A Closed-Form Shave from Occam's Quantum Razor: Exact Results for Quantum Compression

Paul M. Riechers
John R. Mahoney
Cina Aghamohammadi
James P. Crutchfield

SFI WORKING PAPER: 2015-10-041

SANTA FE INSTITUTE

Santa Fe Institute Working Paper 15-07-XXX
arxiv.org:1507.XXXX [physics.gen-ph]

A Closed-Form Shave from Occam's Quantum Razor

Paul M. Riechers,* John R. Mahoney,† Cina Aghamohammadi,‡ and James P. Crutchfield§
Complexity Sciences Center and Department of Physics
University of California at Davis
One Shields Avenue, Davis, CA 95616
(Dated: October 27, 2015)

Recently it was shown that the causal organization of a classical stochastic process can be substantially compressed using the so-called *q-machine* representation. The overlaps between quantum signal states play an important role in this compression. At longer length scales, where more of the process's structure is accounted for, these overlaps become more important and, as it would seem, more difficult to compute. Here we derive useful expressions for these overlaps, including one based on a spectral decomposition, affording us theoretical simplicity as well as greatly improved computational ability. These expressions are based on a new quantum transient structure defined here, the quantum-pairwise-merger-machine (QPMM). Armed with the overlaps, we ultimately proceed to compute the quantum communication cost, defined by the ensemble's von Neumann entropy. We simplify matters further by making use of a surrogate Gram matrix. We also provide more explicit proofs regarding the connection between the cryptic order of the stochastic process and the optimal coding length of the q-machine.

PACS numbers: 03.67.Hk 03.67.-a 03.65.-w 03.65.Db
Keywords: epsilon-machine, information compression, quantum state overlap, crypticity, spectral decomposition, holomorphic functional calculus

CONTENTS

* pmriechers@ucdavis.edu
† jrmahoney@ucdavis.edu
‡ caghamohammadi@ucdavis.edu
§ chaos@ucdavis.edu

I. INTRODUCTION, STATEMENT OF RESULTS

The definition of and quantification of "pattern" has been an active area of research for decades [1–4]. These studies seek to answer questions like, "how unpredictable is this pattern?", and "how much memory is needed to accurately predict the next element?" Computational mechanics [5, 6], an extension of statistical mechanics, is a framework established to answer these very questions. In much of computational mechanics, the emphasis is on stationary stochastic processes. These are also the patterns of interest in the present work.

An important question about a process's structure concerns two observers, Alice and Bob, who wish to synchronize their predictions. How much information must be communicated? The answer is given by the statistical complexity of the process C_μ [7]. One could ask, is it more efficient to do this via a quantum channel? Extending early answers [8, 9], we recently proposed a sequence of constructions (the q-machine) [10] that offer substantial quantum compression. Each word length yields a quantum communication rate $C_q(L)$. It was demonstrated that maximum compression $C_q(\infty) = C_q(k)$ is achieved at a length-scale called the *cryptic order* k [11]—a recently discovered classical, topological property that

is a cousin to the more familiar Markov order.

This article provides the analytical underpinnings for the recent construction and computation of the quantum communication cost $C_q(L)$ [10]. In particular, we present a closed-form expression for the quantum overlaps between quantum signal states, overlaps that result in compressibility. The overlap expression is in terms of the spectrum (and related projection operators) of the Quantum Pairwise Merger Machine, introduced here. While the size of the Hilbert space containing the signal states increases exponentially with word length, these overlaps allow us to construct an equivalent density matrix in a fixed Hilbert space, leading to a significant computational advantage. **The fixed Hilbert space size was already demonstrated. The first big win here is with not having to compute word probabilities for all words of length L. The second medium win is realizing that for finite cryptic order, we can truncate our sum. The third big win is the spectral form which shows how the overlaps come from a finite part and an infinite part. The infinite part is easy to deal with, and the finite part is only as difficult as the zero-index.** We demonstrate that an even greater advantage is obtained using a surrogate matrix that has the same spectrum as the density matrix—the M-by-M Gram matrix. This surrogate matrix is linear in the overlaps, and is straightforward to calculate. The quantum communication cost $C_q(L)$ (and $C_q(\infty)$) is thus calculable with a new level of numerical efficiency, and in some cases analytically.

To illustrate the power of our method, we describe the long-length behavior of the q-machine state entropy for any infinite-cryptic-order process. **I know we are trying to talk about an "application", but this sentence seems disconnected.**

II. TWO REPRESENTATIONS OF A PROCESS

A. ϵ-machine

Consider processes generated by finite hidden Markov models (HMMs). **Do we need to define processes?** While a given process may have many alternative HMMs, there is a unique, canonical presentation—the process's *ϵ-machine*. **Think about presentation vs representation?** The recurrent states $\boldsymbol{\mathcal{S}}$ of a process's ϵ-machine are known as the *causal states* $\sigma \in \boldsymbol{\mathcal{S}}$. The causal states are the minimal sufficient statistic of the past $X_{:0}$ for predicting the future $X_{0:}$.

An ϵ-machine is a type of HMM satisfying three conditions: unifilarity, probabilistically distinct states, and irreducibility. **Don't need prob distinct and irred.**

An ϵ-machine generates output as it transitions from state to state. Beginning in some state σ_i, $\Pr(\sigma_j, x|\sigma_i)$ gives the probability of transitioning to state σ_j and generating the symbol x. A string of output $x_0 x_1 \ldots x_t \ldots$ is simply generated by a sequence of such transitions. These conditional probabilities are contained in the output-labeled transition matrices,

$$T^{(x)}_{ij} = \Pr(\sigma_j, x|\sigma_i) \in [0, 1] \ ,$$

for $i, j = 1, \ldots, M$ where $M = |\boldsymbol{\mathcal{S}}|$.

The flow of state probability is given by the sum over all output symbols,

$$T := \sum_{x \in \mathcal{A}} T^{(x)} \ , \tag{1}$$

The transition probabilities are normalized. That is, the transition matrix T is *row-stochastic*:

$$\sum_{j=1}^{M} T_{i,j} = \sum_{j=1}^{M} \sum_{x \in \mathcal{A}} \Pr(\sigma_j, x|\sigma_i) = 1 \ .$$

Its component matrices $T^{(x)}_{ij}$ are said to be *substochastic*.

Unifilarity means that for each state σ_i, each symbol x may lead to at most one successor state σ_j. In terms of the labeled transition matrices, for each row i, and each symbol x, the row $T^{(x)}_{ij}$ has at most one nonzero entry.

By way of summarizing, we have the main object that generates a process.

Definition 1. *The* ϵ-machine $\mathcal{M}$ *is the set* $\{\boldsymbol{\mathcal{S}}, \{T^{(x)}\}_{x \in \mathcal{A}}, \boldsymbol{\pi}\}$.

$\boldsymbol{\mathcal{S}}$ is the set of causal states, $\mathcal{A}$ is the set of possible symbol outputs, $\{T^{(x)}\}_{x \in \mathcal{A}}$ are the labeled transition matrices and $\boldsymbol{\pi}$ the stationary distribution over states. **We can have $\boldsymbol{\pi}$ as a row vector with elements π_i.** The probability that a word $w = x_0, \ldots, x_{L-1}$ is generated by the ϵ-machine is given by $\Pr(w) = \boldsymbol{\pi} \prod_{i=0}^{L-1} T^{(x_i)} \mathbb{1}$ **where $\mathbb{1} = [1, \ldots, 1]^T$.** Since these probabilities are constructed to agree with those in a process language, the ϵ-machine is said to *generate* or *represent* that process.

One of the most important complexity measures for a process is its statistical complexity.

Definition 2. *The* statistical complexity C_μ *of an ϵ-machine $\mathcal{M}$ is the Shannon entropy of the stationary distribution over its causal states.* $C_\mu = H[\boldsymbol{\pi}] = -\sum \pi_i \log(\pi_i)$.

The statistical complexity has several operational meanings. In particular, it quantifies the communication cost of synchronizing two predicting agents through a classical channel.

B. q-machine

The q-machine is a quantum representation of a classical process. Introduced in [10], this representation is of interest because it offers the most complete quantum compression of this classical process known so far.

The length-L q-machine consists of a set $\{|\eta_i(L)\rangle\}_{i=1}^{M}$ of pure *signal states* which are in one-to-one correspondence with the classical causal states $\sigma_i \in \boldsymbol{\mathcal{S}}$. Each signal state $|\eta_i(L)\rangle$ encodes the set of length-L words that may follow causal state σ_i, as well as each corresponding conditional probability.

We recall the definition of the signal states:

$$|\eta_i(L)\rangle \equiv \sum_{w\in\mathcal{A}^L} \sum_{\sigma_j\in\boldsymbol{\mathcal{S}}} \sqrt{\Pr(w,\sigma_j|\sigma_i)}\ |w\rangle\,|\sigma_j\rangle\ , \quad (2)$$

where $\left\{\,|w\rangle\,\right\}_{w\in\mathcal{A}^L}$ denotes an orthonormal basis in the "word" Hilbert space with one dimension for each possible word of length L. Similarly, $\left\{\,|\sigma_j\rangle\,\right\}_{j=1}^{M}$ denotes an orthonormal basis in the "state" Hilbert space with one dimension for each classical causal state.

The ensemble of length-L quantum signal states is then described by the density matrix,

$$\rho(L) = \sum_i \pi_i\,|\eta_i(L)\rangle\,\langle\eta_i(L)|\ . \quad (3)$$

The von Neumann entropy of an ensemble is defined in terms of its density matrix, $S(\rho) = -\mathrm{tr}(\rho\log(\rho))$ where $\mathrm{tr}(\cdot)$ is the trace of its argument.

In analogy with the statistical complexity, the quantity

$$C_q(L) \equiv S(\rho(L)) = -\mathrm{tr}\big(\rho(L)\log(\rho(L))\big)$$

has the operational meaning of the communication cost of synchronization for predictors of a stationary stochastic process when communicating via a quantum channel using these quantum signal states.

III. QUANTUM OVERLAPS AND CRYPTIC ORDER

Somewhere in here need to show why overlaps have anything to do with merging. This fact is dependent on the form of our signals states.

Computation of the quantum overlaps as a function of L by *direct* use of the signal states $\{\eta_j(L)\}$ suffers from an exponential increase in the size of the Hilbert space. This leads to both theoretical and numerical difficulties. We can do much better by making use of the structure of the process.

Since $\langle w|w'\rangle = \delta_{w,w'}$ and $\langle\sigma|\sigma'\rangle = \delta_{\sigma,\sigma'}$, the quantum overlap $\langle\eta_i(L)|\eta_j(L)\rangle$ only gathers contributions from shared output-histories of the future final causal state. These contributions can be understood, in part, in terms of non-counifilarity. However non-counifilarity does not reveal the entire story. We learn more from considering the cryptic order, finding that quantum overlaps stabilize at precisely this length-scale.

The previous introduction of the q-machine showed numerically that its quantum compression is improved with increasing word length, and explained this through discussion of path merging in some examples. The present article provides more detailed proofs of some of the previous ideas. Here we prove some statements about the behavior of overlaps and $C_q(L)$ and their relation to the cryptic order. Recall the definitions of the Markov and cryptic orders. **We can also define with probs. And this is more typical for Markov..**

Definition 3. *The* Markov order R *of a process is the smallest length* L *for which* $\mathrm{H}[\mathcal{S}_L|X_{0:L}] = 0$.

(Note that we use Python-style indexing: left-inclusive and right-exclusive. Also, for the relative temporal ordering of hidden states and emitted observables, note that $\mathcal{S}_t$ immediately precedes X_t; i.e., X_t takes $\mathcal{S}_t$ to $\mathcal{S}_{t+1}$.)

The Markov order is one standard characterization of a process's historical dependence. It can be understood as the minimum number of symbols observed that guarantees synchronization to the state. Beginning with the naive distribution over causal states (stationary distribution), observation of symbols leads to less state uncertainty (on average). A particular next observation can lead to "termination" of a candidate state path, or to the "merging" of two or more candidate paths. The Markov order is the length at which all such terminations and mergings have occurred.

Definition 4. *The* cryptic order k *of a process is the minimum length* L *for which* $\mathrm{H}[\mathcal{S}_L|X_{0:}] = 0$.

(Note that $X_{0:}$, the "future" denotes the semi-infinite sequence of random variables associated with process observables beginning at $t = 0$.)

Cryptic order, on the other hand, effectively ignores the termination events, and is therefore upper-bounded by the Markov order $k \leq R$. This bound is also easy to see given the extra condition on $X_{L:\infty}$ [12, 13].

Definition 5. *An* L-merge *consists of a word of length* L*, and two state paths each of length* $L+1$ *that each generate the word* w *and merge exactly at length* $L+1$*. We denote the word* $w = (x_0,\ldots,x_{L-1})$ *and state paths* $(a_0,\ldots,a_{L-1},F)$ *and* $(b_0,\ldots,b_{L-1},F)$ *where* $a_i \neq b_i, \forall i \in [0,\,L-1]$ *and, trivially,* $F = F$.

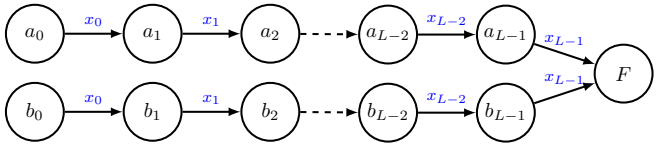

FIG. 1. L-merge. Two paths generate the same word and merge states only on the last output symbol.

The following lemma states an intuitive relation between cryptic order and L-merges.

Lemma 1. *Given an ϵ-machine with cryptic order k: for $L \leq k$, there exists an $L-$merge; for $L > k$, there exists no $L-$merge.*

Proof. *By definition of cryptic order, for any $x_{0:}$, $\mathcal{S}_k$ is unique. Also, there must exist some $x_{0:}$ such that the induced $\mathcal{S}_{k-1}$ is not unique (or else k would not be the minimal length). Call (two of) these distinct states $\mathcal{S}_a$ and $\mathcal{S}_b$. There must exist two state paths SA and SB that lead to $\mathcal{S}_a$ and $\mathcal{S}_b$ on the word $x0 : k-2$. Any two such state paths must remain distinct until length k; unifilarity of ϵ-machine does not allow for path splitting (on the same symbol). Therefore SA and SB combined with $x0 : k-1$ and final state $\mathcal{S}_k$ make up a k-merge.*

Therefore, cryptic order k implies existence of k-merge.

Also, by simply removing states on the left of an L-merge, we see that existence of an L-merge implies the existence of an $(L-1)$-merge.

Therefore, cryptic order k implies existence of L-merge for $L \leq k$.

For the second part of the proof, assume there exists an L-merge for $L > k$.

The final state σ_L leads to a cone of futures x_L : with non-zero measure. Append this cone of futures to the L-merge word $x0 : L-1$. For any of these words $x_{0:}$ there there are then consistent state paths such that the penultimate state $\mathcal{S}_{L-1}$ may be $aL-1$ or $bL-1$.

Therefore, by definition the cryptic order must be at least L, $k \geq L$. This is a contradiction and completes the proof.

Note that a Markov order R process need not have an R-merge, or any non-zero-merge.

Proof. *By definition of cryptic order k,*

$$\mathrm{H}[\mathcal{S}_k|X_{0:}] = 0 \ ,$$

which means that for any given $x_{0:}$ there exist a unique σ_k. For any $L > k$, that instance $x_{0:}$ also induces a unique σ_L (by unifilarity of the ϵ-machine), so

$$\mathrm{H}[\mathcal{S}_L|X_{0:}] = 0 \qquad \text{for } L \geq k \ .$$

But, if there exists an L-merge then there exists some realization of the output $x_{0:}$ for which $\mathrm{H}[\mathcal{S}_{L-1}|x_{0:}] > 0$ because there exist at least two $\mathcal{S}_L$ for $x_{0:}$. So there can be no L-merge for $L > k$. However, by definition of cryptic order,

$$\mathrm{H}[\mathcal{S}_{L-1}|X_{0:}] > 0 \qquad \text{for } L \leq k \ ,$$

which means there exists some $x_{0:}$ that has at least two possible $\mathcal{S}_{L-1}$. From the definition of L-merge this leads to existence of an L-merge for $L \leq k$. That completes the proof.

Lemma 2. *Given a process with cryptic order k, each quantum overlap $\langle \eta_a(L)|\eta_b(L)\rangle$ is a nondecreasing function of L for $L \in [0,k]$. For all remaining L, $k \leq L$, the overlap takes the constant value $\langle \eta_a(k)|\eta_b(k)\rangle$.*

Transition matrices are real. Word and state space basis vectors are orthonormal. Change j_n to j_L. "Group terms where $j_L = l_L$ or not".

Proof.

$$\begin{aligned}\langle \eta_a(L)|\eta_b(L)\rangle &= \sum_{\substack{w,w' \in \mathcal{A}^L \\ j_L, l_L \in \{i\}_{i=1}^M}} \sqrt{T^{(w)}_{al_L}} \sqrt{T^{(w')}_{bj_L}} \langle w|w'\rangle \langle \sigma_{l_L}|\sigma_{j_L}\rangle \\ &= \sum_{w, j_L} \sqrt{T^{(w)}_{aj_L}} \sqrt{T^{(w)}_{bj_L}}.\end{aligned}$$

So we have

$$\begin{aligned}&\langle \eta_a(L+1)|\eta_b(L+1)\rangle \\ &= \sum_{\substack{w' \in \mathcal{A}^{L+1} \\ j_{L+1}}} \sqrt{T^{(w')}_{aj_{L+1}}} \sqrt{T^{(w')}_{bj_{L+1}}} \\ &= \sum_{\substack{w \in \mathcal{A}^L, s \in \mathcal{A} \\ j_L, l_L, j_{L+1}}} \sqrt{T^{(w)}_{aj_n}} \sqrt{T^{(s)}_{j_n j_{L+1}}} \sqrt{T^{(w)}_{bl_L}} \sqrt{T^{(s)}_{l_L j_{L+1}}} \\ &= \sum_{\substack{w \in \mathcal{A}^L, s \in \mathcal{A} \\ j_L, j_{L+1}}} \sqrt{T^{(w)}_{aj_L}} \sqrt{T^{(s)}_{j_L j_{L+1}}} \sqrt{T^{(w)}_{bj_L}} \sqrt{T^{(s)}_{j_L j_{L+1}}} \\ &+ \sum_{\substack{w \in \mathcal{A}^L, s \in \mathcal{A} \\ j_L \neq l_L, j_{L+1}}} \sqrt{T^{(w)}_{aj_L}} \sqrt{T^{(s)}_{j_L j_{L+1}}} \sqrt{T^{(w)}_{bl_L}} \sqrt{T^{(s)}_{l_L j_{L+1}}},\end{aligned}$$

The first sum is shown to represent the overlaps obtained already at length L, here we split the sum to two parts

where the first contains ....

$$\sum_{\substack{w\in\mathcal{A}^L, s\in\mathcal{A}\\ j_L, j_{L+1}}} \sqrt{T^{(w)}_{aj_L}}\sqrt{T^{(s)}_{j_Lj_{L+1}}}\sqrt{T^{(w)}_{bj_L}}\sqrt{T^{(s)}_{j_Lj_{L+1}}}$$
$$= \sum_{\substack{w\in\mathcal{A}^L\\ j_L}} \sqrt{T^{(w)}_{aj_L}}\sqrt{T^{(w)}_{bj_L}}\Big(\sum_{\substack{s\in\mathcal{A}\\ j_{L+1}}} \sqrt{T^{(s)}_{j_Lj_{L+1}}}\sqrt{T^{(s)}_{j_Lj_{L+1}}}\Big)$$
$$= \sum_{\substack{w\in\mathcal{A}^L\\ j_L}} \sqrt{T^{(w)}_{aj_L}}\sqrt{T^{(w)}_{bj_L}} = \langle \eta_a(L)|\eta_b(L)\rangle \,.$$

We use lemma 1 to analyze the second sum, which represents the change in the overlaps, finding that,

$$\sum_{\substack{w\in\mathcal{A}^L, s\in\mathcal{A}\\ j_L\neq l_L, j_{L+1}}} \sqrt{T^{(w)}_{aj_L}}\sqrt{T^{(s)}_{j_Lj_{L+1}}}\sqrt{T^{(w)}_{bl_L}}\sqrt{T^{(s)}_{l_Lj_{L+1}}} \geq 0,$$

with equality when $L \geq k$. In summary,

$$\langle \eta_a(L+1)|\eta_b(L+1)\rangle \geq \langle \eta_a(L)|\eta_b(L)\rangle$$

with equality for $L \geq k$.

Lemma 3. *$C_q(L)$ is a function of only the probabilities π_i's and the quantum overlaps.* ***Isn't this in Jozsa?***

Proof. *The signal states $\{|\eta_i(L)\rangle\}$ span only an $|\boldsymbol{\mathcal{S}}|$-dimensional subspace of the $|\mathcal{A}|^L|\boldsymbol{\mathcal{S}}|$-dimensional Hilbert space in which they are defined. Therefore, we may use a Gram-Schmidt process to construct an orthonormal basis $\{|\Psi_j\rangle\}$ of size $|\boldsymbol{\mathcal{S}}|$ such that,*

$$|\eta_i(L)\rangle = \sum_j a_{ij}(L)\,|\Psi_j\rangle$$

where $a_{ij}(L)$ depend only on the quantum overlaps. (The constructed basis is also a function of the overlaps, but the point is that it is orthonormal.) ***Do we want a more formal statement?*** *The spectrum of the density operator depends only on the coordinates in an orthonormal basis, coordinates that here depend only on $\boldsymbol{\pi}$ and overlaps. Finally, the von Neumann entropy depends only on the spectrum.*

I suggest rethinking lemma vs theorem in this paper. I think this overvalues theorem 1 and undervalues previous lemmas.

Theorem 1. *$C_q(L)$ has constant value $C_q(k)$ for $L \geq k$.*

Proof. *Using the lemma 2 and 3 the proof is complete.*

For $L < k$, going to $L+1$ we must find at least one overlap that actually increases. There is a similar classical statement.

Note that while the set of overlaps continues its augmentation at each length up until the cryptic order, we do not currently have a corresponding statement about the nontrivial change in $C_q(L)$ (or its monotonicity).

The naive computation of these overlaps and $C_q(L)$ is numerically difficult and, in all but the most trivial cases, analytically intractible. The preceding Theorem 1 provides a useful halting criterion, but only for finite-cryptic-order processes. The general (and ubiquitous) infinite-cryptic-order case remains.

More importantly, we now understand the cumulative structure of overlaps and the fact that the relevant subspace is of constant size. This allows us to encapsulate these cumulative non-counifilar contributions within a transient structure we name the Quantum Pairwise Merger Machine (QPMM), described next.

IV. (QUANTUM) PAIRWISE MERGER MACHINE

Here we give a recipe for the construction of the pairwise merger machine (PMM) from the ϵ-machine. The PMM describes how two different states can be lead to "merge" in following a particular word. Adding appropriate weights to the PMM yields the quantum pairwise merger machine (QPMM). Merger is important for understanding quantum overlaps. **I realize we never really state why this is so in this paper. It has to do with the choice of quantum rep - that is the block state rep. Is this the best place?** This structure captures not only which states merge given which words, but also the contribution each merger makes to a quantum overlap.

1. Construct the set of (unordered) pairs of (distinct) ϵ-machine states: (σ_j, σ_k). We call these "pair-states". To this set, add a special state called "SINK" which is the terminal state.

2. For each pair-state (σ_j, σ_k), and for each symbol x in the alphabet, there are three possibilities. If:

   (a) at least one of the two ϵ-machine states σ_j, σ_k has no outgoing transition on symbol x. Then do nothing.

   (b) both ϵ-machine states σ_j, σ_k have a transition on symbol x to the same state σ_m. Then connect pair-state (σ_j, σ_k) to "SINK" with edge labeled x. This represents a "merger".

   (c) both ϵ-machine states σ_j and σ_k have a transition on symbol x to two distinct ϵ-machine states σ_m and σ_n where $m \neq n$. (There are no further restrictions on m, n.) Then connect pair-state (σ_j, σ_k) to pair-state (σ_m, σ_n)

3. Remove all edges (in QPMM) that are not part of some path that leads to "SINK".

4. Remove all pair-states that do not have some path to "SINK".

This is the PMM.

1. Now add information about transition probabilities to this topological structure to obtain the QPMM. For each pair-state (σ_j, σ_k) in our structure, add to each outgoing edge the weight $\sqrt{\Pr(x|\sigma_j)\Pr(x|\sigma_k)}$, where x is the symbol associated with that edge. Note that two states in QPMM may be connected with multiple edges (for different symbols).

The QPMM, is *not* a HMM since the edge weights do not yield a stochastic matrix. However, like a HMM, we can consider its "labeled transition matrices" $\{\zeta^{(x)}\}$, one for each symbol in the alphabet. Just as for their classical counterparts, we index these matrices such that $\zeta^{(x)}_{i,j}$ indicates the edge going from pair-state i to pair-state j. **Need that for equation 4.** The important object is the net state-to-state substochastic matrix $\zeta = \sum_{x\in\mathcal{A}} \zeta^{(x)}$. This matrix is at the heart of our closed-form expression of the quantum overlaps at arbitrary (including infinite) word length. These overlaps then quickly yield the corresponding $C_q(L)$.

V. WHAT TO DO WITH THE QPMM

A. Quantum Overlap

Decide on notation for pair-states. Could use (i,j), or (σ_i,σ_j), or $\delta_{i,j}$ The quantum overlap between distinct quantum states is directly obtainable from the QPMM (or more specifically from ζ) using the following equation,

$$\langle \eta_\sigma(L)|\eta_{\sigma'}(L)\rangle = \langle \delta_{\sigma,\sigma'}| \left(\sum_{n=0}^{L} \zeta^n\right) |\delta_{\substack{\text{sync}\\ \text{sink}}}\rangle \ . \tag{4}$$

Where (σ,σ') denotes the pair-state corresponding to the two quantum states of interest, and "SINK" denotes the terminal state described earlier.

There are two trivial cases:

- For $\sigma = \sigma'$, **we define** $\mu_{\sigma,\sigma} = \delta_{\substack{\text{sync}\\ \text{sink}}}$ yielding $\langle \eta_\sigma(L)|\eta_\sigma(L)\rangle = \langle \delta_{\substack{\text{sync}\\ \text{sink}}}|\delta_{\substack{\text{sync}\\ \text{sink}}}\rangle = 1$ for all L.
- If the state-pair $\{\sigma,\sigma'\}$ does *not* appear in the QPMM, **this can be thought of as padding zeta with zeros. There will be no transitions in or out of this zero block, and so** $\langle \eta_\sigma(L)|\eta_{\sigma'}(L)\rangle = 0$ for all L.

This provides the ability to compute overlaps for any process and for any word length L with complexity $\mathcal{O}(L*|QPMM|^2)$.

Given that ζ is substochastic with a terminal state, we can see that $\lim_{L\to\infty}\zeta^L = 0$. It may be the case that ζ is nilpotent, meaning that there is some exponent n for which $\zeta^n = 0$. The minimum such exponent is referred to as the *degree* of the matrix ζ. We can see from equation BLAH that if ζ has degree n, we may replace the upper sum limit with $min(L,n)$, thus truncating the sum.

What would be the structural significance of this nilpotence?

Jordan Chevalley decomposition allows us to write $\zeta = D + N$ where D is diagonalizable and N is nilpotent. Come back to this

Definition 6. *The* depth *of a (Q)PMM is the longest path through the (Q)PMM ending in SINK.*

Note that when there is only a SINK state, the depth is 1; the longest path is [SINK,]. When there exists a path of length L for any $L \in \mathbb{N}$, we call this infinite depth.

The number of states in the (Q)PMM is bounded by the number of states in the ϵ-machine, $|(Q)PMM| \leq |\boldsymbol{\mathcal{S}}| + 1$. Therefore, the largest finite depth possible in a (Q)PMM.

Infinite depth requires either infinite ϵ-machine states, or a cycle in the (Q)PMM. Infinite depth is implied by a cycle, but not necessarily by infinite ϵ-machine states.

Lemma 4. *The cryptic order of a process is one less than the depth of its (Q)PMM. $k = d - 1$.*

Proof. *The*

Therefore, an infinite-cryptic-order process with a finite ϵ-machine must have a cycle in the (Q)PMM.

Lemma 5. *When the cryptic order is finite, the zero-index ν_0 of ζ is equal to the depth of a QPMM.*

Proof. *When cryptic order is finite k, we have at largest k-merge that end at SINK.*

When the cryptic order is infinite, the zero-index ν_0 of ζ is not generally equal to the depth of a QPMM. For instance, the Nemo process has infinite cryptic order and therefore infinite depth, but has a single zero-eigenvalue, and so $\nu_0 = 1$.

Could we create a different version of depth that agrees with the index even for infinite cryptic order processes? For instance, the Nemo has infinite cryptic order but index 1. How do we get a depth measure to tell us about that "1"? Conjecture: define a new depthC that dies when it hits any node involve in a cycle. So for Nemo, it will go from SINK directly to the cycle giving $depthC = 1$ as desired. Also checked for another

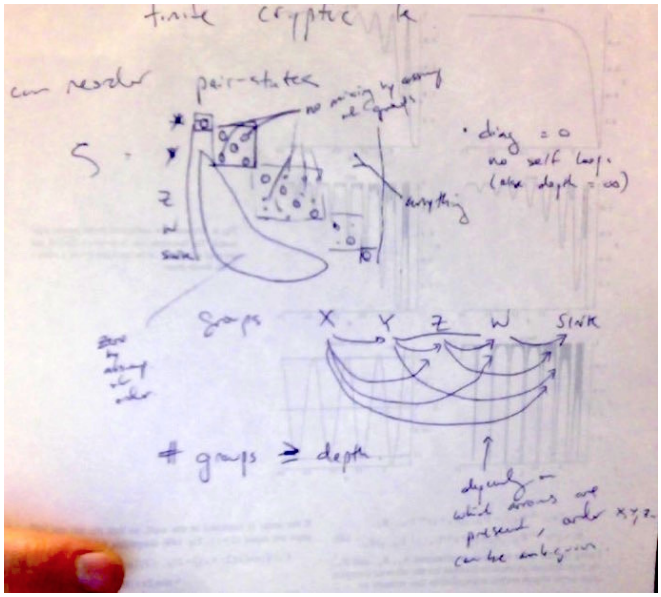

FIG. 2. For finite cryptic order, we can reorder states so that ζ has the above form. Groups of states are defined so that a group does not transition to itself, but only to a group downstream. The number of these groups is an upper bound on the depth.

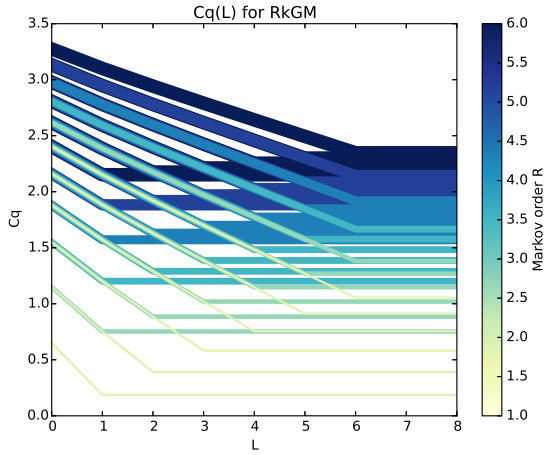

FIG. 3. Cq for RkGM for many Rs and ks. Prob of self-loop $p = 0.8$. Notice how Total computation takes about 3 seconds.

process with $depthC = 2 = index$, defined, "A B 0 0.5; A C 1 0.5; B A 0 0.2; B B 1 0.3; B B 2 0.5; C B 2 1.0"

QPMM for finite cryptic is a DAG. for infinite is DCG.

This means that we may replace the limit in the above sum with k because the $k+1$ sum adds nothing. So, for finite cryptic order processes we find the closed form,

$$\langle \eta_\sigma(L)|\eta_{\sigma'}(L)\rangle = \langle \delta_{\sigma,\sigma'}| \Big(\sum_{n=0}^{\min(L,k)} \zeta^n \Big) |\delta_{\substack{\text{sync}\\ \text{sink}}}\rangle \ . \quad (5)$$

This may be applied, for instance, to example RkGM SEE BELOW.

When the cryptic order is infinite, the above limit remains unchanged, and the sum remains unbounded. This is consistent with the definition of index since $\lim_{L\to\infty} \zeta^L = 0$ due to ζ being substochastic CHECK.

To address this issue, we consider a spectral decomposition of ζ. This allows a closed-form expression for certain infinite cryptic order processes as well.

In the case when ζ is diagonalizable, we have,

$$\langle \eta_\sigma(L)|\eta_{\sigma'}(L)\rangle = \sum_{\lambda\in\Lambda_\zeta} \frac{1-\lambda^{L+1}}{1-\lambda} \langle \mu_{\sigma,\sigma'}| \zeta_\lambda |\delta_{\substack{\text{sync}\\ \text{sink}}}\rangle \ , \quad (6)$$

where ζ_λ is the projection operator of ζ associated with the eigenvalue λ. This encompasses However, for finite cryptic-order larger than one, there will necessarily be nondiagonalizability due to the zero eigenvalue. **Can't we have a zero eigenvalue with degeneracy and all trivial Jordan blocks? The claim about nondiagonizability seems to be based on some additional feature of ζ.** For any (finite or infinite) cryptic order, if the zero eigenvalue is the only source of nondiagonalizability (i.e., the index of all other eigenvalues is unity), then the quantum overlap can be expressed as:

$$\langle \eta_\sigma(L)|\eta_{\sigma'}(L)\rangle = \sum_{\lambda\in\Lambda_\zeta\setminus\{0\}} \frac{1-\lambda^{L+1}}{1-\lambda} \langle \mu_{\sigma,\sigma'}| \zeta_\lambda |\delta_{\substack{\text{sync}\\ \text{sink}}}\rangle + \sum_{m=0}^{\min\{L,\,\nu_0-1\}} \langle \mu_{\sigma,\sigma'}| \zeta_{0,m} |\delta_{\substack{\text{sync}\\ \text{sink}}}\rangle \ , \quad (7)$$

where ν_0 is the index of the zero eigenvalue (which is the size of its largest Jordan block) and $\zeta_{0,m} = \zeta^m \zeta_0$. We refer to this case as the almost-diagonalizable case, common to most processes of interest.

In the case of finite cryptic order, $\nu_0 - 1$ is identifiable with the depth of the transient tree, and the cryptic order is $k = \nu_0 - 1$. But, interestingly, finite (or even zero) ν_0 does not guarantee finite cryptic order. Loops in the transient tree associated with non-zero eigenvalues actually guarantee infinite cryptic order. If ζ has eigenvalues other than zero, $k = \infty$ (where k is the cryptic order). Nevertheless, the methods developed here address these cases just as well—sometimes even with more ease.

The most general spectral decomposition is straightforward, but not obviously important here.

We can also directly find the asymptotic overlap for

the most general case via:

$$\langle \eta_\sigma(\infty)|\eta_{\sigma'}(\infty)\rangle = \langle \mu_{\sigma,\sigma'}| \left(\sum_{n=0}^{\infty} \zeta^n\right) |\delta_{\substack{\text{sync}\\\text{sink}}}\rangle$$
$$= \langle \mu_{\sigma,\sigma'}| \left(I-\zeta\right)^{-1} |\delta_{\substack{\text{sync}\\\text{sink}}}\rangle \ . \quad (8)$$

Note that $I-\zeta$ is invertible since ζ is substochastic: hence its spectral radius is less than unity.

Again, we can recast the asymptotic result in spectral form if we desire.

B. Density Matrix

The density matrix can now be expressed using a fixed M-by-M matrix, valid for all L. Using Gram-Schmidt process one can choose a new orthonormal basis. Let:

$$\begin{aligned}
|\eta_1(L)\rangle &= |e_1^{(L)}\rangle \\
|\eta_2(L)\rangle &= a_{21}^{(L)} |e_1^{(L)}\rangle + a_{22}^{(L)} |e_2^{(L)}\rangle \\
|\eta_3(L)\rangle &= a_{31}^{(L)} |e_1^{(L)}\rangle + a_{32}^{(L)} |e_2^{(L)}\rangle + a_{33}^{(L)} |e_3^{(L)}\rangle \\
&\vdots
\end{aligned}$$

etc. Then

$$a_{21}^{(L)} = \langle \eta_1(L)|\eta_2(L)\rangle = \langle \mu_{\sigma^1,\sigma^2}| \left(\sum_{n=1}^{L} \zeta^n\right) |\delta_{\substack{\text{sync}\\\text{sink}}}\rangle \ ,$$
$$a_{22}^{(L)} = \left(1 - |\langle \eta_1(L)|\eta_2(L)\rangle|^2\right)^{1/2} \ ,$$
$$a_{31}^{(L)} = \langle \eta_1(L)|\eta_3(L)\rangle = \langle \mu_{\sigma^1,\sigma^3}| \left(\sum_{n=1}^{L} \zeta^n\right) |\delta_{\substack{\text{sync}\\\text{sink}}}\rangle \ ,$$

etc. Now it is useful to rewrite what we can in matrix form:

$$\begin{bmatrix} \langle \eta_1(L)| \\ \langle \eta_2(L)| \\ \langle \eta_3(L)| \\ \vdots \\ \langle \eta_M(L)| \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & & & & 0 \\ a_{21}^{(L)} & a_{22}^{(L)} & & & \\ a_{31}^{(L)} & a_{32}^{(L)} & a_{33}^{(L)} & & \\ \vdots & & & \ddots & \\ a_{M1}^{(L)} & & \cdots & & a_{MM}^{(L)} \end{bmatrix}}_{\equiv A_L} \begin{bmatrix} \langle e_1^{(L)}| \\ \langle e_2^{(L)}| \\ \langle e_3^{(L)}| \\ \vdots \\ \langle e_M^{(L)}| \end{bmatrix} \ ,$$

which defines the lower-triangular matrix A_L. Note that the rightmost matrix of orthonormal basis vectors is just the identity matrix since we are working in that basis.

In this new basis, we can construct the M-by-M density matrix as:

$$\rho(L) = \sum_{i=1}^{M} p_i |\eta_i(L)\rangle \langle \eta_i(L)|$$
$$= \begin{bmatrix} |\eta_1(L)\rangle & \cdots & |\eta_M(L)\rangle \end{bmatrix} \underbrace{\begin{bmatrix} \pi_1 & & 0 \\ & \ddots & \\ 0 & & \pi_M \end{bmatrix}}_{\equiv D_\pi} \begin{bmatrix} \langle \eta_1(L)| \\ \langle \eta_2(L)| \\ \langle \eta_3(L)| \\ \vdots \\ \langle \eta_M(L)| \end{bmatrix}$$
$$= A_L^\dagger D_\pi A_L \ .$$

(Since all entries are real, the conjugate transpose is just the transpose, but this more general framework may be useful if we want to eventually consider the effect of adding phase to the quantum states.)

C. von Neumann Entropy

$$C_q(L) = -\text{tr}\left[\rho(L)\log\rho(L)\right]$$
$$= -\text{tr}\left[A_L^\dagger D_\pi A_L \log(A_L^\dagger D_\pi A_L)\right]$$
$$= -\sum_{\lambda\in\Lambda_{A_L^\dagger D_\pi A_L}} \lambda\log\lambda \ .$$

This is relatively easy to calculate now since density matrix is only the M-by-M function of L. Can calculate analytically from spectrum of ρ (which, in a strange way, has already folded in the spectrum of ζ).

D. Gram Matrix

The A_L matrix is burdensome because it contains nonlinear dependence on the quantum overlap of the quantum states. In this section, we will show how to avoid this nonlinearity, and rather obtain the von Neumann entropy from a transformation that yields a linear relationship with the overlaps.

It has been found [14] that the *Gram matrix*, with elements $G_{mn}^{(L)} = \sqrt{\pi_m\pi_n}\,\langle \eta_m(L)|\eta_n(L)\rangle$, can be evaluated instead of the density matrix to evaluate the same von Neumann entropy. In particular, $G^{(L)}$ has the same spectrum as $\rho(L)$, even with the same multiplicities. I.e., $\Lambda_{G^{(L)}} = \Lambda_{\rho(L)}$, while a_λ, g_λ, and ν_λ remain unchanged for all λ in the spectrum. (This is a slightly stronger statement than Josza's original, but is justified since $\rho(L)$ and $G^{(L)}$ are both M-by-M dimensional.)

Here, we briefly explore the relationship between $\rho(L)$ and $G^{(L)}$, and then focus on the closed-form expression

for $G^{(L)}$, which is much nicer than $\rho(L)$. This allows us to calculate and understand the $C_q(L)$ more easily.

Earlier, we found that the density matrix can be written as:

$$\rho(L) = A_L^\dagger D_\pi A_L \ ,$$

which can be rewritten as:

$$\begin{aligned}\rho(L) &= A_L^\dagger D_\pi^{1/2} D_\pi^{1/2} A_L \\ &= \left(D_\pi^{1/2} A_L\right)^\dagger D_\pi^{1/2} A_L \ .\end{aligned}$$

It is easy to show that

$$\begin{aligned}\mathrm{tr}\left[\left(D_\pi^{1/2} A_L\right)^\dagger D_\pi^{1/2} A_L\right] &= \mathrm{tr}\left[D_\pi^{1/2} A_L \left(D_\pi^{1/2} A_L\right)^\dagger\right] \\ &= \mathrm{tr}\left[D_\pi^{1/2} A_L A_L^\dagger D_\pi^{1/2}\right] \ ,\end{aligned}$$

which means that the sum of the eigenvalues is conserved in transforming from $A_L^\dagger D_\pi A_L$ to $D_\pi^{1/2} A_L A_L^\dagger D_\pi^{1/2}$. It is less obvious that the spectrum is also conserved, but this is also true, and even easy to prove. (Observe that $AB\vec{v} = \lambda\vec{v} \Longrightarrow BAB\vec{v} = \lambda B\vec{v} \Longrightarrow BA(B\vec{v}) = \lambda(B\vec{v})$.) Interestingly, the new object turns out to be exactly the Gram matrix, which was previously introduced although without this explicit relationship to the density matrix. We now see that:

$$\begin{aligned}&D_\pi^{1/2} A_L A_L^\dagger D_\pi^{1/2} \\ &\quad= D_\pi^{1/2} \begin{bmatrix} \langle \eta_1(L)| \\ \vdots \\ \langle \eta_M(L)| \end{bmatrix} \left[|\eta_1(L)\rangle \ \cdots \ |\eta_M(L)\rangle\right] D_\pi^{1/2} \\ &\quad= \begin{bmatrix} \sqrt{\pi_1}\langle \eta_1(L)| \\ \vdots \\ \sqrt{\pi_M}\langle \eta_M(L)| \end{bmatrix} \left[\sqrt{\pi_1}|\eta_1(L)\rangle \ \cdots \ \sqrt{\pi_M}|\eta_M(L)\rangle\right] \\ &\quad= \begin{bmatrix} \sqrt{\pi_1\pi_1}\langle \eta_1(L)|\eta_1(L)\rangle & \cdots & \sqrt{\pi_1\pi_M}\langle \eta_1(L)|\eta_M(L)\rangle \\ \vdots & \ddots & \vdots \\ \sqrt{\pi_M\pi_1}\langle \eta_M(L)|\eta_1(L)\rangle & \cdots & \sqrt{\pi_1\pi_M}\langle \eta_M(L)|\eta_M(L)\rangle \end{bmatrix} \\ &\quad= G^{(L)} \ .\end{aligned}$$

Since the spectrum is preserved, we can use the Gram matrix directly to compute the von Neumann entropy:

$$\begin{aligned}C_q(L) &= -\sum_{\lambda \in \Lambda_{G^{(L)}}} \lambda \log \lambda \\ &= -\mathrm{tr}\left[G^{(L)} \log G^{(L)}\right] \ .\end{aligned}$$

E. New Gram matrix

This transformation to the Gram matrix suggests a similar and even more helpful move that can be made while preserving the spectrum: Define the new Gram matrix to be:

$$\begin{aligned}\mathcal{G}^{(L)} &\equiv D_\pi A_L A_L^\dagger \\ &= D_\pi \begin{bmatrix} \langle \eta_{\sigma_1}(L)| \\ \vdots \\ \langle \eta_{\sigma_M}(L)| \end{bmatrix} \left[|\eta_{\sigma_1}(L)\rangle \ \cdots \ |\eta_{\sigma_M}(L)\rangle\right] \\ &= \begin{bmatrix} \pi_1 \langle \eta_1(L)|\eta_1(L)\rangle & \cdots & \pi_1 \langle \eta_1(L)|\eta_M(L)\rangle \\ \vdots & \ddots & \vdots \\ \pi_M \langle \eta_M(L)|\eta_1(L)\rangle & \cdots & \pi_M \langle \eta_M(L)|\eta_M(L)\rangle \end{bmatrix}\end{aligned}$$

Clearly, this preserves the same trace as the density matrix and previous Gram matrix. It also preserves the spectrum. It also has this advantages that we don't have square-roots of two different state probabilities appearing at each element, but rather has just a single probability attached to each element.

Since the spectrum is preserved, we can use the new Gram matrix to compute the von Neumann entropy:

$$C_q(L) = -\sum_{\lambda \in \Lambda_{\mathcal{G}^{(L)}}} \lambda \log \lambda \tag{9}$$

$$= -\mathrm{tr}\left[\mathcal{G}^{(L)} \log \mathcal{G}^{(L)}\right] \ . \tag{10}$$

VI. EXAMPLES

Let us look at several examples.

A. Biased Coin

The Biased Coins Process provides a first, simple case that realizes a nontrivial quantum state entropy [8]. There are two biased coins, named A and B. The first generates 1 with probability q; the second, 0 with probability p. A coin is picked and flipped, generating outputs 0 or 1. With probability q the other coin is used next similarly with different probability. Its two causal-state ϵ-machine is shown in figure 4.

After constructing the QPMM for the Perturbed Coin process, as outlined in Figures 4 through 5, we observe:

$$\zeta^{(0)} = \begin{bmatrix} 0 & \sqrt{p(1-q)} \\ 0 & 0 \end{bmatrix} \ ,$$

$$\zeta^{(1)} = \begin{bmatrix} 0 & \sqrt{q(1-p)} \\ 0 & 0 \end{bmatrix} \ ,$$

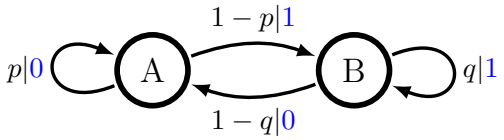

FIG. 4. ϵ-machine for the perturbed coin process.

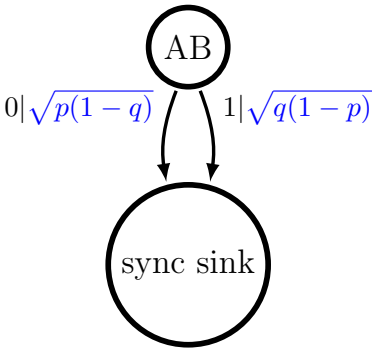

FIG. 5. QPMM for the perturbed coin process.

and so

$$\zeta = \begin{bmatrix} 0 & \beta \\ 0 & 0 \end{bmatrix} ,$$

where we have defined $\beta \equiv \sqrt{p(1-q)} + \sqrt{q(1-p)}$, and will also define the suggestive quantity $\gamma \equiv (1-\beta^2)^{-1/2}$.

The only overlap to consider is $\langle \eta_A(L) | \eta_B(L) \rangle$. For this, we note that $\langle \mu_{A,B}| = \langle \delta_{A,B}| = \begin{bmatrix} 1 & 0 \end{bmatrix}$. Also, $|\delta_{\substack{\text{sync} \\ \text{sink}}}\rangle = \begin{bmatrix} 0 & 1 \end{bmatrix}^\top$.

For a spectral perspective, note that ζ here is a nilpotent matrix with only a zero eigenvalue with index of two: $\Lambda_\zeta = \{0\}$ and $\nu_0 = 2$. Since the projection operators must sum to the identity, we have $\zeta_0 = I$.

ζ^L is the null matrix for $L > 1$, so either by Eq. (5) or by Eq. (7), we have:

$$\langle \eta_A(L) | \eta_B(L) \rangle = \sum_{m=1}^{\min\{L,\,1\}} \langle \delta_{A,B}| \, \zeta^m \, |\delta_{\substack{\text{sync} \\ \text{sink}}}\rangle \ .$$

I.e.,

$$\langle \eta_A(L) | \eta_B(L) \rangle = \begin{cases} 0 & \text{if } L = 0 \\ \beta & \text{if } L \geq 1 \ . \end{cases}$$

1. *Density matrix approach to the VNE*

For the density matrix, we turn to the L-dependent orthonormal basis $\{|e_1^{(L)}\rangle, |e_2^{(L)}\rangle\}$, and use the stationary distribution over $\boldsymbol{\mathcal{S}}$: $\pi = \begin{bmatrix} \frac{p}{p+q} & \frac{q}{p+q} \end{bmatrix}$.

Apparently, for $L = 0$ we have: $|\eta_A(0)\rangle = |e_1^{(0)}\rangle$ and $|\eta_B(0)\rangle = |e_2^{(0)}\rangle$. Hence, $\rho(0) = D_\pi$ and $C_q(0) = H_2(\frac{p}{p+q}) = C_\mu$ qubits.

For $L \geq 1$ we have: $|\eta_A(L)\rangle = |e_1^{(L)}\rangle$ and $|\eta_A(L)\rangle = a_{21}^{(L)} |e_1^{(L)}\rangle + a_{22}^{(L)} |e_2^{(L)}\rangle$, where $a_{21}^{(L)} = \langle \eta_A(L) | \eta_B(L) \rangle = \beta$ and $a_{22}^{(L)} = (1-\beta^2)^{1/2} = \gamma^{-1}$ for $L \geq 1$. I.e.:

$$A_L = \begin{bmatrix} 1 & 0 \\ \beta & \gamma^{-1} \end{bmatrix} \qquad \text{for } L \geq 1 \ .$$

Hence, the density matrix is:

$$\begin{aligned} \rho(L) &= A_L^\dagger D_\pi A_L \\ &= \begin{bmatrix} 1 & \beta \\ 0 & \gamma^{-1} \end{bmatrix} \begin{bmatrix} \frac{p}{p+q} & 0 \\ 0 & \frac{q}{p+q} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \beta & \gamma^{-1} \end{bmatrix} \\ &= \tfrac{1}{p+q} \begin{bmatrix} p & q\beta \\ 0 & q\gamma^{-1} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \beta & \gamma^{-1} \end{bmatrix} \\ &= \tfrac{q}{p+q} \begin{bmatrix} \frac{p}{q} + \beta^2 & \beta/\gamma \\ \beta/\gamma & 1-\beta^2 \end{bmatrix} \qquad \text{for } L \geq 1 \ . \end{aligned}$$

Since $\det(\rho(L) - \lambda I) = \lambda^2 - \lambda + \frac{pq}{(p+q)^2}(1-\beta^2)$, we find the eigenvalues of the density matrix to be:

$$\Lambda_{\rho(L)} = \left\{ \tfrac{1}{2} \pm \tfrac{1}{2(p+q)} \sqrt{4pq\beta^2 + (p-q)^2} \right\} ,$$

which yields the von Neumann entropy:

$$C_q(L) = - \sum_{\lambda \in \{\frac{1}{2} \pm \frac{\sqrt{4pq\beta^2 + (p-q)^2}}{2(p+q)}\}} \lambda \log \lambda \quad \text{for } L \geq 1 \ .$$

2. *New Gram matrix approach to the VNE*

The new Gram matrix for the Perturbed Coin is:

$$\mathcal{G}^{(L)} = D_\pi \begin{bmatrix} \langle \eta_A(L) | \eta_A(L) \rangle & \langle \eta_A(L) | \eta_B(L) \rangle \\ \langle \eta_B(L) | \eta_A(L) \rangle & \langle \eta_B(L) | \eta_B(L) \rangle \end{bmatrix} .$$

Specifically, we have:

$$\mathcal{G}^{(0)} = \tfrac{1}{p+q} \begin{bmatrix} p & 0 \\ 0 & q \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \tfrac{1}{p+q} \begin{bmatrix} p & 0 \\ 0 & q \end{bmatrix}$$

and

$$\mathcal{G}^{(L)} = \tfrac{1}{p+q} \begin{bmatrix} p & 0 \\ 0 & q \end{bmatrix} \begin{bmatrix} 1 & \beta \\ \beta & 1 \end{bmatrix} = \tfrac{1}{p+q} \begin{bmatrix} p & p\beta \\ q\beta & q \end{bmatrix} \quad \text{for } L \geq 1 \ .$$

The eigenvalues of $\mathcal{G}^{(0)}$ are just its diagonal entries here, so, $C_q(0) = H_2(\frac{p}{p+q})$ qubits. For $L \geq 1$, $\det(\mathcal{G}^{(L)} - \lambda I) = \lambda^2 - \lambda + \frac{pq}{(p+q)^2}(1-\beta^2)$, which will give the same answer for eigenvalues and VNE.

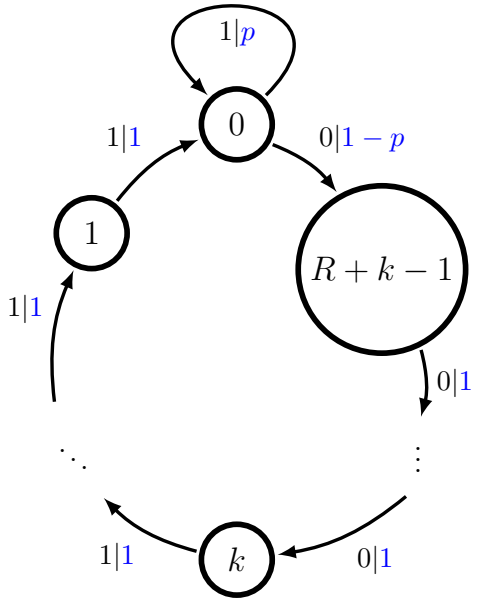

FIG. 6. ϵ-machine for the (R–k)-Golden Mean process.

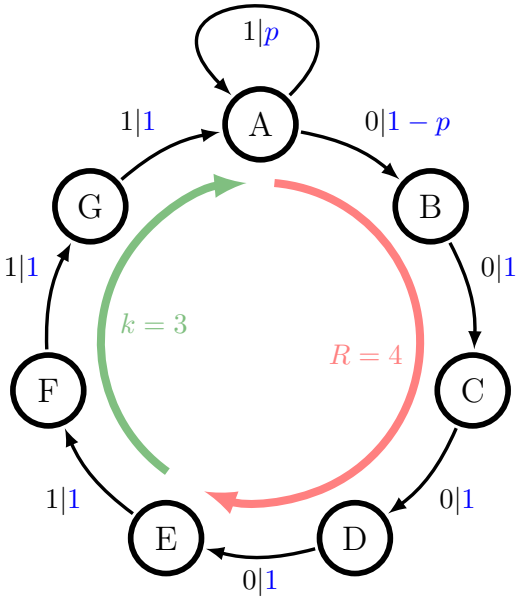

FIG. 7. ϵ-machine for the (4–3)-Golden Mean process.

As it is shown in the new method there is no need to construct density matrix, instead we used new Gram matrix which can be easily calculated having quantum overlaps. Clearly, the new Gram matrix method is more elegant for our purposes. (And this is evident even at $M = 2$. A_L grows as M grows.)

B. (R–k)-Golden Mean

The (R–k)-Golden Mean process is constructed to have Markov-order R and cryptic-order k. Its ϵ-machine is shown in Fig. 6. The 0^{th} state has probability $\pi_0 = \frac{1}{R+k-p(R+k-1)}$ while all other states have probability $\pi_1 = \frac{1-p}{R+k-p(R+k-1)}$.

The QPMM is strictly tree-like of depth k (and maximal width k). All edges have a unit weight except for those edges leaving A-paired states—those edges (numbering k in total) have an associated weight of $\sqrt{p}$.

The eigenvalues of the new Gram matrix can be obtained from:

$$\det(\mathcal{G}^{(L)} - \lambda I) = (\pi_1 - \lambda)^{R+k-\min(L,k)-1} \times \begin{vmatrix} \pi_0 - \lambda & \pi_0\sqrt{p} & \cdots & \pi_0\sqrt{p}^{\min(L,k)} \\ \pi_1\sqrt{p} & \pi_1 - \lambda & & \pi_1\sqrt{p}^{\min(L,k)-1} \\ \vdots & & \ddots & \\ \pi_1\sqrt{p}^{\min(L,k)} & & & \pi_1 - \lambda \end{vmatrix} = 0 \ ,$$

which directly yields the von Neumann entropy. Note

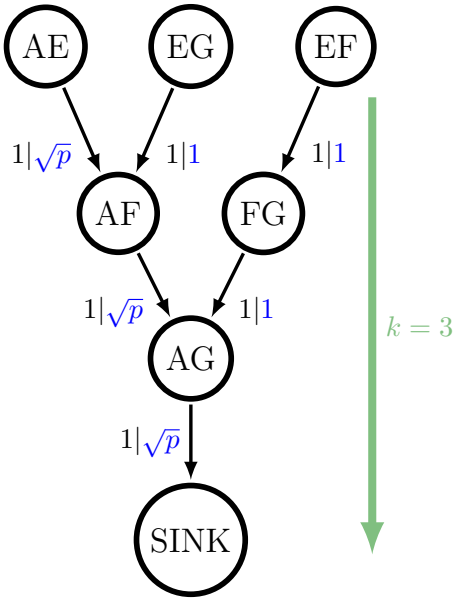

FIG. 8. QPMM for the (4–3)-Golden Mean process.

that although the $H_{\text{vN}}(L)$ is *not* actually linear in L, it appears approximately linear by eye.

To give an explicit example, let's consider the (4–3)-GM process which, by definition, has $R = 4$ and $k = 3$. The A state has probability $\pi_A = \frac{1}{R+k-p(R+k-1)} = \frac{1}{7-6p}$ while all other states have probability $\pi_B = \frac{1-p}{7-6p}$.

- For $L = 0$,

$$\det(\mathcal{G}^{(0)} - \lambda I) = (\pi_B - \lambda)^6(\pi_A - \lambda) \ ,$$

yielding $\Lambda_{\mathcal{G}}^{(0)} = \{\pi_B, \pi_A\}$ (with $a_{\pi_B} = 6$) and

$$H_{\text{vN}}^{(0)} = -6\pi_B \log \pi_B - \pi_A \log \pi_A \ .$$

- For $L = 1$,

$$\begin{aligned}\det(\mathcal{G}^{(1)} - \lambda I) =& (\pi_B - \lambda)^5 \\ & \times \left[\lambda^2 - (\pi_A + \pi_B)\lambda + \pi_A \pi_B (1-p)\right] \ ,\end{aligned}$$

yielding $\Lambda_{\mathcal{G}}^{(1)} = \{\pi_B, c_+, c_-\}$ with $c_\pm = \frac{1}{2}(\pi_A + \pi_B) \pm \frac{1}{2}\left[(\pi_A + \pi_B)^2 - 4\pi_A \pi_B (1-p)\right]^{1/2}$ (and with $a_{\pi_B} = 5$), and

$$H_{\text{vN}}^{(1)} = -5\pi_B \log \pi_B - c_+ \log c_+ - c_- \log c_- \ .$$

- For $L = 2$,

$$\begin{aligned}\det(\mathcal{G}^{(2)} - \lambda I) &= (\pi_B - \lambda)^4 \\ &\times \begin{vmatrix} \pi_A - \lambda & \pi_A p^{1/2} & \pi_A p \\ \pi_B p^{1/2} & \pi_B - \lambda & \pi_B p^{1/2} \\ \pi_B p & \pi_B p^{1/2} & \pi_B - \lambda \end{vmatrix} \ .\end{aligned}$$

- For $L \geq 3$,

$$\begin{aligned}\det(\mathcal{G}^{(L)} - \lambda I) &= \det(\mathcal{G}^{(3)} - \lambda I) \\ &= (\pi_B - \lambda)^3 \\ &\times \begin{vmatrix} \pi_A - \lambda & \pi_A p^{1/2} & \pi_A p & \pi_A p^{3/2} \\ \pi_B p^{1/2} & \pi_B - \lambda & \pi_B p^{1/2} & \pi_B p \\ \pi_B p & \pi_B p^{1/2} & \pi_B - \lambda & \pi_B p^{1/2} \\ \pi_B p^{3/2} & \pi_B p & \pi_B p^{1/2} & \pi_B - \lambda \end{vmatrix} \ .\end{aligned}$$

C. Nemo

Our third example is the Nemo process, which is shown in figure 9. Since the recurrent states simply permute upon observing a 0, the word 0000... never reveals the current state. This once again means that the process is non markovian and has $R = \infty$. In other hand if we know the current state and we observe 000..0 we will never know the starting state and that means infinite cryptic order $K = \infty$.

After constructing the QPMM for the process, as outlined in Figures 9 through 10, we observe:

$$\zeta = \begin{bmatrix} 0 & \sqrt{1-p} & 0 & 0 \\ 0 & 0 & \sqrt{1/2} & 0 \\ \sqrt{(1-p)/2} & 0 & 0 & \sqrt{p/2} \\ 0 & 0 & 0 & 0 \end{bmatrix} \ .$$

Unlike the previous examples, ζ here has nonzero eigenvalues due to a loop in the transient structure, which implies infinite cryptic order. Specifically the eigenvalues are zero and the three cubed roots of $\frac{1}{2}(1-p)$: $\Lambda_\zeta = \left\{0, [\frac{1}{2}(1-p)]^{1/3} e^{i2\pi n/3} : n \in \{0,1,2\}\right\}$. Since ζ is diagonalizable, the index of all these eigenvalues is one. Since the algebraic multiplicity of all of the eigenvalues is one, each of the projection operators can be expressed as $\zeta_\lambda = \frac{1}{\langle \boldsymbol{\lambda} | \boldsymbol{\lambda} \rangle} |\boldsymbol{\lambda}\rangle \langle \boldsymbol{\lambda}|$.

The overlaps to consider are $\langle \eta_A(L) | \eta_B(L) \rangle$, $\langle \eta_B(L) | \eta_C(L) \rangle$, and $\langle \eta_B(L) | \eta_C(L) \rangle$. For this, we note that $\langle \mu_{A,B}| = \langle \delta_{A,B}| = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}$, $\langle \mu_{B,C}| = \langle \delta_{B,C}| = \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix}$, and $\langle \mu_{A,C}| = \langle \delta_{A,C}| = \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix}$. Also, $|\delta_{\substack{\text{sync} \\ \text{sink}}}\rangle = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}^\top$.

For all $L > 0$, the quantum overlap of the L^{th}-order quantum states are:

$$\langle \eta_\sigma(L) | \eta_{\sigma'}(L) \rangle = \sum_{\lambda \in \Lambda_\zeta \setminus \{0\}} \frac{1 - \lambda^{L+1}}{1 - \lambda} \langle \delta_{\sigma,\sigma'} | \zeta_\lambda | \delta_{\substack{\text{sync} \\ \text{sink}}} \rangle \ .$$

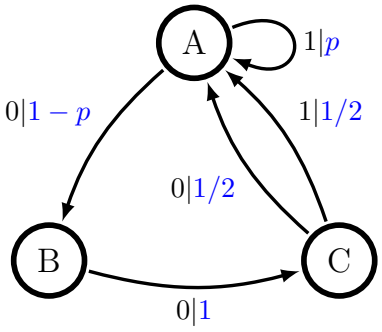

FIG. 9. ϵ-machine for the Nemo process.

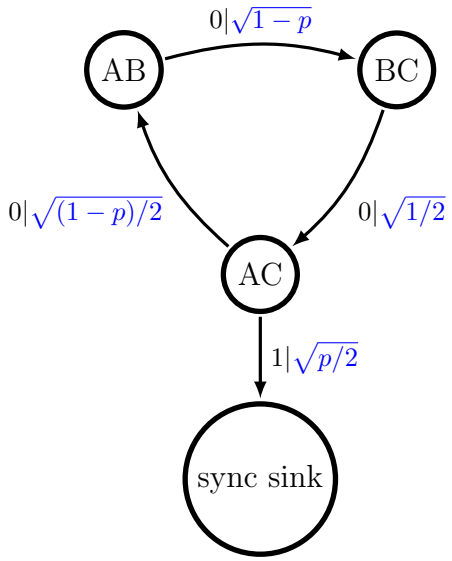

FIG. 10. QPMM for the Nemo process.

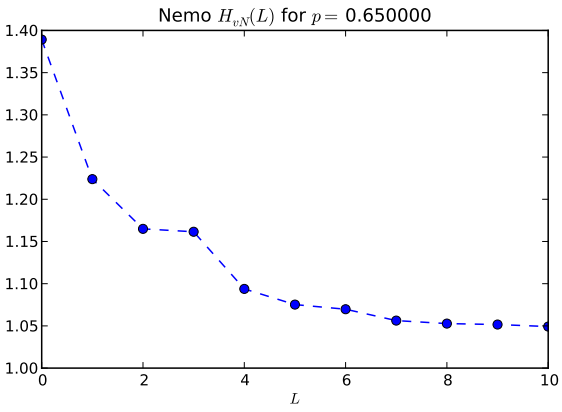

FIG. 11. $C_q(L)$ for the Nemo process at a particular p.

The asymptotic overlap is:

$$\langle \eta_\sigma(\infty)|\eta_{\sigma'}(\infty)\rangle = \langle \delta_{\sigma,\sigma'}| \left(I-\zeta\right)^{-1} |\delta_{\substack{\text{sync}\\ \text{sink}}}\rangle \ .$$

For the density matrix, we could turn to the L-dependent orthonormal basis $\{|e_1^{(L)}\rangle, |e_2^{(L)}\rangle, |e_3^{(L)}\rangle\}$, and use the stationary distribution over $\boldsymbol{\mathcal{S}}$: $\pi = \frac{1}{3-2p}\begin{bmatrix}1 & 1-p & 1-p\end{bmatrix}$. However, it is easier to obtain the eigenvalues of the Gram matrix of D_π-$\times$-the overlap matrix, and we have shown that these eigenvalues are guaranteed to be the same as those for the density matrix. Numerically, this is both efficient and easy to implement. We can go up to arbitrarily high L without any problem.

Recall that the cryptic order of the Nemo process is infinite, which can be seen from the loop in the QPMM transient structure. Spectrally, this can be seen from non-zero eigenvalues of ζ. Therefore, the quantum overlap does not saturate at any finite L. Nevertheless, through the QPMM, we have completely characterized all modes of convergence to the asymptotic coding benefit of high-order representations.

VII. GENERAL BEHAVIOR OF VON NEUMANN ENTROPY FOR INFINITE CYPTIC ORDER PROCESSES AT LARGE LENGTH

For infinite cryptic order processes one question that could come to mind is what would be the behavior of $C_q(L)$ for large L. Using the method that we have proposed, we will study this question analytically.

Lemma 6. *For small variations $\{\delta\lambda\}_{\lambda\in\Lambda_\rho}$ in the eigenvalues of the density matrix ρ, the change in the von Neumann entropy is approximately:*

$$\delta S = -\sum_{\lambda\in\Lambda_\rho} \log(\lambda)\,\delta\lambda \ .$$

Proof.

$$S = -\sum_{\lambda\in\Lambda_\rho} \lambda\log\lambda \ .$$

Directly calculating the first differential yields:

$$\delta S = -\sum_{\lambda\in\Lambda_\rho} \left(\log(\lambda)\,\delta\lambda + \delta\lambda\right) \ .$$

However, knowing

$$\sum_{\lambda\in\Lambda_\rho} \lambda = 1 \ ,$$

we have

$$\sum_{\lambda\in\Lambda_\rho} \delta\lambda = 0 \ , \tag{11}$$

which leads to

$$\delta S = -\sum_{\lambda\in\Lambda_\rho} \log(\lambda)\,\delta\lambda \ .$$

Recall that the eigenvalues of the density matrix $\rho(L)$ are the same as the eigenvalues of the associated Gram matrix $G^{(L)}$. I.e., $\Lambda_{G^{(L)}} = \Lambda_{\rho(L)}$.

Definition 7. *We define:*
$G^{(L)}$ is a Gram matrix at length L corresponding to $\rho(L)$.
$\lambda^{(L)} \in \Lambda_{G^{(L)}}$ is any one of its eigenvalues.
$|\lambda^{(L)}\rangle$ is the right eigenvector of $G^{(L)}$ corresponding to $\lambda^{(L)}$.
$\langle\lambda^{(L)}|$ is the left eigenvector of $G^{(L)}$ corresponding to $\lambda^{(L)}$.

If using the symmetric version of G, the right and left eigenvectors are simply transposes of each other: $\langle\lambda^{(L)}| = \left(|\lambda^{(L)}\rangle\right)^\top$. For simplicity of the proofs, we will assume non-degeneracy of the eigenvalues of $G^{(L)}$, so that the projection operator associated with $\lambda^{(L)}$ is $\frac{|\lambda^{(L)}\rangle\langle\lambda^{(L)}|}{\langle\lambda^{(L)}|\lambda^{(L)}\rangle}$, where the denominator assures normalization. Nevertheless, the eigenbasis of $G^{(L)}$ is always complete, and the final result retains general validity.

When L is large, $\delta(G^{(L)}) \equiv G^{(L)} - G^{(L+1)}$ is small compared to $G^{(L)}$, which means we can look at it as a perturbation to $G^{(L)}$.

Lemma 7. *Perturbing $G^{(L)}$ to $G^{(L)} + \delta(G^{(L)})$, the first order change in its eigenvalues is given by:*

$$\delta(\lambda^{(L)}) = \frac{\langle\lambda^{(L)}|\,\delta(G^{(L)})\,|\lambda^{(L)}\rangle}{\langle\lambda^{(L)}|\lambda^{(L)}\rangle}\ .$$

Proof. *This is the standard first-order non-degenerate perturbation theory familiar in quantum mechanics, with the allowance for non-normalized bras and kets. The first-order variation of each side of the eigen-relation*

$$G^{(L)}\,|\lambda^{(L)}\rangle = \lambda^{(L)}\,|\lambda^{(L)}\rangle$$

yields

$$\delta(G^{(L)})\,|\lambda^{(L)}\rangle + G^{(L)}\delta(|\lambda^{(L)}\rangle) = \delta(\lambda^{(L)})\,|\lambda^{(L)}\rangle + \lambda^{(L)}\delta(|\lambda^{(L)}\rangle)\ .$$

Multiplying by $\langle\lambda^{(L)}|$ gives:

$$\langle\lambda^{(L)}|\,\delta(G^{(L)})\,|\lambda^{(L)}\rangle + \langle\lambda^{(L)}|\,G^{(L)}\delta(|\lambda^{(L)}\rangle) = \delta(\lambda^{(L)})\,\langle\lambda^{(L)}|\lambda^{(L)}\rangle + \lambda^{(L)}\,\langle\lambda^{(L)}|\,\delta(|\lambda^{(L)}\rangle).$$

Now using

$$\langle\lambda^{(L)}|\,G^{(L)} = \lambda^{(L)}\,\langle\lambda^{(L)}|\,,$$

we have

$$\delta(\lambda^{(L)}) = \frac{\langle\lambda^{(L)}|\,\delta(G^{(L)})\,|\lambda^{(L)}\rangle}{\langle\lambda^{(L)}|\lambda^{(L)}\rangle}\ .$$

Lemma 8. *At any L, $\delta(G^{(L)})$ is given exactly by:*

$$\delta(G^{(L)}) = (\mathcal{G}^{(L+1)}) - (\mathcal{G}^{(L)}) = (max_i\{\zeta_i\})^L A$$

which A is independent of L

Proof.

$$\begin{aligned}\delta(G^{(L)}) &= (\mathcal{G}^{(L+1)}) - (\mathcal{G}^{(L)})\\ &= dots\end{aligned}$$

Lemma 9. *For large L, $\delta(G^{(L)})$ scales as:*

$$\delta(G^{(L)}) = (max_i\{\xi_i\})^L A +$$

which A is independent of L

Proof. *Paul's proof*

Lemma 10. *For large L in the first non zero order*

$$\delta(\lambda_{i,L}) = \delta(\lambda_{i,L+1}) - \delta(\lambda_{i,L}) = c_i(max_i\{\zeta_i\})^L$$

Proof. *Using lemma 7 and* **??**

$$\delta(\lambda_{i,L}) = (max_i\{\zeta_i\})^L \frac{\langle L_{i,L}|\,A\,|R_{i,L}\rangle}{\langle L_{i,L}|R_{i,L}\rangle}.$$

now we can rewrite

$$\begin{aligned}|R_{i,L}\rangle &= |R_{i,\infty}\rangle + O(max_i\{\zeta_i\})\\ \langle L_{i,L}| &= \langle L_{i,\infty}| + O(max_i\{\zeta_i\}).\end{aligned}$$

Because we want the first non zero order, this $\frac{\langle L_{i,L}|A|R_{i,L}\rangle}{\langle L_{i,L}|R_{i,L}\rangle}$ should be calculated in zero order which means

$$\delta(\lambda_{i,L}) = (max_i\{\zeta_i\})^L \frac{\langle L_{i,\infty}|\,A\,|R_{i,\infty}\rangle}{\langle L_{i,\infty}|R_{i,\infty}\rangle}.$$

calling

$$c_i = \frac{\langle L_{i,\infty}|\,A\,|R_{i,\infty}\rangle}{\langle L_{i,\infty}|R_{i,\infty}\rangle}.$$

the proof is complete.

Theorem 2. *At large L, $C_q(L)$ decays exponentially*

Proof. *Using lemma 6 and 10 we have*

$$\delta S = S(L) - S(\infty) = -\sum_i (max_i\{\zeta_i\})^L \log {\lambda_{i,L}}^{c_i}$$

now we can rewrite

$$\lambda_{i,L} = \lambda_{i,\infty} + O(max_i\{\zeta_i\}).$$

that means for the first non zero order we have

$$S(L) = S(\infty) - (max_i\{\zeta_i\})^L \log \prod_i {\lambda_{i,\infty}}^{c_i}.$$

Using equation 11 and lemma 10 which means some of c_i's are negative and some positive so we could not be definite about the sign of $(\log \prod_i {\lambda_{i,\infty}}^{c_i})$. So we could have two cases, converges from above or bellow but in both cases exponentially. we suspect that the later case would never happen.

VIII. CONCLUSION

At the beginning of the article the newly introduced $q-machine$ is reviewed which can substantially compress the causal organization of classical information source and could be used for two observer to synchronize their predictions of the process . Introducing quantum pair-wise merger machine we give an instruction to calculate

quantum overlaps. It is known that having quantum overlaps and distribution over states are enough to calculate von Neumann entropy, using Gram matrix.

To get ride of non linearity of the ordinary Gram matrix in terms of distribution over states, a new Gram matrix is introduced which is linear in that sense and it also preserves the spectrum and could be used to calculate von Neumann entropy. Different examples have been investigated to show how the method provides extremely efficient numerical algorithms for finding the quantum overlaps and the von Neumann entropy as a function of sequence length. To show the power of our method analytically the general behavior of von Neumann entropy for large length has been investigated.

Appendix A: Appendix

Definition 8. *The L^{th} concatenation machine $\mathcal{M}^{\otimes L}$ is defined by $\{\boldsymbol{\mathcal{S}}, \{T^{(w)} = T^{(x_0)} \cdots T^{(x_{L-1})}, w = x_0 \cdots x_{L-1} \in \mathcal{A}^L\}, \langle \pi |\}$.*

The concatenation machine has the same causal states and stationary distribution as the ϵ-machine. Its labeled transition matrices correspond to words of length L and corresponding paths through the ϵ-machine.

Intuition might suggest that the q-machine is related to the concatenation machine, given that it is defined in terms of similar (square roots of) products of labeled transition matrices. This intuition is verified in the following lemma.

Lemma 11.

$$\mathscr{Q}(\mathcal{M}, L) = \mathscr{Q}(\mathcal{M}^{\otimes L}, 1), \tag{A1}$$

where the $\mathscr{Q}$ operator maps an ϵ-machine to its $q-$machine at length L

Proof. *Using the definition in equation 2 the left hand side becomes,*

$$|\eta_j(L)\rangle = \sum_{w^L \in \mathcal{A}^L} \sum_{\sigma_k \in \boldsymbol{\mathcal{S}}} \sqrt{\Pr(w^L, \sigma_k | \sigma_j)} \ |w^L\rangle \, |\sigma_k\rangle$$

where

$$\begin{aligned} \Pr(w^L, \sigma_k | \sigma_j) &= \sum_{i_1, \cdots, i_{L-1}} T^{(x_0)}_{j i_1} T^{(x_1)}_{i_1 i_2} \cdots T^{(x_{L-1})}_{i_{L-1} k} \\ &= T^{(w)}_{jk} \end{aligned}$$

using this

$$|\eta_j(L)\rangle = \sum_{w \in \mathcal{A}^L} \sum_{\sigma_k \in \boldsymbol{\mathcal{S}}} T^{(w)}_{jk} \, |w\rangle \, |\sigma_k\rangle$$

and that is the right hand side.

ACKNOWLEDGMENTS

We thank Ryan James for helpful conversations. The authors thank the Santa Fe Institute for its hospitality during visits. JPC is an SFI External Faculty member. This material is based upon work supported by, or in part by, the John Templeton Foundation and U. S. Army Research Laboratory and the U. S. Army Research Office under contract W911NF-13-1-0390.

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

[2] P. Ball. *The Self-Made Tapestry: Pattern Formation in Nature*. Oxford University Press, New York, 1999.

[3] R. Hoyle. *Pattern Formation: An Introduction to Methods*. Cambridge University Press, New York, 2006.

[4] H. Kantz and T. Schreiber. *Nonlinear Time Series Analysis*. Cambridge University Press, Cambridge, United Kingdom, second edition, 2006.

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

[6] C. R. Shalizi and J. P. Crutchfield. Computational mechanics: Pattern and prediction, structure and simplicity. *J. Stat. Phys.*, 104:817–879, 2001.

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

[8] M. Gu, K. Wiesner, E. Rieper, and V. Vedral. Quantum mechanics can reduce the complexity of classical models. *Nature Comm.*, 3(762):1–5, 2012.

[9] P. Gmeiner. Equality conditions for internal entropies of certain classical and quantum models. *arXiv preprint arXiv:1108.5303*, 2011.

[10] C Aghamohammadi J. R. Mahoney and P. Crutchfield J. Occam's quantum strop: Synchronizing and compressing classical cryptic processes via a quantum channel. *Nature Comm.*, 2015.

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

[12] J. R. Mahoney, C. J. Ellison, and J. P. Crutchfield. Information accessibility and cryptic processes. *J. Phys. A: Math. Theo.*, 42:362002, 2009.

[13] J. R. Mahoney, C. J. Ellison, R. G. James, and J. P. Crutchfield. How hidden are hidden processes? a primer on crypticity and entropy convergence. *CHAOS*, 21(3):037112, 2011.

[14] R. Jozsa and J. Schlienz. Distinguishability of states and von neumann entropy. *Phys. Rev. A*, 2000.