Information Theoretic Properties of Quantum Automata

D. Gier
Department of Physics and Astronomy, The University of Kansas, Lawrence, KS 66045 and
Complexity Sciences Center and Physics Department,
University of California Davis, One Shields Avenue, Davis, CA 95616
(Dated: January 31, 2016)

ϵ-machines serve as provably minimal classical models of stochastic processes, but they are often not ideal in that their statistical complexity (C_μ) is demonstrably greater than the excess entropy ($\mathbf{E}$). Quantum models of the same processes are more efficient as their statistical complexity (C_q) obeys the relation $C_\mu \geq C_q \geq \mathbf{E}$. In this report we explore the information theoretic properties of these q-machines, particularly the value for C_q when considering words of increasing length and reverse processes, as well as the state trace entropy, the quantum analog of the classical block state entropy. We also develop a construction of a quantum bi-directional machine and discuss efforts to calculate the entanglement within q-machines to see if it is the mechanism for information compression in q-machines.

Contents

INTRODUCTION

Stochastic processes are physical systems for which there is inherent uncertainty in the output of the system, even given a perfect knowledge of initial conditions. Common examples include the fluctuations of the stock market, orientations of layers of atoms in crystal stacking, measurements of quantum spin-systems and sequences of neural firing. In the case where the process can be represented with a finite set of states and discrete transitions in time or space, the formalism of computational mechanics is able to represent the structure of the system through ϵ-machines [1]. An ϵ-machine serves as a minimal model capable of reproducing the physical process. It can be visualized as a set of finite states (A,B,C...) with transitions between them. Each transition is associated with a certain probability and an output symbol (0,1,2...). By transitioning through the machine and combining successive symbols one can form output strings with associated probabilities. An ϵ-machine representing the Even Process can be seen in Figure 1. The topology of the ϵ-machine completely determines the possible strings and associated probabilities it produces. For example, the Even Process will always generate ones in pairs and never singly. Though ϵ-machines are provably minimal, for many processes their statistical complexity (which can be thought of as the 'memory' of the machine) is greater than the excess entropy, $\mathbf{E}$ (the amount of information it is possible to predict about the future). Thus these classical models store information that is 'wasted.'

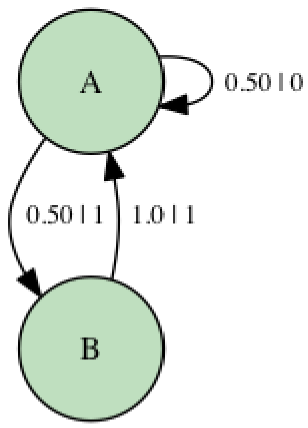

FIG. 1: The Even Process

A recent construction of quantum automata, introduced in [2], presents a more efficient method of modeling stochastic processes. The statistical complexity of these q-machines is generally less than that of their classical counterparts. However their information theoretic properties are not well-studied. In the following, we examine these machines and draw links between the classical and quantum representations of processes. We look specifically at the statistical complexity when considering different length words and the state-trace entropy, which has a comprehensible explanation in terms of classical states and transitions. Finally we discuss ideas on the calculation and role of entanglement in q-machines, especially as a candidate for the q-machine's ability to compress classical information.

ϵ-MACHINES AND Q-MACHINES

Classical ϵ-machines Construction

An ϵ-machine M can be described mathematically using a set of causal states, $\sigma_k \in \mathcal{S}$, symbols $s \in \mathcal{A}$ and a set of transition probabilities T^s_{ij} between states σ_i and σ_j, emitting a symbol s. The way this machine generates strings is governed by the transition probabilities so that for a machine transitioning as $\sigma_A \to \sigma_B \to \sigma_C$ with transition probabilities T^0_{AB} and T^1_{BC}, the word '01' is generated. One important property of ϵ-machines is unifilarity, which says that state transitions are completely determined by the starting state and the output symbol. Thus a state cannot transition to two other states on the same symbol. This ensures that if an observer knows the causal state at one time, they are synchronized with the machine and will know the causal state at any later time if they keep track of the intervening output. Thus they have maximal predictive power.

Classical Information Measures

Certain information theoretic quantities can be calculated for a process directly from its corresponding ϵ-machine. These then give insight into the order, structure and predictability of the process. One such measure, known as the excess entropy $\mathbf{E}$, characterizes the amount of information about the future that can be predicted by knowing the entire past:

$$\mathbf{E} = I[X_{-\infty:0} : X_{0:\infty}] \tag{1}$$

where I[X,Y] is the Shannon mutual information between variables X and Y, which corresponds to the degree to which measurements of the two are correlated. This quantity can be calculated using the ϵ-machine with

$$\mathbf{E} = I[S^+ : S^-] \tag{2}$$

where S^+ are the forward causal states and S^- are the reverse causal states.

A process' cryptic order, k, can also be easily calculated given an ϵ-machine using the equation:

$$k = min(L : H[S_L|\overrightarrow{X}_{0:\infty}]) \tag{3}$$

where S_L is the set of the past L causal states and $\overrightarrow{X}_{0:\infty}$ is the entire future string. For example $k = 1$ for the Golden Mean Process in 2 because, even with the entire future string, it isn't clear whether the process started in state A or B if the first symbol is a 1. After the first state is given, the one can find the causal state at any other time.

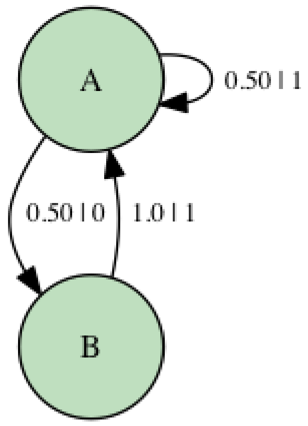

FIG. 2: The Golden Mean Process

Another quantity of particular interest for ϵ-machines is the statistical complexity C_μ:

$$C_\mu = H(\Pi) \tag{4}$$

where $H(X) = -\sum_i p_i \log p_i$ is the Shannon entropy for a probability distribution and Π is the stationary state distribution, defined as the probability that the machine will be each causal state in S when stopped after running for an infinite amount of time. C_μ is understood as the amount of information stored within a machine's causal states or, alternatively, the amount of information necessary to synchronize two machines for the same process.

q-machine Construction

For a given causal state, σ_j, the states of the q-machine are constructed in [2] from the classical ϵ-machine in this fashion:

$$|\eta_j(L)\rangle = \sum_{w^L \in |\mathcal{A}|^L} \sum_{\sigma_k \in \mathcal{S}} \sqrt{Pr(w^L, \sigma_k|\sigma_j)}\, |w^L\rangle\, |\sigma_k\rangle \tag{5}$$

Each state of the q-machine is bipartite because it consists of states of the Hilbert subspace of words of length L,

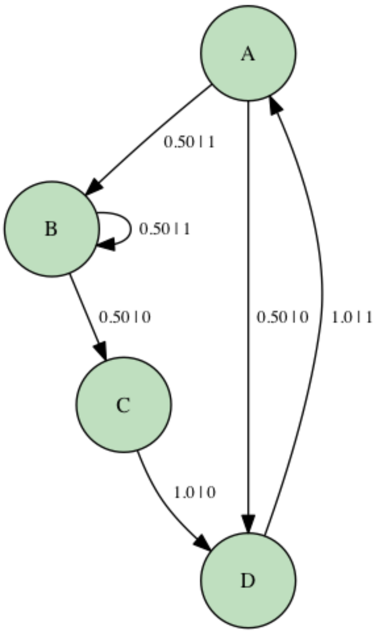

FIG. 3: The Phase Slip Backtrack Process

$|w^L\rangle$, and the Hilbert subspace of classical causal states $|\sigma_k\rangle$. Notably for L = 0 we recover the ϵ-machine and classical state transitions, in which case the state ignores the word subspace and thus is not bipartite. Thus the q-machine is a generalization of the classical ϵ-machine which includes consideration of the words the machine produces.For a given L we can construct a state operator or density matrix to fully characterize the q-machine state with

$$\rho(L) = \sum_j |\eta_j\rangle \langle\eta_j| \tag{6}$$

This quantum representation of a stochastic process opens the door to new methods of analysis based upon studies of bipartite quantum systems and non-classical correlations. There is a degree of overlap in quantum states that does not exist in classical ϵ-machines due to their discrete set of states. In the classical case two states with identical transitions of nearly the same probabilities will contribute separately to the statistical complexity even though there is significant crossover in their output strings. This overlap is considered in q-machines, which allows them to be more efficient. This finding can be quantified by defining the quantum statistical complexity, C_q for q-machines as

$$C_q = S(\rho(L)) = -tr(\rho log(\rho)) \tag{7}$$

where $S(\rho)$ is the von Neumann entropy of the density matrix ρ and $tr(\rho log(\rho))$ is calculated by diagonalizing ρ. In the case that L = 0, $C_q = C_\mu$ as expected.

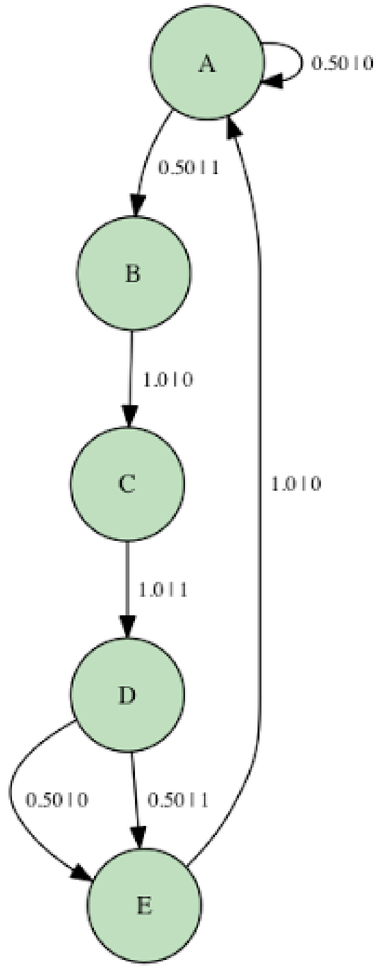

FIG. 4: The Noisy Random Phase Slip Process

PROPERTIES OF C_q

In previous studies it has been noted that $C_q \leq C_\mu$ for single symbols which is the primary motivation for the study of q-machines [2]. Furthermore $C_q(L)$ has been found to decrease monotonically for increasing L [5]. This result has not been rigorously proven, but it has held true for all examples of processes that have been tested. Furthermore [5] finds that $C_q(L)$ reaches a minimum value at L = k, the cryptic order. Examples of this result can be seen in Figure 6 (k = 2) and Figure 7 (k = 3). The associated ϵ-machines can be found in Figures 3 and 4. Many processes have infinite cryptic orders, in which case C_q asymptotically approaches a certain value as in Figure 8, which represents the Nemo Process of Figure 5. This process also has the property that the decrease in C_q is not convex, which somewhat complicates the analysis because a lower bound cannot be established with certainty.

For every process and ϵ-machine there is a corresponding reverse machine that generates the same strings

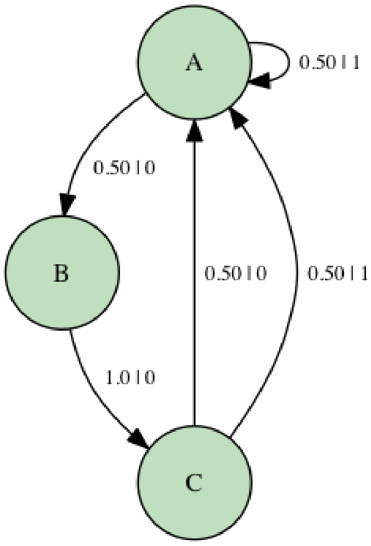

FIG. 5: The Nemo Process

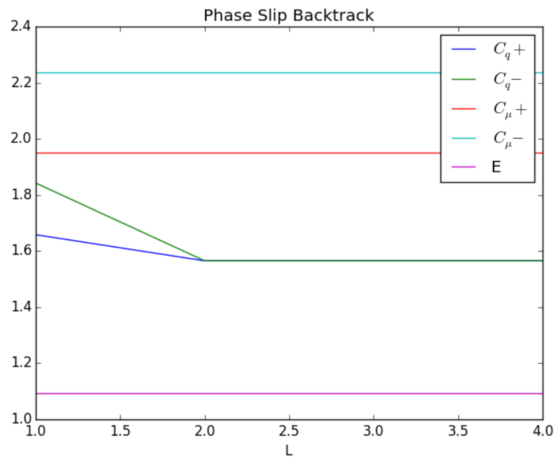

FIG. 6: Comparison of statistical complexity for the Phase Slip Backtrack Process

backwards. While a process and its reverse have the same excess entropy **E**, the corresponding ϵ-machines do not generally have $C_\mu^+ = C_\mu^-$. Using the construction of [2] (in which L = 1), $C_q^+ \neq C_q^-$ generally as well. The two are only equal if the reverse machine is identical to the forward machine (as with the Golden Mean Process in Fig. 2). However, when words with $L > 1$ are considered, the values for C_q^+ and C_q^- converge to some value $C_q^\infty$for $L = k$, where k is the cryptic order. These results can be seen clearly in Figs. 6, 7 and 8. An enumeration over all machines for which the number of states is 5 or less and the alphabet size is 3 or less found no q-machines for which this was false. There is no proof developed, but this survey suggests that the result is general.

The analysis of the Phase Slip Backtrack Process, summarized in Figure 6, is particularly noteworthy. The

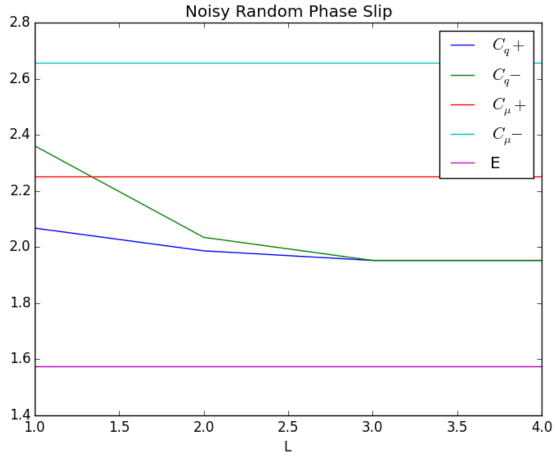

FIG. 7: Comparison of statistical complexity for the Noisy Random Phase Slip Process

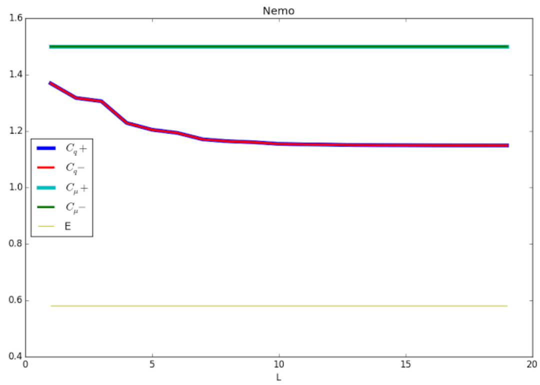

FIG. 8: Statistical complexity for the Nemo Process

forward ϵ-machine consists of four causal states, and the reverse machine has five. The classical statistical complexity is strongly dependent upon the number of causal states, since it is calculated using the stationary state distribution, and $C_\mu^- > C_\mu^+$ as intuition would suggest. Despite this C_q^+ and C_q^- reach the same value at L = k = 2, suggesting that the q-machine is allowing us to view an information property that is more fundamental to the process itself than the ϵ-machine is capable of.

There are two important future research questions : 'Why is it that C_q^+ and C_q^- converge to the same value?' and 'Does the value for C_q^∞ hold in special significance, and can it be calculated in another way?' It would be worthwhile to develop a proof that $C_q^+ = C_q^-$ generically. A possible answer to the second question is that C_q^∞ is (or is related to) the mutual information between the past and present for q-machines and that $I[S^+; S^-] \neq \mathbf{E}$ generally due to the effect of entanglement or other nonclassical correlations. This possibility and its ramifications are further addressed below in the discussion of entanglement in q-machines.

STATE TRACE ENTROPY

Another effort to understand q-machines involves studying the quantum analogs of the classical entropy measures. The block entropy has a quantum counterpart known as the state trace entropy and defined as

$$STE = S(\rho_w) \quad (8)$$

where $S(\rho)$ is the von Neumann entropy and $\rho_w = Tr_S(\rho)$ is a density matrix in the word space resulting from taking a partial trace of the q-machine matrix over the state space. The state trace entropy is calculated by creating the quantum version of a concatenation machine for a certain word length and then doing the partial trace and finding the resulting von Neumann entropy. If the partial trace is instead taken over the word space it results in C_μ regardless of word length. The matrix will be the same dimension as the number of classical states, and the eigenvalues are the probabilities of being in a given classical state. Taking a trace over the diagonalized matrix (as in (7)) is then equivalent to taking the Shannon entropy of the stationary state distribution, which is equal to C_μ by (4). The state trace entropy can be seen for a variety of processes in Figs. 9, 10, 11 and 12.

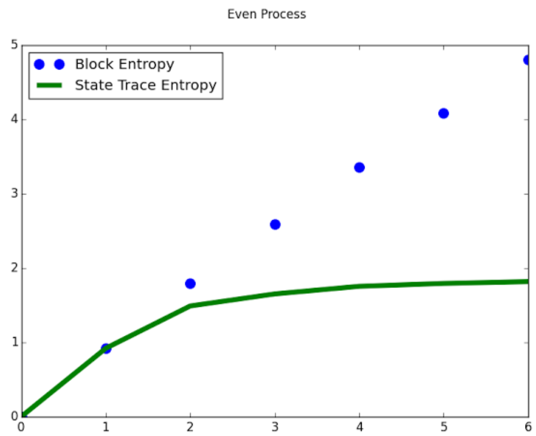

FIG. 9: State Trace Entropies for the Even Process

The behavior of this state trace entropy is interesting for a few reasons. It is bounded by the block entropy for every known process and always seems to asymptotically approach a value for these processes, but it is not yet known if this value is calculable in another way. For L=1 the state trace entropy is equal to the classical block entropy in most cases. The exceptions are when the ϵ-machine had isolated branching (two transitions from one state to another with different output symbols) in which case the block entropy was slightly greater. Unlike C_q, the STE of a process and its result do not approach the same value.

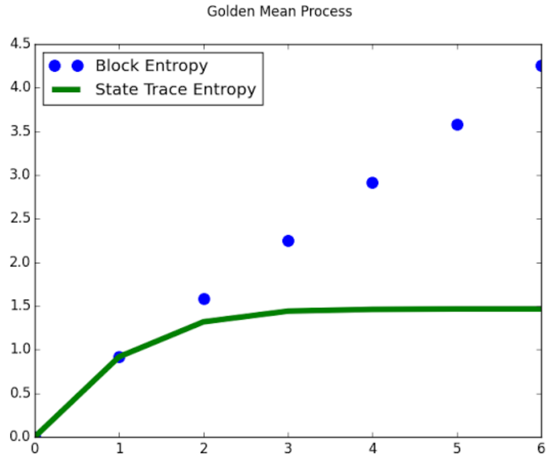

FIG. 10: State Trace Entropies for the Golden Mean Process

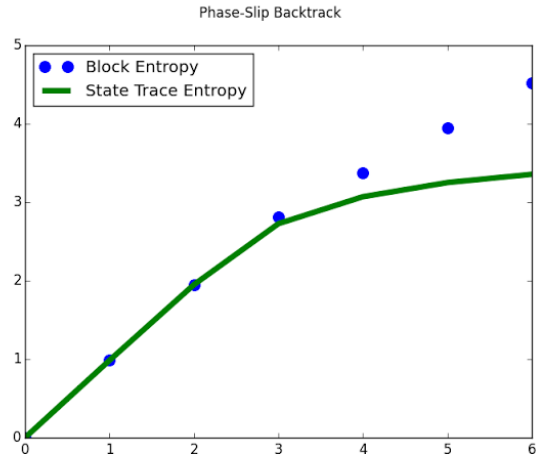

FIG. 11: State Trace Entropies for the Phase Slip Backtrack Process

The interpretation of the state trace entropy can be further illuminated by looking at the way it is calculated by eigenvalue decomposition. Given a quantum state's density matrix ρ, the eigenvalues of $Tr_S(\rho)$ determine the state trace entropy. For a classical machine of $|S|$ states, the number of non-zero eigenvalues increases to $|S|^2$ states as L approaches ∞. Thus a general upper bound on the state trace entropy for any machine is $-log(1/|S|^2)$. Each eigenvalue of $Tr_S(\rho)$ is associated with a state-state path through the classical ϵ-machine. Looking specifically at $Tr_S(\rho)$ diagonal terms that are non-zero correspond to possible words in the machine, and any that are zero correspond to impossible words.

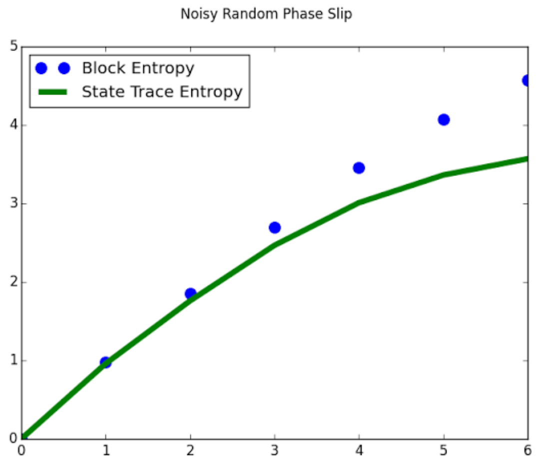

FIG. 12: State Trace Entropies for the Noisy Random Phase Slip Process

Off-diagonal terms appear when two words form separate paths with the same start and end. A machine with few paths will have $Tr_S(\rho)$ be a sparse matrix. For example, $Tr_S(\rho)$ for periodic processes is always a diagonal matrix for all L. Also, from this analysis it is clear that $Tr_S(\rho)$ is always a symmetric matrix.

For larger L, the dimension of $Tr_S(\rho)$ grows and more state-state paths are possible, resulting in more eigenvalues until an L is reached at which a word of that length can serve as a transition between any initial state and any final state. In some instances this is not possible and there are less than $|S|^2$ eigenvalues for every L. Periodic processes are one example due to very restricted transition possibilities. Generally increasing L results in more possible words connecting pairs of states, which explains the asymptotically increasing value of the state trace entropy. From the examples it is also obvious that the rate of increase in the state trace entropy is strongly dependent on the size of the machine. For two state machines (Figures 9 and 10) the state trace entropy tops out quickly as all eigenvalues are accounted for and all state-state paths are possible with words of small L. For larger machines (Figure 11 with four states and 12 with five states) the increase is more gradual as it takes longer words to connect each pair of states. Also machines with more transitions tend to approach their maximum value more quickly because state-state paths are easier to establish with shorter words.

Some interesting research questions going forward are 'Does the state trace entropy always approach a finite value that can be calculated in another way?' and 'Is the state trace entropy and its ability to quantify the 'connectedness' of q-machines useful in physical applications?' Additionally a proof that it increases asymptotically and a more in-depth analysis of that upper bound would be useful.

ENTANGLEMENT IN QUANTUM AUTOMATA

Entanglement Measures

The ability of q-machines to store and communicate information more efficiently than classical ϵ-machines is likely due to their exploitation of non-classical correlations that make quantum systems unique. The Hilbert space of the q-machine is bipartite because it consists of both word and state subspaces. This can be represented as

$$H_{q-machine} = H^L_{word} \otimes H_\sigma. \tag{9}$$

The dimension of H_σ is $|S|$, the number of classical states in the ϵ-machine. The dimension of H_{word} is dependent on the length of words in the q-machine construction but is generally of size $|A|^L$ where $|A|$ is the size of the alphabet. The two subsystems are said to be entangled if the q-machine state cannot be represented as a conjunction of separable states of each.

There are many ways of quantifying the entanglement of bipartite systems, as described in [4], which define entanglement in slightly different ways. Two interesting measures that were considered are concurrence and squashed entanglement. Concurrence is limited because it is only defined for a system of two qubits, and thus it could only account for a very limited set of q-machine states. Squashed entanglement is more generic, but its computation is an NP-hard problem so it cannot be reasonably found for any non-trivial systems.

Entanglement of formation does not suffer from these limitations, though its calculation can be difficult to do generically, as is necessary for q-machine states. It is equal to the number of Bell States necessary to prepare a copy of ρ, which is found by

$$E_f(\rho) = \inf \sum_j p_j E(\Phi_j) \tag{10}$$

where the infimum is taken over all pure state decompositions of ρ, and $E(\Phi) = S(Tr_B |\Phi\rangle \langle\Phi|)$. Because there are an infinite number of pure state decompositions this calculation is not trivial, but it can be done through a minimization procedure.

Calculation of Entanglement in a q-machine

The method of calculating entanglement between the word and state subspaces of the q-machine is adapted from the method of [6], which applies to all qudit systems (those with d possible states). The algorithm randomly creates a pure state decomposition of ρ, finds its entanglement of formation and the gradient and then changes the decomposition to minimize the entanglement. In adapting the code used in [6], we found that it failed for even simple cases where an analytic calculation was valid. To solve this we implemented our own version using the same methods in an iPython notebook. This work was not completed but is a promising avenue for future research. It is also worth noting that the entanglement of formation between the states and words is not the only way that the q-machine could make use of non-classical correlations to increase efficiency. Other measures of entanglement between other components of the q-machine may be an easier and more fruitful way to quantify this information gap between the classical and quantum case, but this is a reasonable and doable first step.

The Quantum Bi-Directional Machine

One major motivation for calculating entanglement within q-machines is the potential applications for a quantum bi-directional machine. A classical bi-directional machine is capable of generating both the forward and reverse process, and each of its states corresponds to one state from the forward machine and one from the reverse. The excess entropy, **E**, can be computed using the bi-directional machine by (1). Thus the quantum bi-directional machine consists of not only the word and forward causal state subspaces but also the reserve state subspace. Its structure is then

$$H_{qbi} = H_{word}^{L} \otimes H_{\sigma^+} \otimes H_{\sigma^-}. \tag{11}$$

By replacing the measures of the classical excess entropy calculation with their quantum counterparts we arrive at the equation

$$S(\sigma^+;\sigma^-) = \mathbf{E} \tag{12}$$

where $S(X;Y)$ is the von Neumann mutual information and **E** is the classical excess entropy. It is possible that this relation will not hold in the quantum case due to non-classical correlations between the forward and reverse causal states. A mismatch would be notable and could be explained by calculating the entanglement between the two subsystems. By comparing this to the excess entropy of the process, it will hopefully be clear how the q-machine makes use of non-classical correlations. A likely candidate for the q-machine's extra efficiency is this entanglement between the forward and reverse causal states. In that case the entanglement of formation between them may be equal to the gap between C_q and C_μ. Simple examples of quantum bi-directional machines have been created but code does not yet exist to calculate them directly from a process' ϵ-machine generically. Additionally the lack of reliable code to calculate the entanglement of formation at present limits the ability to answer this question now. This result will be notable whether or not $S(\sigma^+;\sigma^-) = \mathbf{E}$ for q-machines.

SUMMARY AND CONCLUSIONS

Quantum automata, particularly q-machines, are an exciting emerging new model for viewing stochastic processes. Using some sort of non-classical correlation they perform more efficiently than ϵ-machines as measured by their respective statistical complexities.

In this report we summarize some studies into their properties. The first set of results involve the statistical complexity. C_q is generally lower than its classical counterpart and decreases monotonically with greater word length until it reaches a minimum at the cryptic order. Additionally C_q approaches the same value for the forward and reverse process, a symmetry which is not present in the classical formulation. The second area of study was the state trace entropy which was found to asymptotically approach a finite value for every machine studied and to correspond to the state-state paths within a machine. In a sense it measures a process or machine's 'connectedness' with more transitions resulting in a higher state trace entropy. The final subproject was the calculation of entanglement within the q-machine. Algorithms are partially developed to do this for any machine, and these studies could help answer questions about the source of the q-machine's efficiency. In addition, calculating entanglement and finding the mutual information between the quantum bi-directional machine's forward and reverse causal states has been started, and doing so would enable us to more fully quantify the q-machine's behavior.

These investigations further the idea that q-machines have special promise in data storage, processing and communications. Advances in quantum computing necessitate the development of new and interesting ways to use information skillfully over quantum channels. Because they exploit non-classical correlations for greater efficiency, q-machines could potentially serve many applications in technology related to quantum computing.

ACKNOWLEDGEMENTS

I thank Professor Jim Crutchfield, John Mahoney and Ryan James for their help and guidance with this summer REU project. I also thank the entire the research group at the Complexity Sciences Center in Davis and Rena Zieve, the REU coordinator. Many thanks as well to my fellow REU students and the Davis Student Co-op for making the summer enjoyable. Finally I want to acknowledge Zuni for his encouragement and support.

[1] J. Crutchfield Nature Physics, **8** 17-24 (2012)
[2] M. Gu, et. al. Nature Communications **3** 762 (2012)
[3] K. Wiesner, J. Crutchfield Physica D **237**:9 (2008) 1173-1195
[4] R. Horodecki, et. al. Rev. Mod. Phys. **81**, 865 (2009)
[5] J.R. Mahoney, C. Aghamohammadi, and J.P. Crutchfield Santa Fe Institute Working Paper 15-08-030. arxiv.org:1508.02760 [quant-ph].
[6] S. Ryu, W. Cai, A. Caro Physical Review A **77** 052312 (2008)