for every $\sigma \in \mathcal{S}$ there exists some POVM M_σ such that an observer knowing σ and the outcome of M_σ can uniquely determine the internal state to which the HMCQS transitioned. The generator in Fig. 9 meets this criterion with $M_A = M_{01}$. M_B can be any POVM.

For a classical, unifilar HMC, h_μ can be calculated as:

$$h_\mu = \sum_{\sigma_i, \sigma_j \in \mathcal{S}} \pi_i \sum_{x \in \mathcal{X}} T_{ij}^x \log_2 T_{ij}^x \ , \tag{48}$$

where π is the stationary state distribution. This result dates back to the foundations of information theory [7].

Similarly, for a quantum unifilar HMCQS to which one can synchronize, we can write:

$$s = \sum_{\sigma_i, \sigma_j \in \mathcal{S}} \pi_i \sum_{x \in \mathcal{X}} T_{ij}^x \log_2 T_{ij}^x \ . \tag{49}$$

Let's walk through this logic for the 3-Symbol Quantum Golden Mean. In state A ($\pi_A = \frac{2}{3}$), the generator emits a qubit either in state $|0\rangle$ or $|1\rangle$, each with probability $\frac{1}{2}$. The density matrix describing this qubit is a maximally-mixed state, and any measurement performed on it involves 1 bit of irreducible randomness. If it is in state B ($\pi_B = \frac{1}{3}$), it emits a qubit in state $|+\rangle$ deterministically. Thus, averaging over the states, the von Neumann entropy rate is $s = \frac{2}{3}$ bits/timestep.

Despite this, when restricted to measuring with a repeated rank-one POVM, the entropy rate of the observed classical process cannot reach the lower bound of $\frac{2}{3}$ bits per symbol because the minimum entropy basis when in state A is M_{01} while the minimum entropy basis when in state B is $M_\pm$. However, an experimenter using the adaptive measurement protocol in Fig. 10 would observe symbol sequences with an entropy rate of $\frac{2}{3}$ bits per symbol once they have synchronized.

They start by measuring in the $M_{01} = \{|0\rangle, |1\rangle\}$ basis and use the outcome of the initial measurement to select a new basis. If the outcome is '0', they are able to synchronize to the process generator which is necessarily in state A. They continue in state A measuring with POVM M_{01} until they observe a '1' at which point they transition to B and measure with $M_\pm$, a zero-entropy measurement. They observe a '+' and return to state A.

In this way, when using an adaptive measurement protocol the recurrent part $\overleftrightarrow{Y}_r$ of the measurement sequence may have a lower entropy rate than h_μ^Y for any repeated rank-one POVM measurement, even if the associated DQMP uses only rank-one PVMs, as in this example. These ideas are formalized and expanded upon in the next section.

F. Unifilar and Nonunifilar Qubit Sources

The last two classes of qubit processes to discuss are generated by the unifilar and nonunifilar qubit sources shown in Figs. 11 and 12, respectively. Both of these generators consist of two internal states ($\mathcal{S} = \{A, B\}$) and emit four possible pure-qubit states ($\mathcal{Q} = \{|0\rangle, |1\rangle, |+\rangle, |-\rangle\}$) that form two orthogonal pairs. However, for each internal state of the unifilar qubit source, the two possible emitted states are orthogonal and one can unambiguously determine the next internal state. In other words, it has the

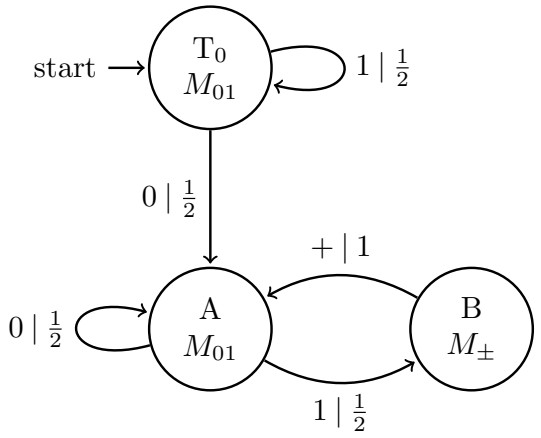


FIG. 10. Adaptive Measurement Protocol (in the form of a DQMP) for the 3-Symbol Quantum Golden Mean process. To synchronize, an observer starts in T_0 (a transient state) and measures with M_{01}. The probability of observing exactly n '1's is $\frac{1}{2^n}$. Upon observing a '0', the observer synchronizes. States A and B correspond exactly to internal states A and B of the source in Fig. 9. The '$-$' transition is not displayed because it has probability 0. The source is quantum unifilar, thus one stays synchronized for future times.

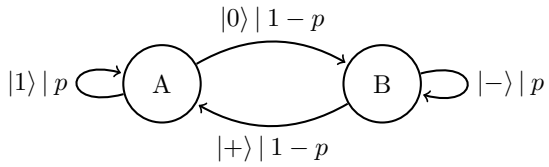


FIG. 11. Unifilar Qubit Source. Each internal state emits one of two orthogonal states and then transitions—e.g., A emits either $|0\rangle$ or $|1\rangle$ that can be distinguished by measurement M_{01}—giving this source the property of quantum unifilarity. p is a parameter that takes values from 0 to 1. Other processes correspond to particular p-values: for example, a nonorthogonal period-2 process ($p = 0$), the maximally-mixed i.i.d. process ($p = \frac{1}{2}$) and a deterministic sequence of either $|1\rangle$ or $|-\rangle$ ($p = 1$).

property of quantum unifilarity. The same is not true of nonunifilar qubit source. The next section illustrates how this difference strongly affects an observer's ability to synchronize.

By varying the parameter p, one can interpolate between one of the several simpler processes already analyzed. Starting with the unifilar qubit source in Fig. 11, for $p = 0$, the generator becomes a period-2 source that emits the word $|\psi_{0+}\rangle$. For $p = \frac{1}{2}$ we obtain a two-state model that generates the i.i.d. maximally-mixed process. And,

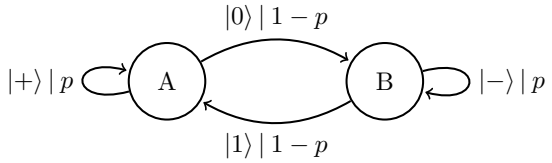


FIG. 12. Nonunifilar Qubit Source. Each internal state emits one of two nonorthogonal states and then transitions. An observer will not be able to determine which state the source transitioned to with any POVM. p takes values from 0 to 1. Other processes correspond to particular p-values: for example an orthogonal period-2 process ($p = 0$), the maximally-mixed i.i.d. process ($p = \frac{1}{2}$) and a deterministic sequence of either $|1\rangle$ or $|-\rangle$ ($p = 1$).

as $p \to 1$ the generator emits longer strings of either only $|1\rangle$ or only $|-\rangle$ qubits, depending on whether the source is in A or B. At $p = 1$ the process becomes nonergodic (here demonstrated by the disconnection between internal states), and the source will only emit either $|1\rangle$ or $|-\rangle$ deterministically.

Similarly, the nonunifilar qubit source in Fig. 12 simplifies for certain p's. For $p = 0$ we obtain a period-2 source, this time with sequence $|\psi_{01}\rangle$, and for $p = \frac{1}{2}$ it also generates the i.i.d. maximally-mixed process. As $p \to 1$ it emits long sequences of either all $|+\rangle$ or all $|-\rangle$ and becomes nonergodic for $p = 1$.

G. Unifilar Qutrit Source

Expanding beyond processes over qubits, we consider a single example of a qutrit ($d = 3$) process whose generator is shown in Fig. 13 and that employs a five-qubit alphabet $\mathcal{Q} = \{|0\rangle, |1\rangle, |2\rangle, |+\rangle, |-\rangle\}$. Using a higher-dimensional Hilbert space makes more measurements available to an observer for synchronization. As a consequence this example process exhibits behavior that is impossible with qubit processes alone: a subspace of Hilbert space (occupied by $|2\rangle$) is always distinguishable from all other states in $\mathcal{Q}$ and can be reserved for synchronization. The other subspace (including $|0\rangle$, $|1\rangle$, $|+\rangle$ and $|-\rangle$) consists of states that cannot be reliably distinguished. Note that this process is quantum unifilar, but one cannot remain synchronized by measuring in a single basis. The next section discusses multiple adaptive measurement protocols that can do so.

	s	$\lvert\mathbf{G}_q\rvert$	$\mathbf{E}_q$	$\mathbf{T}_q$	R_q
	(*bits/timestep*)		(*bits*)	(*bits*×*time steps*)	(*time steps*)
I.I.D. Qubit Process	$S(\rho_{iid})$	$1 - S(\rho_{iid})$	0	0	0
Period-3 Process ($\phi = \pi$)	0	1	$\log_2(3)$	2.33	3
Period-3 Process ($\phi = \pi/2$)	0	1	$\log_2(3)$	4.22	∞
Quantum Golden Mean ($\phi = \pi$)	2/3	1/3	0.2516	1/3	1
Quantum Golden Mean ($\phi = \pi/2$)	0.4495	0.5505	0.1092	0.5687	∞
3-Symbol Quantum Golden Mean	0.6667	0.3333	0.4652	0.8855	∞
Unifilar Qubit Source ($p = 1/3$)	0.9184	0.0816	0.0808	0.1976	∞
Nonunifilar Qubit Source ($p = 1/3$)	0.7306	0.2614	0.3217	0.4090	∞
Unifilar Qutrit Source	0.8002	0.7848	1.290	2.156	∞

TABLE I. Information Properties for Example Quantum Processes/Sources. Decimal values were numerically estimated using Eqs. (38), (45), and (47) with $\ell = 8$ for the Unifilar Qutrit Source, $\ell = 10$ for the Unifilar and Nonunifilar Qubit Sources, and $\ell = 12$ for all other processes. Other values were calculated analytically.

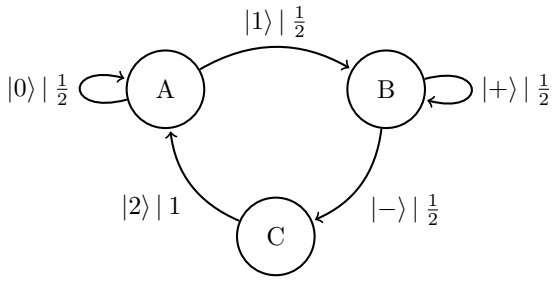


FIG. 13. Unifilar Qutrit Source. When in internal states A and B it emits a qutrit in the subspace of Hilbert space spanned by $|0\rangle$ and $|1\rangle$. When in C it emits $|2\rangle$, which can always be distinguished from all other states in $\mathcal{Q}$. This demonstrates additional opportunities for synchronization in higher-dimensional Hilbert spaces.

H. Discussion

Table I summarizes information properties for the above examples with either analytic results or numerical estimates. Together these examples illustrate a range of different features of separable qudit processes. They demonstrate how the information properties defined and characterized in Section III are both indicative of underlying structural features of quantum information sources and strongly influenced by the distinguishability of states in a process' quantum alphabet. Our analysis of the convergence of the quantum block entropy to its linear asymptote thus gives meaningful and interpretable ways of quantifying the randomness and correlation in a separable quantum process.

We continue by discussing two tasks that an observer might wish to perform when faced with a quantum information source emitting a separable quantum process. First, if they have prior knowledge of the internal structure of the source, they may want to determine the internal state it occupies during a given timestep. This task is *synchronization*, and we discuss it in the next section. Second, if they have no knowledge of the source, they may want to measure the process it produces to infer its internal structure. This task is *system identification*, and we demonstrate how an observer can use a tomographic protocol to perform it in Section VI.

V. SYNCHRONIZING TO A QUANTUM SOURCE

How does an observer of a process with knowledge of its quantum generator determine its internal state? When an observer is certain about the internal state the observer is *synchronized* to the quantum source. The following explores synchronizing to quantum processes—both the manner in which observations lead to inferring the source's state and quantitative measures of partial and full synchronization.

Quantum measurement adds a subtlety to this task in comparison to the task of synchronizing to a classical process given knowledge of its minimal unifilar model—the ϵ-machine—as described in Refs. [12, 47].

A. States of Knowledge

Recall that a hidden Markov chain quantum source (HMCQS) consists of a set of internal states ($\mathcal{S}$), a pure-state alphabet ($\mathcal{Q}$), and a set of labeled transition matrices ($\mathcal{T}$). We assume an observer has complete knowledge of the

HMCQS that generates a process but can only infer the internal state at time $t = \ell$ by applying block measurement $\mathcal{M}_{0:\ell}$ and observing outcome $y_{0:\ell}$. They have no access to the qudits that were emitted before $t = 0$.

An observer's best guess for the internal state of a source given different sequences of observations can be represented as distributions over the source's internal states, known as *mixed states* (not to be confused with mixed quantum states, represented by density matrices). Classically, after observing a particular length-ℓ word $w = x_{0:\ell}$, the observer is in the mixed state $\eta(w) = (p_A, p_B, \ldots)$. After the next observation X_ℓ, they will transition to one of a set of new mixed states depending on the outcome $\eta(w, x_\ell = 0)$, $\eta(w, x_\ell = 1)$, and so on. The word corresponding to the new mixed state is a concatenation of w and the new observation x_ℓ. The set of mixed states for a classical process and the dynamic between them define a process' *mixed state presentation* (MSP), which is unifilar by construction [48].

Using a classical process' MSP rather than a nonunifilar generator of the process has many computational advantages. Two of interest are: it allows one to calculate the entropy rate for processes without finite unifilar presentations [49] and it allows one to calculate the uncertainty an observer experiences while attempting to synchronize to a process' generator [50].

For quantum processes there is no unique MSP but rather a multiplicity of possible MSPs, each corresponding to a different choice of measurement protocol. For a given source and given measurement protocol $\mathcal{M}$ we can define a set of mixed states with each corresponding to the possible measurement sequence one can observe.

Consider the mixed states corresponding to length-ℓ sequences of observations. We restrict $\mathcal{M}_{0:\ell}$ to consist of local POVMs, allowing for adaptive measurement. Given that an observer has applied measurement $\mathcal{M}_{0:\ell}$ and seen measurement outcomes $y_{0:\ell}$, their best guess about the generator's internal state is represented by the conditional distribution $\eta(y_{0:\ell}|\mathcal{M}_{0:\ell}) = \{\Pr(\sigma|y_{0:\ell}, \mathcal{M}_{0:\ell}) \text{ for all } \sigma\}$.

For $t = 0$, an observer has no measurement outcomes with which to inform their prediction about the source's internal state. However, they do know the stationary state distribution π of the model (since it can be calculated directly from $\mathcal{T}$). This serves as a 'best guess' of the source's internal state absent any measurements. If the initial state distribution $\mathcal{S}_0$ is not π and the observer is aware of this fact, the mixed states for that process are $\eta(y_{0:\ell}|\mathcal{M}, \mathcal{S}_0)$. We omit the conditioning on $\mathcal{S}_0$ if $\mathcal{S}_0 = \pi$.

For most repeated PVM measurements of qubit processes there are an uncountably-infinite number of mixed-states [14]. This *measurement-induced complexity* appears even for $|\mathcal{S}| = 2$. The examples in this section step back from this complexity to focus on measurements that are sufficiently informative to allow an observer to synchronize with finite observation sequence $y_{0:\ell}$ and result in only finite or (at most) a countably-infinite set of recurrent mixed states.

B. Average State Uncertainty and Synchronization Information

To compare synchronization behavior for different measurement protocols it is useful to use entropic quantities rather than working with the set of mixed states and their dynamic directly. In particular, we look at the entropy of the possible mixed states for length-ℓ sequences of measurements.

An observer's uncertainty about the source's internal state after applying measurement sequence $\mathcal{M}_{0:\ell}$ (according to some protocol $\mathcal{M}$) and observing outcome $y_{0:\ell}$ is given by the Shannon entropy of the corresponding mixed-state distribution:

$$\begin{aligned}\mathrm{H}\left[\eta(y_{0:\ell}|\mathcal{M})\right] &= \mathrm{H}\left[\Pr(\sigma|y_{0:\ell}, \mathcal{M})\right] \\ &= -\sum_{\sigma\in\mathcal{S}} \Pr(\sigma|y_{0:\ell}, \mathcal{M}) \log_2 \Pr(\sigma|y_{0:\ell}, \mathcal{M}) \ .\end{aligned}$$

One can average over all measurement outcomes on a block of ℓ qudits to find the *average state uncertainty*:

$$\begin{aligned}\mathcal{H}(\ell|\mathcal{M}) &\equiv \mathrm{H}\left[\eta(Y_{0:\ell}|\mathcal{M})\right] \\ &\equiv -\sum_{y_{0:\ell}} \Pr(y_{0:\ell}|\mathcal{M})\, \mathrm{H}\left[\eta(y_{0:\ell}|\mathcal{M})\right] \ .\end{aligned}$$

For a given measurement protocol $\mathcal{M}$, this quantity generally converges to a finite value in the $\ell \to \infty$ limit. This limit does not necessarily exist if is not ergodic; for example, if there is some underlying periodicity. If such a limit exists, the *asymptotic state uncertainty* is:

$$C_\infty(\mathcal{M}) = \lim_{\ell\to\infty} \mathcal{H}(\ell|\mathcal{M}) \ .$$

Both $\mathcal{H}(\ell|\mathcal{M})$ and $C_\infty(\mathcal{M})$ are measured in bits.

If $\mathrm{H}\left[\eta(y_{0:\ell}|\mathcal{M})\right] = 0$, then $y_{0:\ell}$ is a *synchronizing observation*, and an observer who sees outcome $y_{0:\ell}$ can precisely identify the source's internal state. If $\mathcal{H}(\ell|\mathcal{M}) = 0$, then any measurement outcome seen when applying protocol $\mathcal{M}$ to ℓ qudits allows the observer to synchronize to the source. For classical and quantum processes, the Markov order is the first value of ℓ for which $\mathcal{H}(\ell)$ vanishes (for some $\mathcal{M}_{0:\ell}$ in the quantum case). This generally does not occur for HMCQS with nonorthogonal states in $\mathcal{Q}$ except in the $\ell \to \infty$ limit since R_q is generally infinite.

Being synchronized after measuring ℓ qudits does not guarantee the observer remains synchronized after measuring qudit $\ell + 1$. In classical processes for which this occurs, persistent synchronization requires that the HMC of the process that the observer uses satisfies the additional condition of *unifilarity*. In the quantum setting, synchronization persists if and only if the underlying HMCQS is *quantum unifilar* and, when the observer is in internal state σ, their measurement protocol ensures they apply the measurement for which the internal state at time $t + 1$ is completely determined.

This is equivalent to the statement 'There exists some protocol $\mathcal{M}$ with measurements $\mathcal{M}_{0:\ell}$ and M_ℓ such that:

$$\begin{aligned} &\mathrm{H}\left[\Pr(\sigma|y_{0:\ell}, \mathcal{M}_{0:\ell})\right] = 0 \\ &\quad \implies \mathrm{H}\left[\Pr(\sigma|y_{0:\ell}y_\ell, \mathcal{M}_{0:\ell} \otimes M_\ell)\right] = 0 \ , \end{aligned} \tag{50}$$

for all $y_\ell \in \mathcal{Y}$.'

The measurement that maintains synchronization when the source is in one internal state (σ_i) does not need to be the same as the measurement that maintains synchronization when the source is in another internal state (σ_j). When the measurement required to maintain synchronization depends upon the HMCQS's current state, then *adaptive measurement protocols* are capable of maintaining synchronization even when no fixed-basis measurement can, as the following demonstrates with multiple examples.

Finally, the total amount of state uncertainty that an observer encounters while synchronizing to a process using a given measurement protocol $\mathcal{M}$ is the *synchronization information*, given by:

$$\mathbf{S}(\mathcal{M}) = \sum_{\ell=0}^{\infty} \mathcal{H}(\ell|\mathcal{M}) \ . \tag{51}$$

Note that, if the asymptotic state uncertainty $C_\infty(\mathcal{M})$ is greater than 0, $\mathbf{S}(\mathcal{M})$ is infinite. When $C_\infty(\mathcal{M}) = 0$, we can estimate $\mathbf{S}(\mathcal{M})$ by terminating the sum at some finite ℓ.

For classical processes the synchronization information is closely tied to the transient information $\mathbf{T}$. We will similarly connect $\mathbf{S}(\mathcal{M})$ to the quantum transient information $\mathbf{T}_q$ and use it as a way to compare synchronization via different measurement protocols.

The remainder of this section explores synchronization to various models from Section *IV*. We pair each with a variety of measurements, both repeated POVMs and adaptive protocols, to demonstrate the way $\mathcal{H}(\ell|\mathcal{M})$ is affected by this choice.

C. Synchronizing to Quantum Presentations of Classical Processes

A HMCQS with a qudit alphabet consisting entirely of orthogonal states ($\langle\psi_x|\psi_{x'}\rangle = 0$, for all $|\psi_x\rangle$ and $|\psi_{x'}\rangle \in \mathcal{Q}$), is a quantum presentation of the underlying classical process $\overleftrightarrow{X}$. An observer can perform a measurement using orthogonal projectors $P_x = |\psi_x\rangle\langle\psi_x|$ for all $|\psi_x\rangle \in \mathcal{Q}$ that unambiguously discriminates between the pure qudit states the source emits.

The task of synchronizing to such an HMCQS is equivalent to synchronizing to a source emitting classical symbols. If it is quantum unifilar, then the underlying classical HMC is unifilar and synchronization is exponentially fast (on average) [12, 47]. Absent unifilarity, an observer may not be able to synchronize even asymptotically—i.e., $C_\infty(\mathcal{M}) > 0$ for all $\mathcal{M}$—despite measuring with orthogonal projectors.

D. Synchronizing to Periodic Processes

All periodic quantum processes have vanishing von Neumann entropy rate—i.e., $s = 0$—and each state only has one incoming and one outgoing transition. Due to these simplifications, to synchronize an observer simply determines the source's *phase*; i.e., which of the p internal states the source occupies at time t. Once this phase is determined for any t, it is also determined for all other t. The initial phase uncertainty is equivalent to the initial state uncertainty $\mathrm{H}[\pi] = \log_2(p)$ since π is uniform.

Any measurement protocol $\mathcal{M}$ that distinguishes between states in $\mathcal{Q}$ gives information about the phase and allows an observer to synchronize to the source asymptotically. If two states in $\mathcal{Q}$ cannot be distinguished by $\mathcal{M}$, then synchronization may not be possible, even with an infinite number of measurements.

For classical periodic processes, an observer synchronizes at the Markov order $\ell = p$, and the synchronization information and the transient information are equal: $\mathbf{S} = \mathbf{T}$ [11]. In contrast, quantum periodic processes generically have infinite Markov order, if there are nonorthogonal states in $\mathcal{Q}$, and only asymptotically synchronize. Thus, $\mathbf{S}(\mathcal{M}) \geq \mathbf{T}_q$. The condition for equality is that $\mathcal{M}$ is the optimal measurement over the process. For $R_q = \infty$ this means that $\mathcal{M}$ must be a global measurement over the bi-infinite chain of qudits.

Figure 14 shows the average state uncertainty for an observer measuring three different period-5 qubit processes consisting of two alphabet states: $|0\rangle$ and $|\psi(\phi = 3\pi/4)\rangle$. All other period-5 sequences with this qubit alphabet are

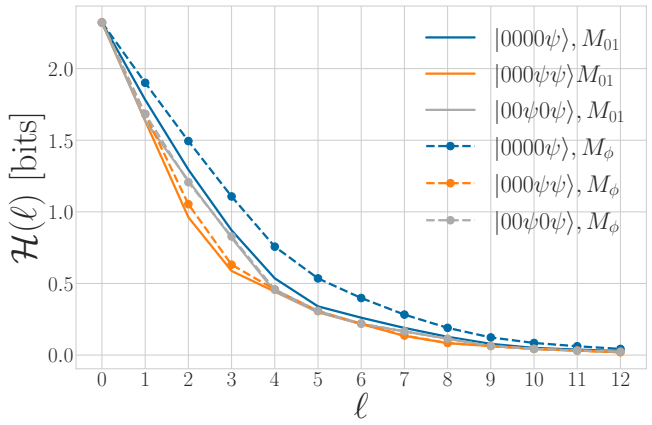


FIG. 14. Average state uncertainty $\mathcal{H}(\ell)$ for period-5 qudit processes: ψ denotes state $|\psi(\phi)\rangle$. The associated PVM is $M_\phi = \{|\psi(\phi)\rangle, |\psi_(\phi+\pi)\rangle\}$. We set $\phi = 3\pi/4$. The area under each curve is the synchronization information $\mathbf{S}$ for that process and measurement.

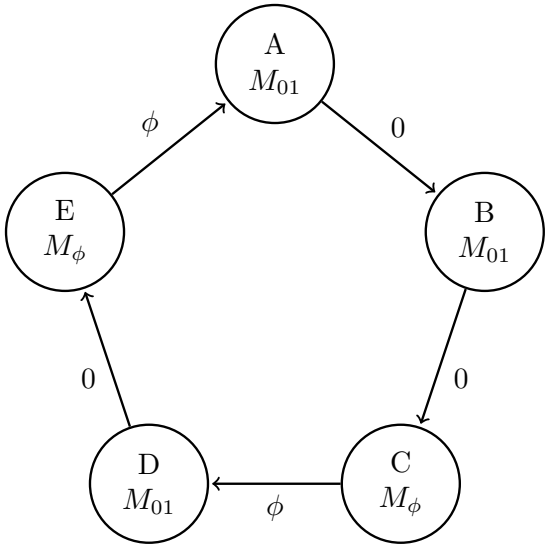


FIG. 15. Adaptive measurement protocol for period-5 sequence $|00\psi0\psi\rangle$. Each state is labeled with the next measurement to perform, either M_{01}, with possible outcomes '0' and '1', or M_ϕ, the PVM with elements $|\psi(\phi)\rangle\langle\psi(\phi)|$ and $|\psi(\phi+\pi)\rangle\langle\psi(\phi+\pi)|$ and possible outcomes ϕ and $(\phi+\pi)$. The five states shown are the measurement protocol's recurrent states, that an observer only encounters when synchronized. An observer who is synchronized and using this protocol sees a measured period-5 process with word '$00\phi0\phi$'.

equivalent to these three under shift and swap symmetries. As expected, $\mathcal{H}(\ell)$ decreases from $\mathcal{H}(0) = \log_2 5$ to $C_\infty = 0$ for all combinations of sequence and measurement basis. That noted, the rate at which synchronization occurs—and the total amount of uncertainty seen by an observer, the synchronization information—depends on both the particular sequence and the particular repeated measurement applied to it.

For period-5 process with word '0000ψ', $\mathbf{T}_q \approx 6.32$ bits×time steps, using Eq. (47) with $\ell = 12$. When measuring it in basis M_{01} (M_ϕ) we find that $\mathbf{S}(M_{01}) \approx 7.92$ bits ($\mathbf{S}(M_\phi) \approx 9.30$ bits). These are also estimated by calculating up to $\ell = 12$. Similarly, the bound between $\mathbf{S}(\mathcal{M})$ and $\mathbf{T}_q$ holds for the other period-5 words. For word '$000\psi\psi$', $\mathbf{T}_q \approx 4.86$ bits×time steps, $\mathbf{S}(M_{01}) \approx 6.84$ bits, and $\mathbf{S}(M_\phi) \approx 7.05$ bits. For word '$00\psi0\psi$', $\mathbf{T}_q \approx 5.51$ bits×time steps, $\mathbf{S}(M_{01}) \approx 7.39$ bits, and $\mathbf{S}(M_\phi) \approx 7.47$ bits.

At this point we wish to emphasize that s, $\mathbf{E}_q$, and R_q are equal for generic period-5 processes with nonorthogonal alphabets $\mathcal{Q}$. Despite this, $\mathbf{T}_q$ and $\mathbf{S}(\mathcal{M})$ identify physically-relevant differences between these processes for the task of synchronization. These differences are intrinsic to the quantum process itself ($\mathbf{T}_q$) and also appear within the measurement outcomes an observer obtains ($\mathbf{S}(\mathcal{M})$).

As a final comment on periodic processes, we note that while $s = 0$ for all periodic quantum process, many measurement protocols (even those for which $C_\infty = 0$) give a measured classical process with a nonzero entropy rate h_μ^Y. In fact, no fixed-basis measurement of a periodic process results in an observed process with $h_\mu^Y = 0$ unless (i) all states in $\mathcal{Q}$ are orthogonal and (ii) the measurement is in an orthogonal basis which includes one projector for each state in $\mathcal{Q}$.

In contrast, an observer using an adaptive measurement protocol can easily have zero uncertainty in measurement outcomes once synchronized. For example, Fig. 15 shows the recurrent states for a DQMP that swaps between two different measurements—M_{01} and M_ϕ—depending on the internal state of the source. The sequence of measurement outcomes observed is deterministic, and the recurrent measured process is a period-5 process with word '$00\phi0\phi$' and $h_\mu^{Y_r} = 0$.

E. Synchronizing with PVMs

The $|0\rangle$-$|+\rangle$ Quantum Golden Mean Process has multiple synchronizing observations; see the generator in Fig. 6. Measuring with M_{01} and seeing a '1' synchronizes the observer to internal state B; measuring with $M_\pm$ and seeing a '$-$' synchronizes the observer to internal state A. Let's consider the mixed states produced by these two repeated PVMs in turn, starting with M_{01}.

Figure 16 shows the mixed states for the measured process obtained by repeatedly applying M_{01}. Observing synchronizing measurement $y =$ '1' means that the source

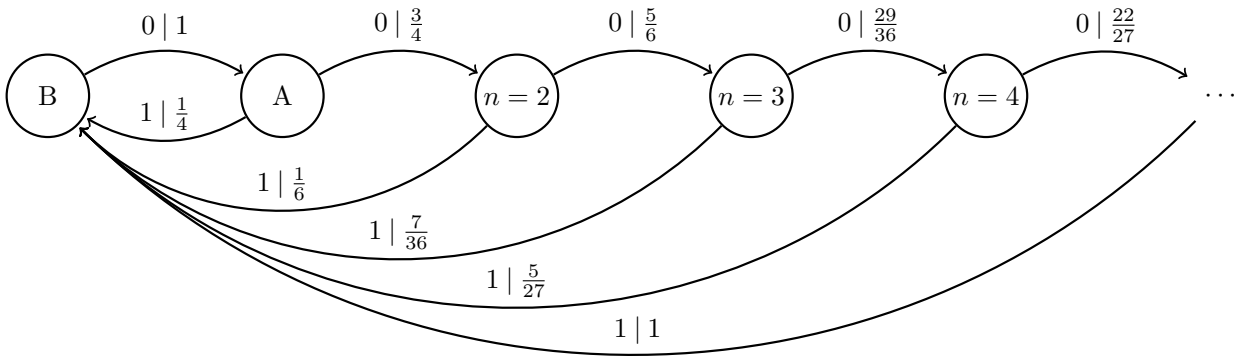


FIG. 16. Mixed state presentation for the $|0\rangle$-$|+\rangle$ Quantum Golden Mean Process measured with M_{01}. n refers to the number of consecutive '0's since the most recent '1'.

just transitioned to state B while emitting a $|+\rangle$ qubit; i.e., $\mathrm{H}\left[\eta(\text{`1'})\right] = 0$. Additionally, the source in state $B$ can only transition to state $A$ while emitting a $|0\rangle$ qubit, so the observer will see a '0', and $\mathrm{H}\left[\eta(\text{`10'})\right] = 0$. However, once the source is in state A an observer easily desynchronizes from the source if they observe another '0', as this is consistent with either of the two transitions to the source. As an observer sees more '0's they transition to mixed states further to the right of Fig. 16. To summarize, measuring with M_{01} makes use of synchronizing observations '1' and '10', but its MSP has a countably-infinite number of recurrent states corresponding to sequences of n '0's.

This measured process is an infinite-state classical renewal process, whose information properties can be estimated using methods from Ref. [51]. After seeing n '0's in a row, the next observation will be '1' with probability:

$$\begin{aligned}\Pr(\text{`1'}|\text{`0'}^n) &= \tfrac{1}{4} * \Pr(A|\text{`0'}^n) \\ &= \tfrac{3}{16}\left(1 - \left(-\tfrac{1}{3}\right)^n\right) \ .\end{aligned}$$

We find $h_\mu^Y \approx 0.60$ bits per symbol, $\mathbf{E}^Y \approx 0.053$ bits, and a single-symbol entropy of $\mathrm{H}\left[1\right] \approx 0.65$ bits. $\mathbf{E}^Y$'s small value and the fact that h_μ^Y is not significantly lower than $\mathrm{H}\left[\ell = 1\right]$ indicate that the infinite-state renewal process presentation provides only a small predictive advantage over using a biased coin with $\Pr(\text{`0'}) = 5/6$. This is further evidenced by how $\Pr(B|\text{`0'}^n) = \frac{1}{4}\left(1 - (-1/3)^{n-1}\right)$ converges exponentially quickly to the its asymptotic value of $\Pr(B|\text{`0'}^n) = 1/4$.

Measuring with $M_\pm$, symbol '$-$' is a synchronizing observation and indicates the source is in state A. The recurrent mixed states for this observed process are shown in Fig. 17, and they also form a classical renewal process. Observing n '+'s since the last '$-$', the probability that the generator is in state A is $\Pr(A|\text{`+'}^n) = \frac{3}{5}\left(1 - (-2/3)^{n+1}\right)$. The measured process obtained with $M_\pm$ has $h_\mu^Y \approx 0.90$ bits per symbol, $\mathbf{E}^Y \approx 0.020$ bits, and $\mathrm{H}\left[\ell = 1\right] \approx 0.91$ bits. Again, the infinite-state MSP provides only a small predictive advantage over a biased coin with $\Pr(\text{`+'}) = 2/3$. Also, $\Pr(A|\text{`+'}^n)$ converges exponentially quickly to $\Pr(A|\text{`+'}^n) = 3/5$.

Applying either of these two fixed-basis measurements gives an average state uncertainty that decreases monotonically with ℓ, as shown in Fig. 18. Since this source is not quantum unifilar, an observer repeatedly synchronizes and desynchronizes while measuring this process regardless of basis, and $\mathcal{H}(\ell)$ does not approach 0 as $\ell \to \infty$. We find that $C_\infty(M_{01}) \approx 0.62$ bits and $C_\infty(M_\pm) \approx 0.54$ bits. Measuring with $M_\pm$ not only results in less state uncertainty asymptotically but also results in less average uncertainty for all values of ℓ.

Figure 18 also displays the average state uncertainty for two other relevant PVMs: M_θ for $\theta = \pi/4$ and $\theta = 3\pi/4$. Recall that M_θ is the PVM consisting of projectors onto orthogonal states $|\psi(\theta)\rangle$ and $|\psi(\theta + \pi)\rangle$. For comparison, M_{01} corresponds to $\theta = 0$ and $M_\pm$ corresponds to $\theta = \pi/2$.

For $\theta = \pi/4$, the symbol states $|0\rangle$ and $|+\rangle$ give the observer the exact same distribution of measurement outcomes, and the observer cannot gain any information about the state of the source beyond the stationary state distribution. This can also been seen as the maximum of the asymptotic state uncertainty in Fig. 19.

For $\theta = 3\pi/4$—and for the majority of values of θ—M_θ has no synchronizing observations. Despite this, Fig. 18

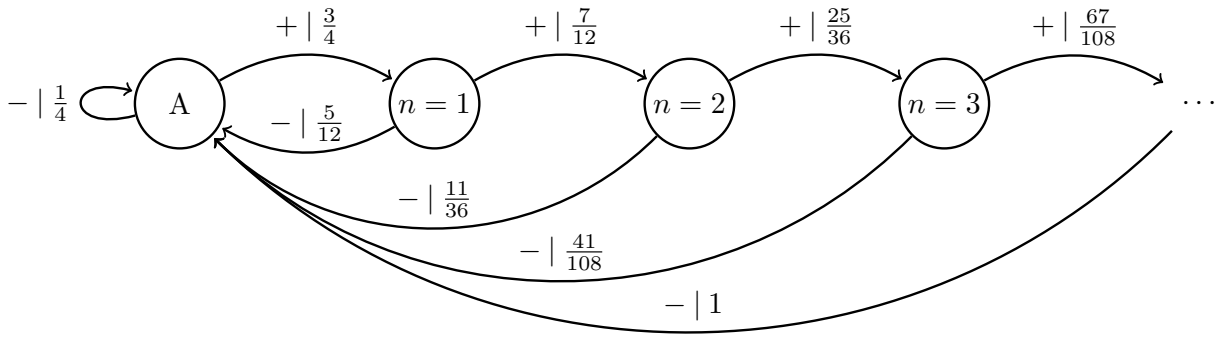


FIG. 17. Mixed state presentation for the $|0\rangle$-$|+\rangle$ Quantum Golden Mean Process measured with $M_{\pm}$. n refers to the number of consecutive '+'s since the most recent '−'.

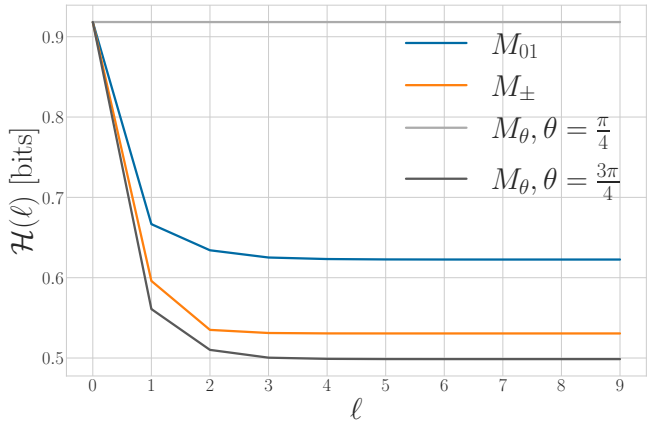


FIG. 18. Average state uncertainty $\mathcal{H}(\ell)$ for the $|0\rangle$-$|+\rangle$ Quantum Golden Mean generator after ℓ measurements.

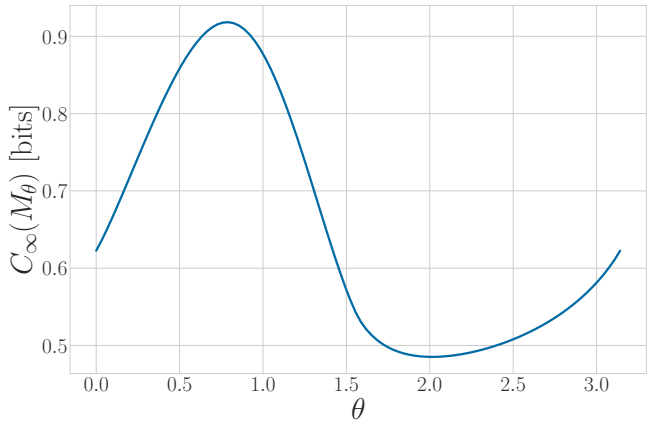


FIG. 19. Asymptotic state uncertainty when applying the PVM M_θ to the $|0\rangle$-$|+\rangle$ Quantum Golden Mean.

demonstrates that $\mathcal{H}(\ell|M_{\theta=3\pi/4})$ is lower for all ℓ than both bases that can exactly synchronize.

A measured process for generic θ does not have the renewal process structure of Figs. 16 and 17. Instead, in the absence of synchronizing observations, the number of mixed states generically grows exponentially with ℓ. If we approximating the MSP using length-ℓ observations, then there will be $|\mathcal{Y}^\ell|$ mixed states—one for each possible sequence of measurements—each corresponding to a different distribution over the source's internal states. These infinite-state MSPs can be characterized by their statistical complexity dimension [52].

Despite the explosive complexity, many of these PVMs give lower average state uncertainties than M_{01} and $M_{\pm}$. $\theta = 3\pi/4$ is a representative example. The asymptotic state uncertainty when applying M_θ is shown in Fig. 19. Note that the maximum value of $C_\infty(\mathcal{M}_\theta) = \mathrm{H}\,[\pi]$ occurs when $\theta = \pi/4$, as discussed. The minimum asymptotic state uncertainty for this set of PVMs is $C_\infty \approx 0.49$ bits, which occurs for $\theta \approx 2.01$. An observer may choose to use a basis with an exponential set of mixed states to lower their uncertainty in the source's internal state, at the cost of having to track the probability of a larger set of mixed states.

F. Maintaining Synchrony with Adaptive Measurement

Adaptive measurement protocols are also capable of maintaining synchronization when no fixed-basis measurement can. To appreciate this, consider Fig. 11's unifilar qubit source.

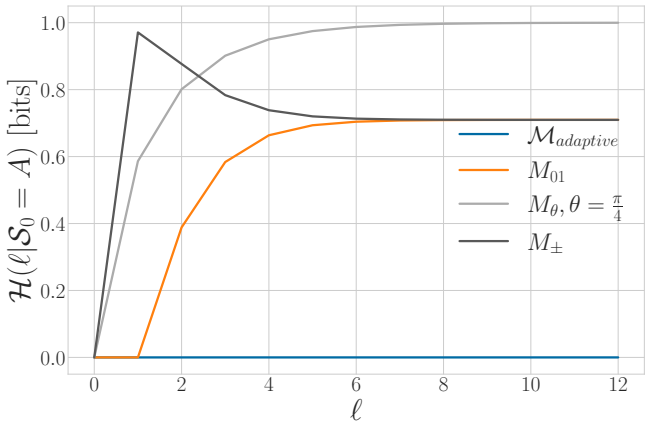


FIG. 20. Average state uncertainty $\mathcal{H}(\ell)$ for the unifilar qubit source ($p = 0.6$) initialized in state A. Only the adaptive measurement protocol (described in the text) is able to maintain synchronization.

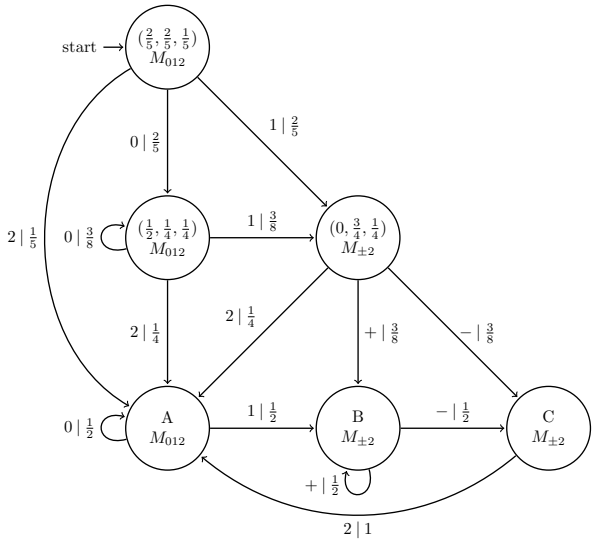


FIG. 21. Adaptive measurement protocol defined for the qutrit process generator in Fig. 13. The three transient mixed states are labeled with the internal source states probabilities (p_A, p_B, p_C) and the three recurrent states correspond exactly to those states. This adaptive protocol permanently synchronizes to the source since the source is quantum unifilar. Transitions with zero probability are omitted.

For $0 < p < 1$ there are no synchronizing observations, however almost any measurement reduces the average state uncertainty below H $[\pi]$. Additionally, if an observer comes to know the source's state by other means (perhaps the source is initialized in state A at time $t = 0$), then they are able to maintain synchronization with a simple adaptive measurement protocol $\mathcal{M}_{adaptive}$.

This protocol, defined here only for recurrent states, is as follows. If the source is in state A, apply measurement M_{01}, and if the source is in state B, apply measurement $M_\pm$. It is only possible to define it this simply and maintain synchronization since the source is quantum unifilar.

Figure 20 shows the behavior of $\mathcal{H}(\ell)$ as the observer desynchronizes from the source initialized in state A. For each repeated PVM measurement, it eventually reaches an asymptotic value, though $\mathcal{H}(\ell)$ may be nonmonotonic.

G. Synchronizing to a Qutrit Source

A qubit source ($d = 2$) presents limited opportunities for unambiguous state discrimination and, hence, for synchronization. Qutrits allow for more general behavior and have been suggested for applications in quantum communication networks where one state is reserved specifically for synchronization [53].

Let's apply this logic to the unifilar qutrit source in Fig. 13. Synchronizing observation sequences for this process are '2', '1+', and '1−' that definitively place the source in states A, B, and C, respectively. Once synchronized an observer may remain synchronized by measuring with $M_{012} = \{|0\rangle\langle 0|, |1\rangle\langle 1|, |2\rangle\langle 2|\}$ when in state A, $M_{\pm 2} = \{|+\rangle\langle +|, |-\rangle\langle -|, |2\rangle\langle 2|\}$ when in state B, and either of the above when in state C.

We explore synchronizing to this source with five different measurement protocols and compare them in Fig. 22. The first two are repeated PVMs in the M_{012} basis and the $M_{\pm 2}$ basis. The other three are adaptive measurement protocols that share a recurrent dynamic, but have different transient states. Consider measuring in the M_{012} ($M_{\pm 2}$) basis until observing a 2 and therefore synchronizing to source state A. Then, use M_{012} when the source is in state A and $M_{\pm 2}$ when the source is in state B or C. We refer to this protocol as $\mathcal{M}_{012,sync}$ ($\mathcal{M}_{\pm 2,sync}$). The fifth protocol $\mathcal{M}_{adaptive}$ is defined by the DQMP in Fig. 21 that uses an adaptive protocol over three transient mixed states in addition to the protocol just defined for the recurrent states.

The average state uncertainties for an observer implementing these five measurement protocols are shown in Fig. 22. The fixed basis measurements do not lead to persistent synchronization and have a nonzero asymptotic state uncertainty ($C_\infty(M_{012}) \approx 0.40$ bits and $C_\infty(M_{\pm 2}) \approx 0.72$ bits. If one measures in a fixed basis until synchronizing (by observing a 2) and then takes advantage of the quantum unifilarity of the source to stay synchronized, then the asymptotic state uncertainty vanishes. This is the case for $\mathcal{M}_{012,sync}$ and $\mathcal{M}_{\pm 2,sync}$, which have synchronization informations of $\mathbf{S}(\mathcal{M}_{012,sync}) \approx 3.91$ bits

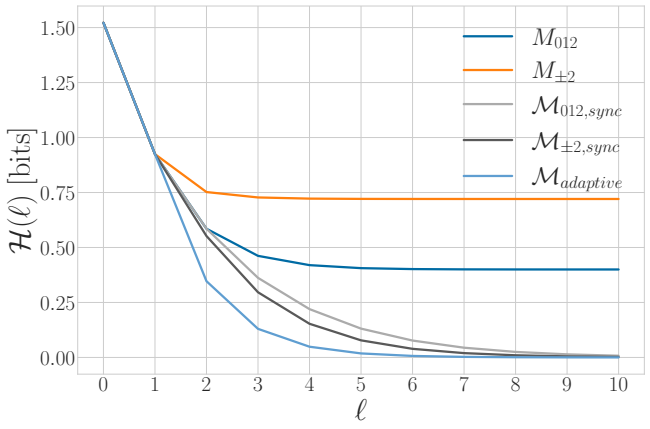


FIG. 22. Average state uncertainty $\mathcal{H}(\ell)$ while measuring the process generated by the unifilar qutrit source. M_{012} and $M_{\pm 2}$ are fixed-basis measurements. $\mathcal{M}_{012,sync}$ and $\mathcal{M}_{\pm 2,sync}$ measure in a fixed basis until they observe a 2 and stay permanently synchronized afterwards. $\mathcal{M}_{adaptive}$ refers to the protocol in Fig. 21.

and $\mathbf{S}(\mathcal{M}_{012,sync}) \approx 3.60$ bits. The extra complexity of the measurement protocol $\mathcal{M}_{adaptive}$ admits additional synchronizing words ('1+' and '1−') and a lower synchronization information ($\mathbf{S}(\mathcal{M}_{adaptive}) \approx 3.00$ bits) than the simpler strategy of waiting to see a '2'. Note that these three synchronization informations are all greater than our estimate of the quantum transient information for this process ($\mathbf{T}_q \approx 2.16$ bits $\times$ symbols) These values of $\mathbf{S}(\mathcal{M})$ were estimated using $\ell = 10$.

H. Discussion

This section has detailed the process of synchronizing to a known qudit source. In contrast to sources of classical random variables, an observer may attempt to synchronize to a qudit source using a variety of measurement protocols. We can compare different protocols with two informational quantities introduced above: the asymptotic state uncertainty $C_\infty(\mathcal{M})$ and the synchronization information $\mathbf{S}(\mathcal{M})$.

For sources that are not quantum unifilar, no protocol can remain synchronized to the source. Nevertheless, for different measurement protocols $\mathcal{M}_0$ and $\mathcal{M}_1$ we can compare $C_\infty(\mathcal{M}_0)$ and $C_\infty(\mathcal{M}_1)$. The protocol with a lower value is better at the task of synchronization in the sense that an observer's uncertainty in the source's internal state will be lower on average.

This leads to a natural question: What is the best measurement protocol for synchronizing to a source? To answer it we introduce a new protocol-independent property of a qudit process, the *minimal asymptotic state uncertainty*:

$$C_{min} = \min_{\mathcal{M}} C_\infty(\mathcal{M}) \ ,$$

where the minimum is taken over all possible measurement protocols defined via DQMP. In practice, determining C_{min} for a quantum process requires a proof that no measurement protocol can achieve a lower value. Fully exploring the space of measurement protocols is beyond the present scope. Nevertheless, we introduced candidates for C_{min} in the above examples and found the minimal asymptotic state uncertainty for a restricted class of repeated PVMs numerically; recall Fig. 19.

For sources that are quantum unifilar, we explored several protocols that are capable of persistent synchronization. For such processes, $C_{min} = 0$ bits. We can compare different synchronizing measurement protocols $\mathcal{M}_0$ and $\mathcal{M}_1$ through their synchronization informations $\mathbf{S}(\mathcal{M}_0)$ and $\mathbf{S}(\mathcal{M}_1)$. A lower value means that the observer experiences less state uncertainty while synchronizing. What is the minimal amount of state uncertainty an observer can experience? We define the *minimal synchronization information* for a quantum process as:

$$\mathbf{S}_{min} = \min_{\mathcal{M}} \mathbf{S}(\mathcal{M}) \ .$$

This minimum is also taken over all DQMPs. Establishing that a given protocol is minimal is nontrivial.

One open question prompted by this work is 'Is it always possible to synchronize to a source that is quantum unifilar?' or equivalently 'Does $C_{min} = 0$ for all processes generated by quantum unifilar sources?'. Answering this question will also require a greater understanding of the space of DQMPs. Progress may involve proving the existence or nonexistence of a protocol that is able to synchronize to the unifilar qubit source in Fig. 11.

Finally, we recount several reasons why synchronization is an important task not only for determining a source's internal state, but also for improving predictions of future measurement outcomes. Generally, HMCQSs have inherent stochasticity. Periodic sources are an exception. Through synchronization we may substantially reduce the uncertainty in measurement outcomes. In the extreme case where a source's internal state has only one possible transition, this uncertainty vanishes, and we can measure the next qudit with a PVM which has a deterministic outcome.

For example, if we know a $|0\rangle$-$|+\rangle$ Quantum Golden Mean generator is in state B the next qudit is $|0\rangle$, and we will always see '0' if we apply the measurement M_{01}. Thus, we consider synchronization as a form of dynamical inference where an observer uses knowledge of both a

source's internal structure and a sequence of measurement outcomes to inform a more accurate prediction of future measurement outcomes. Each mixed state corresponds to a different prediction.

VI. QUANTUM PROCESS SYSTEM IDENTIFICATION

Synchronization is a task that performed when an observer has an accurate model (here, a HMCQS) of the quantum information source. The next natural question is: How does an observer create an accurate model of an *unknown* quantum information source? This section begins to answer this question. It starts by briefly reviewing how one infers a classical information source from data. It then discusses how to identify the state of a single qudit with quantum state tomography. By combining these two ideas we arrive at an inference method for stationary qudit sequences.

A. Classical System Identification

An observer of an unknown stationary classical process is limited to use only the available data—distributions of words over symbol alphabet $\mathcal{X}$—to infer the source's structure. We denote the distribution of length-ℓ words $\mathcal{P}(\ell) = \{w = x_0 \cdots x_{\ell-1} | w \in \mathcal{X}^\ell\}$.

For $\ell = 1$ we obtain a distribution over symbols in $\mathcal{X}$. With $\mathcal{P}(1)$ we can reconstruct a memoryless (i.i.d.) model of the source with one internal state and each $\Pr(x \in \mathcal{X})$ obtained directly from $\mathcal{P}(1)$.

For $\ell = 2$ we use $\mathcal{P}(2)$ to create conditional probability distributions for the next symbol conditioned on the previous one; i.e., $\Pr(x_1|x_0)$, for all x_0, $x_1 \in \mathcal{X}$. From these conditional distributions we may construct a Markov approximation of the source with $|\mathcal{X}|$ states, one corresponding to each symbol. The conditional distributions set the transition matrices for Markov approximation of the source: i.e., $\Pr(x_1|x_0) = T_{x_0,x_1}$.

Similarly, for $\ell > 2$ we obtain a length-ℓ HMM approximation of the source by conditioning on length $\ell - 1$ words, each of which corresponds to a different internal state. In this simplified picture, the number of internal states of the model grows exponentially with ℓ, as the number of possible words of length ℓ is $|\mathcal{X}|^\ell$. This leads to numerical problems when inferring processes with correlations over long periods of time. That said, there are methods for both combining states that reflect identical predictions of future symbols and for performing inference over machine topologies with a certain number of states—known as Bayesian Structural Inference [54]—to determine the most likely ϵ-machine for the source.

B. Tomography of a Qudit

We cannot directly apply the procedure for inferring classical source dynamics from word distributions to qudit processes since observations depend on the measurement basis/protocol one uses. To fully characterize a stationary quantum source we must instead take many measurements in different bases to reconstruct the qudit density matrices. This task is known as *quantum state tomography*. We begin by inferring an individual qudit state ρ_0 before introducing a general method for quantum system identification using separable sequences of qudits.

Tomographic reconstruction of a single unknown qudit density matrix ρ_0 through measurement is challenging for two main reasons:

1. ρ_0's complete description requires a number of parameters that scales exponentially with the state's Hilbert space dimension.
2. Quantum measurement is probabilistic, so one must prepare and measure many copies of ρ_0 to estimate a single parameter.

Specific combinations of measurements are particularly useful for this task. For example ρ_0 can be inferred by measuring with a set of mutually-unbiased bases (MUB) [3] or a single informationally-complete POVM (IC-POVM) [4]. For a qubit, one possible MUB consists of the x-, y-, and z-bases. By measuring many copies of the qubit in each of these three bases one obtains three probabilities—$\Pr(+x)$, $\Pr(+y)$, and $\Pr(+z)$—that uniquely determine the density matrix ρ_0. We do not discuss the necessary number of measurements to determine these parameters to within a desired tolerance for general qudit tomography, a question which is well-studied [33, 55].

An example of an IC-POVM for a qubit consists of the projectors onto states:

$$
\begin{aligned}
|\phi_1\rangle &= |0\rangle\,, \\
|\phi_2\rangle &= \frac{1}{\sqrt{3}}\,|0\rangle + \sqrt{\frac{2}{3}}\,|1\rangle\,, \\
|\phi_3\rangle &= \frac{1}{\sqrt{3}}\,|0\rangle + \sqrt{\frac{2}{3}} e^{i2\pi/3}\,|1\rangle\,, \quad \text{and} \\
|\phi_4\rangle &= \frac{1}{\sqrt{3}}\,|0\rangle + \sqrt{\frac{2}{3}} e^{i4\pi/3}\,|1\rangle\ . \qquad (52)
\end{aligned}
$$

This is also a symmetric IC-POVM, or SIC-POVM, because any combination of two projectors has the same inner product.

By measuring many identical copies of ρ_0 with the same IC-POVM one obtains a probability distribution over the d^2 possible measurement outcomes. This provides $d^2 - 1$ parameters (due to normalization) that uniquely determine the density matrix ρ_0.

The existence and properties of SIC-POVMs in higher dimensional Hilbert spaces is an active area of research [56].

C. Tomography of a Qudit Process

Now that we know how to estimate the density matrix for an individual qudit we can begin analyzing length-ℓ density matrices $\rho_{0:\ell}$. We will measure each qudit in turn rather than performing a joint measurement over the entire length-ℓ block of qudits. This is a luxury afforded to us because we are focusing on separable qudit sequences—the generic case of entangled qudit processes requires measuring in nonlocal bases.

For $\ell = 1$ we reconstruct ρ_0 as described above and obtain a memoryless (i.i.d.) estimate of the source that emits qudit ρ_0 at every timestep, following Eq. (4). Unless ρ_0 is a pure state, there are many single-state HMCQS that generate this process since many different pure-state ensembles correspond to the same density matrix. A unique memoryless model of the source may be obtained by diagonalizing ρ_0 and having the source emit each pure eigenstate $|\psi_i\rangle$ with probability equal to the corresponding eigenvalue λ_i.

For $\ell = 2$ we must reconstruct the two-qudit density matrix $\rho_{0:2}$. (Recall that our indexing is left-inclusive and right-exclusive, therefore $\rho_{0:2}$ is the joint state of the qudits for $t = 0, 1$). Due to stationarity, $\rho_{0:2}$ (the joint state of the two qudits) must be consistent with the one-qudit marginals, i.e. $tr_0(\rho_{0:2}) = \rho_1 = \rho_0 = tr_1(\rho_{0:2})$.

We will describe the iterative procedure for reconstructing $\rho_{0:\ell}$ for *qubits* in detail. When $d = 2$, $\rho_{0:2}$ has 15 real parameters that must be determined via tomography. Tomography on the one-qubit marginals determines 3 parameters, and the condition of stationarity fixes 3 more. The state can be reconstructed fully by considering combinations of the set of mutually-unbiased measurements. For two qubits, this means the 16 combinations of Pauli matrices ($\sigma_{I_0} \otimes \sigma_{x_1}$, $\sigma_{x_0} \otimes \sigma_{x_1}$, $\sigma_{x_0} \otimes \sigma_{y_1}$, and so on.) [57]. To fully characterize $\rho_{0:2}$ 9 of these values must be determined—those not involving the identity operators σ_{I_0} and σ_{I_1} which are fixed by the one-qubit marginals. For $d > 2$, this procedure can be modified by using a set of mutually-unbiased bases in that higher-dimensional Hilbert space.

After determining $\rho_{0:2}$ one can continue on to determine $\rho_{0:3}$ (63 real parameters for qubits). Many of these parameters are fixed by the previous tomography on the one-qubit marginals (3 parameters), the two-qubit marginals (15 parameters), and their stationarity conditions (6 and 15 parameters respectively). Combinations of three one-site Pauli matrices are sufficient for full reconstruction. One may continue this procedure for larger ℓ, typically until the number of measurements becomes experimentally infeasible.

D. Cost of I.I.D.

Quantum information sources are often assumed to be i.i.d. [6, 33]. If an observer performing quantum state tomography assumes that an unknown quantum information source is i.i.d., then they will not go beyond determining the one-qudit marginal ρ_0. If the qudits are instead correlated, this will lead to an overestimate of the source's randomness. They will erroneously conclude that each qudit is in state ρ_0 and that the entropy rate is $S(\rho_0)$ bits per timestep. The latter overestimates the true entropy rate by a factor of $S(1) - s$. An observer can obtain better estimates for the entropy rate and other informational quantities by following the above procedure and tomographically reconstructing blocks of qudits of length-ℓ.

To demonstrate the degree to which this assumption may mislead an observer, consider the process generated by nonunifilar qubit source in Fig. 12. Assume the source begins in its stationary state distribution: $(p_A = 1/2, p_B = 1/2)$. For any p such that $0 \leq p < 1$, ρ_0 is the maximally-mixed state and an observer assuming an i.i.d. process estimates that $s = 1$ bit per timestep. This is only an accurate description of the source for $p = 1/2$. Whereas, for many values of p, this source has significant correlations between subsequent qubits.

Let's take a closer look at the extreme values of p. For $p = 0$ the process is period-2 with $s = 0$ bits per timestep and $\mathbf{E}_q = 1$ bit. This pattern can be easily detected by measuring in the $\{|0\rangle, |1\rangle\}$ basis, where a measurement of 0 (1) immediately synchronizes an observer to state B (A). As $p \to 1$, the two source states become increasingly disconnected, $s \to 0$, and $\mathbf{E}_q \to 1$. An observer measuring in the $\{|+\rangle, |-\rangle\}$ basis observes a + (−) is likely to measure another + (−). This source's rich and varied behavior at different values of p will go entirely unappreciated when considering only the one-qubit density matrix, ρ_0.

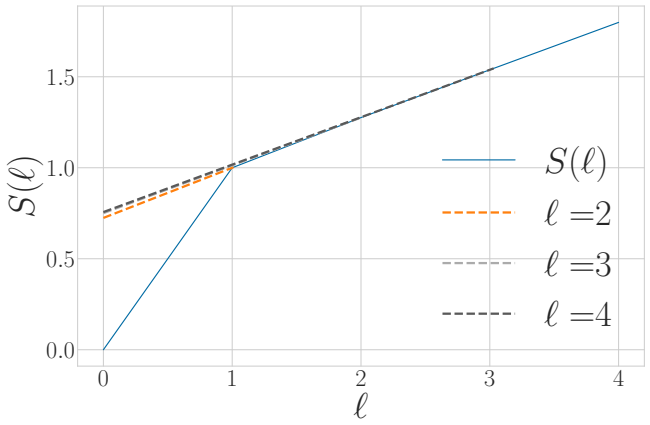


FIG. 23. Block entropy and length-ℓ estimates for information properties of the process generated by the nonunifilar qubit source in Fig. 12 with $p = 0.05$. The slope and y-intercept of the linear estimates are s and $\mathbf{E}_q$, respectively. Note that the estimates do not improve significantly for $\ell > 2$, indicating that the two-qubit correlations are most significant for determining information properties of the process.

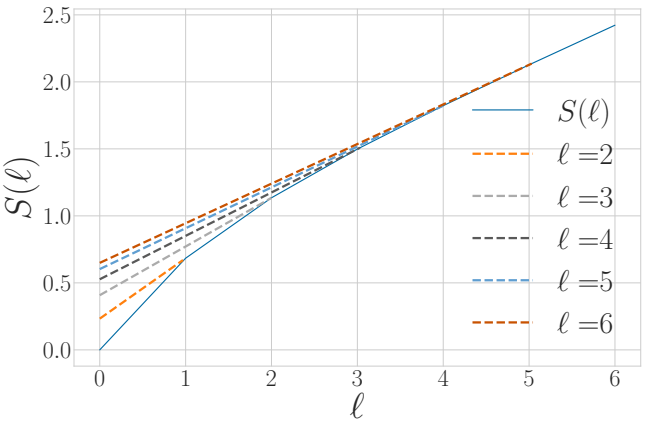


FIG. 24. Block entropy and length-ℓ estimates for information properties of the process generated by the unifilar qubit source in Fig. 11 with $p = 0.05$. The slope and y-intercept of the linear estimates are s and $\mathbf{E}_q$, respectively. Note that the estimates improve steadily for larger ℓ, indicating long-range correlations.

E. Finite Length Estimation of Information Properties

We just saw an example of how, if an observer assumes a source is i.i.d., they will generally underestimate the structure and correlation of the quantum process and will overestimate its entropy rate. The same is true if they only perform tomography on blocks of qudits up to finite length ℓ. Consider now an experimenter who tomographically reconstructs the density matrix $\rho_{0:\ell}$ and then assumes there are no additional (longer-range) correlations within the qudit process.

Their estimates for the quantum entropy rate, quantum excess entropy and quantum transient information are given by Eqs. (38), (45), and (47) respectively.

The difference between the length-ℓ estimate of an information property and its true value depends on the process' internal structure and quantum alphabet. The estimates for the nonunifilar and unifilar qubit sources in Figs. 23 and 24 represent two extremes in this respect. As previously discussed, the single-qubit density matrix for the nonunifilar qubit source is the maximally-mixed state. If an observer instead reconstructs $\rho_{0:2}$, they significantly improve their estimate of s, $\mathbf{E}_q$, and $\mathbf{T}_q$. However, for $\ell > 2$ these estimates do not improve dramatically. There is always a trade-off between the number of experiments necessary to reconstruct the process tomographically and the accuracy of the estimates obtained from that reconstruction. For this source, $\ell = 2$ strikes a balance between those two resources.

In contrast, for the unifilar qubit source the estimates of s, $\mathbf{E}_q$, and $\mathbf{T}_q$ improve steadily as ℓ increases. Longer-range correlations are captured by increasing the length of the reconstructed density matrices $\rho_{0:\ell}$. When observing this source it is likely worth finding $\rho_{0:\ell}$ for the largest ℓ that is experimentally feasible.

When faced with an unknown quantum source, how does one pick an appropriate value of ℓ? One strategy is to increase ℓ until the correction made to the relevant information quantities by going to $\ell + 1$ is below some threshold. Stationarity (and the resulting concavity of the quantum block entropy) ensure that future corrections will also be below that threshold.

F. Tomography with a Known Quantum Alphabet

If an observer has additional knowledge of what possible pure states an HMCQS may emit (i.e., the quantum alphabet $\mathcal{Q}$), they can leverage this knowledge to simplify the task of system identification by inferring the word probabilities of the underlying classical process $\overleftrightarrow{X}$ rather than performing full tomographic reconstruction. For qubits we can represent this simplification geometrically via the Bloch sphere; see Fig. 25. Each point on the surface of the Bloch sphere represents a pure qubit state in $\mathcal{H}^2$—such as, $|0\rangle$ and $|1\rangle$ on the poles of the z-axis—and each interior point represents a possible qubit density matrix. The following assumes each element of $\mathcal{Q}$ is unique.

1. $\ell = 1$

The length-1 density matrix ρ_0 must satisfy the equation

$$\rho_0 = \sum_{|\psi_x\rangle \in \mathcal{Q}} \Pr(|\psi_x\rangle)\, |\psi_x\rangle\langle\psi_x| \ .$$

After finding ρ_0 via tomography one may rearrange this equation to infer the length-1 word distributions of $\overleftrightarrow{X}$ given that $\Pr(X = x) = \Pr(|\psi_x\rangle)$. The feasibility of this task depends on the relationship between $|\mathcal{Q}|$ and d. For the qubit case ($d = 2$) one can uniquely infer $\Pr(X_0)$ if $|\mathcal{Q}| \leq 3$; assuming no degenerate states in $\mathcal{Q}$.

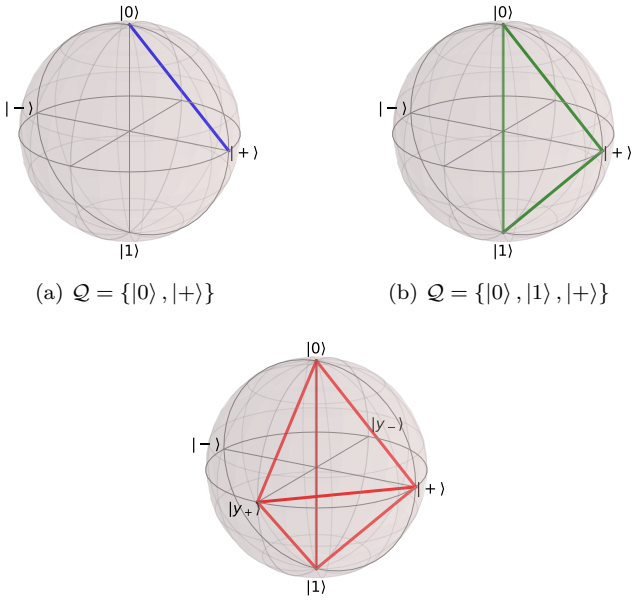
When $\mathcal{Q}$ is known it may not be necessary to fully reconstruct ρ_0. We first present several simple examples before giving a general algorithm for finding $\rho_{0:\ell}$ without performing full state tomography on $\rho_{0:\ell}$.

For $d = 2$ and $|\mathcal{Q}| = 2$, the possible values of ρ_0 are restricted to a chord within the Bloch sphere defined by $\rho_0 = p\,|\psi_0\rangle\langle\psi_0| + (1-p)\,|\psi_1\rangle\langle\psi_1|$ with $0 < p < 1$. One needs only to determine the parameter p rather than reconstruct ρ_0 in its entirety. An observer can also pick a uniquely informative measurement to determine p. The optimal PVM for doing so is one whose antipodal projectors can be connected with the diameter of the Bloch sphere that runs parallel to the line of possible values of ρ_0. For example, if $\mathcal{Q} = \{|0\rangle, |+\rangle\}$ then the set of possible density matrices lies on the line segment in Fig. 25a. The best PVM is then M_θ with $\theta = \frac{3\pi}{4}$ and measurement outcomes y_0 (corresponding to a projection on pure state $|\psi(\theta = \frac{3\pi}{4})\rangle$) and y_1 (corresponding to the orthogonal projector). In this case it can easily be shown that $p = \frac{\sqrt{2}+1}{2} - \sqrt{2}\Pr(y_0|\rho_0)$ and that $\cos^2(\frac{3\pi}{8}) \leq \Pr(y_0|\rho_0) \leq \sin^2(\frac{3\pi}{8})$. M_{01} and $M_\pm$ would also be able to determine p, but require more samples to determine p to within some desired tolerance.

For $d = 2$ and $|\mathcal{Q}| = 3$, the possible values of ρ_0 are confined to a simplex in the Bloch sphere defined by $\rho_0 = p_0\,|\psi_0\rangle\langle\psi_0| + p_1\,|\psi_1\rangle\langle\psi_1| + (1-p_0-p_1)\,|\psi_2\rangle\langle\psi_2|$ with $0 < p_0, p_1 < 1$ and $p_0 + p_1 < 1$. One may determine the parameters p_0 and p_1 rather than the 3 parameters usually required to characterize a qubit mixed state. There are two simple choices of measurements to do so: using a IC-POVM or using two different PVMs.

For the first case, consider measuring ρ_0 with a SIC-POVM with elements $E_y = \frac{1}{2}\,|\phi_y\rangle\langle\phi_y|$, with each $|\phi_y\rangle$ described by Eq. 52. The probability of observing measurement y can be written as:

$$\begin{aligned}\Pr(y|\rho_0) &= \sum_x \Pr(|\psi_x\rangle)\Pr(y|\,|\psi_x\rangle) \\ &= \sum_x \frac{p_x}{2}|\langle\psi_x|\phi_y\rangle|^2 \ .\end{aligned}$$



(a) $\mathcal{Q} = \{|0\rangle, |+\rangle\}$ (b) $\mathcal{Q} = \{|0\rangle, |1\rangle, |+\rangle\}$

(c) $\mathcal{Q} = \{|0\rangle, |1\rangle, |+\rangle, |y_+\rangle\}$

FIG. 25. The Bloch sphere representation of the boundary of the set of possible length-1 density matrices (ρ_0) for the given $\mathcal{Q}$. (a) The set of valid states is a line segment. An observer only needs to determine one parameter. (b) The set of valid states is the interior of a triangle. An observer must determine two parameters. (c) The set of valid states is the interior of a tetrahedron. An observer must determine three parameters, and the decomposition into an ensemble of basis states is not unique. Here $|y_+\rangle = \frac{1}{\sqrt{2}}(|0\rangle - i\,|1\rangle)$.

One can rearrange the system of 4 equations (one for each POVM element) to obtain a unique set of p_x's.

Alternatively, one uses two PVMs whose projectors can be connected by (ideally orthogonal) diameters of the Bloch sphere that are parallel to the simplex of possible ρ_0 values. This will yield 2 parameters that uniquely determine a point on the ρ_0 simplex. An example with $\mathcal{Q} = \{|0\rangle, |1\rangle, |+\rangle\}$ is shown in Fig. 25b, for which the possible values of ρ_0 are confined to the interior of a triangle in the Bloch sphere. One can determine p_0 and p_1 by measuring with orthogonal PVMs M_{01} and $M_\pm$ (among many other combinations), in which case $(1 - p_0 - p_1) = 2\Pr(\text{`+'}|\rho_0, M_\pm) - 1$ and $(p_0 - p_1) = 2\Pr(\text{`0'}|\rho_0, M_{01}) - 1$.

For $d = 2$ and $|\mathcal{Q}| = 4$ the possible values of ρ_0 are confined to a tetrahedron in the Bloch sphere whose vertices are the elements of $\mathcal{Q}$, and one cannot uniquely infer the classical symbol distribution from a fully-reconstructed ρ_0. For example, if $\mathcal{Q} = \{|0\rangle, |1\rangle, |+\rangle, |-\rangle\}$ and ρ_0 is the maximally-mixed state, this could correspond to any mixture of the form $\rho_0 = p\,|0\rangle\langle 0| + p\,|1\rangle\langle 1| + (1-p)\,|+\rangle\langle +| + (1-p)\,|-\rangle\langle -|$ with $0 < p < 1$. Thus, for $d = 2$ and $|\mathcal{Q}| \geq 3$, the tomographic advantage to knowing $\mathcal{Q}$ is reduced but not eliminated as an observer can immediately

exclude any value of ρ_0 that lies outside the convex polyhedron defined by the elements of $\mathcal{Q}$. This is shown in Fig. 25c, where the region of possible ρ_0 values is confined to less than $\frac{1}{4}$ of the volume of the Bloch sphere.

Similar simplifications apply for $d = 3$ (qutrits) when $\mathcal{Q}$ is known. Full tomography of an arbitrary mixed qutrit state requires the determination of 8 parameters, whereas determining the classical distribution given $\mathcal{Q}$ requires $|\mathcal{Q}| - 1$ parameters. This presents an advantage in general when $|\mathcal{Q}| \leq 8$. We do not explicitly construct measurements that can realize this advantage, as a geometric understanding of mixed states over the 8-dimensional space $\mathcal{H}^3$ is significantly more involved than the 3-dimensional Bloch sphere describing mixed states over $\mathcal{H}^2$.

For a generic qudit the number of parameters required for full tomography is $d^2 - 1$. And so, we expect that knowledge of $\mathcal{Q}$ gives a clear tomographic advantage (fewer parameters must be determined) when $|\mathcal{Q}| < d^2$.

We are now prepared to give a general protocol for $|\mathcal{Q}| = n$ and arbitrary d. We wish to recover the underlying distribution of alphabet states ($p_x = \Pr(|\psi_x\rangle)$, $0 \leq x < n$) from measurement statistics alone. First, we construct a POVM with $n + 1$ elements: $E_y = c_y |\psi_x\rangle \langle\psi_x|$ for each $|\psi_x\rangle$ in $\mathcal{Q}$, and $E_n = \mathbb{I} - \sum_{y=0}^{n-1} E_y$. Each c_y is a parameter which can be varied to ensure that E_n is positive semi-definite. By applying this POVM to ρ_0 we obtain a distribution ($\Pr(y|\rho_0)$) over the $n + 1$ possible measurements. The first n are related to our desired distribution by:

$$\begin{aligned}\Pr(y|\rho_0) &= \sum_x p_x \Pr(y|\,|\psi_x\rangle) \\ &= \sum_x p_x c_y |\langle\psi_x|\psi_y\rangle|^2 \ ,\end{aligned}$$

with one equation for each $y < n$. If a set of p_x's is a solution to this system of linear equations, it is consistent with the observed measurements. The solution will be unique for $|\mathcal{Q}| < d^2$.

2. $\ell = 2$

Knowing $\mathcal{Q}$ provides further advantage when considering tomography of multiple qudits. The distribution over classical words of length ℓ has $|\mathcal{Q}|^\ell - 1$ parameters, whereas full tomography of ℓ qudits requires the determination of $d^{2\ell} - 1$ parameters.

For $\ell = 2$:

$$\begin{aligned}\rho_{0:2} = \sum_{|\psi_{x_0}\rangle, |\psi_{x_1}\rangle \in \mathcal{Q}} \Big[&\Pr(|\psi_{x_0}\rangle \otimes |\psi_{x_1}\rangle) \\ &\big(|\psi_{x_0}\rangle \otimes |\psi_{x_1}\rangle\big)\big(\langle\psi_{x_0}| \otimes \langle\psi_{x_1}|\big)\Big] \ .\end{aligned} \tag{53}$$

Once again we consider the case of $d = 2$ and $|\mathcal{Q}| = 2$ explicitly. There are 4 length-2 classical word probabilities, but there are 3 constraints imposed by (i) normalization, (ii) stationarity, and (iii) consistency with the one-qubit marginal. Thus, one only needs to determine a single parameter to reconstruct $\rho_{0:2}$.

Consider the task of reconstructing the length-2 density matrix produced by the $|0\rangle$-$|+\rangle$ Quantum Golden Mean generator in Fig. 6 with the knowledge that $\mathcal{Q} = \{|0\rangle, |+\rangle\}$. One would first analyze the one-qubit density matrix to find that $\Pr(|0\rangle) = \frac{2}{3}$ and $\rho_0 = \frac{2}{3}|0\rangle\langle 0| + \frac{1}{3}|+\rangle\langle +|$.

The word probability $\Pr(|00\rangle)$ is the only necessary additional information to find $\rho_{0:2}$. The following simple measurement protocol can determine $\Pr(|00\rangle)$: Measure two consecutive qubits with $M_\pm$. If the qubit is in state $|+\rangle$, one will never see outcome '$-$'. If the qubit is in state $|0\rangle$, one will see outcome '$-$' with probability $\frac{1}{2}$. And so, $\Pr(|00\rangle) = 4\Pr(\text{`}{-}{-}\text{'}|\rho_{0:2}, M_\pm \otimes M_\pm)$.

This procedure can be generalized to arbitrary 2-element alphabets $\mathcal{Q} = \{|\psi_0\rangle, |\psi_1\rangle\}$. First, measure two consecutive qudits with a PVM $M_{\tilde{0}1}$, where one element is a projector onto $|\psi_1\rangle$ with outcome '1' and the orthogonal projector corresponds to measurement outcome '$\tilde{0}$'. Second:

$$\begin{aligned}&\Pr(\,|\psi_0\rangle \otimes |\psi_0\rangle) \\ &\quad = (\Pr(\text{`}\tilde{0}\text{'}|\,|\psi_0\rangle, M_{\tilde{0}1}))^{-2} \Pr(\text{`}\tilde{0}\tilde{0}\text{'}|\rho_{0:2}, M_{\tilde{0}1} \otimes M_{\tilde{0}1}) \\ &\quad = (1 - |\langle\psi_0|\psi_1\rangle|)^{-2} \Pr(\text{`}\tilde{0}\tilde{0}\text{'}|\rho_{0:2}, M_{\tilde{0}1} \otimes M_{\tilde{0}1}) \ .\end{aligned}$$

This provides a clear advantage over the usual 9 parameters necessary to reconstruct $\rho_{0:2}$ as it takes into account that the one-qubit marginals and stationarity each impose 3 constraints.

For $d = 2$ and $|\mathcal{Q}| \geq 3$ we must determine additional parameters of the underlying classical distribution over $\mathcal{Q}$. We do so by repeatedly applying the SIC-POVM with elements $E_y = \frac{1}{2}|\phi_y\rangle\langle\phi_y|$, with each $|\phi_y\rangle$ described by Eq. 52.

The probability of observing the length-2 measurement

sequence $y_0 y_1$ can be written as:

$$\begin{aligned}\Pr(y_0 y_1|\rho_{0:2}) &= \sum_{x_0,x_1} p_{x_0,x_1} \Pr(y_0 y_1| \, |\psi_{x_0}\rangle \otimes |\psi_{x_1}\rangle) \\ &= \sum_{x_0,x_1} \frac{p_{x_0,x_1}}{4} |\langle\psi_{x_0}|\phi_{y_0}\rangle|^2 |\langle\psi_{x_1}|\phi_{y_1}\rangle|^2 \ .\end{aligned}$$

There are $|\mathcal{Q}|^2$ p_{x_0,x_1}'s. If a set of p_{x_0,x_1}'s is a solution to this system of equations it is consistent with the measurement statistics, and the solution will be unique for $|\mathcal{Q}| = 3$.

For $|\mathcal{Q}| = n$ and arbitrary d we may construct a POVM from $\mathcal{Q}$ as we did for $\ell = 1$. It is then possible infer a set of p_{x_0,x_1} that solves the resulting system of equations from measurement statistics. For length-2 words the equations are:

$$\begin{aligned}\Pr(y_0 y_1|\rho_{0:2}) &= \sum_{i} p_{x_0,x_1} \Pr(y_0 y_1| \, |\psi_{x_0}\rangle \otimes |\psi_{x_1}\rangle) \\ &= \sum_{x_0,x_1} p_{x_0,x_1} c_{y_0} c_{y_1} |\langle\psi_{x_0}|\psi_{y_0}\rangle|^2 |\langle\psi_{x_1}|\psi_{y_1}\rangle|^2 \ .\end{aligned}$$

There are n^2 equations, one for each length-2 measurement sequence corresponding to the POVM elements $E_{y_0} \otimes E_{y_1}$. Calculating the other possible outcomes (corresponding to E_n) is redundant due to normalization.

3. $\ell \geq 3$

Extending this analysis to a length-ℓ density matrix $\rho_{0:\ell}$ takes the form:

$$\rho_{0:\ell} = \sum_{|\psi_w\rangle \in \mathcal{Q}^\ell} \left[\Pr(|\psi_w\rangle) \, |\psi_w\rangle \langle\psi_w| \right] ,$$

where each $|\psi_w\rangle$ has the form of Eq. (5). One can determine the length-ℓ word distributions uniquely for the general case where $|\mathcal{Q}|^\ell < d^{2\ell}$.

The various measurement strategies explored above for $\ell = 2$ can be extended to arbitrary values of ℓ. We will explicitly describe two: using a SIC-POVM for $d = 2$ and using a POVM constructed from $\mathcal{Q}$ for arbitrary d.

Consider repeatedly applying the SIC-POVM with elements $E_y = \frac{1}{2} |\phi_y\rangle \langle\phi_y|$, with each $|\phi_y\rangle$ described by Eq. (52). Taking ℓ measurements, one observes a length-ℓ word $y_{0:\ell}$. $\mathcal{Y}$ is the 4-element alphabet of measured symbols.

The probability of observing the length-ℓ measurement sequence $y_{0:\ell}$ can be written as:

$$\begin{aligned}\Pr(y_{0:\ell}|\rho_{0:\ell}) &= \sum_{|\psi_w\rangle} \Pr(|\psi_w\rangle) \Pr(y_{0:\ell}| \, |\psi_w\rangle) \\ &= \sum_{|\psi_w\rangle} \frac{\Pr(|\psi_w\rangle)}{2^\ell} |\langle\psi_w|\phi_{y_{0:\ell}}\rangle|^2 \ ,\end{aligned}$$

where $|\phi_{y_{0:\ell}}\rangle = \bigotimes_{t=0}^{\ell} |\phi_{y_t}\rangle$ and the factor of 2^ℓ comes from the POVM elements. Each of the 4^ℓ sequences has a probability that can be estimated from measurement. This system of equations can be solved to find underlying length-ℓ word probabilities that are consistent with measurements. If $|\mathcal{Q}| \leq 3$ this solution is unique.

We can also measure ℓ times with a POVM constructed directly from $\mathcal{Q}$. In this case the resulting equations are:

$$\begin{aligned}\Pr(y_{0:\ell}|\rho_{0:\ell}) &= \sum_{|\psi_w\rangle} \Pr(|\psi_w\rangle) \Pr(y_{0:\ell})| \, |\psi_w\rangle) \\ &= \sum_{|\psi_w\rangle} p_w \left(\prod_{t=0}^{\ell-1} c_{y_t} \right) |\langle\psi_w|\psi_{y_{0:\ell}}\rangle|^2 \ .\end{aligned}$$

As before, one infer the $|\mathcal{Q}|^\ell$ underlying word probabilities (p_w's) uniquely in the case where $|\mathcal{Q}| < d^2$.

G. Source Reconstruction

After observing a separable qudit process and finding the length-ℓ density matrix $\rho_{0:\ell}$ can one infer the HMCQS that generated it? The following reconstructs the source that generates $\rho_{0:\ell}$ of an unknown process for different values of ℓ. Note that a source which generates $\rho_{0:\ell}$ may fail to generate $\rho_{0:\ell+1}$.

1. $\ell = 1$

After determining ρ_0 an observer may construct an i.i.d. approximation of the quantum information source. Given any decomposition of ρ_0 into pure states $|\psi_x\rangle$—as in Eq. (3)—the corresponding HMCQS consists of one internal state $\mathcal{Q} = \{|\psi_x\rangle\}$ and the single transition probability for each $|\psi_x\rangle$ is its corresponding probability $\Pr(|\psi_x\rangle)$ in the decomposition of ρ_0. A unique model may be obtained by taking ρ_0's eigendecomposition. In this case the model emits each eigenstate $|\psi_i\rangle$ with probability equal to the corresponding eigenvalue: $\Pr(|\psi_i\rangle) = \lambda_i$.

2. $\ell = 2$

With $\rho_{0:2}$ an observer begins to model a source that generates correlations between qudits. Doing so requires finding a separable decomposition of $\rho_{0:2}$ of Eq. (53)'s form. If $\mathcal{Q}$ is known, multiple procedures for finding such a separable decomposition have been introduced above. For a generic two-qudit density matrix, determining whether it is separable or entangled is generally NP-hard [58]. There are also many necessary and sufficient conditions for separability; for example, the Positive-Partial Transpose (PPT) criterion [59].

The following assumes that the observer has a separable decomposition of $\rho_{0:2}$ into alphabet states $\mathcal{Q}'$ that may or may not be the source's alphabet $\mathcal{Q}$. In general, the sets of basis states in the decomposition of a two-qudit density matrix may differ, but we require a symmetric decomposition such that the basis states of both qudits are $\mathcal{Q}'$. From this decomposition they can construct an HMCQS with an underlying Markov dynamic described by Eq. (9). This HMCQS has $|\mathcal{Q}'|$ internal states and transition probabilities $T_{\sigma_x,\sigma_{x'}} = \Pr(|\psi_{x'}\rangle \,|\, |\psi_x\rangle)$ that can be calculated from $\rho_{0:2}$. Different separable decompositions of $\rho_{0:2}$ yield different HMCQS, whose statistics over longer-length sequences may differ. Determining which is a more accurate model of the source requires performing tomography on $\rho_{0:3}$ to refine the model.

3. $\ell \geq 3$

Finding a separable decomposition becomes more computationally expensive as the number of qudits increases [60]. Nevertheless, if one obtains a separable decomposition of $\rho_{0:3}$ where the basis states for all 3 qudits are $\mathcal{Q}'$, then one can create a length-2 Markov approximation of the source. Each length-2 sequence of qudits in $\mathcal{Q}'$ corresponds to a different internal HMCQS state. Thus, it has $|\mathcal{Q}'|^2$ internal states, unless some length-2 sequences are forbidden. The transition probabilities take the form $T_{\sigma_{x_0,x_1},\sigma_{x_1,x_2}} = \Pr(|\psi_{x_2}\rangle \,|\, |\psi_{x_0}\rangle\,|\psi_{x_1}\rangle)$. Only transition probabilities that obey concatenation are nonzero.

One can continue in this manner, approximating the source given a separable decomposition of $\rho_{0:\ell}$ to obtain an HMCQS with $|\mathcal{Q}'|^{\ell-1}$ states or fewer, if some sequences are forbidden. Each state corresponds to a word of length $\ell-1$ where the pure states composing the word are drawn from $\mathcal{Q}'$. The number of internal states in the model grows exponentially with ℓ, but not all of these states may lead to unique future predictions. If so, they can be combined without a loss of predictivity, as is done in classical computational mechanics. A general algorithm for doing so is beyond the present scope.

H. Discussion

The preceding detailed many aspects of identifying an unknown quantum process by tomographically reconstructing the length-ℓ density matrices $\rho_{0:\ell}$. When correlations exist between qudits, this provides a predictive advantage over the common assumption that sources are i.i.d.. Starting with the method for reconstructing classical processes, we developed a variety of measurement protocols for different values of d and $|\mathcal{Q}|$ and ℓ to find a process' statistics when $\mathcal{Q}$ is known. We then introduced effective models for quantum sources derived entirely from separable decompositions of the density matrices $\rho_{0:\ell}$ reconstructed via tomography.

Since our aim is to harness correlations to improve predictions of future measurement outcomes, it is worth asking, How accurately can future measurement outcomes be predicted? To begin to answer this the following now defines maximally-predictive measurements for the special case of $\ell = 2$ and for arbitrary ℓ.

For a correlated qudit process, an observer with knowledge of the length-2 density matrix $\rho_{0:2}$ may perform a measurement on the first qudit and condition upon the outcome to reduce their uncertainty when measuring the second qudit. They apply M_0 (with possible outcomes y_i) on the first qubit ρ_0. This leaves the joint system in the classical-quantum state:

$$\rho_{0:2}^{M_0} = \sum_{y_i} \Pr(y_i)\, |y_i\rangle \langle y_i| \otimes \rho_1^{y_i} \ ,$$

where the $\rho_1^{y_i}$ are the qudit density matrices conditioned on outcome y_i.

The conditional von Neumann entropy of the second qubit is then:

$$S(\rho_1|M_0(\rho_0)) = \sum_{y_i} \Pr(y_i) S(\rho_1^{y_i}) \ .$$

The rank-one measurement with the minimal uncertainty in measurement outcomes is in the eigenbasis of $\rho_1^{y_i}$. Different y_i values generally correspond to different minimal-entropy measurements on the second qubit.

For two qubits, we can find the PVM for which the conditional von Neumann entropy of the second qubit is minimized. This leads to a basis-independent property of the process:

$$S_{min}(\rho_1|M_{min}(\rho_0)) = \min_{M_0} S(\rho_1|M_0(\rho_0)) \ ,$$

where the minimum is taken over all PVMs on ρ_0.

If an experimenter reconstructs $\rho_{0:\ell}$, a measurement on all but the last qudit in the block with a measurement

protocol $\mathcal{M}$ with possible outcomes $y_{0:\ell-1}$ leaves the block in the classical-quantum state:

$$\rho_{0:\ell}^{\mathcal{M}} = \sum_{y_{0:\ell-1}} \Pr(y_{0:\ell-1}) \, |y_{0:\ell-1}\rangle \langle y_{0:\ell-1}| \otimes \rho_{\ell}^{y_{0:\ell-1}} \ .$$

To minimize the conditional von Neumann entropy of the ℓ-th qubit, we calculate:

$$S_{min}(\rho_\ell|\mathcal{M}_{min}(\rho_{0:\ell-1})) = \min_{\mathcal{M}} S(\rho_\ell|\mathcal{M}(\rho_0)) \ ,$$

where the minimum is taken over all local measurement protocols on $\rho_{0:\ell-1}$.

Further development necessitates exploring the space of measurement protocols to find those with the greatest predictive advantage over i.i.d. models for arbitrary separable qudit processes.

Finally, we note that our procedure for source reconstruction required a number of model states that grows exponentially with ℓ. Future work on finding *minimal* models from density matrices will also require combining states with identical predictions and performing inference over possible model topologies, as with classical Bayesian Structural Inference.

VII. CONCLUSION

Inspired by prior information-theoretic studies of classical stochastic processes, we introduced methods to quantify structure and information production for stationary quantum information sources. We identified properties related to the quantum block entropy $S(\ell)$ that allow one to determine the amount of randomness and structure within a given qudit process. We gave bounds on informational properties of the resulting measured classical processes. In particular, we showed that they cannot have a lower entropy rate or block entropy (at any ℓ) than the original quantum process.

We analyzed a number of hidden Markov chain quantum sources (HMCQSs), explaining how an observer synchronizes to a source's internal states via measuring the emitted qudits. If the source allows synchronizing observations, then we showed that adaptive measurement protocols are capable of synchronizing and maintaining synchronization when fixed-basis measurements cannot.

Sequels will extend these methods and results in a number of ways. Despite focusing here on separable quantum sequences for simplicity, entangled qudit sequences can similarly be studied by combining a HMCQS and a D-dimensional quantum system capable of sequentially generating matrix product states [24]. Doing so will open up the study of entropy convergence of matrix product operators [61].

Many results exist for classical stochastic processes that may be extended to quantum processes. For example, there exist closed-form expressions for informational measures for nondiagonalizable classical dynamics [50, 62–64]. Extending these to quantum dynamics would allow for more accurate determination of the quantum information properties introduced here. Similarly, the preceding lays the groundwork for fluctuation theorems and large deviation theory of separable quantum processes. Finally, it will be worthwhile to develop a causal equivalence relation for quantum stochastic processes and develop quantum ϵ-machines by extending classical results [65].

Separable quantum sequences also serve as a resource for information processing by finite-state quantum information transducers that transform one quantum process to another. Beyond interest in their own right, such operations have thermodynamic consequences, either requiring work to operate (as overcoming dissipation induced by Landauer erasure [66]) or acting as a quantum version of information-powered engines capable of leveraging environmental correlations to perform useful work [26, 67, 68]. This behavior has already been demonstrated for certain quantum processes [69].

Finally, though spatially-extended, ground and thermal states of spin chains under various Hamiltonians are quantum processes. And so, the quantum information measures introduced here can serve to classify these states according to the source complexity required to generate them.

ACKNOWLEDGMENTS

We thank Fabio Anza, Sam Loomis, Alex Jurgens, and Ariadna Venegas-Li and participants of the Telluride Science Research Center Information Engines Workshops for helpful discussions. JPC acknowledges the kind hospitality of the Telluride Science Research Center, Santa Fe Institute, and California Institute of Technology for their hospitality during visits. This material is based upon work supported by, or in part by, Grant Nos. FQXi-RFP-IPW-1902 and FQXi-RFP-1809 from the Foundational Questions Institute and Fetzer Franklin Fund (a donor-advised fund of Silicon Valley Community Foundation) and grants W911NF-18-1-0028 and W911NF-21-1-0048 from the U.S. Army Research Laboratory and the U.S. Army Research Office.

Appendix A: Information in Classical Processes

This section briefly recounts properties that quantify the randomness and correlation of classical stochastic processes as developed in Ref. [11]. Details and complete proofs can be found there. The main text here introduces versions appropriate to quantum processes.

1. Shannon Entropy

We quantify the amount of uncertainty in a discrete random variable X with possible outcomes $\{x_1, x_2, ...x_n\}$ by its *Shannon entropy*: [7]:

$$\mathrm{H}[X] \equiv -\sum_{i=1}^{n} \Pr(x_i) \log_2 \Pr(x_i) \ .$$

We use $\log_2$, in which case the units for Shannon entropy are *bits*.

To study correlations between multiple random variables we use several additional information quantities related to the Shannon entropy. First, the *joint entropy* of two discrete random variables—X and Y, with possible outcomes $\{x_1, x_2, ...x_n\}$ and $\{y_1, y_2, ...y_m\}$, respectively—is defined as:

$$\mathrm{H}[X, Y] \equiv -\sum_{i=1}^{n} \sum_{j=1}^{m} \Pr(x_i, y_j) \log_2 \Pr(x_i, y_j) \ .$$

X and Y are statistically independent if and only if the joint entropy decomposes as $\mathrm{H}[X, Y] = \mathrm{H}[X] + \mathrm{H}[Y]$.

Second, the *conditional entropy* of X conditioned on Y is:

$$\mathrm{H}[X|Y] \equiv -\sum_{i=1}^{n} \sum_{j=1}^{m} \Pr(x_i, y_j) \log_2 \Pr(x_i|y_j) \ .$$

$\mathrm{H}[X|Y]$ is the uncertainty in the value of X after already knowing the value of Y. Note that $\mathrm{H}[X|Y] \geq 0$ and is not symmetric: $\mathrm{H}[X|Y] \neq \mathrm{H}[Y|X]$. If X and Y are uncorrelated, then $\mathrm{H}[X|Y] = \mathrm{H}[X]$.

From the above definitions one can derive the following identity, linking conditional and joint entropies:

$$\mathrm{H}[X|Y] = \mathrm{H}[X, Y] - \mathrm{H}[Y] \ .$$

Third and finally, the *mutual information* between two random variables is:

$$\mathrm{I}[X : Y] \equiv \mathrm{H}[X] - \mathrm{H}[X|Y] \ .$$

This is the amount of information one can gain about X by having complete knowledge of Y. The mutual information is symmetric, and $\mathrm{I}[X : Y] = 0$ if and only if X and Y are statistically-independent.

2. Block Entropy

For a classical stochastic process we quantify the amount of uncertainty in a block of ℓ consecutive random variables by taking the joint entropy $\mathrm{H}[X_{0:\ell}]$. This is the *block entropy*:

$$\begin{aligned} \mathrm{H}[\ell] &\equiv \mathrm{H}[X_{0:\ell}] \\ &= -\sum_{x_{0:\ell}} \Pr(x_{0:\ell}) \log_2 \Pr(x_{0:\ell}) \ , \end{aligned}$$

where the sum is taken over all words of length ℓ and $\mathrm{H}[0] \equiv 0$.

$\mathrm{H}[\ell]$ is monotonically increasing and concave down, and its behavior as $\ell \to \infty$ is indicative of a process' correlations and randomness [11]. The generic behavior of $\mathrm{H}[\ell]$ and its relation to other information properties that we will define can be seen in Fig. 2.

3. Shannon Entropy Rate

The following briefly summarizes Ref. [11]'s results. Refer there for a more thorough exploration of information-theoretic quantities related to multivariate systems and the block entropy.

First among these is the *Shannon entropy rate* h_μ:

$$h_\mu = \lim_{\ell \to \infty} \frac{\mathrm{H}[\ell]}{\ell} \ . \tag{A1}$$

The limit in Eq. (A1) is guaranteed to exist for all stationary processes [70]. h_μ is irreducible randomness produced by an information source. Its units are *bits per symbol*.

The Shannon entropy rate can equivalently be written using the *conditional entropy*:

$$h_\mu = \lim_{\ell \to \infty} \mathrm{H}[X_0|X_{-\ell:0}] \ , \tag{A2}$$

Therefore, h_μ can equivalently be thought of as the average uncertainty in the next symbol if all preceding symbols are known.

To better appreciate h_μ we consider a few simple cases.

For an i.i.d. process the block entropy trivially is $\mathrm{H}[\ell] = \ell\,\mathrm{H}[X_0]$ and, therefore, $h_\mu = \mathrm{H}[X_0]$. Since there are no correlations between variables, knowledge of past symbols cannot reduce the uncertainty of the next.

If a process is periodic (for example consisting of alternating 0's and 1's), then a keen observer will note this pattern and be able to predict with certainty all future symbols of the process. In this case $h_\mu = 0$.

For stationary Markov and hidden Markov processes, an observer can leverage past observations to reduce their uncertainty about succeeding symbols. h_μ will be their average uncertainty about the next symbol once they have accounted for all of the correlations with past symbols.

Graphically, h_μ corresponds to the slope of the block entropy curve as $\ell \to \infty$ as shown in Fig. 2.

4. Redundancy

For a stochastic process with alphabet $\mathcal{X}$ the maximum entropy rate is $\log_2 |\mathcal{X}|$, corresponding to i.i.d. random variables X_t with uniform distributions over all measurement outcomes x_t. Any other stationary process can be *compressed* down to its entropy rate $h_\mu < \log_2 |\mathcal{X}|$.

The amount that a particular source can be compressed is known as its *redundancy*, defined as:

$$\mathbf{R} \equiv \log_2 |\mathcal{X}| - h_\mu \ .$$

R includes two very different effects: bias within individual random variables and correlations between different random variables. To determine the relative importance of those two factors requires closer examination of $\mathrm{H}[\ell]$.

5. Block Entropy Derivatives and Integrals

Since the limit in Eq. (A1) exists, $\mathrm{H}[\ell]$ scales (at most) linearly. We are interested in how $\mathrm{H}[\ell]$ converges to its linear asymptote, and we will see that taking discrete derivatives of $\mathrm{H}[\ell]$ (and integrals of those derivatives) provides us with useful quantities for classifying processes.

Consider applying a discrete derivative operator Δ to a function $F : \mathbb{Z} \to \mathbb{R}$:

$$\Delta F(\ell) = F(\ell) - F(\ell - 1) \ .$$

We can apply Δ to F multiple times to obtain higher-order derivatives:

$$\Delta^n F(\ell) = (\Delta \circ \Delta^{n-1}) F(\ell) \ .$$

Taking discrete derivatives of $\mathrm{H}[\ell]$ yields a set of functions $\Delta^n \mathrm{H}[\ell]$. We then study how these discrete derivatives themselves converge to their asymptotic values. To do so, we take "integrals" of a discrete function $\Delta F(\ell)$ in the following manner:

$$\sum_{\ell=A}^{B} \Delta F(\ell) = F(B) - F(A-1) \ . \tag{A3}$$

To study the convergence properties of each $\Delta^n \mathrm{H}[\ell]$, we compare it at each ℓ to its asymptotic value $\lim_{\ell\to\infty} \Delta^n \mathrm{H}[\ell]$. We do so with the following general integral form:

$$\mathcal{I}_n \equiv \sum_{\ell=\ell_0}^{\infty} [\Delta^n H(\ell) - \lim_{\ell\to\infty} \Delta^n \mathrm{H}[\ell]] \ , \tag{A4}$$

where ℓ_0 is the first value of ℓ for which $\Delta^n \mathrm{H}[\ell]$ is defined.

6. Entropy Gain

The first derivative of $\mathrm{H}[\ell]$ is known as the *entropy gain*. It is defined as:

$$\Delta \mathrm{H}[\ell] \equiv \mathrm{H}[\ell] - \mathrm{H}[\ell - 1] \ ,$$

for $\ell > 0$. We set $\Delta \mathrm{H}[0] \equiv \log_2(\mathcal{X})$. The entropy gain is the amount of additional uncertainty introduced by increasing the block size by one random variable, and its units are *bits per symbol*. Note that, because $\mathrm{H}[\ell]$ is monotone increasing and concave, $\Delta \mathrm{H}[\ell] \geq \Delta \mathrm{H}[\ell + 1] \geq 0$, for all ℓ. This behavior is shown in Fig. 3.

The entropy gain can be rewritten as a conditional entropy:

$$\Delta \mathrm{H}[\ell] = \mathrm{H}[X_0 | X_{-\ell:0}] \ , \tag{A5}$$

in which case its relation to the entropy rate becomes clear. Using Eq. (A2) we see that:

$$h_\mu = \lim_{\ell\to\infty} \Delta \mathrm{H}[\ell] \ .$$

For an observer with no prior knowledge of an information source, it will often be necessary to estimate the entropy rate of a source using finite sequences of data. In this case the entropy gain can serve as a finite-ℓ approximation of the true entropy rate:

$$\begin{aligned} h_\mu(\ell) &\equiv \Delta \mathrm{H}[\ell] \\ &\equiv \mathrm{H}[\ell] - \mathrm{H}[\ell - 1] \ . \end{aligned} \tag{A6}$$

7. Predictability Gain

The second derivative of $\mathrm{H}[\ell]$ is the *predictability gain*, defined as:

$$\begin{aligned}\Delta^2 \mathrm{H}[\ell] &\equiv \Delta h_\mu(\ell) \\ &= h_\mu(\ell) - h_\mu(\ell-1) \ ,\end{aligned}$$

where $\ell > 0$. Note that $\Delta^2 \mathrm{H}[\ell] \leq 0$.

$|\Delta^2 \mathrm{H}[\ell]|$ is the average amount of additional predictive information an observer obtains when expanding their observations from blocks of length $\ell-1$ to blocks of length ℓ, and its units are *bits per symbol*[2]. Large values of $|\Delta^2 \mathrm{H}[\ell]|$ imply that the ℓ-th measurement is particularly informative to an observer and therefore greatly improves their estimate of the entropy rate as given by Eq. (A6).

One can also calculate higher-order discrete derivatives of $\mathrm{H}[\ell]$. For our purposes this is not necessary except to note that, for stationary processes:

$$\lim_{\ell\to\infty} \Delta^n \mathrm{H}[\ell] = 0, n \geq 2 \ .$$

For $n = 2$ this follows from convergence of $h_\mu(\ell)$, and the argument for all $n > 2$ is similar.

8. Total Predictability

We can now integrate the functions $\Delta^n \mathrm{H}[\ell]$. As a general heuristic, the larger the magnitude of these integrals, the more correlation or statistical bias exists within the process.

We begin by studying how the predictability gain $\Delta^n \mathrm{H}[\ell]$ converges to its asymptotic value $\lim_{\ell\to\infty} \Delta^2 \mathrm{H}[\ell] = 0$. Since $\Delta^2 \mathrm{H}[0]$ is undefined, we will integrate using Eq. (A4) with $\ell_0 = 1$ to obtain the total predictability $\mathbf{G}$:

$$\mathbf{G} \equiv \mathcal{I}_2 = \sum_{\ell=1}^{\infty} \Delta^2 H(\ell) \ . \tag{A7}$$

Since $\Delta^2 \mathrm{H}[\ell] < 0$ for all ℓ, $\mathbf{G} < 0$ as well. The units of $\mathbf{G}$ are *bits per symbol*. Graphically, it is the the area between the predictability gain curve and its linear asymptote of 0, as seen in Fig. 4.

To interpret $\mathbf{G}$'s value, we apply Eq. (A3) to get:

$$\begin{aligned}\mathbf{G} &= -\Delta \mathrm{H}[0] + \lim_{\ell\to\infty} \Delta \mathrm{H}[\ell] \\ &= -\log_2(|\mathcal{X}|) + h_\mu \\ &= -\mathbf{R} \ .\end{aligned}$$

A process' total predictability is then equal in magnitude to its redundancy, and we can interpret $|\mathbf{G}|$ as the amount of predictable information per symbol for a process. Here, we emphasize once more that for a given process with alphabet $\mathcal{X}$, any random variable has a maximum entropy of $\log_2(|\mathcal{X}|)$, that consists of two kinds of information: h_μ, the irreducible randomness, and $|\mathbf{G}|$, the amount of information that an observer can possibly predict about it.

The total predictability is thus a function of the entropy rate for a given process and shares h_μ's weakness: it cannot identify statistical correlations between random variables. A large value of $|\mathbf{G}|$ could be the result of either an i.i.d. process with heavily-biased random variables or strong correlations between subsequent variables. Fortunately, our next quantity does distinguish between these cases.

9. Excess Entropy

Investigating the convergence of the entropy gain to the asymptotic entropy rate h_μ leads to a well-studied process property, the excess entropy:

$$\begin{aligned}\mathbf{E} \equiv \mathcal{I}_1 &= \sum_{\ell=1}^{\infty} \left[\Delta \mathrm{H}[\ell] - \lim_{\ell\to\infty} \Delta \mathrm{H}[\ell]\right] \\ &= \sum_{\ell=1}^{\infty} \left[\Delta \mathrm{H}[\ell] - h_\mu\right] \ .\end{aligned}$$

Since $\Delta \mathrm{H}[\ell] \geq h_\mu$, $\mathbf{E} \geq 0$ and its units are *bits*.

Our interpretation of $\mathbf{E}$ becomes clearer after applying Eq. (A3) to obtain:

$$\mathbf{E} = \lim_{\ell\to\infty} \left[\mathrm{H}[\ell] - h_\mu \ell\right] \ . \tag{A8}$$

While h_μ determines the linear asymptotic behavior of $\mathrm{H}[\ell]$, $\mathbf{E}$ encapsulates all sublinear effects. We largely discuss processes for which $\mathbf{E}$ is finite, known as *finitary* processes. A process for which $\mathbf{E}$ is infinite (for example if $\mathrm{H}[\ell]$ scales logarithmically with ℓ), is known as an *infinitary* process. Graphically, $\mathbf{E}$ corresponds to the $\ell = 0$ intercept of the linear asymptote to the block entropy curve, as shown in Fig. 2, as well as the area between the area between the entropy gain curve and its asymptote h_μ as seen in Fig. 3.

The excess entropy can also be written as a mutual information between two halves of a process' chain of random variables:

$$\mathbf{E} = \lim_{\ell\to\infty} \mathrm{I}\left[X_{-\ell:0} : X_{0:\ell}\right] \ . \tag{A9}$$

This suggests it is the total amount of information in a process' past useful for predicting the future. It is

therefore considered an indicator of the amount of process *memory*. A more structural approach reveals that $\mathbf{E}$ is a lower bound on the actual amount of memory required to predict a stochastic process [65].

Importantly, $\mathbf{E}$ easily distinguishes between i.i.d. processes (for which $\mathbf{E} = 0$) and processes with correlations between random variables ($\mathbf{E} > 0$). For all processes that are periodic with period p, $h_\mu = 0$ and $\mathrm{H}[\ell]$ reaches a maximum value of $\log_2 p$ for $\ell = p$. Therefore, all period-p processes have $\mathbf{E} = \log_2 p$.

As with h_μ, it is often useful for an observer to make an approximation of the true excess entropy using only finite length-ℓ symbol sequences. Using Eq. (A8) we can estimate $\mathbf{E}$ as:

$$\mathbf{E}(\ell) \equiv \mathrm{H}[\ell] - \ell h_\mu(\ell) \ . \tag{A10}$$

10. Transient Information

Based upon our analysis of the excess entropy, we can now say that as $\ell \to \infty$ the block entropy curve has a linear asymptote:

$$\mathrm{H}[\ell] \sim \mathbf{E} + h_\mu \ell \ . \tag{A11}$$

We capture the way that $\mathrm{H}[\ell]$ converges to this asymptote by taking another discrete integral to obtain the transient information:

$$\mathbf{T} \equiv -\mathcal{I}_0 = \sum_{\ell=0}^{\infty} \left[\mathbf{E} + h_\mu \ell - \mathrm{H}[\ell]\right] \ .$$

The units of $\mathbf{T}$ are *bits*×*symbol*. $\mathbf{T}$ is represented graphically in Fig. 2 as the area between the $\mathrm{H}[\ell]$ curve and its linear asymptote for $\ell \to \infty$.

The transient information can be rewritten as:

$$\mathbf{T} = \sum_{\ell=1}^{\infty} \ell \left[h_\mu(\ell) - h_\mu\right] \ ,$$

indicating that $\mathbf{T}$ is a measure of how difficult it is to *synchronize* to a process. An observer is considered "synchronized" when they are able to infer exactly which internal state the source currently occupies. If they remain synchronized, then they can optimally predict future measurement outcomes. That is, their estimate of the source's entropy rate $h_\mu(\ell)$ is equal to its actual entropy rate h_μ.

$\mathbf{T}$ is a notable quantity since, unlike $\mathbf{E}$, it is capable of distinguishing between different period-p processes.

11. Markov Order

The final property we introduce for classical stochastic processes is the Markov order. A process has Markov order R if R is the minimum value for which the following condition holds:

$$\Pr(X_0 | X_{R:0}) = \Pr(X_0 | X_{-\infty:0}) \ .$$

From Eq. (2) we conclude that a Markov process is one for which $R \leq 1$. More generally, for a process with Markov order R the distribution for X_t depends only on the previous R symbols. Equivalently, R is the value of ℓ for which the block entropy reaches its linear asymptote:

$$\mathrm{H}[R] = \mathbf{E} + h_\mu R \ .$$

This fact is represented graphically in Fig. 2.

The overwhelming majority of finite-state hidden Markov processes have infinite Markov order, meaning that $\mathrm{H}[\ell] < \mathbf{E} + h_\mu \ell$ for all ℓ [71]. Nevertheless, an observer can still make use of finite-ℓ estimates of the information properties defined here. In fact, these estimates typically converge exponentially fast with ℓ [72].

Appendix B: Quantum Channels for Preparation and Measurement

The main text claims that a separable qudit process is the result of passing realizations of a classical process $\overleftrightarrow{X}$ with symbol alphabet $\mathcal{X}$ through a classical-quantum channel that takes $x \in \mathcal{X} \to |\psi_x\rangle \in \mathcal{H}$. It also claims that the act of measurement can be described as a quantum-classical channel taking $\rho_{0:\ell} \to y_{0:\ell}$. The following elaborates on both claims, adopting the formalism of Ref. [6].

We first describe passing realizations of a classical process $\overleftrightarrow{X}$ through a *conditional quantum encoder*. Consider a classical-quantum state composed of a classical register of dimension $|\mathcal{X}|$ and a qudit in $\mathcal{H}$ initialized in the state $|0\rangle$. Furthermore, let the classical register store the outcome of some classical random variable X_t with outcomes $x \in \mathcal{X}$ such that:

$$\rho_{XA} = \sum_x \Pr(x) \left|x\right\rangle \left\langle x\right|_X \otimes \left|0\right\rangle \left\langle 0\right|_A \ . \tag{B1}$$

This state will serve as input to a conditional quantum encoder $\mathcal{E}_{XA\to B}$ that consists of a set $\{\mathcal{E}^x_{A\to B}\}$ of $|\mathcal{X}|$ completely-positive trace-preserving (CPTP) maps. We construct each map such that it transforms the initial quantum state $|0\rangle$ to the desired pure qudit state $|\psi_x\rangle$;

i.e.:

$$\mathcal{E}^x_{A\to B}(|0\rangle\langle 0|_A) = |\psi_x\rangle\langle\psi_x|_B \ , \tag{B2}$$

with $\mathcal{E}^x_{A\to B}$ being a unitary operation. (For $d = 2$ we can consider rotations on the Bloch sphere.)

The encoding is:

$$\begin{aligned}\rho_B &= \mathcal{E}_{XA\to B}(\rho_{XA}) \\ &= tr_X\left(\sum_x \Pr(x)\,|x\rangle\langle x| \otimes \mathcal{E}^x_{A\to B}(|0\rangle\langle 0|)\right) \\ &= \sum_x \Pr(x)\,|\psi_x\rangle\langle\psi_x|_B \ .\end{aligned}$$

Now, taking ℓ classical registers that consist of length-ℓ words $x_{0:\ell}$ of a classical process $\overleftrightarrow{X}$, we can use $\mathcal{E}_{XA\to B}$ to encode $X_{0:\ell}$ and obtain:

$$\begin{aligned}\rho_{0:\ell} &= tr_{X_{0:\ell}}\left(\sum_{x_{0:\ell}} \Pr(x_{0:\ell}) \prod_{t=0}^{\ell-1} |x_t\rangle\langle x_t| \otimes \mathcal{E}^{x_t}_{A\to B}(|0\rangle\langle 0|)\right) \\ &= \sum_w \Pr(w)\,|\psi_w\rangle\langle\psi_w| \ ,\end{aligned}$$

where $|\psi_w\rangle$ take the separable form of Eq. (5) and $\rho_{0:\ell}$ therefore matches Eq. (6).

Measurement of qudits is described similarly. Let M be a POVM with elements $\{E_y\}$ that acts on a qudit state ρ in such a way that it records each measurement outcome in a classical register Y. The distribution over values in Y is determined by:

$$\begin{aligned}Y &= M(\rho) \\ &= \sum_y tr(E_y\rho)\,|y\rangle\langle y| \ .\end{aligned}$$

Likewise consider $\mathcal{M}_{0:\ell}$ to be a length-ℓ sequence of measurements with possible measurement outcome sequences $y_{0:\ell}$ determined according to some protocol $\mathcal{M}$. When applying $\mathcal{M}_{0:\ell}$ to ℓ consecutive qudits in the joint state $\rho_{0:\ell}$ we assume $\mathcal{M}_{0:\ell}$ consists of local measurements in the form of Eq. (10). In this case we factor the POVM elements corresponding to particular sequences of measurement outcomes: $E_{y_{0:\ell}} = \bigotimes_{t=0}^{\ell-1} E_{y_t}$. A length-$\ell$ measurement outcome can be stored in $Y_{0:\ell}$, a set of ℓ classical registers, with its associated probability $tr(E_{y_{0:\ell}}\rho_{0:\ell})$ so that:

$$\begin{aligned}Y_{0:\ell} &= \mathcal{M}_{0:\ell}(\rho_{0:\ell}) \\ &= \sum_{y_{0:\ell}} tr(E_{y_{0:\ell}}\rho_{0:\ell})\,|y_{0:\ell}\rangle\langle y_{0:\ell}| \\ &= \sum_{y_{0:\ell}} tr(\bigotimes_{t=0}^{\ell-1} E_{y_t}\rho_{0:\ell})\,|y_{0:\ell}\rangle\langle y_{0:\ell}| \ ,\end{aligned}$$

where the last line assumes local measurements. Each $|y_{0:\ell}\rangle$ is then a separable state and all $|y_{0:\ell}\rangle$ are orthogonal.

[1] W. Heisenberg. Über den anschaulichen inhalt der quantentheoretischen kinematik und mechanik. *Z. Phys.*, 43(3-4):172–198, March 1927. 2

[2] A. Peres. Two simple proofs of the Kochen-Specker theorem. *J. Phys. A*, 24(4), 1991. 2

[3] W. K. Wootters and B. D. Fields. Optimal state-determination by mutually unbiased measurements. *Ann. Phys. (N.Y.)*, 191(2):363–381, 1989. 2, 32

[4] J. M. Renes, R. Blume-Kohout, A. J. Scott, and C. M. Caves. Symmetric informationally complete quantum measurements. *J. Math. Phys.*, 45(6):2171–2180, 2004. 2, 32

[5] J. von Neumann. *Mathematical foundations of quantum mechanics: New edition.* Princeton university press, 2018. 2, 7

[6] M. Wilde. *Quantum Information Theory.* Cambridge University Press, 2nd edition, Dec 2017. 2, 9, 10, 33, 43

[7] C. E. Shannon. A mathematical theory of communication. *Bell Sys. Tech. J.*, 27:379–423, 623–656, 1948. 2, 22, 40

[8] B. Schumacher. Quantum coding. *Phys. Rev. A*, 51(4):2738–2747, 1995. 2

[9] J. P. Crutchfield. Between order and chaos. *Nature Physics*, 8:17–24, 2012. 3

[10] J. P. Crutchfield and K. Young. Inferring statistical complexity. *Phys. Rev. Let.*, 63:105–108, 1989. 3

[11] J. P. Crutchfield and D. P. Feldman. Regularities unseen, randomness observed: Levels of entropy convergence. *Chaos*, 13(1):25–54, 2003. 3, 14, 18, 20, 26, 40

[12] N. F. Travers and J. P. Crutchfield. Exact synchronization for finite-state sources. *J. Stat. Phys.*, 145:1181–1201, 2011. 3, 24, 26

[13] N. F. Travers and J. P. Crutchfield. Asymptotic synchronization for finite-state sources. *J. Stat. Phys.*, 145:1202–1223, 2011. 3

[14] A. E. Venegas-Li, A. M. Jurgens, and J. P. Crutchfield. Measurement-induced randomness and structure in controlled qubit processes. *Phys. Rev. E*, 102:040102, Oct 2020. 3, 7, 21, 25

[15] A. S. Holevo. Quantum coding theorems. *Usp. Mat. Nauk*, 53(6):193–230, 1998. 3

[16] N. Datta and Y. Suhov. Data Compression Limit for an Information Source of Interacting Qubits. *Quantum Inf. Process.*, 1(4):257–281, 2002.

[17] D. Petz and M. Mosonyi. Stationary quantum source

coding. *J. Math. Phys.*, 42(10):4857–4864, 2001. 3, 6, 12

[18] Y. Nagamatsu, A. Mizutani, R. Ikuta, T. Yamamoto, N. Imoto, and K. Tamaki. Security of quantum key distribution with light sources that are not independently and identically distributed. *Phys. Rev. A*, 93(4):042325, 2016. 3

[19] F. A. Pollock, C. Rodriguez-Rosario, T. Frauenheim, M. Paternostro, and K. Modi. Operational Markov condition for quantum processes. *Phys. Rev. Lett.*, 120:040405, 2018. 3

[20] F. A. Pollock, C. Rodriguez-Rosario, T. Frauenheim, M. Paternostro, and K. Modi. Non-markovian quantum processes: Complete framework and efficient characterization. *Phys. Rev. A*, 97:012127, 2018.

[21] P. Taranto, F. A. Pollock, S. Milz, M. Tomamichel, and K. Modi. Quantum markov order. *Phys. Rev. Lett.*, 122:14041, 2019.

[22] P. Taranto, F. A. Pollock, and K. Modi K. Non-Markovian memory strength bounds quantum process recoverability. *npj Quantum Inf.*, 7(1):1–8, 2021.

[23] P. Taranto, S. Milz, F. A. Pollock, and K. Modi. Structure of quantum stochastic processes with finite Markov order. *Phys. Rev. A*, 99:042108, Apr 2019. 3

[24] C. Schon, E. Solano, F. Verstraete, J. I. Cirac, and M. M. Wolf. Sequential generation of entangled multiqubit states. *Phys. Rev. Lett.*, 95:110503, 2005. 3, 39

[25] C. Schon, K. Hammerer, M. M. Wolf, J. O. Cirac, and E. Solano. Sequential generation of matrix-product states in cavity QED. *Phys. Rev. A*, 55(3):032311, 2007. 3

[26] A. B. Boyd, D. Mandal, and J. P. Crutchfield. Correlation-powered information engines and the thermodynamics of self-correction. *Phys. Rev. E*, 95(1):012152, 2017. 3, 39

[27] A. Chapman and A. Miyake. How an autonomous quantum Maxwell demon can harness correlated information. *Phys. Rev. E*, 92(6):1–12, 2015. 3

[28] M. Gu, K. Wiesner, E. Rieper, and V. Vedral. Occam's quantum razor: How quantum mechanics can reduce the complexity of classical models. *Nat. Commun.*, 3, 2012. 5

[29] 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.

[30] W. Y. Suen, T. J. Elliot, J. Thompson, A. J. Garner, J. R. Mahoney, V. Vedral, and M. Gu. Surveying structural complexity in quantum many-body systems. *J. Stat. Phys.*, 187, 04 2022.

[31] F. C. Binder, J. Thompson, and M. Gu. Practical unitary simulator for non-Markovian complex processes. *Phys. Rev. Lett.*, 120:240502, Jun 2018.

[32] S. P. Loomis and J. P. Crutchfield. Strong and Weak Optimizations in Classical and Quantum Models of Stochastic Processes. *J. Stat. Phys.*, 176(6):1317–1342, 2019. 5

[33] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, USA, 10th edition, 2011. 6, 9, 10, 11, 32, 33

[34] S. Gudder. Quantum Markov chains. *J. Math. Phys.*, 49(7), 2008. 7

[35] A. Monras, A. Beige, and K. Wiesner. Hidden Quantum Markov Models and non-adaptive read-out of many-body states. 2010. 7

[36] K. Wiesner and J. P. Crutchfield. Computation in finitary stochastic and quantum processes. *Physica D*, 237(9):1173–1195, 2008.

[37] S. Srinivasan, G. Gordon, and B. Boots. Learning hidden quantum Markov models. *AISTATS 2018*, pages 1979–1987, 2018. 7

[38] D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac. Matrix product state representations. *arXiv preprint quant-ph/0608197*, 2006. 7

[39] F. Verstraete, J. J. García-Ripoll, and J. I. Cirac. Matrix product density operators: Simulation of finite-temperature and dissipative systems. *Phys. Rev. Lett.*, 93(20):1–5, 2004. 7

[40] C. Moore and J. P. Crutchfield. Quantum automata and quantum grammars. *Theoret. Comp. Sci.*, 237:1-2:275–306, 2000. 8

[41] D. Qiu, L. Li, P. Mateus, and J. Gruska. Quantum Finite Automata. In *Handbook of Finite State Based Models and Applications*, pages 113–144. 10 2012.

[42] S. Zheng, D. Qiu, L. Li, and J. Gruska. One-way finite automata with quantum and classical states. In *Lecture Notes in Computer Science*, volume 7300 LNAI, pages 273–290. 2012. 8

[43] M. Junge, R. Renner, D. Sutter, M. M. Wilde, and A. Winter. Universal recovery maps and approximate sufficiency of quantum relative entropy. *Ann. Henri Poincaré*, 19(10):2955–2978, 2018. 10

[44] O. E. Lanford and D. W. Robinson. Mean entropy of states in quantum-statistical mechanics. *J. Math. Phys.*, 9(7):1120–1125, 1968. 12

[45] M. Ohya and D. Petz. *Quantum Entropy and Its Use*. 1993. 12

[46] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience, New York, 1991. 12

[47] N. F. Travers and J. P. Crutchfield. Equivalence of history and generator ϵ-machines. arxiv.org:1111.4500 [math.PR]. 24, 26

[48] C. J. Ellison, J. R. Mahoney, and J. P. Crutchfield. Prediction, retrodiction, and the amount of information stored in the present. *J. Stat. Phys.*, 136(6):1005–1034, 2009. 25

[49] A. M. Jurgens and J. P. Crutchfield. Shannon entropy rate of hidden Markov processes. *Journal of Statistical Physics*, 183(2):1–18, 2021. 25

[50] J. P. Crutchfield, P. Riechers, and C. J. Ellison. Exact complexity: Spectral decomposition of intrinsic computation. *Phys. Lett. A*, 380(9-10):998–1002, 2016. 25, 39

[51] S. Marzen and J. P. Crutchfield. Informational and causal architecture of discrete-time renewal processes. *Entropy*, 17(7):4891–4917, 2015. 28

[52] A. Jurgens and J. P. Crutchfield. Divergent predictive states: The statistical complexity dimension of stationary, ergodic hidden Markov processes. *Chaos*, 31(8):0050460, 2021. 29

[53] Y. Fujiwara. Parsing a sequence of qubits. *IEEE Trans. Inf. Theory*, 59(10):6796–6806, 2013. 30

[54] C. C. Strelioff and J. P. Crutchfield. Bayesian structural inference for hidden processes. *Physical Review E*, 89(4):042119, 2014. 32

[55] Rob T Thew, Kae Nemoto, Andrew G White, and William J Munro. Qudit quantum-state tomography. *Phys. Rev. A*, 66(1):012303, 2002. 32

[56] I. Bengtsson and K. Życzkowski. *Geometry of quantum states: An introduction to quantum entanglement*. Cambridge University Press, 2017. 33

[57] J. Lawrence, Č. Brukner, and A. Zeilinger. Mutually unbiased binary observable sets on n qubits. *Phys. Rev. A*, 65:032320, Feb 2002. 33

[58] Sevag Gharibian. Strong np-hardness of the quantum separability problem. *arXiv preprint arXiv:0810.4507*, 2008. 38

[59] Pawel Horodecki. Separability criterion and inseparable mixed states with positive partial transposition. *Phys. Lett. A*, 232(5):333–339, 1997. 38

[60] M. Horodecki, P. Horodecki, and R. Horodecki. Separability of n-particle mixed states: necessary and sufficient conditions in terms of linear maps. *Phys, Letters A*, 283(1-2):1–7, 2001. 38

[61] B. Pirvu, V. Murg, J. I. Cirac, and F. Verstraete. Matrix product operator representations. *New J. Phys.*, 12(2):025012, 2010. 39

[62] P. Riechers and J. P. Crutchfield. Spectral simplicity of apparent complexity, Part I: The nondiagonalizable metadynamics of prediction. *Chaos*, 28:033115, 2018. 39

[63] P. M. Riechers and J. P. Crutchfield. Beyond the spectral theorem: Decomposing arbitrary functions of nondiagonalizable operators. *AIP Advances*, 8:065305, 2018.

[64] P. Riechers and J. P. Crutchfield. Spectral simplicity of apparent complexity, Part II: Exact complexities and complexity spectra. *Chaos*, 28:033116, 2018. 39

[65] C. R. Shalizi and J. P. Crutchfield. Computational mechanics: Pattern and prediction, structure and simplicity. *J. Stat. Phys.*, 104(3-4):817–879, 2001. 39, 43

[66] R. Landauer. Irreversibility and heat generation in the computing process. *IBM J. Res. Develop.*, 5(3):183–191, 1961. 39

[67] A. B. Boyd, D. Mandal, and J. P. Crutchfield. Leveraging environmental correlations: The thermodynamics of requisite variety. *J. Stat. Phys.*, 167(6):1555–1585, 2016. 39

[68] A. M. Jurgens and J. P. Crutchfield. Functional thermodynamics of Maxwellian ratchets: Constructing and deconstructing patterns, randomizing and derandomizing behaviors. *Phys. Rev. Res.*, 2(3):033334, 2020. 39

[69] R. Huang, P. Riechers, M. Gu, and V. Narasimhachar. Engines for predictive work extraction from memoryfull quantum stochastic processes. 07 2022. 39

[70] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience, New York, second edition, 2006. 40

[71] R. G. James, J. R. Mahoney, C. J. Ellison, and J. P. Crutchfield. Many roads to synchrony: Natural time scales and their algorithms. *Physical Review E*, 89:042135, 2014. 43

[72] N. F. Travers. Exponential bounds for convergence of entropy rate approximations in hidden Markov models satisfying a path-mergeability condition. *Stochastic Proc. Appln.*, 124(12):4149–4170, 2014. 43