An Introduction to Algorithmic Information Theory and Quantum Mechanics

Samuel Epstein
samepst@jptheorygroup.org

April 17, 2023

Abstract

This article presents a survey of published and unpublished material of the intersection of Algorithmic Information Theory and Quantum Mechanics. It is, to the author's knowledge, the first of its type. Three different notions of the algorithmic content of quantum states are reviewed. Notions of algorithmic quantum typicality and mutual information are introduced. The relationship between algorithmic information and quantum measurements is explored. One of the surprising results is that an overwhelming majority of quantum (pure and mixed) states, when undertaking decoherence, will result in a classical probability with no algorithmic information. Thus most quantum states decohere into white noise. A quantum analouge of Martin Löf random sequence is reviewed. Algorithmic Information Theory presents new complications for the Many Worlds Theory, as it conflicts with the Independence Postulate. When algorithmically complicated processes are ruled out, measurements are required to produce distributions over quantum states that have cloneable information.

Contents

1 Introduction **3**

2 Background **5**
- 2.1 Algorithmic Information Theory 5
  - 2.1.1 Algorithmic Information Between Probabilities 6
- 2.2 Quantum Mechanics 7
  - 2.2.1 Quantum Operations 8

3 Quantum Complexity **9**
- 3.1 Three Meausres of the Algorithmic Content of Individual Quantum States 9
  - 3.1.1 BVL Complexity 9
  - 3.1.2 Vitányi Complexity 9
  - 3.1.3 Gács Complexity 10
- 3.2 Properties of Universal Matrix $\boldsymbol{\mu}$ and Gács Complexity 10
- 3.3 No Cloning Theorem 11
- 3.4 Addition Inequality 12
- 3.5 Subadditivity, Strong Subadditivity, Strong Superadditivity 13
- 3.6 Relation Between Complexities 14
  - 3.6.1 Vitányi Complexity and Gács Complexity 14
  - 3.6.2 BVL Complexity and Gács Complexity 15
- 3.7 Quantum EL Theorem 16
  - 3.7.1 Computable Projections 17
  - 3.7.2 Quantum Data Compression 17

4 Quantum Typicality **19**
- 4.1 Definition of Quantum Randomness Deficiency 19
  - 4.1.1 Uncomputable Mixed States 20
- 4.2 Conservation Over Quantum Operations 21
- 4.3 A Quantum Outlier Theorem 21
  - 4.3.1 Computable Projections 23

5 Quantum Information **25**
- 5.1 Definition of Quantum Algorithmic Information 25
- 5.2 Paucity of Self-Information 26
  - 5.2.1 Pure States 26
  - 5.2.2 Mixed States 27
- 5.3 Information Nongrowth 28
  - 5.3.1 Algorithmic No-Cloning Theorem 29

5.3.2 Purification 30
5.3.3 Decoherence 30

6 Quantum Measurements **32**
6.1 Typicality and Measurements 32
6.2 Information and Measurements 32
6.3 Algorithmic Contents of Measurements 33
6.3.1 Algorithmic Content of Decoherence 35
6.4 PVMs 35

7 Infinite Quantum Spin Chains **36**
7.1 Infinite Quantum Bit Sequences 36
7.1.1 NS Randomness 37
7.2 Closure Properties 37
7.3 Gács Complexity and NS Random Sequences 38
7.4 Encodings of States 39
7.5 Quantum Operation Complexity 39
7.6 Initial Segment Incompressibility 39
7.7 Quantum Ergodic Sources 41
7.8 Measurement Systems 43
7.9 NS Solovay States 44

8 The Many Worlds Theory **46**
8.1 Many Worlds Theory 46
8.1.1 Branching Worlds 47
8.1.2 Deriving the Born Rule 47
8.2 Violating the Independence Postulate 48
8.3 Conclusion 49

9 Conclusion **50**
9.1 Signals from Classical and Quantum Sources 50
9.2 Apriori Distribution 51
9.3 Measurement Before Information Cloning 52

A Information Between Basis States **54**

Chapter 1

Introduction

Classical information theory studies the communication of bits across a noisy channel. Quantum Information Theory (QIT) studies the kind of information ("quantum information") which is transmitted by microparticles from a preparation device (sender) to a measuring apparatus (receiver) in a quantum mechanical experiment—in other words, the distinction between carriers of classical and quantum information becomes essential. The notion of a qubit can be defined at an abstract level, without giving preference to any particular physical system such as a spin-12 particle or a photon. Qubits behave very differently than bits. To start, qubits can be in a linear superposition between 0 and 1. Qubits can have entanglement, where two objects at a distance become a single entity. The study of entanglement and in particular the question how it can be quantified is therefore a central topic within quantum information theory. However, due to the no-cloning theorem [WZ82], instant communication is not possible. Some other aspects of QIT are as follows.

1. **Quantum Computing**: includes hardware (quantum computers), software, algorithm such as Shor's factoring algorithm or Grover's algorithm, and applications.

2. **Quantum Communication**: quantum networking, quantum internet, quantum cryptography.

3. **Applications in Physics**: applications to convex optimizations, black holes, and exotic quantum phases of matter.

4. **Quantum Shannon Theory**: quantum channels, quantum protocols, quantum information and entropy.

One aspect of Quantum Shannon Theory (QST) that has had relatively little study is its relationship to Algorithmic Information Theory (AIT). AIT, in part, is the study of the information content of individual strings. A string is random if it cannot be compressed with respect to a universal Turing machine. This paper surveys the current state of research of QST and AIT and provides unpublished results from the author. Hopefully it will convince the reader that there is a fruitful area of research of QST and AIT. Some areas of this intersection include algorthmic content of quantum states, how typical a quantum state is with respect to a quantum source, and how to quantify the algorithmic content of a measurement. One can also gain further insight into quantum transformations, such as purification, decoherence, and approximations to quantum cloning.

As this survey will show, there are some aspects of AIT that directly transfer over to quantum mechanics. This includes comparable definitions of complexity, and conservation inequalities. In addition, there exist quantum versions of the EL Theorem, [Lev16, Eps19c] and the Outlier Theorem, [Eps21b]. However there are some aspects of AIT that are different in the context of quantum mechanics. This includes the fact the self information of most quantum pure states is zero, with $\mathbf{I}(|\psi\rangle : \psi) \approx 0$. This has implications on the algorithmic content of measurements and decoherence. The main areas covered in this article are

- **Chapter 2:** This chapter covers the background material needed for the article. A new definition, the algorithmic information between probabilities, is introduced and shown to have information non-growth properties with respect to randomized processing.
- **Chapter 3:** Three different algorithmic measures of quantum states are covered. Their properties are described, including an addition inequality, a Quantum EL Theorem, and a generalized no-cloning theorem. Multiple relationships between the complexities are proven.
- **Chapter 4:** The notion of the algorithmic typicality of one quantum state with respect to another quantum state is introduced. Typicality is conserved with respect to quantum operations. A quantum outlier theorem is proven. This states that non-exotic projections must have atypical pure states in their images.
- **Chapter 5:** The definition of quantum algorithmic information is introduced. Quantum information differs from classical algorithmic information in that an overwhelming majority of pure states have negligible self-information. Information is conserved over quantum operations, with implications to quantum cloning, quantum decoherence, and purification.
- **Chapter 6:** Quantum algorithmic information upper bounds the amount of classical information produced by quantum measurements. Given a quantum measurement, for an overwhelming majority of pure states, the measurement will be random noise. An overwhelming majority of quantum pure states , when undertaking decoherence, will result in a classical probability with no algorithmic information.
- **Chapter 7:** A quantum equivalent to Martin Löf random sequence is introduced. Such quantum random states have incompressible initial segments with respect to a new measure quantum complexity called Quantum Operation Complexity. This complexity term measures the cost of approximating a state with a classical and quantum component.
- **Chapter 8:** This chapter shows the Many Worlds Theory and AIT are in conflict, as shown through the existence of a finite experiment that measures the spin of a large number of electrons. After the experiment there are branches of positive probability which contain forbidden sequences that break the Independence Postulate, a postulate in AIT.
- **Chapter 9:** This chapter concludes the paper with a discussion of the boundary between quantum information and classical information. We show that measurements are necessary to produce distributions over quantum states that have cloneable information.
- **Appendix A:** Properties of the quantum information of basis states are proven.

Chapter 2

Background

The reader is assumed to be familiar with both algorithmic information theory and quantum information theory, but we review the core terms. We use $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\{0,1\}$, $\{0,1\}^*$, and $\{0,1\}^\infty$ to denote natural numbers, integers, rational numbers, reals, bits, finite strings, and infinite sequences. $\{0,1\}^{*\infty} \stackrel{\text{def}}{=} \{0,1\}^* \cup \{0,1\}^\infty$. $\|x\|$ denotes the length of the string. Let $X_{\geq 0}$ and $X_{>0}$ be the sets of non-negative and of positive elements of X. $[A] \stackrel{\text{def}}{=} 1$ if statement A holds, else $[A] \stackrel{\text{def}}{=} 0$. The set of finite bit-strings is denoted by $\{0,1\}^*$. For set of strings $A \subseteq \{0,1\}^*$, $[\![A]\!] = \{x\alpha : x \in A, \alpha \in \{0,1\}^\infty\}$. When it is clear from the context, we will use natural numbers and other finite objects interchangeably with their binary representations.

The ith bit of $\alpha \in \{0,1\}^{*\infty}$ is denoted α_i, and its n bit prefix is denoted $\alpha_{\leq n}$. $\langle x \rangle \in \{0,1\}^*$ for $x \in \{0,1\}^*$ is the self-delimiting code that doubles every bit of x and changes the last bit of the result. For positive real functions f, by $<^+ f$, $>^+ f$, $=^+ f$, and $<^{\log} f$, $>^{\log} f$, $\sim f$ we denote $\leq f+O(1)$, $\geq f-O(1)$, $= f\pm O(1)$ and $\leq f+O(\log(f+1))$, $\geq f - O(\log(f+1))$, $= f\pm O(\log(f+1))$. Furthermore, $\stackrel{*}{<} f$, $\stackrel{*}{>} f$ denotes $< O(1)f$ and $> f/O(1)$. The term and $\stackrel{*}{=} f$ is used to denote $\stackrel{*}{>} f$ and $\stackrel{*}{<} f$.

A probability measure Q over $\{0,1\}^*$ is elementary if it has finite support and range that is a subset of rationals. Elementary probability measures can be encoded into finite strings $\langle Q \rangle$ in the standard way.

2.1 Algorithmic Information Theory

$T_y(x)$ is the output of algorithm T (or $\perp$ if it does not halt) on input $x \in \{0,1\}^*$ and auxiliary input $y \in \{0,1\}^{*\infty}$. T is prefix-free if for all $x, s \in \{0,1\}^*$ with $s \neq \emptyset$, either $T_y(x) = \perp$ or $T_y(xs) = \perp$. The complexity of $x \in \{0,1\}^*$ with respect to T_y is $\mathbf{K}_T(x|y) \stackrel{\text{def}}{=} \inf\{\|p\| : T_y(p) = x\}$.

There exist optimal for $\mathbf{K}$ prefix-free algorithms U, meaning that for all prefix-free algorithms T, there exists $c_T \in \mathbb{N}$, where $\mathbf{K}_U(x|y) \leq \mathbf{K}_T(x|y) + c_T$ for all $x \in \{0,1\}^*$ and $y \in \{0,1\}^{*\infty}$. For example, one can take a universal prefix-free algorithm U, where for each prefix-free algorithm T, there exists $t \in \{0,1\}^*$, with $U_y(tx) = T_y(x)$ for all $x \in \{0,1\}^*$ and $y \in \{0,1\}^{*\infty}$. $\mathbf{K}(x|y) \stackrel{\text{def}}{=} \mathbf{K}_U(x|y)$ is the Kolmogorov complexity of $x \in \{0,1\}^*$ relative to $y \in \{0,1\}^{*\infty}$.

The algorithmic probability is $\mathbf{m}(x|y) = \sum\{2^{-\|p\|} : U_y(p) = x\}$. By the coding theorem $\mathbf{K}(x|y) =^+ -\log \mathbf{m}(x|y)$. The amount of mutual information between two strings x and y is $\mathbf{I}(x : y) = \mathbf{K}(x) + \mathbf{K}(y) - \mathbf{K}(x,y)$. By the chain rule $\mathbf{K}(x,y) =^+ \mathbf{K}(x) + \mathbf{K}(y|x, \mathbf{K}(x))$. The halting sequence $\mathcal{H} \in \{0,1\}^\infty$ is the infinite string where $\mathcal{H}_i \stackrel{\text{def}}{=} [U(i)$ halts$]$ for all $i \in \mathbb{N}$. The amount of information that $\mathcal{H}$ has about $x \in \{0,1\}^*$ is $\mathbf{I}(x;\mathcal{H}) = \mathbf{K}(x) - \mathbf{K}(x|\mathcal{H})$. The randomness deficiency of $x \in \{0,1\}^*$ with respect to elementary probability P over $\{0,1\}^*$ is $\mathbf{d}(x|P) =$

$\lfloor -\log P(x) - \mathbf{K}(x|\langle P\rangle)\rfloor$. we say $t : \{0,1\}^* \to \mathbb{R}_{\geq 0}$ is a P-test, for some probability P, if $\sum_x t(x)P(x) \leq 1$. Let $\mathbf{t}_P$ be a universal lower computable P-test, where for any other lower computable P-test t, $\mathbf{t}_P(x) \stackrel{*}{>} \mathbf{m}(t)t(x)$. Then by the universality of the deficiency of randomness, [Gó1], $\mathbf{d}(x|P) =^+ \log \mathbf{t}_P(x)$. The transform of a probability Q by $f : \{0,1\}^* \to \{0,1\}^*$, is the probability fQ, where $fQ(x) = \sum_{f(y)=x} Q(y)$. Both randomness deficiency and information enjoy conservation inequalities.

Theorem 1 (See [Gó1]) $\mathbf{d}(f(x)|fQ) <^+ \mathbf{d}(x|Q)$.

Theorem 2 ([Lev84]) $\mathbf{I}(f(x) : y) <^+ \mathbf{I}(x : y)$.

Lemma 1 ([Eps22]) $\mathbf{I}(f(a); \mathcal{H}) <^+ \mathbf{I}(a; \mathcal{H}) + \mathbf{K}(f)$.

Lemma 2 *Let ψ_a be an enumerable semi-measure, semi-computable relative to a.* $\sum_c 2^{\mathbf{I}(\langle a,c\rangle:b)}\psi_a(c) \stackrel{*}{<} 2^{\mathbf{I}(a:b)}/\mathbf{m}(\psi)$.

Proof. This requires a slight modification of the proof of Proposition 2 in [Lev84], by requiring ψ to have a as auxilliary information. For completeness, we reproduce the proof. We need to show $\mathbf{m}(a,b)/(\mathbf{m}(a)\mathbf{m}(b)) \stackrel{*}{>} \sum_c(\mathbf{m}(a,b,c)/(\mathbf{m}(b)\mathbf{m}(a,c)))\mathbf{m}(\psi)\psi_a(c)$, or $\sum_c(\mathbf{m}(a,b,c)/\mathbf{m}(a,c))\mathbf{m}(c|a) \stackrel{*}{<} \mathbf{m}(a,b)/\mathbf{m}(a)$, since $\mathbf{m}(c|a) \stackrel{*}{>} \mathbf{m}(\psi)\psi_a(c)$. Rewrite it $\sum_c \mathbf{m}(c|a)\mathbf{m}(a,b,c)/\mathbf{m}(a,c) \stackrel{*}{<} \mathbf{m}(a,b)/\mathbf{m}(a)$ or $\sum_c \mathbf{m}(c|a)\mathbf{m}(a)\mathbf{m}(a,b,c)/\mathbf{m}(a,c) \stackrel{*}{<} \mathbf{m}(a,b)$. The latter is obvious since $\mathbf{m}(c|a)\mathbf{m}(a) \stackrel{*}{<} \mathbf{m}(a,c)$ and $\sum_c \mathbf{m}(a,b,c) \stackrel{*}{<} \mathbf{m}(a,b)$. $\square$

The stochasticity of a string $x \in \{0,1\}^*$ is $\mathbf{Ks}(x) = \min_{\text{Elementary } Q} \mathbf{K}(Q) + 3\log\max\{\mathbf{d}(x|Q), 1\}$. Strings with high stochsticity measures are exotic, in that they have high mutual information with the halting sequence.

Lemma 3 ([Lev16, Eps21b]) $\mathbf{Ks}(x) <^{\log} \mathbf{I}(x; \mathcal{H})$.

2.1.1 Algorithmic Information Between Probabilities

We can generalize from information from strings to information between arbitrary probability measures over strings.

Definition 1 (Information, Probabilities)
For semi-measures p and q over $\{0,1\}^*$, $\mathbf{I}_{\text{Prob}}(p : q) = \log \sum_{x,y\in\{0,1\}^*} 2^{\mathbf{I}(x:y)}p(x)q(y)$.

Definition 2 (Channel) *A channel $f : \{0,1\}^* \times \{0,1\}^* \to \mathbb{R}_{\geq 0}$ has $f(\cdot|x)$ being a probability measure over $\{0,1\}^*$ for each $x \in \{0,1\}^*$. For probability p, channel f, $fp(x) = \sum_z f(x|z)p(z)$.*

Theorem 3 *For probabilities p and q over $\{0,1\}^*$, computable channel f,* $\mathbf{I}_{\text{Prob}}(fp : q) <^+ \mathbf{I}_{\text{Prob}}(p : q)$.

Proof. Using Lemma 1,

$$\mathbf{I}_{\text{Prob}}(fp : q) = \log \sum_{x,y} 2^{\mathbf{I}(x:y)} \sum_z f(x|z)p(z)q(y) <^+ \log \sum_{y,z} q(y)p(z) \sum_x 2^{\mathbf{I}((x,z):y)} f(x|z).$$

Using Lemma 2,

$$\mathbf{I}_{\mathrm{Prob}}(fp : q) <^{+} \log \sum_{z,y} q(y)p(z)2^{\mathbf{I}(z:y)} =^{+} \mathbf{I}_{\mathrm{Prob}}(p : q).$$

□

Thus processing cannot increase information between two probabilities. If the the probability measure is concentrated at a single point, then it contains self-information equal to the complexity of that point. If the probability measure is spread out, then it is white noise, and contains no self-information. Some examples are as follows.

Example 1

- *In general, a probability p, will have low $\mathbf{I}_{\mathrm{Prob}}(p : p)$ if it has large measure on simple strings, or low measure on a large number of complex strings, or some combination of the two.*
- *If probability p is concentrated on a single string x, then $\mathbf{I}_{\mathrm{Prob}}(p : p) = \mathbf{K}(x)$.*
- *The uniform distribution over strings of length n has self information equal to (up to an additive constant) $\mathbf{K}(n)$.*
- *There are semi-measures that have infinite self information. Let α_n be the n bit prefix of a Martin Löf random sequence α and $n \in [2, \infty)$. Semi-measure $p(x) = [x = \alpha_n]n^{-2}$ has $\mathbf{I}_{\mathrm{Prob}}(p : p) = \infty$.*
- *The universal semi-measure $\mathbf{m}$ has no self information.*

Example 2 (Uniform Spread) *An example channel f has $f(\cdot|x)$ be the uniform distribution over strings of length $\|x\|$. This is a cannonical spread function. Thus if p is a probability measure concentrated on a single string, then $\mathbf{I}_{\mathrm{Prob}}(p : p) = \mathbf{K}(x)$, and $\mathbf{I}(fp : fp) =^{+} \mathbf{K}(\|x\|)$. Thus f results in a decrease of self-information of p. This decrease of information occurs over all probabilities and computable channels.*

2.2 Quantum Mechanics

We use the standard model of qubits used throughout quantum information theory. We deal with finite N dimensional Hilbert spaces $\mathcal{H}_N$, with bases $|\alpha_1\rangle, |\alpha_2\rangle, \dots, |\alpha_n\rangle$. We assume $\mathcal{H}_{n+1} \supseteq \mathcal{H}_n$ and the bases for $\mathcal{H}_n$ are the beginning of that of $\mathcal{H}_{n+1}$. An n qubit space is denoted by $\mathcal{Q}_n = \bigotimes_{i=1}^{n} \mathcal{Q}_1$, where qubit space $\mathcal{Q}_1$ has bases $|0\rangle$ and $|1\rangle$. For $x \in \Sigma^n$ we use $|x\rangle \in \mathcal{Q}_n$ to denote $\bigotimes_{i=1}^{n} |x[i]\rangle$. The space $\mathcal{Q}_n$ has 2^n dimensions and we identify it with $\mathcal{H}_{2^n}$.

A pure quantum state $|\phi\rangle$ of length n is represented as a unit vector in $\mathcal{Q}_n$. Its corresponding element in the dual space is denoted by $\langle\phi|$. The tensor product of two vectors is denoted by $|\phi\rangle \otimes |\psi\rangle = |\phi\rangle |\psi\rangle = |\phi\psi\rangle$. The inner product of $|\psi\rangle$ and $\langle\phi|$ is denoted by $\langle\psi|\phi\rangle$.

The symbol Tr denotes the trace operation. The conjugate transpose of a matrix M is denoted by M^*. For Hermitian matrix with eigenvalue decomposition $A = \sum a_i |\psi_i\rangle \langle\psi_i|$, $|A| = \sum |a_i| |\psi_i\rangle \langle\psi_i|$. The tensor product of two matrices is denoted by $A \otimes B$. Projection matrices are Hermitian matrices with eigenvalues in $\{0, 1\}$. For tensor product space $\mathcal{G}_X \otimes \mathcal{G}_Y$, the partial trace is denoted by Tr_Y. For $B^X = \mathrm{Tr}_Y B$, $\mathrm{Tr}(A \cdot B^X) = \mathrm{Tr}((A \otimes I) \cdot B)$, which is used frequently throughout the paper. For positive positive semidefinite matrices, σ and ρ we say $\sigma \leq \rho$ if $\rho - \sigma$ is positive semidefinite. For positive semidefinite matrices A, B, C, if $A \leq B$ then $\mathrm{Tr}AC \leq \mathrm{Tr}BC$. Mixed states are represented by density matrices, which are, self adjoint, positive semidefinite, operators of trace 1. A semi-density matrix has non-negative trace less than or equal to 1. The von Neumann entropy of a density matrix σ with orthogonal decomposition $\sum p_i |\psi_i\rangle\langle\psi_i|$ is $S(\sigma) = -\sum p_i \log p_i$.

A pure quantum state $|\phi\rangle$ and (semi)density matrix σ are called *elementary* if their real and imaginary components have rational coefficients. Elementary objects can be encoded into strings or integers and be the output of halting programs. Therefore one can use the terminology $\mathbf{K}(|\phi\rangle)$ and $\mathbf{K}(\sigma)$, and also $\mathbf{m}(|\phi\rangle)$ and $\mathbf{m}(\sigma)$.

We say program $q \in \{0,1\}^*$ lower computes positive semidefinite matrix σ if, given as input to universal Turing machine U, the machine U reads $\leq \|q\|$ bits and outputs, with or without halting, a sequence of elementary semi-density matrices $\{\sigma_i\}$ such that $\sigma_i \leq \sigma_{i+1}$ and $\lim_{i\to\infty} \sigma_i = \sigma$. A matrix is lower computable if there is a program that lower computes it.

2.2.1 Quantum Operations

A quantum operation is the most general type of operation than can be applied to a quantum state. In Chapters 4 and 5, conservation inequalities will be proven with respect to quantum operations. A map transforming a quantum state σ to $\varepsilon(\sigma)$ is a quantum operation if it satisfies the following three requirements

1. The map of ε is positive and trace preserving, with $\mathrm{Tr}(\sigma) = \mathrm{Tr}(\varepsilon(\sigma))$.
2. The map is linear with $\varepsilon(\sum_i p_i\sigma_i) = \sum_i p_i\varepsilon(\sigma_i)$.
3. The map is completely positive, were any map of the form $\varepsilon \otimes \mathbf{M}$ acting on the extended Hilbert space is also positive.

Another means to describe quantum operations is through a series of operators. A quantum operation ε on mixed state σ_A can be seen as the appending of an ancillia state σ_b, applying a unitary transform U, then tracing out the ancillia system with

$$\varepsilon(\sigma_A) = \mathrm{Tr}_B\left(U(\sigma_A \otimes \sigma_B)U^*\right). \tag{2.1}$$

A third way to characterize a quantum operation is through Kraus operators, which can be derived using an algebraic reworking of Equation 2.1. Map ε is a quantum operation iff it can be represented (not necessarily uniquely) using a set of matrices $\{M_i\}$ where $\varepsilon(\sigma) = \sum_i M_i \varepsilon M_i^*$ and $\sum_i M_i^* M_i \leq I$, where I is the identity matrix.

A quantum operation ε is elementary iff it admits a represented of the form in Equation 2.1 where B, U, and σ_B are each elementary, in that they each can be encoded with a finite string. The encoding of an elementary quantum operation is denoted by $\langle\varepsilon\rangle = \langle B\rangle\langle U\rangle\langle\sigma_B\rangle$. Each elementary quantum operation admits an elementary Kraus operator representation $\{M_i\}$, in that each M_i is an elementary matrix. This elementary Kraus operator is computable from $\langle\varepsilon\rangle$.

Chapter 3

Quantum Complexity

3.1 Three Meausres of the Algorithmic Content of Individual Quantum States

The formal study of Algorithmic Information Theory and Quantum Mechanics began with the introduction of three independent measures of the algorithmic content of a mixed or pure quantum state, detailed in the papers [BvL01, G0́1, Vit01]. BVL complexity [BvL01] measures the complexity of a pure quantum state $|\psi\rangle$ by the length of the smallest input to a universal quantum Turing machine that outputs a good approximation of $|\psi\rangle$. Vitányi complexity [Vit01] measures the entropy of a pure state $|\psi\rangle$ as the amount of classical information needed to reproduce a good approximation of $|\psi\rangle$. Gács complexity measures the entropy of a pure or mixed quantum state by using a quantum analouge of of the universal semi-measure $\mathbf{m}$.

3.1.1 BVL Complexity

BVL complexity, introduced in [BvL01], uses a universal quantum Turing machine to define the complexity of a pure quantum state. The input and output tape of this machine consists of symbols of the type 0, 1, and #. The input is an ensemble $\{p_i\}$ of pure states $|\psi_i\rangle$ of the same length n, where $p_i \geq 0$ and $\sum_i p_i = 1$. This ensemble can be represented as a mixed state of n qubits. If, during the operation of the quantum Turing machine, all computational branches halt at a time t, then the contents on the output tape are considered the output of the quantum Turing machine. The BVL Complexity of a pure state, $\mathbf{Hbvl}[\epsilon](|\psi\rangle)$ is the size of the smallest (possibly mixed state) input to a universal quantum Turing machine such that fidelity between the output and $|\psi\rangle$ is at least ϵ. The fidelity between a mixed state output σ and $|\psi\rangle$ is $\langle\psi|\,\sigma\,|\psi\rangle$. We require that the input quantum state be elementary. We also require that universal quantum Turing machine be conditioned on the number of qubits n, on a classical auxillary tape.

3.1.2 Vitányi Complexity

Vitányi complexity [Vit01] is a measure of the algorithmic information content of a pure state $|\psi\rangle$. It is equal to the minimum of the Kolmogorov complexity of an approximating elementary pure state $|\phi\rangle$ summed with a score of their closeness. We use a slightly different definition than the original [Vit01], in that we use a classical machine and not a quantum machine for the approximation. Let N be the dimension of the Hilbert space.

$$\mathbf{Hv}(|\psi\rangle) = \min_{\text{Elementary } |\theta\rangle \in \mathcal{H}_N} \mathbf{K}(|\theta\rangle\,|N) - \log|\langle\psi|\theta\rangle|^2.$$

3.1.3 Gács Complexity

Gács complexity [GÓ1] is defined using the following universal lower computable semi-density matrix, parametered by $x \in \{0,1\}^*$, with

$$\boldsymbol{\mu}_x = \sum_{\text{Eementary } |\phi\rangle \in \mathcal{H}_N} \mathbf{m}(|\phi\rangle\,|x, N)\,|\phi\rangle\langle\phi|.$$

The parameter N represents the dimension of the Hilbert space. We use $\boldsymbol{\mu}_X$ to denote the matrix $\boldsymbol{\mu}$ over the Hilbert space denoted by symbol X. The Gács entropy of a mixed state σ, conditioned on $x \in \{0,1\}^*$ is defined by

$$\mathbf{Hv}(\sigma|x) = \lceil -\log \mathrm{Tr}\boldsymbol{\mu}_x\sigma \rceil.$$

We use the following notation for pure states, with $\mathbf{Hg}(|\phi\rangle\,|x) = \mathbf{Hg}(\,|\phi\rangle\langle\phi|\,|x)$. For empty x we use the notaion $\mathbf{Hg}(\sigma)$. This definition generalizes $\underline{H}$ in [GÓ1] to mixed states. Note than in [GÓ1], there is another measure of quantum algorithmic entropy, $\overline{H}$, which we will not cover in this paper. An infinite version of algorithmic entropy can be found at [BOD14].

3.2 Properties of Universal Matrix $\boldsymbol{\mu}$ and Gács Complexity

The matrix $\boldsymbol{\mu}$ is important in algorithmic information theory and quantum mechanics, as it is the foundation for the information term defined in Chapter 5. The following theorem shows that the lower computable semi-density matrix $\boldsymbol{\mu}$ is universal. It is greater than any other lower computable matrix, weighted by their complexity. This parallels the classical case, where universal measure $\mathbf{m}$ majorizes lower computable semi measure p, with $\mathbf{m}(x) \overset{*}{>} \mathbf{m}(p)p(x)$. This theorem is used throughout the paper, and will not be explicitly cited.

Theorem 4 **([GÓ1])** *Let $q \in \{0,1\}^*$, and the dimension of the Hilbert space, N, compute lower compute semi-density matrix A. Then $\boldsymbol{\mu} \overset{*}{>} \mathbf{m}(q|N)A$.*

Proof. A can be composed into a sum $\sum_i p(i)\,|\psi_i\rangle\,\langle\psi_i|$, where each $|\psi_i\rangle$ is elementary, p is a semi-measure, with $\sum_i p(i) \le 1$, and p is computable from q. Thus since p is computable from q and N,

$$A = \sum_i p(i)\,|\psi_i\rangle\,\langle\psi_i| \overset{*}{<} \mathbf{m}(p|N)^{-1}\sum_i \mathbf{m}(i|N)\,|\psi_i\rangle\,\langle\psi_i| \overset{*}{<} \mathbf{m}(q|N)^{-1}\sum_i \mathbf{m}(i|N)\,|\psi_i\rangle\,\langle\psi_i| \overset{*}{<} \boldsymbol{\mu}/\mathbf{m}(q|N).$$

□

Theorem 5 **([GÓ1])** $\boldsymbol{\mu}_{ii} \overset{*}{=} \mathbf{m}(i|N)$.

Proof. The matrix $\rho = \sum_i \mathbf{m}(i|N)\,|i\rangle\,\langle i|$ is lower computable, so $\rho \overset{*}{<} \boldsymbol{\mu}$ so $\boldsymbol{\mu}_{ii} \overset{*}{>} \mathbf{m}(i|N)$. Furthermore, $f(i) = |i\rangle\,\boldsymbol{\mu}\,\langle i|$ is a lower computable semi-measure, so $\mathbf{m}(i|N) \overset{*}{>} \boldsymbol{\mu}_{ii}$. □

Theorem 6 **([GÓ1])** $\mathrm{Tr}_Y\boldsymbol{\mu}_{XY} \overset{*}{=} \boldsymbol{\mu}_X$.

Proof. Let $\rho = \mathrm{Tr}_Y \boldsymbol{\mu}_{XY}$, which is a lower computable semi-density matrix because one can enumerate elementary pure states $|\psi\rangle\langle\psi|$ in the space XY, take their partial trace, $\mathrm{Tr}_T |\psi\rangle\langle\psi|$, and add the resulting pure or mixed state to the sum ρ. Thus $\rho \overset{*}{<} \boldsymbol{\mu}_X$. Let $\sigma = \boldsymbol{\mu}_X \otimes |\psi\rangle\langle\psi|$, where $|\psi\rangle$ is a reference elementary state. Thus $\sigma \overset{*}{<} \boldsymbol{\mu}_{XY}$ so

$$\boldsymbol{\mu}_X = \mathrm{Tr}_Y \sigma \overset{*}{<} \mathrm{Tr}_Y \boldsymbol{\mu}_{XY}.$$

$\square$

Theorem 7 ([Gó1]) *$\mathbf{Hg}(\sigma) <^+ \mathbf{Hg}(\sigma \otimes \rho)$.*

Proof. Note that this theorem is not less general than that of Theorem 9, because both σ and ρ can be non-elementary. Using Theorem 6 and the properties of partial trace,

$$2^{-\mathbf{Hg}(\sigma)} \overset{*}{>} \mathrm{Tr}\sigma\boldsymbol{\mu}_X \overset{*}{>} \mathrm{Tr}\sigma\mathrm{Tr}_Y\boldsymbol{\mu}_{XY} \overset{*}{>} \mathrm{Tr}(\sigma \otimes I)\boldsymbol{\mu}_{XY} \overset{*}{>} \mathrm{Tr}(\sigma \otimes \rho)\boldsymbol{\mu}_{XY} \overset{*}{=} 2^{-\mathbf{Hg}(\sigma\otimes\rho)}.$$

3.3 No Cloning Theorem

In classical algorithmic information theory, one can easily reproduce a string x, with

$$\mathbf{K}(x) =^+ \mathbf{K}(x, x).$$

However the situation is much different in the quantum case. Due to the no-cloning theorem, [WZ82] one cannot clone a quantum state. The following theorem generalizes this no-go result, by showing there exist tensor products $|\psi\rangle^m$ that has signifigantly more $\mathbf{Hg}$ measure than $|\psi\rangle$. The following theorem presents a new proof to this result.

Theorem 8 ([Gó1]) *$\log\binom{m+N-1}{m} <^+ \max_{|\psi\rangle} \mathbf{Hg}(|\psi\rangle^{\otimes m}) <^+ \mathbf{K}(m) + \log\binom{m+N-1}{m}$*

Proof. Let $\mathcal{H}_N$ be an N dimensional Hilbert space and let $\mathcal{H}_N^m$ be an m-fold tensor space of $\mathcal{H}_N$. Let $\mathrm{Sym}(\mathcal{H}_N^m)$ be the subspace of $\mathcal{H}_N^m$ consisting of all pure states of the form $|\psi\rangle^{\otimes m}$. The subspace $\mathrm{Sym}(\mathcal{H}_N^m)$ is spanned by M basis vectors, where M is the number of multisets of size m from the set $\{1, \ldots, N\}$. This is because for each such multiset $S = \{i_1, \ldots, i_m\}$, one can construct a basis vector $|\psi_S\rangle$ that is the normalized superposition of all basis vectors of $\mathrm{Sym}(\mathcal{H}_N^m)$ that are permutations of S. If $S' \neq S$, then $|\psi_S\rangle$ is orthogonal to $|\psi_{S'}\rangle$. Thus the dimension of $\mathrm{Sym}(\mathcal{H}_N^m)$ M, is $\binom{m+N-1}{m}$ because choosing a multiset is the same as splitting an interval of size m into N intervals. For the upper bounds, let P_S be the projector onto $\mathrm{Sym}(\mathcal{H}_N^m)$. If $|\psi\rangle \in \mathrm{Sym}(\mathcal{H}_N^m)$, then $\langle\psi| P_S |\psi\rangle = 1$ so

$$\mathbf{Hg}(|\psi\rangle) <^+ \mathbf{K}(P_S/M|N^m) - \log\langle\psi|\frac{1}{M}P_S/M|\psi\rangle <^+ \mathbf{K}(m) + \log\binom{m+N-1}{m}.$$

For the lower bound, let $c = \max_{|\psi\rangle\in\mathcal{H}_N} \mathbf{Hg}(|\psi\rangle^{\otimes m})$. We have for all $|\psi\rangle \in \mathcal{H}_N$

$$\mathrm{Tr}\boldsymbol{\mu}|\psi\rangle^m\langle\psi|^m \overset{*}{>} 2^{-c}. \tag{3.1}$$

Let Λ be the uniform distribution on the unit sphere of $\mathcal{H}_N$. And let

$$\rho = \int |\psi\rangle^m \langle\psi|^m \, d\Lambda.$$

$\mathrm{Tr}\rho = \int \mathrm{Tr}\,|\psi\rangle^m \langle\psi|^m \, d\Lambda = \int d\Lambda = 1$. Furthermore for $|\phi\rangle^m, |\nu\rangle^m \in \mathrm{Sym}(\mathcal{H}_N^m)$, with unitary transform U such that $U^m |\psi\rangle^m = |\rho\rangle^m$, we have

$$\langle\nu|^n \rho\, |\nu\rangle^n = \int \langle\phi|^m \left(U^{*m} |\psi\rangle^m \langle\psi|^m U^m\right) |\phi\rangle^m \, d\Lambda = \int \langle\phi^m|\psi^m\rangle \langle\psi^m|\phi^m\rangle^m \, d\Lambda = \langle\phi|^n \rho\, |\phi\rangle^n .$$

For any pure state $|\psi\rangle \in \mathcal{H}_N^m$, such that $\langle\psi| P_S |\psi\rangle = 0$, then $\langle\psi| \rho |\psi\rangle = 0$. Thus $\rho = P_S/M$. Integrating Equation 3.1, by $d\Lambda$ results in

$$2^{-c} \overset{*}{<} \mathrm{Tr}\boldsymbol{\mu}\rho \overset{*}{=} \mathrm{Tr}\boldsymbol{\mu}P_S/M \overset{*}{=} \binom{m+N-1}{m}^{-1}$$
$$c >^+ \log \binom{m+N-1}{m}.$$

□

3.4 Addition Inequality

The addition theorem for classical entropy asserts that the joint entropy for a pair of random variables is equal to the entropy of one plus the conditional entropy of the other, with $\mathcal{H}(\mathcal{X}) + \mathcal{H}(\mathcal{Y}/\mathcal{X}) = \mathcal{H}(\mathcal{X}, \mathcal{Y})$. For algorithmic entropy, the chain rule is slightly more nuanced, with $\mathbf{K}(x) + \mathbf{K}(y|x, \mathbf{K}(x)) =^+ \mathbf{K}(x, y)$. An analogous relationship cannot be true for Gács entropy, $\mathbf{Hg}$, since as shown in Theorem 8, there exists elementary $|\phi\rangle$ where $\mathbf{Hg}(|\phi\rangle\,|\phi\rangle) - \mathbf{Hg}(|\phi\rangle)$ can be arbitrarily large, and $\mathbf{Hg}(|\phi\rangle\,/\,|\phi\rangle) =^+ 0$. However, the following theorem shows that a chain rule inequality does hold for $\mathbf{Hg}$.

For $n^2 \times n^2$ matrix A, let $A[i, j]$ be the $n \times n$ submatrix of A starting at position $(n(i-1)+1, n(j-1)+1)$. For example for $n = 2$ the matrix

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{bmatrix}$$

has $A[1,1] = \begin{bmatrix} 1 & 2 \\ 5 & 6 \end{bmatrix}$, $A[1,2] = \begin{bmatrix} 3 & 4 \\ 7 & 8 \end{bmatrix}$, $A[2,1] = \begin{bmatrix} 9 & 10 \\ 13 & 14 \end{bmatrix}$, $A[2,2] = \begin{bmatrix} 11 & 12 \\ 15 & 16 \end{bmatrix}$.

For $n^2 \times n^2$ matrix A and $n \times n$ matrix B, let M_{AB} be the $n \times n$ matrix whose (i, j) entry is equal to $\mathrm{Tr}A[i, j]B$. For any $n \times n$ matrix C, in can be seen that $\mathrm{Tr}A(C \otimes B) = \mathrm{Tr}M_{AB}C$. Furthermore if A is lower computable and B is elementary, then M_{AB} is lower computable.

For elementary semi density matrices ρ, we use $\langle\rho, \mathbf{Hg}(\rho)\rangle$ to denote the encoding of the pair of an encoded ρ and an encoded natural number $\mathbf{Hg}(\rho)$.

Theorem 9 **([Eps19a])** *For semi-density matrices σ, ρ, elementary ρ,*
$\mathbf{Hg}(\rho) + \mathbf{Hg}(\sigma|\langle\rho, \mathbf{Hg}(\rho)\rangle) <^+ \mathbf{Hg}(\sigma \otimes \rho)$.

Proof. Let $\boldsymbol{\mu}_{2n}$ be the universal lower computable semi density matrix over the space of 2n qubits, $\mathcal{Q}_{2n} = \mathcal{Q}_n \otimes \mathcal{Q}_n = \mathcal{Q}_A \otimes \mathcal{Q}_B$. Let $\boldsymbol{\mu}_n$ be the universal matrix of the space over n qubits. We define the following bilinear function over complex matrixes of size $n \times n$, with $T(\nu, \delta) = \mathrm{Tr}\boldsymbol{\mu}_{2n}(\nu \otimes \delta)$. For fixed ρ, $T(\nu, \rho)$ is of the form $T(\nu, \rho) = \mathrm{Tr} M_{\boldsymbol{\mu}_{2n}\rho}\nu$. The matrix $M_{\boldsymbol{\mu}_{2n}\rho}$ has trace equal to

$$\begin{aligned}\mathrm{Tr} M_{\boldsymbol{\mu}_{2n}\rho} &= T(\rho, I)\\ &= \mathrm{Tr}\boldsymbol{\mu}_{2n}(\rho \otimes I)\\ &= \mathrm{Tr}\left((\mathrm{Tr}_{\mathcal{Q}_B}\boldsymbol{\mu}_{2n})\rho\right)\\ &\stackrel{*}{=} \mathrm{Tr}\boldsymbol{\mu}_n\rho\\ &\stackrel{*}{=} 2^{-\mathbf{Hg}(\rho)},\end{aligned}$$

using Theorem 6, which states $\mathrm{Tr}_Y\boldsymbol{\mu}_{XY} \stackrel{*}{=} \boldsymbol{\mu}_X$. By the definition of M, since $\boldsymbol{\mu}_{2n}$ and ρ are positive semi-definite, it must be that $M_{\boldsymbol{\mu}_{2n}\rho}$ is positive semi-definite. Since the trace of $M_{\boldsymbol{\mu}_{2n}\rho}$ is $\stackrel{*}{=} 2^{-\mathbf{Hg}(\rho)}$, it must be that up to a multiplicative constant, $2^{\mathbf{Hg}(\rho)}M_{\boldsymbol{\mu}_{2n}\rho}$ is a semi-density matrix.

Since $\boldsymbol{\mu}$ is lower computable and ρ is elementary, by the definition of M, $2^{\mathbf{Hg}(\rho)}M_{\boldsymbol{\mu}_{2n}\rho}$ is lower computable relative to the string $\langle \rho, \mathbf{Hg}(\rho)\rangle$. Therefore we have that $2^{\mathbf{Hg}(\rho)}M_{\boldsymbol{\mu}_{2n}\rho} \stackrel{*}{<} \boldsymbol{\mu}_{\langle \rho, \mathbf{Hg}(\rho)\rangle}$. So we have that $-\log \mathrm{Tr} 2^{\mathbf{Hg}(\rho)}M_{\boldsymbol{\mu}_{2n}\rho}\sigma = -\mathbf{Hg}(\rho) - \log T(\sigma, \rho) =^+ \mathbf{Hg}(\sigma \otimes \rho) - \mathbf{Hg}(\rho) >^+ -\log \boldsymbol{\mu}_{\langle \rho, \mathbf{Hg}(\rho)\rangle}\sigma =^+ \mathbf{Hg}(\sigma | \langle \rho, \mathbf{Hg}(\rho)\rangle)$. □

3.5 Subadditivity, Strong Subadditivity, Strong Superadditivity

Theorem 10 ([GÓ1]) $\mathbf{Hg}(\sigma)$ *is subadditive, with* $\mathbf{Hg}(\sigma \otimes \rho) <^+ \mathbf{Hg}(\sigma) + \mathbf{Hg}(\rho)$.

Proof.

$$\begin{aligned}&2^{-\mathbf{Hg}(\sigma)-\mathbf{Hg}(\rho)}\\ =&\left(\mathrm{Tr}\boldsymbol{\mu}_X\sigma\right)\left(\mathrm{Tr}\boldsymbol{\mu}_Y\rho\right)\\ =&\mathrm{Tr}(\sigma \otimes \rho)(\boldsymbol{\mu}_X \otimes \boldsymbol{\mu}_Y)\\ \stackrel{*}{>}&\mathrm{Tr}(\sigma \otimes \rho)(\boldsymbol{\mu}_{XY})\\ \stackrel{*}{=}&2^{-\mathbf{Hg}(\sigma\otimes\rho)}.\end{aligned}$$

□

A function $\mathbf{L}$ from quantum mixed states to whole numbers is strongly subadditive if there exists a constant $c \in \mathbb{N}$ such that for all mixed states ρ_{123}, $\mathbf{L}(\rho_{123}) + \mathbf{L}(\rho_2) < \mathbf{L}(\rho_{12}) + \mathbf{L}(\rho_{23}) + c$. Similarly $\mathbf{L}$ is strongly superadditive if there exists a constant $c \in \mathbb{N}$ such that for all mixed states ρ_{123}, $\mathbf{L}(\rho_{12}) + \mathbf{L}(\rho_{23}) < \mathbf{L}(\rho_{123}) + \mathbf{L}(\rho_2) + c$.

Theorem 11 $\mathbf{Hg}$ *is not strongly subadditive.*

Proof. We fix the number of qubits n, and for $i \in [1..2^n]$, $|i\rangle$ is the ith basis state of the n qubit space. Let $|\psi\rangle = \sum_{i=1}^{2^n} 2^{-n/2}|i\rangle|i\rangle$. The pure state $|\psi\rangle$ is elementary, with $\mathbf{K}(|\psi\rangle|2^{2n}) =^+ 0$. We define the the $3n$ qubit mixed state $\rho_{123} = .5|\psi\rangle\langle\psi| \otimes |1\rangle\langle 1| + .5|1\rangle\langle 1| \otimes |\psi\rangle\langle\psi|$. $\rho_{12} = .5|\psi\rangle\langle\psi| + .5|1\rangle\langle 1| \otimes 2^{-n}I$. $\rho_{23} = .5 * 2^{-n}I \otimes |1\rangle\langle 1| + .5|\psi\rangle\langle\psi|$. $\rho_2 = 2^{-n}I$. $\mathbf{Hg}(\rho_{12}) =^+ -\log \mathrm{Tr}\boldsymbol{\mu}^{2n}\rho_{12} <^+ -\log \mathrm{Tr}\boldsymbol{\mu}^{2n}|\psi\rangle\langle\psi| <^+ -\log \mathbf{m}(|\psi\rangle|2^{2n})|\langle\psi|\psi\rangle|^2 <^+ 0$. Similarly, $\mathbf{Hg}(\rho_{23}) =^+ 0$. $\mathbf{Hg}(\rho_2) =^+ n$.

So $\mathbf{Hg}(\rho_{123}) + \mathbf{Hg}(\rho_2) >^+ n$ and $\mathbf{Hg}(\rho_{12}) + \mathbf{Hg}(\rho_{23}) =^+ 0$, proving that $\mathbf{Hg}$ is not strongly subadditive. □

Theorem 12 $\mathbf{Hg}$ *is not strongly superadditive.*

Proof. We fix the number of qubits n, and for $i \in [1..2^n]$, $|i\rangle$ is the ith basis state of the n qubit space. Let $|\phi\rangle = \sum_{i=1}^{2^n} 2^{-n/2} |i\rangle |i\rangle |i\rangle$, with $\mathbf{K}(|\phi\rangle |2^{3n}) = 0$. Let $\sigma_{123} = |\phi\rangle\langle\phi|$. $\sigma_{12} = \sigma_{23} = \sum_{i=1}^{2^n} 2^{-n} |i\rangle\langle i| \otimes |i\rangle\langle i|$. $\mathbf{Hg}(\sigma_{123}) =^+ -\log \mathrm{Tr}\sigma_{123}\boldsymbol{\mu}^{3n} <^+ -\log \mathrm{Tr}\,\mathbf{m}(|\phi\rangle |2^{3n})| \langle\phi|\phi\rangle|^2 <^+ 0$. Let D be a unitary transform where $D|i\rangle|i\rangle = |i\rangle|1\rangle$ and $\mathbf{K}(D|2^{2n}) =^+ 0$. So $\mathbf{Hg}(\sigma_{12}) =^+ \mathbf{Hg}(D\sigma_{12}D^*) =^+ \mathbf{Hg}(2^{-n}I \otimes |1\rangle\langle 1|) =^+ n - \log \mathrm{Tr}(I \otimes |1\rangle\langle 1|)\boldsymbol{\mu}^{2n}$. By Theorem 5 and properties of partial trace, $\mathbf{Hg}(2^{-n}I \otimes |1\rangle\langle 1|) =^+ n - \log \mathrm{Tr}\,|1\rangle\langle 1|\,\boldsymbol{\mu}^n =^+ n$. So $\mathbf{Hg}(\sigma_{12}) = \mathbf{Hg}(\sigma_{23}) =^+ n$. So $\mathbf{Hg}(\sigma_{123}) + \mathbf{Hg}(\sigma_2) <^+ n$, and $\mathbf{Hg}(\sigma_{12}) + \mathbf{Hg}(\sigma_{23}) >^+ 2n$, proving that $\mathbf{Hg}$ is not strongly superadditive. □

3.6 Relation Between Complexities

3.6.1 Vitányi Complexity and Gács Complexity

By definition $\mathbf{Hg}(|\psi\rangle) <^+ \mathbf{Hv}(|\psi\rangle)$. In fact, as shown in the following theorem, Vitányi complexity is bounded with respect to Gács complexity.

Theorem 13 **([Gó1])** $\mathbf{Hg}(|\psi\rangle) <^+ \mathbf{Hv}(|\psi\rangle) <^{\log} 4\mathbf{Hg}(|\psi\rangle)$.

Proof. For semi-density matrix A with eigenvectors $\{|a_i\rangle\}$ and decreasing eigenvectors $\{a_i\}$ with $\langle\psi| A |\psi\rangle \geq 2^{-k}$ and $|\psi\rangle = \sum c_i |a_i\rangle$, let A_m be a projector onto the m largest eigenvectors. Let m be the first i where $a_i \leq 2^{-k-1}$. Since $\sum a_i \leq 1$, we have $m \leq 2^{k+1}$. Since

$$\sum_{i\geq m} a_i |c_i|^2 < 2^{-k-1} \sum_i |c_i|^2 = 2^{-k-1},$$

we have

$$\langle\psi| A_m |\psi\rangle \geq \sum_{i<m} |c_i|^2 \geq \sum_{i<m} a_i |c_i|^2 \geq 2^{-k} - \sum_{i\geq m} a_i |c_i|^2 > 2^{-k-1}.$$

Thus there is some $i \leq m$ such that $|\langle\psi|a_i\rangle|^2 \geq 2^{-2k-2}$. Let $\nu = \mathrm{Tr}\boldsymbol{\mu}$ and $\nu_k \in \mathbb{Q}$ be a rational created from the first k digits of ν. Let $\hat{\boldsymbol{\mu}}$ be a lower approximation of $\boldsymbol{\mu}$, with trace greater than ν_k. So $\mathbf{K}(\hat{\boldsymbol{\mu}}) <^{\log} k$. Thus if $\langle\psi|\,\boldsymbol{\mu}\,|\psi\rangle \geq 2^{-k}$, then $\langle\psi|\,\hat{\boldsymbol{\mu}}\,|\psi\rangle \geq 2^{-k-1}$. Thus there is an eigenvector $|u\rangle$ of $\hat{\boldsymbol{\mu}}$ of complexity $\mathbf{K}(|u\rangle |N) <^{\log} 2k$ and $|\langle\psi|u\rangle|^2 \overset{*}{>} 2^{-2k}$, so

$$\mathbf{Hv}(|\psi\rangle) \leq \mathbf{K}(|u\rangle |N) - \log |\langle\psi|u\rangle|^2 <^{\log} 4k <^{\log} 4\mathbf{Hg}(|\psi\rangle).$$

□

We now describe an infinite encoding scheme for an arbitrary (not necessarily elementary) quantum pure state $|\psi\rangle$. This scheme is defined as an injection between the set of pure states and $\{0,1\}^\infty$. We define $\langle\langle|\psi\rangle\rangle\rangle$ to be an ordered list of the encoded tuples $\langle\langle|\theta\rangle\rangle, q, [|\langle\psi|\theta\rangle|^2 \geq q]\rangle$, over all elementary states $|\theta\rangle$ and rational distances $q \in \mathbb{Q}_{>0}$. We use the following definition of the mutual information between sequences.

Definition 3 (**[Lev74]**) *For $\alpha, \beta \in \{0,1\}^\infty$, $\mathbf{I}(\alpha : \beta) = \log \sum_{x,y\in\{0,1\}^*} \mathbf{m}(x|\alpha)\mathbf{m}(y|\beta)2^{\mathbf{I}(x:y)}$.*

The following theorem states that only exotic pure states will have a Vitányi complexity much greater than Gács complexity. States are exotic if they have high mutual information, $\mathbf{I}$, with the halting sequence $\mathcal{H} \in \{0,1\}^\infty$.

Theorem 14 (**[Eps20]**) $\mathbf{Hg}(|\psi\rangle) <^+ \mathbf{Hv}(|\psi\rangle) <^{\log} \mathbf{Hg}(|\psi\rangle) + \mathbf{I}(\langle\langle|\psi\rangle\rangle\rangle : \mathcal{H})$.

The proof to this theorem is quite extensive. We encourage the readers to review [Eps20] if they are interested in learning it.

3.6.2 BVL Complexity and Gács Complexity

Theorem 15 (**[Eps20]**) *For pure state $|\psi\rangle \in \mathcal{Q}_n$, $\mathbf{Hg}(|\psi\rangle) <^+ \mathbf{Hbvl}[\epsilon](|\psi\rangle) + \mathbf{K}(\mathbf{Hbvl}^\epsilon(|\psi\rangle) - \log \epsilon$.*

Proof. For each k and t in $\mathbb{N}$, let $\mathcal{H}_{k,t}$ be the smallest linear subspace spanning elementary k-qubit inputs to the universal quantum Turing machine M of size k that halt in t steps, outputing a n qubit mixed state. As shown in [Mul08], if $t \neq t'$, then $\mathcal{H}_{k,t}$ is perpendicular to $\mathcal{H}_{k,t'}$. Let $P_{k,t}$ be the projection onto $\mathcal{H}_{k,t}$. For each k and t, the universal quantum Turing machine defines a completely positive map $\Psi_{k,t}$ over $\mathcal{H}_{k,t}$, where $\Psi_{k,t}(\nu) = \rho$ implies that the quantum Turing machine, with semi-density matrix ν of length k on the input tape will output the n qubit mixed state ρ and halt in time t. Let ρ be a k qubit mixed state that minimizes $k = \mathbf{Hbvl}[\epsilon](|\psi\rangle)$ in time t.

$$
\begin{aligned}
\rho &\leq P_{k,t} \\
2^{-k}\rho &\leq 2^{-k}P_{k,t} \\
\Psi_{k,t}2^{-k}\rho &\leq \Psi_{k,t}2^{-k}P_{k,t} \\
\Psi_{k,t}2^{-k}\rho &\leq \sum_t \Psi_{k,t}2^{-k}P_{k,t}
\end{aligned}
$$

The semi density matrix $\sum_t \Psi_{k,t}2^{-k}P_{k,t}$ is lower computable relative to k, so

$$
\mathbf{m}(k|N)2^{-k}\Psi_{k,t}\rho \leq \mathbf{m}(k|N)\sum_t \Psi_{k,t}2^{-k}P_{k,t} \overset{*}{<} \boldsymbol{\mu}
$$

$$
\begin{aligned}
\mathbf{m}(k|N)2^{-k}\langle\psi|\,\Psi_{k,t}(\rho)\,|\psi\rangle &\overset{*}{<} \langle\psi|\,\boldsymbol{\mu}\,|\psi\rangle \\
k + \mathbf{K}(k|N) - \log\epsilon &>^+ \mathbf{Hg}(|\psi\rangle).
\end{aligned}
$$

□

Theorem 16 (**[Eps20]**) $\mathbf{Hbvl}[2^{-\mathbf{Hg}(|\psi\rangle)-O(\log\mathbf{Hg}(|\psi\rangle))}](|\psi\rangle) <^{\log} \mathbf{Hg}(|\psi\rangle)$.

Proof. We use reasoning from Theorem 7 in [Ǵ01]. From Theorem 7 in [Eps20] there exists a ρ such that $\mathbf{K}(\rho|N) - \log\langle\psi|\,\rho\,|\psi\rangle <^{\log} \mathbf{Hg}(|\psi\rangle)$. Let $\lceil -\log\langle\psi|\,\rho\,|\psi\rangle\rceil = m$. Let $|u_1\rangle, |u_2\rangle, |u_3\rangle, \ldots$ be the eigenvectors of ρ with eigenvalues $u_1 \geq u_2 \geq u_3 \ldots$ For $y \in \mathbb{N}$, let $\rho_y = \sum_{i=1}^{y} u_i |u_i\rangle\langle u_i|$.

We expand $|\psi\rangle$ in the basis of $\{|u_i\rangle\}$ with $|\psi\rangle = \sum_i c_i |u_i\rangle$. So we have that $\sum_i u_i|c_i|^2 \geq 2^{-m}$. Let $s \in \mathbb{N}$ be the first index i with $u_i < 2^{-m-1}$. Since $\sum_i u_i \leq 1$, it must be that $s \leq 2^{m+2}$. So

$$\sum_{i\geq s} u_i|c_i|^2 < 2^{-m-1}\sum_i |c_i|^2 \leq 2^{-m-1},$$

$$\langle\psi|\,\rho_{2^{m+2}}\,|\psi\rangle \geq \mathrm{Tr}\,\langle\psi|\,\rho_s\,|\psi\rangle > \sum_{i<s} u_i|c_i|^2 \geq 2^{-m} - \sum_{i\geq s} u_i|c_i|^2 > 2^{-m-1}.$$

We now describe a program to the universal quantum Turing machine that will construct $\rho_{2^{m+2}}$. The input is an ensemble $\{u_i\}_{i=1}^{2^{m+2}}$ of vectors $\{|cB(i)\rangle\}$, where $B(i)$ is the binary encoding of index $i \in \mathbb{N}$ which is of length $m+2$. Helper code c of size $=^+ \mathbf{K}(p|N)$ transforms each $|cB(i)\rangle$ into $|u_i\rangle$. Thus the size of the input is $<^+ \mathbf{K}(p|N) + m <^{\log} \mathbf{Hg}(|\psi\rangle)$. The fidelity of the approximation is $\langle\psi|\,\rho_{2^{m+2}}\,|\psi\rangle > 2^{-m-1} \geq 2^{-\mathbf{Hg}(|\psi\rangle)-O(\log \mathbf{Hg}(|\psi\rangle))}$. □

3.7 Quantum EL Theorem

In this paper we prove a Quantum EL Theorem. In AIT, the EL Theorem [Lev16, Eps19d] states that sets of strings that contain no simple member will have high mutual information with the halting sequence.

Theorem 17 ([Lev16, Eps19c])
For finite set $D \subset \{0,1\}^*$, $\min_{x\in D}\mathbf{K}(x) <^{\log} -\log\sum_{x\in D}\mathbf{m}(x) + \mathbf{I}(D;\mathcal{H})$.

It has many applications, including that all sampling methods produce outliers [Eps21b]. The Quantim EL Theorem states that non exotic projections of large rank must have simple quantum pure states in their images. By non exotic, we mean the coding of the projection has low information with the halting sequence. The Quantum EL Theorem has the following consequence.

Claim. *As the von Neumann entropy associated with the quantum source increases, the lossless quantum coding projectors have larger rank and thus must have simpler (in the algorithmic quantum complexity sense) pure states in their images.*

Theorem 18 (Quantum EL Theorem [Eps23b]) *Fix an n qubit Hilbert space. Let P be a elementary projection of rank $> 2^m$. Then, relativized to (n,m), $\min_{|\phi\rangle\in\mathrm{Image}(P)}\mathbf{Hv}(|\phi\rangle) <^{\log} 3(n-m) + \mathbf{I}(\langle P\rangle;\mathcal{H})$.*

Proof. We assume P has rank 2^m. Let Q be the elementary probability measure that realized the stochasticity, $\mathbf{Ks}(P)$, of an encoding of P. We can assume that every string in the support of Q encodes a projection of rank 2^m. We sample N independent pure states according to the uniform distribution Λ on the n qubit space. For each pure state $|\psi_i\rangle$ and projection R in the support of Q, the expected value of $\langle\psi_i|\,R\,|\psi_i\rangle$ is

$$\int \langle\psi_i|\,R\,|\psi_i\rangle\, d\Lambda = \mathrm{Tr}R\int |\psi_i\rangle\langle\psi_i|\, d\Lambda = 2^{-n}\mathrm{Tr}RI = 2^{m-n}.$$

Let random variable $X_R = \frac{1}{N}\sum_{i=1}^N \langle\psi_i|\,R\,|\psi_i\rangle$ be the average projection size of the random pure states onto the projection R. Since $\langle\psi_i|\,R\,|\psi_i\rangle \in [0,1]$ with expectation 2^{m-n}, by Hoeffding's inequality,

$$\Pr(X_R \leq 2^{m-n-1}) < \exp\left[-N2^{-2(m-n)-1}\right]$$

Let $d = \mathbf{d}(P|Q)$. Thus if we set $N = d2^{2(m-n)+1}$, we can find N elementary n qubit states such that $Q(\{R : X_R \leq 2^{m-n-1}\}) \leq \exp(-d)$, where X_R is now a fixed value and not a random variable. Thus $X_P > 2^{m-n-1}$ otherwise one can create a Q-expectation test, t, such that $t(R) = \exp d$. This is a contradiction because

$$1.44d <^+ \log(P) <^+ \mathbf{d}(P|Q) <^+ d,$$

for large enough d which we can assume without loss of generality. Thus there exists i such that $\langle\psi_i| P |\psi_i\rangle \geq 2^{m-n-1}$. Thus $|\phi\rangle = P|\psi_i\rangle / \sqrt{\langle\psi_i| P |\psi_i\rangle}$ is in the image of P and $|\langle\psi_i|\phi\rangle|^2 = \langle\psi_i| P |\psi_i\rangle \geq 2^{m-n-1}$. The elementary state $|\psi_i\rangle$ has classical Kolmogorov complexity $\mathbf{K}(|\psi_i\rangle) <^{\log} \log N + \mathbf{K}(Q,d) <^{\log} 2(m-n) + \mathbf{Ks}(P)$. Thus by Lemma 3,

$$\begin{aligned}
&\min\{\mathbf{Hv}(|\psi\rangle) : |\psi\rangle \in \mathrm{Image}(P)\} \\
&\leq \mathbf{Hv}(|\phi\rangle) \\
&<^{\log} \mathbf{K}(|\psi_i\rangle) + |\langle\psi_i|\phi\rangle|^2 \\
&<^{\log} 3(n-m) + \mathbf{Ks}(P) \\
&<^{\log} 3(n-m) + \mathbf{I}(P;\mathcal{H}).
\end{aligned}$$

□

3.7.1 Computable Projections

Theorem 23 is in terms of elementary described projecctions and can be generalized to arbitrarily computable projections. For a matrix M, let $\|M\| = \max_{i,j} |M_{i,j}|$ be the max norm. A program $p \in \{0,1\}^*$ computes a projection P of rank ℓ if it outputs a series of rank ℓ projections $\{P_i\}_{i=1}^{\infty}$ such that $\|P - P_i\| \leq 2^{-i}$. For computable projection operator P, $\mathbf{I}(P;\mathcal{H}) = \min\{\mathbf{K}(p) - \mathbf{K}(p|\mathcal{H}) : p \text{ is a program that computes } P\}$.

Corollary 1 ([Eps23b]) *Fix an n qubit Hilbert space. Let P be a computable projection of rank $> 2^m$. Then, relativized to (n,m), $\min_{|\phi\rangle \in \mathrm{Image}(P)} \mathbf{Hv}(|\phi\rangle) <^{\log} 3(n-m) + \mathbf{I}(P;\mathcal{H})$.*

Proof. Let p be a program that computes P. There is a simply defined algorithm A, that when given p, outputs P_n such that $\min_{|\psi\rangle \in \mathrm{Image}(P)} \mathbf{Hv}(|\psi\rangle) =^+ \min_{|\psi\rangle \in \mathrm{Image}(P_n)} \mathbf{Hv}(|\psi\rangle)$. Thus by Lemma 1, one gets that $\mathbf{I}(P_n;\mathcal{H}) <^+ \mathbf{I}(P;\mathcal{H})$. The corollary follows from Theorem 23. □

3.7.2 Quantum Data Compression

A quantum source consists of a set of pure quantum states $\{|\psi_i\rangle\}$ and their corresponding probabilities $\{p_i\}$, where $\sum_i p_i = 1$. The pure states are not necessarily orthogonal. The sender, Alice wants to send the pure states to the receiver, Bob. Let $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$ be the density matrix associated with the quantum source. Let $S(\rho)$ be the von Neumann entropy of ρ. By Schumacher compression, [Sch95], in the limit of $n \to \infty$, Alice can compress n qubits into $S(\rho)n$ qubits and send these qubits to Bob with fidelity approaching 1. For example, if the message consists of n photon polarization states, we can compress the inital qubits to $nS(\rho)$ photons. Alice cannot compress the initial qubits to $n(S(\rho)-\delta)$ qubits, as the fidelity will approach 0. The qubits are compressed by projecting the message onto a typical subspace of rank $nS(\rho)$ using a projector P. The projection occurs by using a quantum measurement consisting of P and a second projector $(I-P)$, which projects onto a garbage state.

The results of this paper says that as $S(\rho)$ increases, there must be simple states in the range of P. There is no way to communicate a quantum source with large enough $S(\rho)$ without using simple quantum states.

Chapter 4

Quantum Typicality

4.1 Definition of Quantum Randomness Deficiency

In [G01́], the quantum notion of randomness deficiency was introduced. This quantum randomness deficiency measures the algorithmic atypicality of a pure or mixed quantum state ρ with respect to a second quantum mixed state σ. Mixed states σ are used to model random mixtures $\{p_i\}$ of pure states $\{|\psi_i\rangle\}$, so quantum randomness deficiency is a score of how atypical a quantum state is with respect to a mixture. We first describe typicality with respect to computable σ, and then generalize to uncomputable σ.

Given a density matrix σ, a σ-test is a lower computable matrx T such that $\mathrm{Tr} T\sigma = 1$. Let $\mathcal{T}_\sigma$ be the set of all σ-tests. If σ is computable, there exists a universal σ test $\mathbf{t}_\sigma$, that is lower computable relative to the number of qubits n, $\mathrm{Tr}\sigma\mathbf{t}_\sigma \leq 1$, and for every lower computable σ test T, $O(1)\mathbf{t}_\sigma > \mathbf{m}(T|\sigma)T$.

This universal test can be computed the following manner, analagously to the classical case (see [G21́]). A program enumerates all strings p and lower computes $\mathbf{m}(p|\sigma)$. The program then runs p and continues with the outputs as long as p outputs a series of positive semi-definite matrices T_i such that $\mathrm{Tr} T_i\sigma \leq 1$ and $T_i \leq T_{i+1}$. If p outputs something other than this sequence or does not halt, the sequence is frozen. $\mathbf{t}_\sigma = \sum_p \mathbf{m}(p|\sigma) \lim_i T_i$ is the weighted sum of all such outputs of progams p.

Definition 4 (Quantum Randomness Deficiency) *For mixed states σ and ρ, computable σ, $\mathbf{d}(\rho|\sigma) = \log \mathrm{Tr}\, \mathbf{t}_\sigma \rho$.*

The quantum randomness deficiency, among other interpretations, is score of how typical a pure state is with respect to an algorithmically generated quantum source. Indeed, suppose there is a computable probability P over encodings of elementary orthogonal pure states $\{\langle|\psi_i\rangle\rangle\}$ of orthogonal pure states $\{|\psi_i\rangle\}$, with corresponding density matrix $\sigma = \sum_i P(\langle|\psi_i\rangle\rangle) |\psi_i\rangle\langle\psi_i|$. Then there is a lower-computable σ-test $T = \sum_i 2^{\mathbf{d}(\langle|\psi_i\rangle\rangle|P)} |\psi_i\rangle\langle\psi_i|$ with $O(1)\mathbf{t}_\sigma > T$. Thus $\mathbf{d}(|\psi_i\rangle|\sigma) >^+ \mathbf{d}(\langle|\psi_i\rangle\rangle|P)$, giving high scores to pure states $|\psi_i\rangle$ which are atypical of the source. In general the $\mathbf{d}(|\phi\rangle|\sigma)$ score for arbitrary $|\phi\rangle$ will be greater than a combination of $\mathbf{d}(\cdot|P)$ scores, with $\mathbf{d}(|\phi\rangle|\sigma) >^+ \log\sum 2^{\mathbf{d}(\langle|\psi_i\rangle\rangle|P)}|\langle\phi|\psi_i\rangle|^2$. In fact $\mathbf{d}$ is equivalent to the classical definition of randomness deficiency when σ is purely classical, i.e. only diagonal.

Theorem 19 *For diagonal $\sigma = \sum_i p(i)|i\rangle\langle i|$, $\mathbf{d}(|i\rangle|\sigma) =^+ \mathbf{d}(i|p)$.*

Proof. The positive semi-definite matrix $T = \sum_i 2^{\mathbf{d}(i|p)} |i\rangle \langle i|$ is a σ-test, so $T \overset{*}{<} \mathbf{t}_\sigma$ and thus $\mathbf{d}(|i\rangle |\sigma) >^+ \log \langle i| T |i\rangle =^+ \mathbf{d}(i|p)$. The function $t(i) = \langle i| \mathbf{t}_\sigma |i\rangle$ is a lower computable p-test, so $\mathbf{d}(i|P) >^+ \mathbf{d}(|i\rangle |\sigma)$. □

The following theorem shows that randomness deficiency $\mathbf{d}(\rho|\sigma)$ parallels the classical definition of randomness decificiency, $\mathbf{d}(x|P) = \log \mathbf{m}(x)/P(x)$.

Theorem 20 ([G0́1]) *Relativized to elementary σ, $\log \mathbf{d}(\rho|\sigma) =^+ \log \mathrm{Tr}\rho\sigma^{-1/2}\boldsymbol{\mu}\sigma^{-1/2}\sigma$*

Proof. The matrix $\sigma^{1/2}\mathbf{t}_\sigma\sigma^{1/2}$ is a lower-computable semi density matrix, so $\mathbf{t}_\sigma \overset{*}{<} \sigma^{-1/2}\boldsymbol{\mu}\sigma^{-1/2}$. This implies $\mathrm{Tr}\mathbf{t}_\sigma\rho \overset{*}{<} \mathrm{Tr}\rho\sigma^{-1/2}\boldsymbol{\mu}\sigma^{-1/2}$. □

4.1.1 Uncomputable Mixed States

We now extend $\mathbf{d}$ to uncomputable σ. For uncomputable σ, $\mathcal{T}_\sigma$ is not necessarily enumerable, and thus a universal lower computable randomness test does not necessarily exist, and cannot be used to define the σ deficiency of randomness. So in this case, the deficiency of randomness is instead defined using an aggregation of σ-tests, weighted by their lower algorithmic probabilities. The lower algorithmic probability of a lower computable matrix σ is $\underline{\mathbf{m}}(\sigma|x) = \sum\{\mathbf{m}(q|x) : q \text{ lower computes } \sigma\}$. Let $\mathfrak{T}_\sigma = \sum_{\nu\in\mathcal{T}_\sigma} \underline{\mathbf{m}}(\nu|n)\nu$.

Definition 5 (Quantum Randomness Deficiency (Uncomputable States)) *The randomness deficiency of ρ with respect to σ is $\mathbf{d}(\rho|\sigma) = \log \mathrm{Tr}\mathfrak{T}_\sigma\rho$.*

If σ is computable, then Definition 5 equals Definition 4. By definition, $\mathfrak{T}_\sigma$ is universal, since for every lower computable σ-test ν, $\underline{\mathbf{m}}(\nu)\nu < \mathfrak{T}_\sigma$.

Theorem 21 *For n qbit density matrices σ, ρ, ν, and ξ, $\mathbf{d}(\sigma|\rho) + \mathbf{d}(\nu|\xi) <^+ \mathbf{d}(\sigma\otimes\nu|\rho\otimes\xi)$.*

Proof.

$$\begin{aligned}
\mathbf{d}(\sigma|\rho) + \mathbf{d}(\nu|\xi) &= \log \mathrm{Tr} \sum_{\rho'\in\mathcal{T}_\rho} \underline{\mathbf{m}}(\rho')\rho'\sigma + \log \mathrm{Tr} \sum_{\xi'\in\mathcal{T}_\xi} \underline{\mathbf{m}}(\xi')\xi'\nu \\
&= \log \mathrm{Tr} \left(\left(\sum_{\rho'\in\mathcal{T}_\rho} \underline{\mathbf{m}}(\rho')\rho' \right) \otimes \left(\sum_{\xi'\in\mathcal{T}_\xi} \underline{\mathbf{m}}(\xi')\xi' \right) \right) (\sigma\otimes\nu) \\
&= \log \mathrm{Tr} \left(\sum_{\rho'\in\mathcal{T}_\rho, \xi'\in\mathcal{T}_\xi} \underline{\mathbf{m}}(\rho')\underline{\mathbf{m}}(\xi')\rho'\otimes\xi' \right) (\sigma\otimes\nu) \\
&<^+ \log \mathrm{Tr} \left(\sum_{\rho'\in\mathcal{T}_\rho, \xi'\in\mathcal{T}_\xi} \underline{\mathbf{m}}(\rho'\otimes\xi')\rho'\otimes\xi' \right) (\sigma\otimes\nu) \\
&<^+ \log \mathrm{Tr} \left(\sum_{\kappa\in\mathcal{T}_{\rho\otimes\xi}} \underline{\mathbf{m}}(\kappa)\kappa \right) (\sigma\otimes\nu) \\
&=^+ \mathbf{d}(\sigma\otimes\nu|\rho\otimes\xi).
\end{aligned}$$

□

4.2 Conservation Over Quantum Operations

Proposition 1 *For semi-density matrix ν, relativized to a finite set of elementary matrices $\{M_i\}$, $\underline{\mathbf{m}}(\sum_i M_i^* \nu M_i|N) \stackrel{*}{>} \underline{\mathbf{m}}(\nu|N)$.*

Proof. For every string q that lower computes ν, there is a string q_M of the form rq, that lower computes $\sum_i M_i^* \nu M_i$. This string q_M uses the helper code r to take the intermediary outputs ξ_i of q and outputs the intermediary output $\sum_i M_i^* \xi_i M_i$. Since q_M has access to $\{M_i\}$ on the auxilliary tape, $\mathbf{m}(q_M|N) \stackrel{*}{>} \mathbf{m}(q|N)$.

$$\begin{aligned}
\underline{\mathbf{m}}(\nu|N) &= \sum\{\mathbf{m}(q|N) : q \text{ lower computes } \nu\} \\
&\stackrel{*}{<} \sum\{\mathbf{m}(q_M|N) : q \text{ lower computes } \nu\} \\
&\stackrel{*}{<} \sum\{\mathbf{m}(q'|N) : q' \text{ lower computes } \sum_i M_i^* \nu M_i\} \\
&\stackrel{*}{<} \underline{\mathbf{m}}\left(\sum_i M_i^* \nu M_i/n\right).
\end{aligned}$$

□

The following theorem shows conservation of randomness with respect to elementary quantum operations. It generalizes Theorems 2 and 3 from [Eps19c].

Theorem 22 (Randomness Conservation) *Relativized to elementary quantum operation ε, for semi-density matrices ρ, σ, $\mathbf{d}(\varepsilon(\rho)|\varepsilon(\sigma)) <^+ \mathbf{d}(\rho|\sigma)$.*

Proof. Since the universal Turing machine is relativized to ε, there is an elementary Kraus operator $\{M_i\}$ that can be computed from ε where $\varepsilon(\xi) = \sum_i M_i \xi M_i^*$. If ν is a $\sum_i M_i \rho M_i^*$-test, with $\nu \in \mathcal{T}_{\sum_i M_i \rho M_i^*}$, then $\sum_i M_i^* \nu M_i$ is a ρ-test, with $\sum_i M_i^* \nu M_i \in \mathcal{T}_\rho$. This is because by assumption $\mathrm{Tr}\nu \sum_i M_i \rho M_i^* \leq 1$. So by the cyclic property of trace $\mathrm{Tr} \sum_i M_i^* \nu M_i \rho \leq 1$. Therefore since $\sum_i M_i^* \nu M_i$ is lower computable, $\sum_i M_i^* \nu M_i \in \mathcal{T}_\rho$. From Proposition 1, $\underline{\mathbf{m}}(\sum_i M_i^* \nu M_i|n) \stackrel{*}{>} \underline{\mathbf{m}}(\nu|n)$. So we have the following inequality

$$\begin{aligned}
\mathbf{d}\left(\sum_i M_i \sigma M_i^* | \sum_i M_i \rho M_i^*\right) &= \log \sum_{\nu \in \mathcal{T}_{\sum_i M_i \rho M_i^*}} \underline{\mathbf{m}}(\nu|N) \mathrm{Tr}\nu \sum_i M_i \sigma M_i^* \\
&<^+ \log \sum_{\nu \in \mathcal{T}_{\sum_i M_i \rho M_i^*}} \underline{\mathbf{m}}\left(\sum_i M_i^* \nu M_i|N\right) \mathrm{Tr} \sum_i M_i^* \nu M_i \sigma \\
&<^+ \mathbf{d}(\sigma|\rho).
\end{aligned}$$

□

4.3 A Quantum Outlier Theorem

One recent result in the classical randomness deficiency case is that sampling methods produce outliers [Eps21b]. There are several proofs to this result, with one of them derived from the fact that large sets of natural numbers with low randomness deficiencies are exotic, in that they have high mutual information with the halting sequence.

In this paper, we prove a quantum version of this result. Projections of large rank must contain pure quantum states in their images that are outlying states. Otherwise, the projections are exotic, in that they have high mutual information with the halting sequence. Thus quantum coding schemes that use projections, such as Schumacher compression, must communicate using outlier quantum states. The classical and quantum theorems are analogous, but their proofs are very different!

Theorem 23 ([Eps23c]) *Relativized to an n qubit mixed state σ, for elementary 2^m rank projector P, $3m-2n <^{\log} \max_{|\phi\rangle\in\mathrm{Image}(P)} \mathbf{d}(|\phi\rangle\,|\sigma) + \mathbf{I}(\langle P\rangle;\mathcal{H})$.*

Proof. We relativize the universal Turing machine to $\langle\sigma\rangle$ and $(3m-2n)$. Thus it is effectively relativized to m, n, and σ. Let elementary probability measure Q and $d\in\mathbb{N}$ realize $\mathbf{Ks}(P)$, where $d=\max\{\mathbf{d}(P|Q),1\}$. Without loss of generality we can assume that the support of Q is elementary projections of rank 2^m. There are $d2^{n-m+2}$ rounds. For each round we select an σ-test T, that is of dimension 1, $\mathrm{Tr}\sigma T\leq 1$, and for a certain Q-probability of projections B, $\mathrm{Tr}TB$ is large. We now describe the selection process.

Select a random test T to be $2^{m-2}\,|\psi\rangle\,\langle\psi|$, where $|\psi\rangle$ is an n qubit state chosen uniformly from the unit sphere, with distribution Λ.

$$\mathbf{E}[\mathrm{Tr}T\sigma] = 2^{m-2}\int \mathrm{Tr}\,\langle\psi|\,\sigma\,|\psi\rangle\,d\Lambda = 2^{m-2}\mathrm{Tr}\sigma\int |\psi\rangle\,\langle\psi|\,d\Lambda = 2^{m-n-2}\mathrm{Tr}\sigma = 2^{m-n-2}.$$

Thus the probability that T is a σ-test is $\geq 1-2^{m-n-2}$. Let I_m be an n-qubit identity matrix with only the first 2^m diagonal elements being non-zero. Let $K_m = I - I_m$. Let $p=2^{m-n}$ and $\hat{T}=T/2^{m-2}$. For any projection B of rank 2^m,

$$\begin{aligned}
&\Pr(\mathrm{Tr}B\hat{T}\leq .5p)\\
&=\Pr(\mathrm{Tr}I_m\hat{T}\leq .5p)\\
&=\Pr(\mathrm{Tr}K_m\hat{T}\geq 1-.5p)\\
\mathbf{E}[\mathrm{Tr}K_m\hat{T}] &= 1-p\\
\Pr(\mathrm{Tr}K_m\hat{T}\geq 1-.5p) &\leq (1-p)/(1-.5p)\\
\Pr(\mathrm{Tr}B\hat{T}\geq .5p) &= 1-\Pr(\mathrm{Tr}K_m\hat{T}\geq 1-.5p)\\
&\geq 1-(1-p)/(1-.5p)\\
&= .5p/(1-.5p)\geq .5p\\
\Pr(\mathrm{Tr}BT\geq 2^{2m-n-3}) &\geq .5p.
\end{aligned}$$

Let Ω be the space of all matrices of the form $2^{m-2}\,|\phi\rangle\,\langle\phi|$. Let R be the uniform distribution over Ω. Let $[A,B]$ be 1 if $\mathrm{Tr}AB > 2^{2m-n-3}$, and 0 otherwise. By the above equations, for all $A\in\mathrm{Support}(Q)$, $\int_\Omega[A,B]dR(B)\geq .5p$. So $\sum_A\int_\Omega[A,B]Q(A)dR(B)\geq .5p$. For Hermitian matrix A, $\{A\}$ is 1 if $\mathrm{Tr}A\sigma\leq 1$, and 0 otherwise. So $\int_\Omega\{A\}dR(A)\geq(1-p2^{-2})$. Let $f=\max_T\{T\}\sum Q(A)[T,A]$.

So

$$
\begin{aligned}
.5p &\leq \sum_A \int_\Omega [A,B]Q(A)dR(B)\\
&= \sum_A \int_\Omega \{B\}Q[A,B](A)dR(B) + \sum_A \int_\Omega (1-\{B\})[A,B]Q(A)dR(B)\\
&\leq \sum_A \int_\Omega \{B\}[A,B]Q(A)dR(B) + \int_\Omega (1-\{B\})dR(B)\\
&\leq \sum_A \int_\Omega \{B\}[A,B]Q(A)dR(B) + p2^{-2}\\
p/4 &\leq \sum_A \int_\Omega \{B\}[A,B]Q(A)dR(B) = \int_\Omega \left(\{B\}\sum_A [A,B]Q(A)\right) dR(B) \leq \int_\Omega f dR(B)\\
p/4 &\leq f.
\end{aligned}
$$

Thus for each round i, the lower bounds on f proves there exists a one dimensional matrix $T_i = 2^{m-2}\left|\psi\right\rangle\left\langle\psi\right|$ such that $\mathrm{Tr}T_i\sigma \leq 1$ and $\sum_R\{Q(R) : \mathrm{Tr}T_iR \geq 2^{2m-n-3}\} \geq p/4 = 2^{m-n-2}$. Such a T_i is selected, and the the Q probability is conditioned on those projections B for which $[T_i, B] = 0$, and the next round starts. Assuming that there are $d2^{n-m+2}$ rounds, the Q measure of projections B such there does not exist a T_i with $[T_i, B] = 1$ is

$$\leq (1 - p/4)^{d2^{n-m+2}} \leq e^{-d}.$$

Thus there exists a T_i such that $[T_i, P] = 1$, otherwise one can create a Q test t that assigns e^d to all projections B where there does not exist T_i with $[T_i, B] = 1$, and 0 otherwise. Then $t(P) = e^d$ so

$$1.44d < \log t(P) <^+ \mathbf{d}(P|Q) <^+ d.$$

This is a contradiction, because without loss of generality, one can assume d is large. Let $T_i = 2^{m-2}\left|\psi\right\rangle\left\langle\psi\right|$ with $[T_i, P] = 1$. Let $\left|\phi\right\rangle = P\left|\psi\right\rangle/\sqrt{\left\langle\psi\right|P\left|\psi\right\rangle}$. So $\left\langle\phi\right|T_i\left|\phi\right\rangle \geq 2^{2m-n-3}$ and $\left|\phi\right\rangle$ is in the image of P. Thus by Lemma 3,

$$
\begin{aligned}
2m-n &<^+ \log\left\langle\phi\right|T_i\left|\phi\right\rangle\\
2m-n &<^+ \log \max_{\left|\phi\right\rangle\in\mathrm{Image}(P)} \left\langle\phi\right|T_i\left|\phi\right\rangle\\
2m-n &<^+ \max_{\left|\phi\right\rangle\in\mathrm{Image}(P)} \mathbf{d}(P|\sigma) + \mathbf{K}(T_i)\\
2m-n &<^+ \max_{\left|\phi\right\rangle\in\mathrm{Image}(P)} \mathbf{d}(P|\sigma) + (n-m) + \log d + \mathbf{K}(d) + \mathbf{K}(Q)\\
2m-n &<^+ \max_{\left|\phi\right\rangle\in\mathrm{Image}(P)} \mathbf{d}(P|\sigma) + (n-m) + \mathbf{Ks}(P)\\
3m-2n &<^{\log} \max_{\left|\phi\right\rangle\in\mathrm{Image}(P)} \mathbf{d}(P|\sigma) + \mathbf{I}(P;\mathcal{H}).
\end{aligned}
$$

Note that due to the fact that the left hand side of the equation is $(3m-2n)$ and it has log precision, this enables one to condition the universal Turing machine to $(3m-2n)$. □

4.3.1 Computable Projections

Theorem 23 is in terms of elementary described projecctions and can be generalized to arbitrarily computable projections. For a matrix M, let $\|M\| = \max_{i,j}|M_{i,j}|$ be the max norm. A program

$p \in \{0,1\}^*$ computes a projection P of rank ℓ if it outputs a series of rank ℓ projections $\{P_i\}_{i=1}^{\infty}$ such that $\|P - P_i\| \leq 2^{-i}$. For computable projection operator P, $\mathbf{I}(P;\mathcal{H}) = \min\{\mathbf{K}(p) - \mathbf{K}(p|\mathcal{H}) :$ p is a program that computes $P\}$.

Corollary 2 ([Eps23c]) *Relativized to an n qubit mixed state σ, for computable 2^m rank projector P, $3m - 2n <^{\log} \max_{|\phi\rangle \in \mathrm{Image}(P)} \mathbf{d}(|\phi\rangle\,|\sigma) + \mathbf{I}(\langle P\rangle;\mathcal{H})$.*

Proof. Let p be a program that computes P. There is a simply defined algorithm A, that when given p and σ, outputs P_n such that $\max_{|\psi\rangle \in \mathrm{Image}(P)} \mathbf{d}(|\psi\rangle\,|\sigma) =^{+} \max_{|\psi\rangle \in \mathrm{Image}(P_n)} \mathbf{d}(|\psi\rangle\,|\sigma)$. Thus by Lemma 1, one gets that $\mathbf{I}(P_n;\mathcal{H}) <^{+} \mathbf{I}(P;\mathcal{H})$. The corollary follows from Theorem 23. □

Chapter 5

Quantum Information

5.1 Definition of Quantum Algorithmic Information

For a pair of random variables, $\mathcal{X}$, $\mathcal{Y}$, their mutual information is defined to be $\mathbf{I}(\mathcal{X} : \mathcal{Y}) = \mathcal{H}(\mathcal{X}) + \mathcal{H}(\mathcal{Y}) - \mathcal{H}(\mathcal{X}, \mathcal{Y}) = \mathcal{H}(\mathcal{X}) - \mathcal{H}(\mathcal{X}/\mathcal{Y}) = \sum_{x,y} p(x,y) \log p(x,y)/p(x)p(y)$. This represents the amount of correlation between $\mathcal{X}$ and $\mathcal{Y}$. Another intrepretation is that the mutual information between $\mathcal{X}$ and $\mathcal{Y}$ is the reduction in uncertainty of $\mathcal{X}$ after being given access to $\mathcal{Y}$.

Quantum mutual information between two subsystems described by states ρ_A and ρ_B of a composite system described by a joint state ρ_{AB} is $I(A : B) = S(\rho_A) + S(\rho_B) - S(\rho_{AB})$, where S is the Von Neumman entropy. Quantum mutual information measures the correlation between two quantum states.

As stated in Chapter 2, The algorithmic information between two strings is defined to be $\mathbf{I}(x : y) = \mathbf{K}(x) + \mathbf{K}(y) - \mathbf{K}(x, y)$. By definition, it measures the amount of compression two strings achieve when grouped together.

The three definitions above are based off the difference between a joint aggregate and the separate parts. Another approach is to define information between two semi-density matrices as the deficiency of randomness over $\boldsymbol{\mu} \otimes \boldsymbol{\mu}$, with the mutual information of σ and ρ being $\mathbf{d}(\sigma \otimes \rho|\boldsymbol{\mu} \otimes \boldsymbol{\mu})$. This is a counter argument for the hypothesis that the states are independently chosen according to the universal semi-density matrix $\boldsymbol{\mu}$. This parallels the classical algorithmic case, where $\mathbf{I}(x : y) =^+ \mathbf{d}((x, y)|\mathbf{m} \otimes \mathbf{m}) =^+ \mathbf{K}(x) + \mathbf{K}(y) - \mathbf{K}(x, y)$. However to achieve the conservation inequalities, a further refinement is needed, with the restriction of the form of the $\boldsymbol{\mu} \otimes \boldsymbol{\mu}$ tests. Let $\mathcal{C}_{C\otimes D}$ be the set of all lower computable matrices $A \otimes B$, such that $\mathrm{Tr}(A \otimes B)(C \otimes D) \leq 1$. Let $\mathfrak{C}_{C\otimes D} = \sum_{A\otimes B \in \mathcal{C}_{C\otimes D}} \underline{\mathbf{m}}(A \otimes B|N) A \otimes B$.

Definition 6 (Information) *The mutual information between two semi-density matrices σ, ρ is defined to be* $\mathbf{I}(\sigma : \rho) = \log \mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}(\sigma \otimes \rho)$.

Up to an additive constant, information is symmetric.

Theorem 24 $\mathbf{I}(\sigma : \rho) =^+ \mathbf{I}(\rho : \sigma)$.

Proof. This follows from the fact that for every $A \otimes B \in \mathcal{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}$, the matrix $B \otimes A \in \mathcal{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}$. Furthermore, since $\underline{\mathbf{m}}(A \otimes B|N) \stackrel{*}{=} \underline{\mathbf{m}}(B \otimes A|N)$, this guarantees that $\mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}(\sigma \otimes \rho) \stackrel{*}{=} \mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}(\rho \otimes \sigma)$, thus proving the theorem. □

5.2 Paucity of Self-Information

5.2.1 Pure States

For classical algorithmic information, for all $x \in \{0,1\}^*$,

$$\mathbf{I}(x:x) =^+ \mathbf{K}(x).$$

As shown in this section, this property differs from the quantum case, where there exists quantum states with high descriptional complexity and negligible self information. In fact this is the case for most quantum states, where for most n qubit pure states $|\psi\rangle$,

$$\mathbf{Hg}(|\psi\rangle) \approx n, \quad \mathbf{I}(|\psi\rangle : |\psi\rangle) \approx 0.$$

The following theorem states that the information between two elementary states is not more than the combined length of their descriptions.

Theorem 25 *For elementary ρ and σ, $\mathbf{I}(\rho:\sigma) <^+ \mathbf{K}(\rho|N) + \mathbf{K}(\sigma|N)$.*

Proof. Assume not. Then for any positive constant c, there exists semi-density matrices ρ and σ, such that

$$c\mathbf{m}(\rho|N)\mathbf{m}(\sigma|N)2^{\mathbf{I}(\rho:\sigma)} = c\mathrm{Tr}\mathbf{m}(\rho|N)\mathbf{m}(\sigma|N)\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}(\rho\otimes\sigma) > 1.$$

By the definition of $\boldsymbol{\mu}$, $\mathbf{m}(\rho|N)\rho \overset{*}{<} \boldsymbol{\mu}$ and $\mathbf{m}(\sigma|N)\sigma \overset{*}{<} \boldsymbol{\mu}$. Therefore by the definition of the Kronecker product, there is some positive constant d such that for all ρ and σ, $d\mathbf{m}(\rho|N)\mathbf{m}(\sigma|N)(\rho\otimes\sigma) < (\boldsymbol{\mu}\otimes\boldsymbol{\mu})$, and similarly

$$d\mathrm{Tr}\mathbf{m}(\rho|N)\mathbf{m}(\sigma|N)\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}(\rho\otimes\sigma) < \mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}(\boldsymbol{\mu}\otimes\boldsymbol{\mu}).$$

By the definition of $\mathfrak{C}$, it must be that $\mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}\boldsymbol{\mu}\otimes\boldsymbol{\mu} \leq 1$. However for $c = d$, there exists a ρ and a σ, such that

$$\mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}\boldsymbol{\mu}\otimes\boldsymbol{\mu} > d\mathrm{Tr}\mathbf{m}(\rho|N)\mathbf{m}(\sigma|N)\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}(\rho\otimes\sigma) > 1,$$

causing a contradiction. $\square$

Theorem 26 ([Eps19b]) *Let Λ be the uniform distribution on the unit sphere of $\mathcal{H}_N$.*

1. $\mathbf{Hg}(I/N) =^+ \log N$,
2. $\mathbf{I}(I/N : I/N) <^+ 0$,
3. $\int 2^{-\mathbf{Hg}(|\psi\rangle)} d\Lambda \overset{*}{=} N^{-1}$,
4. $\int 2^{\mathbf{I}(|\psi\rangle : |\psi\rangle)} d\Lambda <^+ 0$.

Proof. (1) follows from $\mathbf{Hg}(I/N) =^+ -\log \mathrm{Tr}\boldsymbol{\mu} I/N =^+ \log N - \log \mathrm{Tr}\boldsymbol{\mu} =^+ \log N$. (2) is due to Theorem 25, with $\mathbf{I}(I/N : I/N) <^+ 2\mathbf{K}(I/N|N) <^+ 0$. (3) uses the fact that $\rho = \int |\psi\rangle\langle\psi| \, d\Lambda = I/N$, because $\mathrm{Tr}\rho = 1$, and $\langle\psi|\,\rho\,|\psi\rangle = \langle\phi|\,\rho\,|\phi\rangle$. Thus $\int 2^{-\mathbf{Hg}(|\psi\rangle)} d\Lambda \overset{*}{=} \int \mathrm{Tr}\boldsymbol{\mu}\,|\psi\rangle\langle\psi|\,d\Lambda \overset{*}{=} \mathrm{Tr}\boldsymbol{\mu}\int|\psi\rangle\langle\psi|\,d\Lambda \overset{*}{=} N^{-1}$. (4) uses the proof of Theorem 8, which states $\int |\psi\rangle\langle\psi| \otimes |\psi\rangle\langle\psi|\,d\Lambda = \int |\psi\psi\rangle\langle\psi\psi|\,d\Lambda = \binom{N+1}{2}^{-1}P$, where P is the projection onto the space of pure states $|\psi\psi\rangle$. So

$$
\begin{aligned}
\int 2^{\mathbf{I}(|\psi\rangle\,:\,|\psi\rangle)} d\Lambda &= \int \mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}\,|\psi\rangle\langle\psi| \otimes |\psi\rangle\langle\psi|\,d\Lambda \\
&= \mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}} \int |\psi\rangle\langle\psi| \otimes |\psi\rangle\langle\psi|\,d\Lambda \\
&= \mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}} \binom{N+1}{2}^{-1} P \\
&\overset{*}{<} \mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}} N^{-2} I \\
&\overset{*}{=} 2^{\mathbf{I}(I/N:I/N)} \\
&<^+ 0.
\end{aligned}
$$

□

5.2.2 Mixed States

The results of the previous section can be extended to mixed states. Given a uniform measure over mixed states, an overwhelming majority of such states contain no algorithmic self information. Let Λ be the uniform distribution of the unit sphere of $\mathcal{H}_N$. Fix any number $M \in \mathbb{N}$. Let the M-simplex be

$$\Delta_M = \{(p_i)_{1\le i\le M} | p_i \ge 0, p_1 + \cdots + p_M = 1\}.$$

We use the uniform distribution over Δ_M, defined as η. However the results work with any distribution over Δ_M. For an integrable function over matrices f we define its M-uniform integral to be

$$\int f(\sigma) d\mu(\sigma) = \int_{\Delta_M}\int_{\Lambda_1}\cdots\int_{\Lambda_M} f\left(\sum_{i=1}^M p_i\,|\psi_i\rangle\langle\psi_i|\right) d\Lambda_1 \ldots d\Lambda_M d\eta(p_1,\ldots,p_M).$$

It is straightforward to see the following theorem holds for the more general distribution $\Pi(\sigma) = \sum_M \pi(M)\mu_M(\sigma)$, for any distribution π over $\mathbb{N}$.

Theorem 27 $\int 2^{\mathbf{I}(\sigma:\sigma)} d\mu(\sigma) <^+ 0$.

Proof.

$$\begin{aligned}
&\int 2^{\mathbf{I}(\sigma:\sigma)} d\mu(\sigma) \\
=&\mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}} \int_{\Delta_M}\int_{\Lambda_1}\cdots\int_{\Lambda_M}\left(\sum_{i=1}^M p_i \left|\psi_i\right\rangle\left\langle\psi_i\right|\right)\otimes\left(\sum_{i=1}^M p_i \left|\psi_i\right\rangle\left\langle\psi_i\right|\right) d\Lambda_1\dots d\Lambda_M d\eta(p_1,\dots,p_M) \\
=&\mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}} \int_{\Delta_M}\int_{\Lambda_1}\cdots\int_{\Lambda_M}\left(\sum_{i,j=1}^M p_i p_j \left|\psi_i\right\rangle\left\langle\psi_i\right|\otimes\left|\psi_j\right\rangle\left\langle\psi_j\right|\right) d\Lambda_1\dots d\Lambda_M d\eta(p_1,\dots,p_M) \\
=&\mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}} \int_{\Delta_M}\int_{\Lambda}\sum_{i=1}^M p_i^2 \left|\psi\psi\right\rangle\left\langle\psi\psi\right| d\Lambda d\eta(p_1,\dots,p_M) \\
&+\mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}} \int_{\Delta_M}\int_{\Lambda_1}\int_{\Lambda_2}\sum_{i,j\in\{1,\dots,M\},i\neq j} 2p_i p_j \left|\psi_1\right\rangle\left\langle\psi_1\right|\otimes\left|\psi_2\right\rangle\left\langle\psi_2\right| d\Lambda_1 d\Lambda_2 d\eta(p_1,\dots,p_M).
\end{aligned}$$

The first term is not greater than

$$\begin{aligned}
&\mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}} \int_{\Delta_M}\int_{\Lambda}\sum_{i=1}^M \left|\psi\psi\right\rangle\left\langle\psi\psi\right| d\Lambda d\eta(p_1,\dots,p_M) \\
=&\mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}} \int_{\Lambda}\sum_{i=1}^M \left|\psi\psi\right\rangle\left\langle\psi\psi\right| d\Lambda.
\end{aligned}$$

At this point, reasoning from the proof of Theorem 26 can be used to show that this term is $O(1)$. The second term is not greater than

$$\begin{aligned}
&\mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}} \int_{\Delta_M}\int_{\Lambda_1}\int_{\Lambda_2}\left(\sum_i p_i\right)\left(\sum_i p_i\right)\left|\psi_1\right\rangle\left\langle\psi_1\right|\otimes\left|\psi_2\right\rangle\left\langle\psi_2\right| d\eta(p_1,\dots,p_M) \\
=&\mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}} \int_{\Delta_M}\int_{\Lambda_1}\int_{\Lambda_2}\left|\psi_1\right\rangle\left\langle\psi_1\right|\otimes\left|\psi_2\right\rangle\left\langle\psi_2\right| d\eta(p_1,\dots,p_M) \\
=&\mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}} \int_{\Lambda_1}\int_{\Lambda_2}\left|\psi_1\right\rangle\left\langle\psi_1\right|\otimes\left|\psi_2\right\rangle\left\langle\psi_2\right| \\
=&\mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}(I/N\otimes I/N).
\end{aligned}$$

Again, at this point, reasoning from the proof of Theorem 26 can be used to show that this term is $O(1)$

5.3 Information Nongrowth

Classical algorithmic information non-growth laws asserts that the information between two strings cannot be increased by more than a constant depending on the computable transform f, with $\mathbf{I}(f(x):y) < \mathbf{I}(x:y) + O_f(1)$ (Theorem 2). Conservation inequalities have been extended to probabilistic transforms and infinite sequences. The following theorem shows information non-growth in the quantum case; information cannot increase under quantum operations, the most general type of transformation that a mixed or pure quantum state can undergo. The following theorem shows information nongrowth with respect to elementary quantum operations. It generalizes Theorems 5 and 10 from [Eps19c].

Theorem 28 (Information Conservation) *Relativized to elementary quantum operation ε, for semi-density matrices ρ, σ, $\mathbf{I}(\varepsilon(\rho):\sigma) <^+ \mathbf{I}(\rho:\sigma)$.*

Proof. Since the universal Turing machine is relativized to ε, there is an elementary Kraus operator $\{M_i\}$ that can be computed from ε where $\varepsilon(\xi) = \sum_i M_i \xi M_i^*$. Given density matrices A, B, C and D, we define $\mathbf{d}'(A\otimes B|C\otimes D) = \log \mathfrak{C}_{C\otimes D} A\otimes B$. Thus $\mathbf{I}(\sigma:\rho) = \mathbf{d}'(\sigma\otimes\rho|\boldsymbol{\mu}\otimes\boldsymbol{\mu})$. The semi-density matrix $\sum_i M_i\boldsymbol{\mu}M_i^*$ is lower semicomputable, so therefore $\sum_i M_i\boldsymbol{\mu}M_i^* \overset{*}{<} \boldsymbol{\mu}$ and also $(\sum_i M_i\boldsymbol{\mu}M_i^*\otimes\boldsymbol{\mu}) \overset{*}{<} \boldsymbol{\mu}\otimes\boldsymbol{\mu}$. So if $E\otimes F\in\mathcal{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}$ then $\mathrm{Tr}(E\otimes F)(\boldsymbol{\mu}\otimes\boldsymbol{\mu}) \le 1$, implying that $\mathrm{Tr}(E\otimes F)(\sum_i M_i\boldsymbol{\mu}M_i^*\otimes\boldsymbol{\mu}) < O(1)$. Thus there is a positive constant c, where $c(E\otimes F)\in\mathcal{C}_{(\sum_i M_i\boldsymbol{\mu}M_i^*)\otimes\boldsymbol{\mu}}$. So we have

$$\begin{aligned}
\mathbf{d}'\left(\sum_i M_i\sigma M_i^*\otimes\rho|\boldsymbol{\mu}\otimes\boldsymbol{\mu}\right) &= \log \sum_{E\otimes F\in\mathcal{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}} \underline{\mathbf{m}}(E\otimes F|N)\mathrm{Tr}(E\otimes F)(\sum_i M_i\sigma M_i^*\otimes\rho)\\
&<^+ \log \sum_{E\otimes F\in\mathcal{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}} \underline{\mathbf{m}}(c(E\otimes F)|N)\mathrm{Tr}c(E\otimes F)\left(\sum_i M_i\sigma M_i^*\otimes\rho\right)\\
&<^+ \mathbf{d}'\left(\sum_i M_i\sigma {M_i}^*\otimes\rho|\sum_i M_i\boldsymbol{\mu}M_i^*\otimes\boldsymbol{\mu}\right).
\end{aligned}$$

Using the reasoning of the proof of Theorem 22 on the elementary Kraus operator $\{M_i\otimes I\}$ and $\mathbf{d}'$, where $\mathcal{C}$ replaces $\mathcal{T}$, we have that

$$\mathbf{d}'\left(\sum_i M_i\sigma M_i^*\otimes\rho|\sum_i M_i\boldsymbol{\mu}M_i^*\otimes\boldsymbol{\mu}\right) <^+ \mathbf{d}'(\sigma\otimes\rho|\boldsymbol{\mu}\otimes\boldsymbol{\mu}).$$

Therefore we have that

$$\begin{aligned}
\mathbf{I}\left(\sum_i M_i\sigma M_i^*:\rho\right) &= \mathbf{d}'\left(\sum_i M_i\sigma M_i^*\otimes\rho|\boldsymbol{\mu}\otimes\boldsymbol{\mu}\right)\\
&<^+ \mathbf{d}'\left(\sum_i M_i\sigma M_i^*\otimes\rho|\sum_i M_i\boldsymbol{\mu}M_i^*\otimes\boldsymbol{\mu}\right)\\
&<^+ \mathbf{d}'(\sigma\otimes\rho|\boldsymbol{\mu}\otimes\boldsymbol{\mu}) =^+ \mathbf{I}(\sigma:\rho).
\end{aligned}$$

□

5.3.1 Algorithmic No-Cloning Theorem

The no-cloning theorem states that every unitary transform cannot clone an arbitrary quantum state. Hiowever some unitary transforms can clone a subset of pure quantum states. For example, given basis states $|1\rangle, |2\rangle, |3\rangle, \ldots$ there is a unitary transform that transforms each $|i\rangle|0\rangle$ to $|i\rangle|i\rangle$. In addition, there exists several generalizations to the no-cloning theorem, showing that imperfect clones can be made. In [BH96], a universal cloning machine was introduced that can clone an arbitrary state with the fidelity of 5/6. Theorem 8 shows a generalization of the no-cloning theorem using Gács complexity.

Given the information function introduced in this chapter, a natural question to pose is whether a considerable portion of pure states can use a unitary transform to produce two states that share a large amount of shared information. The following theorem answers this question in the negative. It states that the amount of information created between states with a unitary transform is bounded by the self information of the original state.

Theorem 29 ([Eps19b]) *Let $C|\psi\rangle|0^n\rangle = |\phi\rangle|\varphi\rangle$, where C is an elementary unitary transform. Relativized to C, $\mathbf{I}(|\phi\rangle : |\varphi\rangle) <^+ \mathbf{I}(|\psi\rangle : |\psi\rangle)$.*

Proof. We have the inequalities

$$\mathbf{I}(|\phi\rangle : |\varphi\rangle) <^+ \mathbf{I}(|\phi\rangle|\varphi\rangle) : |\phi\rangle|\varphi\rangle) <^+ \mathbf{I}(|\psi\rangle|0^n\rangle : |\psi\rangle|0^n\rangle) <^+ \mathbf{I}(|\psi\rangle : |\psi\rangle),$$

where the first inequality is derived using partial trace, the second inequality is derived using the unitary transform C, and the third inequality is derived by appending of an environment, all constituting quantum operations, whose conservation of information is proven in Theoerem 28. □

Theorem 29, combined with the paucity of self-information in pure states (Theorem 26) shows that only a very sparse set of pure states can, given any unitary transform, duplicate algorithmic information.

5.3.2 Purification

Every mixed state can be considered a reduced state of a pure state. The purification process is considered physical, so the extended Hilbert space in which the purified state resides in can be considered the existing environment. It should therefore be possible to regard our system with its mixed state as part of a larger system in a pure state. In this section we proof that the purifications of two mixed states will contain more information than the reduced states.

Purification occurs in the following manner, starting with a density matrix $\rho = \sum_{i=1}^n p_i |i\rangle\langle i|$. A copy of the space is defined with orthonormal basis $\{|i'\rangle\}$. In this instance the purification of ρ is $|\psi\rangle = \sum_{i=1}^n \sqrt{p_i}|i\rangle \otimes |i'\rangle$. For a density matrix ρ of size n, let $\mathcal{P}_\rho^m$ be the set of purifications of ρ of dimension $m \geq 2n$.

Corollary 3 *For all $|\psi_\sigma\rangle \in \mathcal{P}_\sigma^n$, $|\psi_\rho\rangle \in \mathcal{P}_\rho^n$, $\mathbf{d}(\sigma|\rho) <^+ \mathbf{d}(|\psi_\sigma\rangle \,|\, |\psi_\rho\rangle)$.*

Corollary 4 *For all $|\psi_\sigma\rangle \in \mathcal{P}_\sigma^n$, $|\psi_\rho\rangle \in \mathcal{P}_\rho^n$, $\mathbf{I}(\sigma : \rho) <^+ \mathbf{I}(|\psi_\sigma\rangle : |\psi_\rho\rangle)$.*

This all follows from conservation of randomness (Theorem 22) and information (Theorem 28) over quantum operations, which includes the partial trace function.

5.3.3 Decoherence

In quantum decoherence, a quantum state becomes entangled with the environment, losing decoherence. The off diagonal elements of the mixed state become dampened, as the state becomes more like a classical mixture of states.

The single qubit example is as follows. The system is in state $|\psi_Q\rangle = \alpha|0\rangle + \beta|1\rangle$ and the environment is in state $|\psi_E\rangle$. The initial state is $|\psi_{QE}\rangle|\psi_Q\rangle \otimes |\psi_E\rangle = \alpha|0,\psi_E\rangle + \beta|1,\psi_E\rangle$. The combined system undergoes a unitary evolution U, becoming entangled, with the result $U|\psi_{QE}\rangle = \alpha|0,E_1\rangle + \beta|1,E_2\rangle$. The density matrix is $\rho_{QE} = |\alpha|^2|0,E_1\rangle\langle 0,E_1| + |\beta|^2|1,E_2\rangle\langle 1,E_2| + \alpha^*\beta|1,E_2\rangle\langle 0,E_1| + \alpha\beta^*|0,E_1\rangle\langle 1,E_2|$. The partial trace over the environment yields

$$\rho_Q = |\alpha|^2|0\rangle\langle 0|\langle E_1|E_1\rangle + |\beta|^2|1\rangle\langle 1|\langle E_2|E_2\rangle + \alpha^*\beta|1\rangle\langle 0|\langle E_2|E_1\rangle + \alpha\beta^*|0\rangle\langle 1|\langle E_1|E_2\rangle.$$

We have $\langle E_1|E_1\rangle = \langle E_2|E_2\rangle = 1$. Two environment-related terms are time dependent and can be described by an exponential decay function

$$\langle E_1|E_2\rangle = e^{-\gamma(t)}.$$

The larger the decay, the more off diagonal terms are suppressed. So

$$\rho_Q \approx \begin{pmatrix} |\alpha|^2 & \alpha^*\beta e^{-\gamma(t)} \\ \alpha\beta^* e^{-\gamma(t)} & |\beta|^2 \end{pmatrix}.$$

The above example can be generalzied to n qubit density matrices. Let $\mathrm{Decohere}(\sigma, t)$ be a decoherence operation that dampens the off-diagonal elements of σ with decay t. By definition, Decohere is a quantum operation. Randomness is conserved over decoherence. Thus if two states decohere, the first state does not increase in algorithmic atypicality with respect to the second state.

Corollary 5 $\mathbf{d}(\mathrm{Decohere}(\sigma, t)|\mathrm{Decohere}(\rho, t)) <^+ \mathbf{d}(\sigma|\rho)$.

This is a corollary to Theorem 22. When a state loses coherence into the environment will not gain information with any other state.

Corollary 6 *For semi-density matrices σ and ρ,* $\mathbf{I}(\mathrm{Decohere}(\sigma, t) : \mathrm{Decohere}(\rho, t)) <^+ \mathbf{I}(\sigma : \rho)$.

Chapter 6

Quantum Measurements

In quantum mechanics, measurements are modeled by POVMs. A POVM E is a finite or infinite set of positive definite matrices $\{E_k\}$ such that $\sum_k E_k = I$. For a given semi-density matrix σ, a POVM E induces a semi measure over integers, where $E\sigma(k) = \mathrm{Tr}E_k\sigma$. This can be seen as the probability of seeing measurement k given quantum state σ and measurement E. An elementary POVM E has a program q such that $U(q)$ outputs an enumeration of $\{E_k\}$, where each E_k is elementary. A quantum instrument with respect to POVM E, is a quantum operation Φ_E that takes a state σ to a set of outcomes and their probabilities, $\Phi_E(\sigma) = \sum_k E(\sigma(k)) |k\rangle \langle k|$.

6.1 Typicality and Measurements

Theorem 30 shows that measurements can increase only up to a constant factor, the deficiency of randomness of a quantum state with respect to another quantum state. The classical deficiency of randomness of a probability with respect to a another probability is denoted as follows.

Definition 7 (Deficiency, probabilities (Folkflore)) *For probabilities p and q over $\{0,1\}^\infty$, $\mathbf{d}(q|p) = \log \sum_x q(x)\mathbf{m}(x)/p(x)$.*

Note that in the following theorem, $\mathbf{d}(E\sigma|E\rho)$ term represents the classical deficiency of randomness of a semimeasure $E\sigma$ with respect to a computable probability measure $E\rho$,.

Theorem 30 ([Eps19b]) *For density matrices σ, ρ, relativized to elementary ρ and POVM E, $\mathbf{d}(E\sigma|E\rho) <^+ \mathbf{d}(\sigma|\rho)$.*

Proof. $2^{\mathbf{d}(E\sigma|E\rho)} = \sum_k (\mathrm{Tr}E_k\sigma)\mathbf{m}(k|N)/(\mathrm{Tr}E_k\rho) = \mathrm{Tr}(\sum_k(\mathbf{m}(k|N)/\mathrm{Tr}E_k\rho)E_k)\sigma = \mathrm{Tr}\nu\sigma$, where the matrix $\nu = (\sum_k(\mathbf{m}(k|N)/\mathrm{Tr}E_k\rho)E_k)$ has $\nu \in \mathcal{T}_\rho$, since ν is lower computable and $\mathrm{Tr}\nu\rho \leq 1$. So $2^{\mathbf{d}(\sigma|\rho)} \geq \underline{\mathbf{m}}(\nu|N)\mathrm{Tr}\nu\sigma = \underline{\mathbf{m}}(\nu|N)2^{\mathbf{d}(E\sigma|E\rho)}$. Since $\underline{\mathbf{m}}(\nu|N) \stackrel{*}{>} 1$, $\mathbf{d}(E\sigma|E\rho) <^+ \mathbf{d}(\sigma|\rho)$.

6.2 Information and Measurements

Given two mixed states σ and ρ and POVM E, the mutual information between the probabilities of $E\sigma$ and $E\rho$, from Definition 1, is $\mathbf{I}_{\mathrm{Prob}}(E\sigma : E\rho)$. The following theorem states that given two states, the classical (algorithmic) information between the probabilities generated by two quantum measurements is less, up to a logarithmic factor, than the information of the two states. Thus $\mathbf{I}$ represents an upper bound on the amount of classical algorithmic information that can be extracted between two states.

Theorem 31 *Relative to POVMS E and F, $\mathbf{I}_{\mathrm{Prob}}(E\sigma : F\rho) <^{\log} \mathbf{I}(\sigma : \rho)$.*

Note than since the universal Turing machine is relativized to E and F, all $\mathbf{K}$ and $\mathbf{m}$ are conditioned to the number of qubits N. Quantum instruments with respect to POVMs E and F produces two mixed states $\Psi_E(\sigma) = \sum_{i=1}^m E_i(\sigma)\left|i\right\rangle\left\langle i\right|$ and $\Psi_F(\rho) = \sum_{j=1}^m F_j(\rho)\left|j\right\rangle\left\langle j\right|$, where, without loss of generality, m can be considered a power of 2. By Theorem 5, the (i,i)th entry of $\boldsymbol{\mu}$ is $\stackrel{*}{=} \mathbf{m}(i)$, so $\mathcal{T}_{ij} = 2^{\mathbf{K}(i)+\mathbf{K}(j)-O(1)}\left|i\right\rangle\left\langle i\right|\left|j\right\rangle\left\langle j\right|$ is a $\boldsymbol{\mu}\otimes\boldsymbol{\mu}$ test, with $\mathrm{Tr}\mathcal{T}_{i,j}(\boldsymbol{\mu}\otimes\boldsymbol{\mu}) < 1$. So, using the fact that $x/\log x$ is convex,

$$\begin{aligned}
\mathbf{I}(\sigma : \rho) &>^+ \mathbf{I}(\Psi_E(\sigma) : \Psi_F(\rho)) \\
&>^+ \log \sum_{i,j} \mathbf{m}(\mathcal{T}_{i,j})\mathcal{T}_{i,j}\Psi_E(\sigma)\otimes\Psi_F(\rho) \\
&>^+ \log \sum_{i,j} 2^{\mathbf{K}(i)+\mathbf{K}(j)}\mathbf{m}(i,j,\mathbf{K}(i)+\mathbf{K}(j))E_i(\sigma)F_j(\rho) \\
&>^+ \log \sum_{i,j} 2^{\mathbf{I}(i:j)-\mathbf{K}(\mathbf{I}(i:j))}E_i(\sigma)F_j(\rho) \\
&>^+ \log \sum_{i,j} 2^{\mathbf{I}(i:j)}\mathbf{I}(i:j)^{-O(1)}E_i(\sigma)F_j(\rho) \\
&>^{\log} \log \sum_{i,j} 2^{\mathbf{I}(i:j)}E_i(\sigma)F_j(\rho) \\
&>^{\log} \mathbf{I}_{\mathrm{Prob}}(E\sigma : F\rho).
\end{aligned}$$

Corollary 7 *For density matrices ρ and σ, and $i, j \in \mathbb{N}$, relativized to POVMS E and F, $\mathbf{I}(i : j) + \log E_i(\rho)F_j(\sigma) <^{\log} \mathbf{I}(\rho : \sigma)$.*

6.3 Algorithmic Contents of Measurements

This sections shows the limitations of the algorithmic content of measurements of pure quantum states. Theorem 32 says that given a measurement apparatus E, the overwhelming majority of pure states, when measured, will produce classical probabilities with no self-information, i.e. random noise. Theorem 3 shows that there is no randomized way to process the probabilities to produce more self-information, i.e. process the random noise. This is independent of the number of measurement outcomes of E.

To prove this result, we need to define an upper-information term $\mathcal{I}$ that is defined using *upper computable* tests. We say a semi-density matrix ρ is upper computable if there a program $q \in \{0,1\}^*$ such that when given to the universal Turing machine U, outputs, with or without halting, a finite or infinite sequence of elementary matrices ρ_i such that $\rho_{i+1} \preceq \rho_i$ and $\lim_{i\to\infty}\rho_i = \rho$. If U reads $\leq \|q\|$ bits on the input tape, then we say p upper computes ρ. The upper probability of an upper computable mixed state σ is defined by $\overline{\mathbf{m}}(\sigma/x) = \sum\{\mathbf{m}(q/x) : q \text{ upper computes } \sigma\}$.

Let $\mathcal{G}_{C\otimes D}$ be the set of all upper computable matrices (tests) of the form $A\otimes B$, where $\mathrm{Tr}(A\otimes B)(C\otimes D) \leq 1$. Let $\mathfrak{G}_{C\otimes D} = \sum_{A\otimes B\in\mathcal{C}_{C\otimes D}} \overline{\mathbf{m}}(A\otimes B/n)(A\otimes B)$ be an aggregation of upper computable $C\otimes D$ tests of the form $A\otimes B$, weighted by their upper probability.

Definition 8 *The upper information between semi-density matrices A and B is $\mathcal{I}(A : B) = \log \mathrm{Tr}\mathfrak{G}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}(A\otimes B)$.*

Proposition 2 $\mathcal{I}(I/N : I/N) = O(1)$.

Proof. $1 \geq \mathrm{Tr}\mathfrak{G}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}(\boldsymbol{\mu}\otimes\boldsymbol{\mu}) \stackrel{*}{>} \mathrm{Tr}\mathfrak{G}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}(I/N\otimes I/N) \stackrel{*}{>} 2^{\mathbf{I}(I/N:I/N)}$. □

Lemma 4

- *Let Λ be the uniform distribution on the unit sphere of an n qubit space.* $\int 2^{\mathcal{I}(|\psi\rangle\,:\,|\psi\rangle)}d\Lambda = O(1)$,
- $\int 2^{\mathcal{I}(\sigma\,:\,\sigma)}d\mu(\sigma) = O(1)$.

Proof. The proof follows identically to that of Theorems 26 and 27, with reference to Proposition 2.

Lemma 5 **([Eps21a])** *Relativized to POVM E, $\mathbf{I}_{\mathrm{Prob}}(E\sigma{:}E\sigma) <^+ \mathcal{I}(\sigma{:}\sigma)$.*

Proof. Note that all complexity terms are relativized to N, due to the relativization of E. Since $z(k) = \mathrm{Tr}\boldsymbol{\mu}E_k$ is lower semi-computable and $\sum_k z(k) < 1$, $\mathbf{m}(k) \stackrel{*}{>} \mathrm{Tr}\boldsymbol{\mu}E_k$, and so $1 > 2^{\mathbf{K}(k)-O(1)}\mathrm{Tr}\boldsymbol{\mu}E_k$. So $\nu_{i,j} = 2^{\mathbf{K}(i)+\mathbf{K}(j)-O(1)}(E_i\otimes E_j) \in \mathcal{G}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}$, with $\overline{\mathbf{m}}(\nu_{i,j}) \stackrel{*}{>} \mathbf{m}(i,j)$.

$$\begin{aligned}
\mathcal{I}(\sigma{:}\sigma) &= \log \sum_{A\otimes B\in\mathcal{G}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}} \overline{\mathbf{m}}(A\otimes B)(A\otimes B)(\sigma\otimes\sigma)\\
&>^+ \log \mathrm{Tr}\sum_{ij}\nu_{i,j}\overline{\mathbf{m}}(\nu_{i,j})(\sigma\otimes\sigma)\\
&>^+ \log\sum 2^{\mathbf{K}(i)+\mathbf{K}(j)}\mathbf{m}(i,j)E\sigma(i)E\sigma((j)\\
&>^+ \mathbf{I}_{\mathrm{Prob}}(E\sigma{:}E\sigma).
\end{aligned}$$

□

Theorem 32 **([Eps21a])** *Let Λ be the uniform distribution on the unit sphere of an n qubit space. Relativized to POVM E, $\int 2^{\mathbf{I}_{\mathrm{Prob}}(E|\psi\rangle:E|\psi\rangle)}d\Lambda = O(1)$.*

Proof. By Lemma 5, $2^{\mathcal{I}(|\psi\rangle:|\psi\rangle)} \stackrel{*}{>} 2^{\mathbf{I}_{\mathrm{Prob}}(E|\psi\rangle:E|\psi\rangle)}$. From Lemma 4, $\int 2^{\mathcal{I}(|\psi\rangle:|\psi\rangle)}d\Lambda = O(1)$. The integral $\int 2^{\mathbf{I}_{\mathrm{Prob}}(E|\psi\rangle:E|\psi\rangle)}d\Lambda$ is well defined because $2^{\mathbf{I}_{\mathrm{Prob}}(E|\psi\rangle:E|\psi\rangle)} = \mathrm{Tr}\sum_{i,j}\mathbf{m}(i,j)\nu_{i,j}(|\psi\rangle\langle\psi|\otimes|\psi\rangle\langle\psi|)$, for some matrices $\nu_{i,j}$ which can be integrated over Λ. □

Theorem 33 *Relativized to POVM E, $\int 2^{\mathbf{I}_{\mathrm{Prob}}(E\sigma:E\sigma)}d\mu(\sigma) = O(1)$.*

Proof. By Lemma 5, $2^{\mathcal{I}(\sigma:\sigma)} \stackrel{*}{>} 2^{\mathbf{I}_{\mathrm{Prob}}(E\sigma:E\sigma)}$. From Lemma 4, $\int 2^{\mathcal{I}(\sigma:\sigma)}d\mu(\sigma) = O(1)$. □

An implication of Theorems 32 and 33 is that for an overwhelming majority of quantum states, the probabilities induced by a measurement will have negligible self information.

6.3.1 Algorithmic Content of Decoherence

Decoherence was explained in Section 5.3.3. In the idealized case, decoherence transforms an arbitrary density matrix σ into a classical probability, with the off-diagonal terms turned to 0. Let p_σ be the classical probability that σ decoheres to, with $p_\sigma(i) = \sigma_{ii}$. The following corollary to Theorem 32, for an overwhelming majority of pure or mixed states σ, p_σ is noise, that is, has negligible self-information. The corollary follows from the fact that there is a POVM E, where $E_i = |i\rangle\langle i|$ with $E_i|\psi\rangle = p_{|\psi\rangle}(i)$.

Corollary 8 *Let Λ be the uniform distribution on the unit sphere of an n qubit space.*

- $\int 2^{\mathbf{I}_{\mathrm{Prob}}\left(p_{|\psi\rangle}:p_{|\psi\rangle}\right)} d\Lambda = O(1)$,
- $\int 2^{\mathbf{I}_{\mathrm{Prob}}(p_\sigma:p_\sigma)} d\mu(\sigma) = O(1)$.

6.4 PVMs

Quantum measurements is also of the form of PVMs, or projection value measures. A PVM $P = \{\Pi_i\}$ is a collection of projectors Π_i with $\sum_i \Pi_i = I$, and $\mathrm{Tr}\Pi_i\Pi_j = 0$ when $i \neq j$. When a measurement occurs, with probability $\langle\psi|\Pi_i|\psi\rangle$, the value i is measured, and the state collapses to

$$|\psi'\rangle = \Pi_i|\psi\rangle / \sqrt{\langle\psi|\Pi_i|\psi\rangle}.$$

Further measurements of $|\psi'\rangle$ by P will always result in the i measurement, so $P|\psi'\rangle(i) = 1$. To look at the effects of a measurement operation on the algorithmic information theoretic properties of a state, take a PVM, $F = \{\Pi_i\}_{i=1}^{2^{n-c}}$, where n is the number of qubits of the Hilbert space. Let Λ_F be the distribution of pure states when F is measured over the uniform distribution Λ over n qubit spaces. Thus Λ_F represents the F-collapsed states from Λ.

Theorem 34 $n - 2c <^+ \log \int 2^{\mathbf{I}_{\mathrm{Prob}}(F:|\psi\rangle:F|\psi\rangle)} d\Lambda_F$.

Proof. Note that $\int \langle\psi|\Pi_i|\psi\rangle d\Lambda = \mathrm{Dim}(\Pi_i)2^{-n}$. Furthermore, let $\kappa \subset \{1,\dots,2^{n-c}\}$ be the set of numbers $a \in \kappa$ such that $\mathbf{K}(a) >^+ n - c$. So $|\kappa| \overset{*}{>} 2^{n-c}$. We have that if $\langle\psi|\Pi_i|\psi\rangle = 1$ then $\mathbf{I}_{\mathrm{Prob}}(F|\psi\rangle : F|\psi\rangle) = \mathbf{I}_{\mathrm{Prob}}(j \mapsto [i=j] : j \mapsto [i=j]) = \mathbf{I}(i:i) =^+ \mathbf{K}(i)$.

$$\begin{aligned}
&\int 2^{\mathbf{I}(F:|\psi\rangle:F|\psi\rangle)} d\Lambda_F \\
&= \sum_{i=1}^{2^{n-c}} \mathrm{Dim}(\Pi_i)2^{-n}2^{\mathbf{K}(i)} \\
&\overset{*}{>} \sum_{i\in\kappa} \mathrm{Dim}(\Pi_i)2^{-n}2^{n-c} \\
&\overset{*}{>} |\kappa|2^{-n}2^{n-c} \\
&\overset{*}{>} 2^{n-2c}.
\end{aligned}$$

Chapter 7

Infinite Quantum Spin Chains

A qubit abstracts the properties of a single spin 1/2 particle. A complex system can be described by the collection of qubits, which model properties of superposition and entanglement. It can be convenient to consider a system's *thermodynamic limit*, which is the limit of infinite system size. This model is an infinite quantum spin chain. In the study of infinite quantum spin chains one can make a distinction between local and global effects. In addition, one does not need to consider boundary conditions.

A Martin Löf random sequence is the accepted definition in AIT for a random infinite sequence. Can one define a quantum Martin Löf infinite quantum state? This chapter shows that this can be answered in the affirmative, and even landmark theorems in AIT like the Levin-Schnorr theorem can transfer over to the quantum domain.

We first review Martin Löf random sequences. A Martin Löf test is an effective null set of the form $\bigcap_n G_n$, where the measure of open set G_n of the Cantor space goes toward zero. An infinite sequence passes a Martin Löf test if it is not contained in its null set. A Martin Löf random infinite sequence passes all Martin Löf tests. Let MLR be the set of Martin Löf random sequences.

In [NS19], the set of random infinite quantum states was introduced, which we call a NS random state. Just like the classical setting, a NS random state passes allso-called NS tests. An NS test is a quantum analog to Martin Löf tests, and it is defined by projections instead of open sets.

7.1 Infinite Quantum Bit Sequences

Before we introduce NS random sequences, we revisit the notion of C^* algebras and functional states. A C^* algebra, $\mathcal{M}$, is a Banach algebra and a function $*:\mathcal{M}\to\mathcal{M}$ such that

- For every $x\in\mathcal{M}$, $x^{**}=x$.
- For every $x,y\in\mathcal{M}$, $(x+y)^*=x^*+y^*$ and $(xy)^*=y^*x^*$.
- For every $\lambda\in\mathbb{C}$ and $x\in\mathcal{M}$, $(\lambda x)^*=\overline{\lambda}x^*$.
- For all $x\in\mathcal{M}$, $\|x^*x\|=\|x\|\|x^*\|$.

A C^* algebra $\mathcal{M}$ is unital if it admits a multiplicative identity $\mathbf{1}$. A state over unital $\mathcal{M}$ is a positive linear functional $Z:\mathcal{M}\to\mathbb{C}$ such that $Z(\mathbf{1})=1$. States are used to define NS random sequences. The set of states of $\mathcal{M}$ is denoted by $S(\mathcal{M})$. A state is tracial if $Z(x^*x)=Z(xx^*)$, for all $x\in\mathcal{M}$.

The C^* algebra over matrices of size 2^k over $\mathbb{C}$ is denoted by $\mathcal{M}_k$. Each state $\rho\in S(\mathcal{M}_k)$, can be identified with a density matrix S such that $\rho(X)=\mathrm{Tr}SX$, for all $X\in\mathcal{M}$. States that cannot be represented as the convex combination of other states are called pure states. Otherwise they are called mixed states. States are used interchangeably with density matrices, depending on the context. The tracial state $\tau_n \in S(\mathcal{M}_n)$ corresponds to the matrix $2^{-n}I_{2^n}$. The algebra $\mathcal{M}_\infty$ is the direct limit of the ascending sequence of $\mathcal{M}_n$. . A state $Z \in S(\mathcal{M}_\infty)$ over $\mathcal{M}_\infty$ can be seen as a sequence of density matrices $\{\rho_n\}$ that are coherent under partial traces, with $\mathrm{Tr}_{\mathcal{M}_{n+1}}\rho_{n+1} = \rho_n$. We use $Z\lceil n$ to denote the restriction of state Z to the algebra $\mathcal{M}_n$. There is a unique tracial state $\tau \in S(\mathcal{M}_\infty)$, where $\tau\lceil n = \tau_n$. A projection $p \in \mathcal{M}_\infty$ is a self adjoint positive element such that $p = p^2$. A special projection $p \in \mathcal{M}_n$ is a projection represented by an elementary matrix.

7.1.1 NS Randomness

An NS Σ^0_1 set is a computable sequence of special projections $\{p_i\}$ in $\mathcal{M}_\infty$ with $p_i \leq p_{i+1}$ over all i. For state ρ and NS Σ^0_1 set G, $\rho(G) = \sup_i \rho(p_i)$.

We define NS tests. But initially, we will provide the definition for the classical Martin Löf random sequence, to provide a point of reference. A classical Martin Löf test, is a sequence $\{U_n\}$ of uniformly Σ^0_1 sets of infinite sequences $U_n \subseteq \{0,1\}^\infty$ such that $\mu(U_n) \leq 2^{-n}$. An infinite sequence $\alpha \in \{0,1\}^\infty$ is Martin-Löf random if there is no Martin Löf test $\{U_n\}$ such that $\alpha \in \bigcap_n U_n$. There is a universal Martin Löf test $\{V_n\}$ such that if $\alpha \notin \bigcap_n V_n$, then α is random.

Mirroring the classical case, a NS test is an effective sequence of NS Σ^0_1 sets $\langle G^r \rangle$ such that $\tau(G^r) \leq 2^{-r}$. Unlike a classical test, which can either pass or fail a sequence, a NS test can pass a quantum state up to a particular order. For $\delta \in (0,1)$, state $Z \in S(\mathcal{M}_\infty)$ fails test $\langle G^r \rangle$ at order δ if $Z(G^r) > \delta$ for all r. Otherwise Z passes the test at order δ. We says Z passes a NS test if it passes it at all orders $\delta \in (0,1)$.

A state is NS random if it passes every NS test at every order.

Theorem 35 ([NS19]) *There exists a universal NS test $\langle R^n \rangle$, where for each NS test $\langle G^k \rangle$ and each state Z and for each n there exists a k such that $Z(R^n) \geq Z(G^k)$.*

Proof. Let $\langle G^k_n \rangle_{n=1}^\infty$ be an enumeration of NS tests, performed analgously to the classical case (see [GÓ1]). Furthermore let $G^e_m = \langle p^e_{m,r} \rangle_{r\in\mathbb{N}}$. For each $k, n \in \mathbb{N}$, let $q^n_k = \bigvee_{e+n+1\leq k} p^e_{e+n+1,k}$. Thus $q^n_k \leq q^n_{k+1}$ and $\tau q^n_k \leq \sum_e \tau(p^e_{e+n+1,k}) \leq 2^{-n}$. The universal NS test is $R^n = \langle q^n_k \rangle_{k\in\mathbb{N}}$. Since $\tau(R^r) \leq 2^{-n}$, $\rangle R^n \rangle$ is a NS test. For a set e,

$$\rho(R^n) = \sup_k \rho(q^n_k) \geq sup_k \rho(p^e_{n+e+1,k}) = \rho(G^e_{n+e+1}).$$

□

A state Z is NS random if it passes the test $\langle R^n \rangle$. . More information about $\langle R^r \rangle$ can be found in [NS19].

7.2 Closure Properties

The set of NS random sequences has closure properties over (possibly noncomputable) convex combinations, as shown in the following theorem.

Theorem 36 *Every convex combination $Z = \sum_i \alpha Z_i$ of NS random states Z_i, with $\sum_i \alpha_i = 1$ and $\alpha_i \geq 0$, is NS random.*

Proof. Given an NS test $\langle G^r\rangle = \langle p_t^r\rangle$, there exists a NS test $\langle H^r\rangle$ such that for all states Z, $\inf_r Z(H^r) \geq \inf_r Z(G^r)$ and $H^r \supseteq H^{r+1}$. This is by setting H^r equal to $\bigvee_{i\geq r} G^i$. More formally, $\langle H^r\rangle = \langle q_t^r\rangle$, where $q_t^r = \bigvee_{i=1}^t p_t^{r+i}$. Thus there exists a universal NS test $\langle L^r\rangle$ such that $L^r \supseteq L^{r+1}$. Assume that Z is not NS random. Then

$$\lim_{r\to\infty} Z(L^r) > 0$$
$$\lim_{r\to\infty} \sum_i \alpha_i Z_i(L^r) > 0$$
$$\sum_i \alpha_i \lim_{r_i\to\infty} Z_i(L^{r_i}) > 0.$$

So there exists an i such that $\lim_{r\to\infty} Z_i(L^r) > 0$, and thus Z_i is not NS random.

7.3 Gács Complexity and NS Random Sequences

In this section, we characterize NS random states in terms of Gács complexity, $\mathbf{Hg}$.

Theorem 37 *Given state $Z \in \mathcal{M}_\infty$, and program p that enumerates infinite set $A \subseteq \mathbb{N}$, then $\sup_{n\in\mathbb{N}} n - \mathbf{Hg}(Z{\restriction}n) <^+ \sup_{n\in A} n - \mathbf{Hg}(Z{\restriction}n) + \mathbf{K}(p)$.*

Proof. There exists a program p' of size $\|p\|+O(1)$ that outputs a list $\{a_n\} \subseteq A$ such that $n < a_n$. For a given a_n, $\sigma = 2^{n-a_n}\boldsymbol{\mu}_n \otimes I_{a_n - n}$ is a lower computable $2^{a_n} \times 2^{a_n}$ semi-density matrix. There is a program $q = q'\langle a_n, n\rangle$ that lower computes σ where q' is helper code that uses the encodings of a_n and n. By the universal properties of $\boldsymbol{\mu}$, we have the inequality $\mathbf{m}(q|a_n)\sigma \overset{*}{<} \boldsymbol{\mu}_{a_n}$. So, using properties of partial trace,

$$\begin{aligned}
a_n + \log \mathbf{m}(q|a_n)\mathrm{Tr}\sigma Z{\restriction}a_n &<^+ a_n + \log \mathrm{Tr}\boldsymbol{\mu}(Z{\restriction}a_n)\\
a_n + \log \mathrm{Tr}2^{n-a_n}(\boldsymbol{\mu}_n \otimes I_{a_n-n})Z{\restriction}a_n - \mathbf{K}(q|a_n) &<^+ a_n + \log \mathrm{Tr}\boldsymbol{\mu}(Z{\restriction}a_n)\\
n + \log \mathrm{Tr}(\boldsymbol{\mu}_n \otimes I_{a_n-n})Z{\restriction}a_n - \mathbf{K}(\langle n, a_n\rangle|a_n) &<^+ a_n + \log \mathrm{Tr}\boldsymbol{\mu}(Z{\restriction}a_n)\\
n + \log \mathrm{Tr}(\boldsymbol{\mu}_n \mathrm{Tr}_n Z{\restriction}a_n) - \mathbf{K}(p'|a_n) &<^+ a_n + \log \mathrm{Tr}\boldsymbol{\mu}(Z{\restriction}a_n)\\
n - \mathbf{Hg}(Z{\restriction}n) &<^+ a_n - \mathbf{Hg}(Z{\restriction}a_n) + \mathbf{K}(p).
\end{aligned}$$

So $\sup_{n\in\mathbb{N}} n - \mathbf{Hg}(Z{\restriction}n) <^+ \sup_{a_n\in\{a_n\}} a_n - \mathbf{Hg}(Z{\restriction}a_n) + \mathbf{K}(p) <^+ \sup_{n\in A} n - \mathbf{Hg}(Z{\restriction}n) + \mathbf{K}(p)$.

Theorem 38 *Suppose for state Z, and for infinite enumerable set $A \subseteq \mathbb{N}$, $\sup_{n\in A} n - \mathbf{Hg}(Z \restriction n) < \infty$. Then Z is NS random.*

Proof. Suppose Z is not NS random. Let $L^r = \langle p_t^r\rangle$ be the universal NS test. So $\mathrm{Rank}(p_n^r) \leq 2^{n-r}$. Thus $\inf_r Z(L^r) = \delta > 0$. For each r, there exists an n such that $\mathrm{Tr}(p_n^r z_n) \geq \delta$, where $z_n = Z \restriction n$. Since $2^{r-n}p_n^r$ is a computable semi-density matrix given n and r, $\mathbf{m}(r|n)2^{r-n}p_n^r \overset{*}{<} \boldsymbol{\mu}$. So $\mathbf{m}(r|n)2^{r-n}\delta \overset{*}{<} \mathrm{Tr}\boldsymbol{\mu}z_n$, which implies that $\mathbf{Hg}(Z \restriction n) <^+ n - r + \mathbf{K}(r|n)$. Since this property holds for all $r \in \mathbb{N}$, $\sup_n n - \mathbf{Hg}(Z \restriction n) = \infty$. From Theorem 37, $\sup_{n\in A} n - \mathbf{Hg}(Z \restriction n) = \infty$.

7.4 Encodings of States

Let $[Z] \in \{0,1\}^\infty$ be an encoding of the state Z described as follows. For each n, let $e(n,m)$ be the mth enumeration of a pair (p,k) consisting of a special projection p of $\mathcal{M}_n$ and a rational $0 \leq k \leq 1$. For $[Z]$, the ith bit, where $i = 2^n m$ for maximum n, corresponds to 1 if and only if $\mathrm{Tr} pZ{\restriction}n > k$, where (p,k) is the pair enumerated by $e(n,m)$. We say that state $Z \in \mathcal{QH}$ if and only if the halting sequence can be computed from $[Z]$.

7.5 Quantum Operation Complexity

In a canonical algorithmic information theory example, Alice wants to send a single text message x to Bob. Alice sends a program q to Bob such that $x = U(q)$, where U is a fixed universal Turing machine. The cost of the transmission is the length of q. Alice can minimize cost by sending $\mathbf{K}(x)$ bits to Bob, where $\mathbf{K}$ is the Kolmogorov complexity function.

We now look at the quantum case. Suppose that Alice wants to send a (possibly mixed) n qubit quantum state σ to Bob, represented as an density matrix over $\mathbb{C}^{2^n}$, or an element of $S(\mathcal{M}_n)$. Alice has access to two channels, a quantum channel and a classical channel. Alice can choose to send $m \leq n$ qubits ρ on the quantum channel and classical bits $q \in \{0,1\}^*$ on the classical channel, describing an elementary quantum operation η, where $U(q) = [\eta]$. Bob then applies η to ρ to produce $\sigma' = \eta(\rho)$. Bob is not required to produce σ exactly. Instead the fidelity of the attempt is measured by the trace distance between σ and σ'. The trace distance D between two matrices A and B is $D(A,B) = \frac{1}{2}\|A-B\|_{\mathrm{Tr}}$, with $\|A\|_{\mathrm{Tr}} = \mathrm{Tr}|A|$. We use $\mathcal{O}_{m,n}$ to denote the set of elementary quantum operations that take m qubit quantum states to $n \geq m$ qubit quantum states.

Definition 9 *For n qubit density matrix σ, the quantum operation complexity at accuracy ϵ is* $\mathbf{Hoc}^\epsilon(\sigma) = \min\{\mathbf{K}([\eta]) + m : \eta \in \mathcal{O}_{m,n}, \xi \in S(\mathcal{M}_m), D(\sigma, \eta(\xi)) < \epsilon\}$.

7.6 Initial Segment Incompressibility

Due to Levin and Schnorr, [Lev74, Sch71] α is random iff there is an r such that for all n, $\mathbf{K}(\alpha_{\leq n}) \geq n - r$, where $\alpha_{\leq n}$ is a prefix of α of size n, and $\mathbf{K}$ is prefix free Kolmogorov complexity. In this section, we prove a quantum analog to this result. We show that NS states that are NS random have incompressible prefixes with respect to quantum operation complexity. Theorem 39 builds upon the proof of the Theorem 4.4 in [NS19] using quantum operation complexity $\mathbf{Hoc}$.

Theorem 39 *Let Z be a state on $\mathcal{M}_\infty$.*

1. *Let $1 > \epsilon > 0$, and suppose Z passes each NS test at order $1-\epsilon$. Then there is an r where for all n, $\mathbf{Hoc}^\epsilon(Z{\restriction}_n) > n - r$.*

2. *Let $1 > \epsilon > 0$ be lower computable and Z fails some NS test at order $1-\epsilon$. Then either $Z \in \mathcal{QH}$ or for all r, there is an n where $\mathbf{Hoc}^{\sqrt{\epsilon}}(Z{\restriction}_n) < n - r$.*

Proof. (1). Let $\mathbf{K}_t(x)$ be the smallest program to produce x in time t. Let $s(n,r,t)$ be the set of pure n qubit states $\rho \in S(\mathcal{M}_n)$ such that there exists a quantum operation $\eta \in \mathcal{O}_{z,n}$ and pure state $\sigma \in S(\mathcal{M}_z)$ such that $\rho = \eta(\sigma)$ and $\mathbf{K}_t([\eta]) + z \leq n - r$. Let $p(n,r,t)$ be the orthogonal projection in $\mathcal{M}_n$ with minimum $\tau(p(n,r,t))$ such that $\rho(p(n,r,t)) = 1$ for all $\rho \in s(n,r,t)$. Let $p(r,t) = \sup_{n \leq t} p(n,r,t)$. So $p(r,t)$ is in $\mathcal{M}_t$ and $p(r,t)$ is computable from r and t and $p(r,t) \leq p(r,t+1)$.

Let $b(y,n,z)$ be the number of programs of length y which outputs an encoding of an elementary quantum operation $\eta \in \mathcal{O}_{z,n}$. Let $b(y,n)$ be the number of programs of length y which outputs an encoding of an elementary quantum operation $\eta \in \mathcal{O}_{z,n}$, for any $z \leq n$. So

$$\begin{aligned}
\mathrm{Range}(p(n,r,t)) &\leq \sum_{y+z\leq n-r} b(y,n,z)2^{z} \\
\tau(p(n,r,t)) &\leq \sum_{y+z\leq n-r} b(y,n,z)2^{z-n} \\
&\leq \sum_{y+z\leq n-r} b(y,n,z)2^{-y-r} \\
&\leq \sum_{y=1}^{n-r} b(y,n)2^{-y-r} \\
\tau(p(r,t)) &\leq \sum_{n=1}^{\infty} \tau(p(n,r,t)) \\
&\leq \sum_{n=1}^{\infty}\sum_{y=1}^{n-r} b(y,n)2^{-y-r} \\
&= 2^{-r}\sum_{n=1}^{\infty}\sum_{y=1}^{n-r} b(y,n)2^{-y} \leq 2^{-r}.
\end{aligned}$$

So for NS Σ^0_1 set G^r enumerated by the sequence $\{p(r,t)\}_t$, $\langle G^r\rangle$ is a NS test. For each r suppose there is an n such that $\mathbf{Hoc}^{\epsilon}(Z\lceil n) \leq n-r$. So there is an elementary quantum operation $\eta \in \mathcal{O}_{z,n}$ and input $\rho \in S(\mathcal{M}_z)$ such that $\mathbf{K}([\eta]) + z \leq n-r$ and $D(Z\lceil_n, \eta(\rho)) < \epsilon$. So $\eta(\rho)$ is in the range $p(n,r,t)$ for some t and so $\mathrm{Tr}\eta(\rho)p(n,r,t) = 1$. This implies $1-\epsilon < Z(p(n,r,t)) \leq Z(G^r)$. Since this is for all r, Z fails the test at order $1-\epsilon$.

(2). Let $\mathbf{bb}(n)$ be the longest running time of a halting program of length $\leq n$. Let $\langle L^r\rangle$ be the universal NS test, where each L^r is enumerated by $\{p^r_t\}$, with $p^r_t \in \mathcal{M}_{n(r,t)}$. Assume there is an infinite number of r where $\mathrm{Tr}Z\lceil n(r,\mathbf{bb}(r/2))p^r_{\mathbf{bb}(r/2)} > 1-\epsilon$. Fix one such r and let $n = n(r,\mathbf{bb}(r/2))$, and $p = p^r_{\mathbf{bb}(r/2)}$. Projection p has eigenvectors $\{u_i\}$ and kernel spanned by $\{v_i\}$. Thus $2^{-r} \geq \tau(p)$. Let $p' \geq p$ with $p' \in \mathcal{M}_n$ such that each u_i is in the range of p' and $\{v_i\}_{i=1}^k$ is in the range of p' such that k is minimized such that $\tau(p') = 2^{-r}$. Thus $\mathrm{Tr}Z\lceil n(p') > 1-\epsilon$. The eigenvectors of p' are $\{w_i\}_{i=1}^{2^{n-r}}$ and its kernel is spanned by the vectors $\{y_i\}_{i=1}^{2^n-2^{n-r}}$. Let $z' = \mathrm{Proj}(Z\lceil n; p')$ be a density matrix with eigenvalues $v_i \in \mathbb{R}$ corresponding to eigenvectors w_i. For $i \in [1,2^n]$, let $B(i) \in \{0,1\}^*$ be an encoding of n bits of the number i, with $B(1) = 0^{(n)}$, $B(2) = 10^{(n-1)}$, and $B(2^n) = 1^{(n)}$. Let U be a $2^n \times 2^n$ unitary matrix, of the form $U = \sum_{i=1}^{2^{n-r}} |B(i)\rangle\langle w_i| + \sum_{i=1}^{2^n-2^{n-r}} |B(i+2^{n-r})\rangle\langle y_i|$.

Proposition 3 ([NS19]) *Let* $\mathrm{Proj}(s;h) = \frac{1}{\mathrm{Tr}[sh]}hsh$. *Let p be a projection in M_n and σ be a density matrix in M_n. If $\alpha = \mathrm{Tr}p\sigma$ and $\sigma' = \mathrm{Proj}(\sigma;p)$ then $D(\sigma,\sigma') \leq \sqrt{1-\alpha}$.*

Proof. Let $|\psi_\sigma\rangle$ be a purification of σ. Then $\alpha^{-\frac{1}{2}}p|\psi_\sigma\rangle$ is a purification of σ'. Uhlmann's theorem states $F(\sigma,\sigma') \geq \alpha^{-\frac{1}{2}}\langle\psi_\sigma|p|\psi_\sigma\rangle = \alpha^{\frac{1}{2}}$, where F is fidelity, with $F(\sigma,\sigma') = \mathrm{Tr}\sqrt{\sqrt{\sigma'}\sigma\sqrt{\sigma'}}$. Thus the proposition follows from $D(\sigma,\sigma') \leq \sqrt{1-F(\sigma,\sigma')}$. □

Thus for the diagonal $2^{n-r} \times 2^{n-r}$ matrix σ with entries $\{v_i\}_{i=1}^{2^{n-r}}$, $z' = U(\sigma \otimes |0^r\rangle\langle 0^r|)U^*$. By Proposition 3, since $1-\epsilon < \mathrm{Tr}(p'Z{\upharpoonright}n)$ and $z' = \mathrm{Proj}(z_n;p')$, it must be that $D(z', Z{\upharpoonright}n) < \sqrt{\epsilon}$. Thus using quantum operation $\eta = (U, |0^r\rangle\langle 0^r|, \emptyset)$ and input σ,

$$\begin{aligned}
\mathbf{Hoc}^\epsilon(Z{\upharpoonright}n) &\leq \mathrm{Dim}(\sigma) + \mathbf{K}([\eta]) \\
&\leq n - r + \mathbf{K}([(U, |0^r\rangle\langle 0^r|, \emptyset)]) \\
&<^+ n - r + \mathbf{K}(n, r) \\
&<^+ n - r + \mathbf{K}(\mathbf{bb}(r/2), r) \\
&<^+ n - r + r/2 + \mathbf{K}(r) \\
&<^+ n - r/3.
\end{aligned}$$

Thus for every r there exists an n where $\mathbf{Hoc}^\epsilon(Z{\upharpoonright}n) < n - r$. This is because the additive constant of the above equation is not dependent on r.

Otherwise there is some R where for all $r \geq R$, and $q < \mathbf{bb}(r/2)$, $\mathrm{Tr} Z_{n(r,q)} p^r_{n(r,q)} \leq 1-\epsilon$. Thus given R, $\langle L^r\rangle$, $[Z]$, and a lower enumeration of ϵ, one can iterate through each $r \geq R$ and return an s such that $\mathrm{Tr} Z_{n(r,s)} p^r_{n(r,s)} > 1-\epsilon$. This is because the set of rational numbers Q such that $q > 1-\epsilon$ for all $q \in Q$ can be enumerated and the set $V = \{v : \mathrm{Tr} Z_{n(r,v)} p^r_{n(r,v)} > q, q \in Q\}$ can be enumerated using the infinite encoding $[Z] \in \{0,1\}^\infty$. The returned s is the first enumerated element of V. This number s has the property that $s \geq \mathbf{bb}(r/2)$, and can be used to compute the prefix of the halting sequence over all programs of length $\leq r/2$ as every such program that will halt will do so in less than s steps. Thus the halting sequence is computable relative to $[Z]$ and thus $Z \in \mathcal{QH}$.

Corollary 9 *Let state $Z \notin \mathcal{QH}$. Then Z is NS random iff for all $0 < \epsilon < 1$, there is an r, where for all n, $\mathbf{Hoc}^\epsilon(Z \upharpoonright n) > n - r$.*

Proof. Assume Z is NS random. Then for all $0 < \epsilon < 1$, Z passes each NS test at order $1-\epsilon$. Then by Theorem 39 (1), for all $0 < \epsilon < 1$ there is an r where for all n, $\mathbf{Hoc}^\epsilon(Z{\upharpoonright}n) > n-r$. Assume Z is not NS random. Then there is some rational $0 < \delta < 1$ such that Z fails some NS test at order $1-\delta$. Then by Theorem 39 (2), for $\epsilon = \sqrt{\delta}$, for all r, there is an n where $\mathbf{Hoc}^\epsilon(Z{\upharpoonright}n) < n - r$.

7.7 Quantum Ergodic Sources

In [Bru78], Brudno proved that for ergodic measures η over bi-infinite sequences, for η-almost all sequences, the rate of the Kolmogorov complexity of their finite prefixes approaches the entropy rate of η. Therefore the average compression rate of sequences produced by η is not more than its entropy rate. In [BKM+06], a quantum version of Brudno's theorem was introduced relating, in a similar fashion, Von Neumann entropy and BVL complexity (using the fidelity measure). The results provide two bounds with respect to two variants of **Hbvl**: approximate-scheme complexity and finite accuracy complexity.

In this subsection we provide a quantum variant of Brudno's theorem with respect to quantum communication complexity $\mathbb{R}^\epsilon$. Differently from the **Hbvl** results, the bounds provided below are for almost all n, invariant to the accuracy term ϵ.

We define the quasilocal C^* algebra $\mathcal{A}_\infty$, which differs only from $\mathcal{M}_\infty$ in that it is a doubly infinite product space over $\mathbb{Z}$. In particular, $\mathcal{A}$ is the C^* algebra of qbits, i.e. 2×2 matrices acting on $\mathbf{C}^2$. For finite $\Lambda \subset \mathbb{Z}$, $\mathcal{A}_\Lambda = \bigotimes_{z\in\Lambda} \mathcal{A}_z$.

The quasilocal C^* algebra $\mathcal{A}_\infty$ is defined to be the norm closure of $\bigcup_{\Lambda\subset\mathbb{Z}}\mathcal{A}_\Lambda$. For states Ψ over $\mathcal{A}_\infty$, we use $\Psi\lceil n$ to denote Ψ restricted to the finite subalgebra $\mathcal{A}_{\{1,\dots,n\}}$ of $\mathcal{A}_\infty$. The right shift T is a $*$-automorphism on $\mathcal{A}_\infty$ uniquely defined by its actions on local observables $T : a \in \mathcal{A}_{\{m,\dots,n\}} \mapsto \mathcal{A}_{\{m+1,\dots,n+1\}}$. A quantum state Ψ is stationary if for all $a \in \mathcal{A}_\infty$, $\Psi(a) = \Psi(T(a))$. The set of shift-invariant states on $\mathcal{A}_\infty$ is convex and compact in the weak* topology. The extremal points of this set are called ergodic states.

Lemma 6 *Let R_j be the smallest subspace spanned by pure states produced by elementary quantum operations $\eta \in \mathcal{O}_{z,n}$ with $\mathbf{K}(\eta) + z < j$. Then $Dim(R_j) < 2^j$.*

Proof. Let $b(y,z)$ be the number of programs of length y that outputs an elementary quantum operation $\eta \in \mathcal{O}_{x,z}$ over the Hilbert space $\mathcal{H}_{2^n}$. Let $b(y)$ be the number of programs of length y that outputs an elementary quantum operation $\mathcal{O}_{z,n}$ over the Hilbert space $\mathcal{H}_{2^n}$.

$$\begin{aligned}\mathrm{Dim}(R_j) &\le \sum_{y+z<j} b(y,z)2^z\\ &= 2^j \sum_{y+z<j} b(y,z)2^{z-j}\\ &< 2^j \sum_{y,z} b(y,z)2^{-y}\\ &= 2^j \sum_{y} b(y)2^{-y}\\ &\le 2^j.\end{aligned}$$

Theorem 40 *Let Ψ be an ergodic state with mean entropy h. For all $\delta > 0$, for almost all n, there is an orthogonal projector $P_n \in \mathcal{A}_n$ such that for all $\epsilon > 0$,*

1. *$\Psi\lceil n(P_n) > 1 - \delta$.*
2. *For all one dimensional projectors $p \le P_n$, $\mathbf{Hoc}^\epsilon(p)/n \in (h-\delta, h+\delta)$.*

Proof. Let $\delta' < \delta'' < \delta$. From [BDK+05], there is a sequence of projectors $P'_n \in \mathcal{A}_n$ where for almost all n, $\Psi\lceil n(P'_n) > 1-\delta'$, for all one dimensional projectors $p' \le P'_n$, $2^{-n(h+\delta')} < \Psi\lceil n(p') < 2^{-n(h-\delta')}$, and $2^{n(h-\delta')} < \mathrm{Tr}P'_n < 2^{n(h+\delta')}$. Let S'_n be the subspace that P'_n projects onto. Let R_n be the smallest subspace spanned by all pure states produced by an elementary quantum operation $\eta \in \mathcal{O}_{g,n}$, where $\mathbf{K}(\eta) + g < n(h-\delta'')$. Let Q_n be the projector onto R_n. By Lemma 6, $\mathrm{Dim}(R_n) < 2^{n(h-\delta'')}$. Let S_n be the largest subspace of S'_n that is orthogonal to R_n. Let P_n be the orthogonal projector onto S_n. So for sufficiently large n, $\Psi\lceil n(P_n) \ge \Psi\lceil n(P'_n) - \mathrm{Dim}(R_n)2^{-n(h-\delta')} > 1-\delta'-2^{n(h-\delta'')}2^{-n(h-\delta')} = 1-\delta'-2^{n(\delta'-\delta'')} > 1-\delta$, for large enough n.

By definition, since P_n is orthogonal to R_n, for all ϵ, for all one dimensional projectors $p \le P_n$, $\mathbf{Hoc}^\epsilon(p) \ge n(h-\delta'') > n(h-\delta)$. Furthermore, all such p can be produced from an elementary quantum operation η that maps $\lceil n(h+\delta')\rceil$ length pure states into S_n. Therefore for large enough n, $\mathbf{Hoc}^\epsilon(p) \le \mathbf{K}(\eta) + \lceil n(h+\delta')\rceil <^+ \mathbf{K}(n,h) + \lceil n(h+\delta')\rceil < n(h+\delta)$.

7.8 Measurement Systems

We note that pre-measures are of the form $\gamma : \{0,1\}^* \to \mathbb{R}_{\geq 0}$, where $\gamma(x) = \gamma(x0) + \gamma(x1)$. By the Carathéodory's Extension Theorem, each such pre-measure can be uniquely extended to a measure Γ over $\{0,1\}^\infty$. in Chapter 6, measurements of finite collections of qubits are studied. This section deals with measurement measurement systems, which can be applied to infinite quantum states.

Definition 10 (Meaurement System ([Bho21])) *An α-computable measurement system $B = \{(|b_0^n\rangle, |b_1^n\rangle)\}$ is a sequence of orthonormal bases for $\mathcal{Q}_1$ such that each $|b_i^n\rangle$ is elementary and the sequence $\langle |b_1^n\rangle, |b_0^n\rangle\rangle_{n=1}^\infty$ is α-computable.*

Note that the above defiinition can be generalized to a sequence of PVMs. We now define the application of a measurement system B to an infinite quantum state Z which produces a pre-measure p. Let ρ_n be the density matrix associated with $Z \restriction n$. For the first bit, we use the standard definition of measurement, where

$$p(i) = \mathrm{Tr}\, |b_i^1\rangle \langle b_i^1| \rho_1.$$

Given ρ_2, if i is measured on the first bit, then the resulting state would be

$$\rho_2^i = \frac{(|b_i^1\rangle \langle b_i^1| \otimes I)\rho_2(|b_i^1\rangle \langle b_i^1| \otimes I)}{\mathrm{Tr}(|b_i^1\rangle \langle b_i^1| \otimes I)\rho_2}$$

So

$$\begin{aligned} p(ij) &= p(i)p(j|i) \\ &= \left(\mathrm{Tr}\, |b_i^1\rangle \langle b_i^1| \rho_1\right) \mathrm{Tr}\left(I \otimes |b_j^2\rangle \langle b_j^2|\right) \left(\frac{\left(|b_i^1\rangle \langle b_i^1| \otimes I\right) \rho_2 \left(|b_i^1\rangle \langle b_i^1| \otimes I\right)}{\left(|b_i^1\rangle \langle b_i^1| \otimes I\right) \rho_2} \right) \end{aligned}$$

Since $\mathrm{Tr}_2\rho_2 = \rho_1$, $\mathrm{Tr}\, |b_i^1\rangle \langle b_i^1| \rho_1 = \mathrm{Tr}\left(|b_i^1\rangle \langle b_i^1| \otimes I\right) \rho_2$. Therefore

$$p(ij) = \mathrm{Tr}\rho_2 \left(|b_i^1 b_j^2\rangle \langle b_i^1 b_j^2|\right).$$

More generally for $x \in \{0,1\}^n$, we define the pre-measure p to be

$$p(x) = \mathrm{Tr}\rho_n \, |\otimes_{i=1}^n b_{x_i}^i\rangle \langle \otimes_{i=1}^n b_{x_i}^i| \, ;$$

It is straightforward to see that p is a pre-measure, with $p(x) = p(x0) + p(x1)$. Let μ_Z^B be the measure over $\{0,1\}^\infty$ derived from the described pre-measure, using measurement system B and state Z. We recall that MLR is the set of Martin Löf random sequences.

Definition 11 (Bhojraj Random) *A state Z is Bhojraj Random if for any computable measurement system B, $\mu_Z^B(\mathrm{MLR}) = 1$.*

Theorem 41 ([Bho21]) *All NS Random states are Bhojraj Random states,*

Proof. Let state Z be NS random. Let $\{\rho_n\}$ be the density matrices associated with Z. Suppse not. Then there is $\delta \in (0,1)$ and computable measurement system $B = \{|b_0^n\rangle, |b_1^n\rangle\}_{n=1}^{\infty}$ where $\mu_Z^B(\{0,1\}^\infty \setminus \mathrm{MLR}) > \delta$. Let $\{S^m\}$ be a universal ML test. Without loss of generality, this test is of the form

$$S^m = \bigcup_{m \le i} [\![A_i^m]\!],$$

where $[\![A_i^m]\!] \subseteq [\![A_{i+1}^m]\!]$, and $A_i^m\{\tau_1^{m,i}, \ldots, \tau_{k^{m,i}}^{m,i}\} \subset \{0,1\}^i$ for some $0 \le k^{m-i} \le 2^{i-m}$. Thus $\mu(S^m) \le 2^{-m}$, where μ is the uniform distribution over $\{0,1\}^\infty$. We define an NS test as follows. For all m and i, with $m \le i$, let $\tau_a = \tau_a^{m,i}$ and define the special projection

$$p_i^m = \sum_{a \le k^{m,i}} |\otimes_{q=1}^i b_{\tau_\alpha[q]}^q\rangle \langle \otimes_{q=1}^i b_{\tau_\alpha[q]}^q| \,.$$

We define $P^m = \{p_i^m\}_{m \le i}$ we have that $\langle P^m \rangle$ is an NS Test. The special tests p_i^m is uniformly computable in i and m since B and A_i^m are uniformly computable in i and m. Since $[\![A_i^m]\!] \subseteq [\![A_{i+1}^m]\!]$, $\mathrm{Range}(p_i^m) \subseteq \mathrm{Range}(p_{i+1}^m)$. So P^m is an NS Σ_1^0 set for all m. Since $k^{m,i} \le 2^{i-m}$ for all m and i, this implies $\tau(P^m) \le 2^{-m}$ for all m.

For all m, $\{0,1\}^\infty \setminus \mathrm{MLR} \subseteq S^m$. Since by assumption $\mu_Z^B(\{0,1\}^\infty \setminus \mathrm{MLR}) > \delta$, for all m there exists $i(m) > m$ such that

$$\mu_Z^B([\![A_{i(m)}^m]\!]) > \delta.$$

Fix an m and $i = i(m)$ and let $A_i^m = \{\tau_1, \ldots, \tau_{k^{m,i}}\}$, where $k^{m,i} \le 2^{i-m}$. Let p be the pre-measure associated witth μ_Z^B. So we have

$$\delta < \sum_{a \le k^{m,i}} p(\tau_a) = \sum_{a \le k^{m,i}} \mathrm{Tr}\rho_i \, |\otimes_{q=1}^i b_{\tau[q]}^q\rangle \langle \otimes_{q=1}^i b_{\tau[q]}^q| = \mathrm{Tr}\rho_i \sum_{a \le k^{m,i}} |\otimes_{q=1}^i b_{\tau[q]}^q\rangle \langle \otimes_{q=1}^i b_{\tau[q]}^q|$$

So we see that for all m there is an i such that

$$\delta < \mathrm{Tr}\rho_i p_i^m \le Z(P^m).$$

So $\inf_m Z(P^m) > \delta$, contradicting that Z is NS random. □

Theorem 42 ([Bho21]) *There are states that are Bhojraj random and not NS Random.*

7.9 NS Solovay States

A NS Solovay test is a sequence of NS Σ_0^1 sets $\langle G_n \rangle$ such that $\sum_n \tau(G^n) < \infty$. A state Z fails a quantum NS test $\langle G^r \rangle$ at order $\delta \in (0,1)$ if there is an infinite number of $r \in R$ such that $\inf_{r \in R} Z(G^r) > \delta$. Otherwise state Z passes the quantum NS test at order δ. A quantum state Z is NS Solovay random if it passes all NS Solovay tests at all orders. The following theorem shows the equivalence of NS randomness and NS Solovay randomness with respect to every order δ. Given a special projection p, NS Σ_0^1 set $Q = \{q_n\}$, and state Z, we define $Z(p \setminus Q) = \inf_n Z(p \setminus q_n)$. In [Bho21], it was proven that NS randomness is equivalent to NS Solovay randomness.

Proposition 4 *Given a special projection p, NS Σ_0^1 set Q, and state Z, $Z(p) - Z(Q) \le Z(p \setminus Q) \le Z(p)$.*

The proof is straightforward.

Theorem 43 *If a state Z fails an NS test at order δ then it fails an NS Solovay test at order δ.*

Proof. Assume that state Z fails a NS test $\langle G^r \rangle$ at order δ. Since $\sum_r \tau(G^r) \leq 1$, and each G^r is an NS Σ_1^0 set, $\langle G^r \rangle$ is a NS Solovay test. Furthermore since $\inf_r Z(G^r) \geq \delta$, there exists an infinite number of r such that $Z(G^r) > \delta$. Thus Z fails a NS Solovay test at order δ. □

Theorem 44 *For all $\delta' < \delta$, if a state Z fails an NS Solovay test at order δ then it fails an NS test at order δ'.*

Proof. Assume state Z fails NS Solovay test $\langle G^r \rangle$ at order δ. Given $\langle G^r \rangle$, where $G^r = \langle p_n^r \rangle_{n \in \mathbb{N}}$, we construct an NS test $\langle H^r \rangle$ as follows. There exists an m such that $\sum_{n>m} \tau(G^n) \leq 1$. Fix r. Enumerate all unordered sets of $r+1$ natural numbers $\{D_n^r\}_{n \in \mathbb{N}}$, $D_n^r \subset \mathbb{N}$, with infinite repetition.

$$H^r = \{q_n^r\}, q_n^r = \bigvee_{m<n} q_m^r \bigvee \left(\bigwedge_{t \in D_n^r} p_n^t \right).$$

Each H^r can be seen to be an NS Σ_1^0 set. Furthermore $\tau(H^r) \leq \sum_{t>r} Z(G^t) \leq \sum_{t>r} 2^{-t} = 2^{-r}$. So $\langle H^r \rangle$ is an NS test. For each r, $Z(H^r) > \delta'$. Assume not. Then there exists a k such that $Z(H^k) \leq \delta'$. Since Z fails $\langle G^r \rangle$ at order δ, there exists an infinite number of $r \in R$ and $n_r \in \mathbb{N}$ such that $Z(p_{n_r}^r) \geq \delta''$, for some $\delta' < \delta'' < \delta$. We reorder the NS Solovay test $\langle G^r \rangle$ such that r ranges over solely R. Let $z_r = p_{n,r}^r$. Let $D_{n,k}$ be the set of all unordered subsets of $\{1, \ldots, n\}$ of size k. For $k > n$ let $F_{n,k} = \emptyset$. Let

$$F_{n,k} = \left(\bigvee_{A \in D_{n,k}} \bigwedge_{r \in A} z^r \right) \setminus \bigvee_{s>k} F_{n,s}.$$

So for all $n \in \mathbb{N}$, using Proposition 4,

$$\begin{aligned}
& n(\delta'' - \delta') \\
\leq & \sum_{r=1}^{n} (Z(z^r) - Z(H^k)) \\
\leq & \sum_{r=1}^{n} Z(z^r \setminus H^k) \\
\leq & \sum_{r=1}^{n} Z\left(\bigvee_{s=1}^{k} F_{n,s} \wedge z^r \right) \qquad (7.1)
\end{aligned}$$

Equation 7.1 is due to the fact that for $s > k$ there is a t where we have $\mathrm{Range}(F_{n,s}) \leq \mathrm{Range}(q_s^k)$. Let $F_{n,s,r} = F_{n,s} \wedge z^r$, with for a fixed $s \leq k$, $\sum_{i=1}^{n} Z(F_{n,s,i}) \leq s$.

$$\begin{aligned}
& n(\delta'' - \delta') \\
\leq & \sum_{r=1}^{n} Z\left(\bigvee_{s=1}^{k} F_{n,s,r} \right) \\
= & \sum_{s=1}^{k} \sum_{r=1}^{n} Z(F_{n,s,r}) \\
\leq & \sum_{s=1}^{k} s = O(k^2).
\end{aligned}$$

This is a contradiction for large enough n. □

Corollary 10 *A quantum state is NS random if and only if it is NS Solovay random.*

Chapter 8

The Many Worlds Theory

The Many Worlds Theory (**MWT**) was formulated by Hugh Everett [Eve57] as a solution to the measurement problem of Quantum Mechanics. Branching (a.k.a splitting of worlds) occurs during any process that magnifies microscopic superpositions to the macro-scale. This occurs in events including human measurements such as the double slit experiments, or natural processes such as radiation resulting in cell mutations.

One question is if **MWT** causes issues with the foundations of computer science. The physical Church Turing Thesis (**PCTT**) states that any functions computed by a physical system can be simulated by a Turing machine. A straw man argument for showing **MWT** and **PCTT** are in conflict is an experiment that measures the spin of an unending number of electrons, with each measurement bifurcating the current branch into two sub-branches. This results in a single branch in which the halting sequence is outputted. However this branch has Born probability converging to 0, and can be seen as a deviant, atypical branch.

In fact, conflicts do emerge between **MWT** and Algorithmic Information Theory. In particular, the Independence Postulate (**IP**) is a finitary Church-Turing thesis, postulating that certain infinite and *finite* sequences cannot be found in nature, a.k.a. have high "addresses". If a forbidden sequence is found in nature, an information leak will occur. However **MWT** represents a theory in which such information leaks can occur. This blog entry covers the main arguments of this conflict.

8.1 Many Worlds Theory

Some researchers believe there is an inherent problem in quantum mechanics. On one hand, the dynamics of quantum states is prescribed by unitary evolution. This evolution is deterministic and linear. On the other hand, measurements result in the collapse of the wavefunction. This evolution is non-linear and nondeterministic. This conflict is called the measurement problem of quantum mechanics.

The time of the collapse is undefined and the criteria for the kind of collapse are strange. The Born rule assigns probabilities to macroscopic outcomes. The projection postulate assigns new microscopic states to the system measured, depending on the the macroscopic outcome. One could argue that the apparatus itself should be modeled in quantum mechanics. However it's dynamics is deterministic. Probabilities only enter the conventional theory with the measurement postulates.

MWT was proposed by Everett as a way to remove the measurement postulate from quantum mechanics. The theory consists of unitary evolutions of quantum states without measurement collapses. For **MWT**, the collapse of the wave function is the change in dynamical influence of one part of the wavefunction over another, the decoherence of one part from the other. The result is a branching structure of the wavefunction and a collapse only in the phenomenological sense.

8.1.1 Branching Worlds

An example of a branching of universes can be seen in an idealized Stern-Gerlarch experiment with a single electron with spin $|\phi_\uparrow\rangle$ and $|\phi_\downarrow\rangle$. This description can be found in [SBKW10]. There is a measuring apparatus $\mathcal{A}$, which is in an initial state of $|\psi^{\mathcal{A}}_{\text{ready}}\rangle$. After $\mathcal{A}$ reads spin-up or spin-down then it is in state $|\psi^{\mathcal{A}}_{\text{reads spin }\uparrow}\rangle$ or $|\psi^{\mathcal{A}}_{\text{reads spin }\downarrow}\rangle$, respectively. The evolution for when the electron is solely spin-up or spin-down is

$$
\begin{aligned}
|\phi_\uparrow\rangle \otimes |\psi^{\mathcal{A}}_{\text{ready}}\rangle &\xrightarrow{\text{unitary}} |\phi_{\text{absorbed}}\rangle \otimes |\psi^{\mathcal{A}}_{\text{reads spin }\uparrow}\rangle \\
|\phi_\downarrow\rangle \otimes |\psi^{\mathcal{A}}_{\text{ready}}\rangle &\xrightarrow{\text{unitary}} |\phi_{\text{absorbed}}\rangle \otimes |\psi^{\mathcal{A}}_{\text{reads spin }\downarrow}\rangle .
\end{aligned}
$$

Furthermore, one can model the entire quantum state of an observer $\mathcal{O}$ of the apparatus, with

$$
\begin{aligned}
&|\phi_\uparrow\rangle \otimes |\psi^{\mathcal{A}}_{\text{ready}}\rangle \otimes |\xi^{\mathcal{O}}_{\text{ready}}\rangle \\
\xrightarrow{\text{unitary}} &|\phi_{\text{absorbed}}\rangle \otimes |\psi^{\mathcal{A}}_{\text{reads spin }\uparrow}\rangle \otimes |\xi^{\mathcal{O}}_{\text{ready}}\rangle \\
\xrightarrow{\text{unitary}} &|\phi_{\text{absorbed}}\rangle \otimes |\psi^{\mathcal{A}}_{\text{reads spin }\uparrow}\rangle \otimes |\xi^{\mathcal{O}}_{\text{reads spin }\uparrow}\rangle
\end{aligned}
$$

$$
\begin{aligned}
&|\phi_\downarrow\rangle \otimes |\psi^{\mathcal{A}}_{\text{ready}}\rangle \otimes |\xi^{\mathcal{O}}_{\text{ready}}\rangle \\
\xrightarrow{\text{unitary}} &|\phi_{\text{absorbed}}\rangle \otimes |\psi^{\mathcal{A}}_{\text{reads spin }\downarrow}\rangle \otimes |\xi^{\mathcal{O}}_{\text{ready}}\rangle \\
\xrightarrow{\text{unitary}} &|\phi_{\text{absorbed}}\rangle \otimes |\psi^{\mathcal{A}}_{\text{reads spin }\downarrow}\rangle \otimes |\xi^{\mathcal{O}}_{\text{reads spin }\downarrow}\rangle .
\end{aligned}
$$

For the general case, the electron is in a state $|\phi\rangle = a\,|\phi_\uparrow\rangle + b\,|\phi_\downarrow\rangle$, where $|a|^2 + |b|^2 = 1$. In this case, the final superposition would be of the form:

$$
\begin{aligned}
&a\,|\phi_{\text{absorbed}}\rangle \otimes |\psi^{\mathcal{A}}_{\text{reads spin }\uparrow}\rangle \otimes |\xi^{\mathcal{O}}_{\text{reads spin }\uparrow}\rangle \\
&+b\,|\phi_{\text{absorbed}}\rangle \otimes |\psi^{\mathcal{A}}_{\text{reads spin }\downarrow}\rangle \otimes |\xi^{\mathcal{O}}_{\text{reads spin }\downarrow}\rangle .
\end{aligned}
$$

This is a superposition of two branches, each of which describes a perfectly reasonable physical story. This bifurcation is one method on how the quantum state of universe bifurcates into two branches.

8.1.2 Deriving the Born Rule

In my opinion, one of the main problems of **MWT** is its reconciliation of the Born rule, for which no proposed solution has universal consensus. In standard quantum mechanics, measurements are probabilistic operations. Measurements on a state vector $|\psi\rangle$, which is a unit vector over Hilbert space $\mathcal{H}$, are self-adjoint operators $\mathcal{O}$ on $\mathcal{H}$. Observables are real numbers that are the spectrum $\mathrm{Sp}(\mathcal{O})$ of $\mathcal{O}$. A measurement outcome is a subset $E \subseteq \mathrm{Sp}(\mathcal{O})$ with associated projector P_E on $\mathcal{H}$. Outcome E is observed on measurement of $\mathcal{O}$ on $|\psi\rangle$ with probability $P(E) = \langle\psi|\,P_E\,|\psi\rangle$. This is known as the Born rule. After this measurement, the new state becomes $P_E\,|\psi\rangle / \sqrt{\langle\psi|\,P_E\,|\psi\rangle}$. This is known as the projection postulate.

However, the Born rule and the projection postulate are not assumed by **MWT**. The dynamics are totally deterministic. Each branch is equally real to the observers in it. To address these issues, Everett first derived a typicality-measure that weights each branch of a state's superposition. Assuming a set of desirable constraints, Everett derived the typicality-measure to be equal to the norm-squared of the coefficients of each branch, i.e. the Born probability of each branch. Everett then drew a distinction between typical branches that have high typicality-measure and exotic atypical branches of decreasing typicality-measure. For the repeated measurements of the spin of an electron $|\phi\rangle = a\,|\phi_\uparrow\rangle + b\,|\phi_\downarrow\rangle$, the relative frequencies of up and down spin measurements in a typical branch converge to $|a|^2$ and $|b|^2$, respectively. The notion of typicality can be extended to measurements with many observables.

In a more recent resolution to the relation between **MWT** and probability, Deutsch introduced a decision theoretic interpretation [Deu99] that obtains the Born rule from the non-probabilistic axioms of quantum theory and non-probabilistic axioms of decision theory. Deutsch proved that rational actors are compelled to adopt the Born rule as the probability measure associated with their available actions. This approach is highly controversial, as some critics say the idea has circular logic.

Another attempt uses subjective probability [Vai98]. The experimenter puts on a blindfold before he finishes performing the experiment. After he finishes the experiment, he has uncertainty about what world he is in. This uncertainty is the foundation of a probability measure over the measurements. However, the actual form of the probability measure needs to be postulated:

Probability Postulate. *An observer should set his subjective probability of the outcome of a quantum experiment in proportion to the total measure of existence of all worlds with that outcome.*

Whichever explanation of the Born rule one adopts, the following section shows there is an issue with **MWT** and **IP**. There exist branches of substantial Born probability where information leaks occurs.

8.2 Violating the Independence Postulate

In [Lev84, Lev13], the Independence Postulate, **IP**, was introduced:

> *Let $\alpha \in \{0,1\}^{*\infty}$ be a sequence defined with an n-bit mathematical statement (e.g., in Peano Arithmetic or Set Theory), and a sequence $\beta \in \{0,1\}^{*\infty}$ can be located in the physical world with a k-bit instruction set (e.g., ip-address). Then $\mathbf{I}(\alpha:\beta) < k+n+c_{\mathrm{IP}}$, for some small absolute constant c_{IP}.*

The $\mathbf{I}$ term is an information measure in Algorithmic Information Theory. For this blog, the information term we use is $\mathbf{I}(x:y) = \mathbf{K}(x)+\mathbf{K}(y)-\mathbf{K}(x,y)$, where $\mathbf{K}$ is the prefix-free Kolmogorov complexity. We can use this definition of $\mathbf{I}$ because we only deal with finite sequences.

IP can be violated in the following idealized Stern-Gerlarch experiment measuring the spin $|\phi_\uparrow\rangle$ and $|\phi_\downarrow\rangle$ of N isolated electrons. We denote $|\phi_0\rangle$ for $|\phi_\uparrow\rangle$ and $|\phi_1\rangle$ for $|\phi_\downarrow\rangle$. The "address" (in the sense of **IP**) of this experiment is $< O(\log n)$. There is a measuring apparatus $\mathcal{A}$ with initial state of $|\psi^{\mathcal{A}}\rangle$, and after reading N spins of N electrons, it is in the state $|\psi^{\mathcal{A}}[x]\rangle$, where $x \in \{0,1\}^N$, whose ith bit is 1 iff the ith measurement returns $|\phi_1\rangle$. The experiment evolves according to the following unitary transformation:

$$\bigotimes_{i=1}^{N} |\phi\rangle \otimes |\psi^{\mathcal{A}}\rangle \otimes \stackrel{\text{unitary}}{\longrightarrow} \sum_{a_1,\ldots,a_N \in \{0,1\}^N} 2^{-N/2} \bigotimes_{i=1}^{N} |\phi_{a_i}\rangle \otimes |\psi^{\mathcal{A}}[a_1 a_2 \ldots a_n]\rangle \,.$$

If the bits returned are the first N bits of Chaitin's Omega, then a memory leak of size $n - O(\log n)$ has occurred. Thus

$$\text{Born-Probability(a memory leak of size } n - O(\log n) \text{ occurred)} \geq 2^{-n}.$$

8.3 Conclusion

There are multiple variations of **MWT** when it comes to consistency across universes. In one formulation, all universes conform to the same physical laws. In another model, each universe has its own laws, for example different values of gravity, etc. However, the experiment in the previous section shows that mathematics itself is different between universes, regardless of which model is used. In some universes, **IP** holds and there is no way to create information leaks. In other universes information leaks occur, and there are tasks where randomized algorithms fail but non-algorithmic physical methods succeeds. One such task is finding new axioms of mathematics. This was envisioned as a possibility by Gödel [G6̈1], but there is a universal consensus of the impossibility of this task. Not any more! In addition, because information leaks are finite events, the Born probability of worlds containing them is not insignificant. In such worlds, **IP** cannot be formulated, and the the foundations of Algorithmic Information Theory itself become detached from reality.

Formulated another way, let us suppose the Born probability is derived from the probability postulate. We have a "blindfolded mathematician" who performs the experiment described above. Before the mathematician takes off her blindfold, she states the Independence Postulate. By the probability postulate, with measure 2^{-n} over all worlds, there is a memory leak of size $n - O(\log n)$ and **IP** statement by the mathematician is in error.

As a rebuttal, one can, with non-zero probability, just flip a coin n times and get n bits of Chaitin's Omega. One difference between the classical world and **MWT** is the interpretation of probability. In the classical world, probability can be seen as a score of the likelihood the event. Thus if Bob threatens to flip n coins, Alice is $1 - 2^{-n}$ confident that the bits won't correspond to Chaitin's Omega. **IP** postulates away all such memory leaks that are created probabilistically. This is a very reasonable task, given large enough constant c, which results in an overwhelmingly large "confidence score" that a probabilistic memory leak doesn't occur.

In **MWT**, assuming the *probability postulate*, probability is a measure over the space of possible worlds. Thus when Bob now threatens to measure the spin of particles, Alice now knows 2^{-n} of the resultant worlds will contain bits of Chaitin's Omega, violating **IP**.

Chapter 9

Conclusion

9.1 Signals from Classical and Quantum Sources

Information non-growth laws say information about a target source cannot be increased with randomized processing. In classical information theory, we have

$$I(g(X):Y) \leq I(X:Y).$$

where g is a randomized function, X and Y are random variables, and I is the mutual information function. Thus processing a channel at its output will not increase its capacity. Information conservation carries over into the algorithmic domain, with the inequalities

$$\mathbf{I}(f(x):y) <^+ \mathbf{I}(x:y); \qquad \mathbf{I}(f(a);\mathcal{H}) <^+ \mathbf{I}(a;\mathcal{H}).$$

These inequalities ensure target information cannot be obtained by processing. If for example the second inequality was not true, then one can potentially obtain information about the halting sequence $\mathcal{H}$ with simple functions. Obtaining information about $\mathcal{H}$ violates the Independence Postulate, discussed in Chapter 8. Information non growth laws can be extended to signals [Eps23a] which can be modeled as probabilities over $\mathbb{N}$ or Euclidean space[1]. The "signal strength" of a probability p over $\mathbb{N}$ is measured by its self information.

$$\mathbf{I}_{\mathrm{Prob}}(p:p) = \log \sum_{i,j} 2^{\mathbf{I}(i:j)} p(i)p(j).$$

A signal, when undergoing randomized processing f (see Section 2.1.1), will lose its cohesion. Thus any signal going through a classical channel will become less coherent.

$$\mathbf{I}_{\mathrm{Prob}}(f(p):f(p)) <^+ \mathbf{I}_{\mathrm{Prob}}(p:p).$$

In Euclidean space, probabilities that undergo convolutions with probability kernels will lose self information. For example a signal spike at a random position will spread out when convoluted with the Gaussian function, and lose self information. The above inequalities deal with classical transformations. One can ask, is whether, quantum information processing can add new surprises to how information signals occur and evolve.

One can start with the prepare-and-measure channel, also known as a Holevo-form channel. Alice starts with a random variable X that can take values $\{1,\dots,n\}$ with corresponding probabilities $\{p_1,\dots,p_n\}$. Alice prepares a quantum state, corresponding to density matrix ρ_X, chosen

[1] In [Eps23a] probabilities over $\{0,1\}^\infty$ and T_0 second countable topologies were also studied.

from $\{\rho_1, \ldots, \rho_n\}$ according to X. Bob performs a measurement on the state ρ_X, getting a classical outcome, denoted by Y. Though it uses quantum mechanics, this is a classical channel $X \to Y$. So using the above inequality, cohesion will deteriorate regardless of $X's$ probability, with

$$\mathbf{I}_{\mathrm{Prob}}(Y:Y) <^+ \mathbf{I}_{\mathrm{Prob}}(X:X).$$

There remains a second option, constructing a signal directly from a mixed state. This involves constructing a mixed state, i.e. density matrix σ, and then performing a measurement E on the state, inducing the probability $E\sigma(k) = \mathrm{Tr}\sigma E_k$. However from [Eps23a], for elementary (even enumerable) probabilities $E\sigma$,

$$\mathbf{I}_{\mathrm{Prob}}(E\sigma : E\sigma) <^+ \mathbf{K}(\sigma, E).$$

Thus for simply defined density matrices and measurements, no signal will appear. So experiments that are simple will result in simple measurements, or white noise. However it could be that a larger number of uncomputable pure or mixed states produce coherent signals. Theorems 32 and 33 say otherwise, in that the POVM measurement E of a vast majority of pure and mixed states will have negligible self-information. Thus for uniform distributions Λ and μ over pure and mixed states (see Section 5.2.2),

$$\int 2^{\mathbf{I}_{\mathrm{Prob}}(E|\psi\rangle : E|\psi\rangle)} d\Lambda = O(1); \qquad \int 2^{\mathbf{I}_{\mathrm{Prob}}(E\sigma : E\sigma)} d\mu(\sigma) = O(1).$$

This can be seen as a consequence of the vastness of Hilbert spaces as opposed to the limited discriminatory power of quantum measurements. In addition, there could be non-uniform distributions of pure or mixed states that could be of research interest. In quantum decoherence, a quantum state becomes entangled with the environment, losing decoherence. The off diagonal elements of the mixed state become dampened, as the state becomes more like a classical mixture of states. Let p_σ be the idealized classical probability that σ decoheres to, with $p_\sigma(i) = \sigma_{ii}$. Corollary 8 states that for an overwhelming majority of pure or mixed states σ, p_σ is noise, that is, has negligible self-information.

$$\int 2^{\mathbf{I}_{\mathrm{Prob}}\left(p_{|\psi\rangle} : p_{|\psi\rangle}\right)} d\Lambda = O(1); \qquad \int 2^{\mathbf{I}_{\mathrm{Prob}}(p_\sigma : p_\sigma)} d\mu(\sigma) = O(1).$$

This is to be expected, with one supporting fact being for an n qubit space, $i \in \{1, \ldots, 2^n\}$, $\mathbf{E}_\Lambda[p_{|\psi\rangle}(i)] = 2^{-n}$. With Algorithmic Information Theory, we've taken this fact one step further, showing that $p_{|\psi\rangle}$ has no (in the exponential) self-algorithmic information and cannot be processed by deterministic or randomized means to produce a more coherent signal. In addition, it appears a more direct proof of the first decoherence inequality could be possible.

However the measurement process has a surprising consquence, in that the wave function collapse causes an massive uptake in algorithmic signal strength. Let F be a PVM, of size 2^{n-c}, of an n qubit space and let Λ_F be the distribution of pure states when F is measured over the uniform distribution Λ. Thus Λ_F represents the F-collapsed states from Λ. Theorem 34 states

$$n - 2c <^+ \log \int 2^{\mathbf{I}_{\mathrm{Prob}}(F|\psi\rangle : F|\psi\rangle)} d\Lambda_F.$$

9.2 Apriori Distribution

To avoid the pitfall of a signalless distribution that only produces white noise, we can conjecture a new apriori distribution for quantum states that is not signalless. Note that we are dealing with measures over the density operator space and not directly with density operators because we are measuring properties, such as self-information, over all possible (pure or mixed) states. Properties of this apriori distribution can be discerned by working backwards. Indeed, suppose there are a set of (possibly infinite) systems $\{|\psi_i\rangle\}$, where for each system $|\psi_i\rangle$, a measurement occurs, producing a discernable signal. By Theorem 31, this implies the states $|\psi_i\rangle$ have high $\mathbf{I}(|\psi_i\rangle : |\psi_i\rangle)$, where $\mathbf{I}$ is the information function between mixed states introduced in Definition 6. Thus any universal quantum apriori distribution over these systems must be weighted toward states with high self information. One candidate is an probability measure ξ over pure states where

$$\xi(|\psi\rangle) \propto 2^{\mathbf{I}(|\psi\rangle : |\psi\rangle)}.$$

However this area of research is still ongoing. Another clue to this universal quantum apriori distribution is the measurement operation, which as shown above, causes an uptake in signal strength. Take a PVM measurement F, which procures a value i from a state $|\psi\rangle$, projecting to a new state $|\psi'\rangle$. $P\,|\psi'\rangle\,(i) = 1$. By Corollary 7, this new state $|\psi'\rangle$ has self information

$$\mathbf{K}(i) <^{\log} \mathbf{I}(|\psi'\rangle : |\psi'\rangle).$$

The error term is on the order of $\mathbf{K}(P)$. Most of the measurement values i of P will be random, i.e. have large $\mathbf{K}(i)$ (just look at the Kolmorogov complexity of the first 2^n numbers!). Thus simple quantum measurements increase the self information of most measured quantum states. So this fact, and Theorem 26, leads us to the following conclusion.

> *Take a distribution over density operators, such as Λ, where an overwhelming majority of states have negligible self-information. When each such state in its support is measured with a simple apparatus, the result is new a distribution where most of the states have substantial self-information.*

However, the situation is reversed for quantum channels. A quantum state that is transformed by a quantum operation will not increase in self-information. So by Theorem 28, we get the following claim, where equality occurs if the quantum operation is a unitary transform.

> *Given any distribution over density operators, if all the density matrices its support are transformed by a simple quantum operation, then the resultant distribution will give more measure to mixed states with less self-information.*

Thus simple measurements with many operators can only increase self-information, simple quantum operations can only decrease self-information, and simple unitary transforms leave the self-information unaltered. If the operation is complex, then nothing so far has been proven.

9.3 Measurement Before Information Cloning

The no-cloning theorem states that every unitary transform cannot clone an arbitrary quantum state. However there is the possibility of copying information from a subset of states. By "copying information", we mean that two measurements of two states will produce two values that are similar. More formally, the information cloned from a state $|\psi\rangle$ relative to unitary transform U, and POVMs E and F is,

$$\mathbf{I}_{\text{Clone}}(|\psi\rangle) = \mathbf{I}_{\text{Prob}}(E\,|1\rangle : F\,|2\rangle), \text{ where } U\,|\psi\rangle\,|0\rangle = |1\rangle\,|2\rangle\,.$$

This represents the shared signal strength between $|1\rangle$ and $|2\rangle$ when the states[2] were created from a unitary transform U of $|\psi\rangle$ tensored with an ancillia state $|0\rangle$. Note that by Theorems 29 and 31,

$$\mathbf{I}_{\text{Clone}}(|\psi\rangle) <^{\log} \mathbf{I}(|\psi\rangle : |\psi\rangle)\,.$$

The question is, given a distribution Γ over density operators with low expected $\mathbf{I}_{\text{clone}}$, what sort of transform is required to increase this expectation. In this section, we discuss necessary conditions of this transform. We require the following two assumptions.

1. Γ has low expected self-information $\mathbf{I}$.
2. All the transforms and operators have low algorithmic complexity.

Assumption (1) is not restrictive as there still is a large set of natural distributions that have this property. For example any distribution Ω that is less than $2^c\Lambda$ will have $\log\int 2^{\mathbf{I}(|\psi\rangle:|\psi\rangle)}d\Omega <^+ c$. Assumption (2) still allows for a large class of systems. One such example is systems that scale in size, uniformly computable in the number of qubits, n. Thus the initial distribution $\Lambda^{(n)}$, measurements $E^{(n)}$, $F^{(n)}$, and unitary transform $U^{(n)}$ (and any transforms that are applied) are all computable relative to n. Under these conditions, the error terms of the relevant theorems are $O(\log n)$.

How do you create a distribution with high expected $\mathbf{I}_{\text{Clone}}$, where most states can have cloneable information? Any transform that increases cloneable information must increase self-information. However Theorem 28, along with assumption (2) bars quantum operations as a means to create self-information. The only way to potentially increase self-information is to perform a measurement, which as Theorems 31 and 34 show, oftentimes cause an uptake in self-information. Thus we get the following claim.

> *When algorithmically complicated processes are ruled out, measurements are required to produce distributions over quantum states that have cloneable information.*

For example, take the starting distribution to be the uniform measure over pure states, Λ. Let $E = F = \{|i\rangle\langle i|\}$ be POVM measurements over projectors to the basis states and let U be any unitary transform such that $U|i\rangle|0\rangle = |i\rangle|i\rangle$ for $i \in \{1,\ldots,2^n\}$. By Theorems 26 and 29, we have that

$$\int 2^{\mathbf{I}_{\text{Clone}}(|\psi\rangle)}d\Lambda = O(1).$$

Now suppose we apply the measurement $G = E$ to Λ, producing a new distribution Λ_G concentrated evenly among the basis states, where $\Lambda_G(|i\rangle) = 2^{-n}$. Thus we have that $\mathbf{I}_{\text{Clone}}(|i\rangle) = \mathbf{I}_{\text{Prob}}(E|i\rangle : F|i\rangle) = \mathbf{K}(i)$. Since there are $2^{n-O(1)}$ basis states $|i\rangle$ where $n <^+ \mathbf{K}(i)$, we have the following uptake in cloneable information.

$$n <^+ \log\int 2^{\mathbf{I}_{\text{Clone}}(|\psi\rangle)}d\Lambda_G.$$

Other such applications can be seen as generalizations from this extreme example.

[2]Note the thi definition can be generalized to arbitrary (i.e. entangled) states, with $\mathbf{I}_{\text{Prob}}(E\text{Tr}_2|\phi\rangle : F\text{Tr}_1|\phi\rangle)$, where $|\phi\rangle = U|\psi\rangle|0\rangle$.

Appendix A

Information Between Basis States

Special information inequalities and conservation inequalities can be achieved for orthogonal sequences of pure basis states $|1\rangle, |2\rangle, |3\rangle, \ldots$ We use Theorem 5 that $\mathbf{K}(i|n) =^+ \mathbf{Hg}(|i\rangle)$. Let $S(\rho)$ be the von Neumann entropy of σ.

Theorem 45 *Relativized to an orthogonal sequence of elementary states $|1\rangle, |2\rangle, |3\rangle, \ldots$, enumerated by strings $i, j, k \in \{0,1\}^n$.*

1. $\mathbf{d}(|i\rangle \,|\, |j\rangle) = \infty$ *for* $i \neq j$.
2. $\mathbf{Hg}(\sigma \otimes |i\rangle\langle i|) =^+ \mathbf{Hg}(|i\rangle) + \mathbf{Hg}(\sigma| \, |i\rangle, \mathbf{Hg}(|i\rangle))$.
3. $\mathbf{I}(|k\rangle : |i\rangle) <^+ \mathbf{I}(|j\rangle : |i\rangle) + \mathbf{K}(k|j, N)$.
4. $\mathbf{I}(i : j|N) <^{\log} \mathbf{I}(|i\rangle : |j\rangle)$.
5. $\mathbf{I}(|i\rangle : |j\rangle) <^+ \mathbf{I}(i : j|N) + \mathbf{I}(i, j : \mathcal{H}|N)$.
6. $\mathbf{K}(i|N) <^+ \mathbf{I}(|i\rangle : |i\rangle) <^+ \mathbf{K}(i|N) + \mathbf{I}(i : \mathcal{H}|N)$.
7. $\mathbf{I}(|i\rangle : |i\rangle) <^+ 4n|3$.
8. $\mathbf{I}(|i\rangle : |j\rangle) <^+ \mathbf{I}(|i\rangle : |i\rangle) + \mathbf{I}(i, j : \mathcal{H}|n)$.

Proof.

(1) This is due to the fact that $\mathfrak{T}_{|i\rangle\langle i|} \overset{*}{>} \sum_n \mathbf{m}(n)n\,|j\rangle\langle j|$. Thus $\log \mathrm{Tr}\mathfrak{T}_{|i\rangle\langle i|}|j\rangle\langle j| = \infty$.

(2) We use the lower semicomputable matrix $\rho = \boldsymbol{\mu}_{(|i\rangle, \mathbf{Hg}(|i\rangle))} \otimes |i\rangle\langle i|$. so we have that $\boldsymbol{\mu}_{2n} \overset{*}{>} \mathbf{m}(|i\rangle, \mathbf{Hg}(|i\rangle)|2^{2n})\rho \overset{*}{=} \mathbf{m}(i, \mathbf{K}(i|n)|2^{2n})\rho \overset{*}{=} \mathbf{m}(i|2^n)\rho \overset{*}{=} 2^{-\mathbf{Hg}(|i\rangle)}\rho$. So $\mathbf{Hg}(\sigma \otimes |i\rangle\langle i|) <^+ \mathbf{Hg}(|i\rangle) - \log \mathrm{Tr}\rho(\sigma \otimes |i\rangle\langle i|) <^+ \mathbf{Hg}(|i\rangle) + \mathbf{Hg}(\sigma| \, |i\rangle, \mathbf{Hg}(|i\rangle))$. The other direction is given by Theorem 9.

(3) We let $T \in \mathcal{T}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}$ vary over $\boldsymbol{\mu} \otimes \boldsymbol{\mu}$ tests. Since $\boldsymbol{\mu} \overset{*}{>} \sum_i |i\rangle \mathbf{m}(i|N)\langle i|$ it must be that $1 > \mathrm{Tr}T\boldsymbol{\mu} \otimes \boldsymbol{\mu} \overset{*}{>} \mathrm{Tr}T\sum_{i,j} \mathbf{m}(i|N)\mathbf{m}(j|N)\,|i\rangle|j\rangle\langle i|\langle j| \overset{*}{>} \sum_{i,j} \mathbf{m}(i|N)\mathbf{m}(j|N)\langle i|\langle j|T|i\rangle|j\rangle$.

From each T, let $T' = \sum_{i,j}\left(\sum_k \mathbf{m}(k|j,N)\langle i|\langle k| T|i\rangle|k\rangle\right)|i\rangle|j\rangle\langle i|\langle j|$. Since $\mathbf{m}$ and T are lower computable, then so is T'. In addition, $O(1)T' \in \mathcal{T}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}$, because

$$\begin{aligned}
\mathrm{Tr}T'\boldsymbol{\mu}\otimes\boldsymbol{\mu} &= \sum_{i,j}\langle i|\langle j|\boldsymbol{\mu}\otimes\boldsymbol{\mu}|i\rangle|j\rangle\sum_k \mathbf{m}(k|j,N)(\langle i|\langle k|T|i\rangle|k\rangle)\\
&\overset{*}{=}\sum_{i,j}\mathbf{m}(i|N)\mathbf{m}(j|N)\sum_k \mathbf{m}(k|j,N)(\langle i|\langle k|T|i\rangle|k\rangle)\\
&\overset{*}{<}\sum_{i,j}\mathbf{m}(i|N)\sum_k \mathbf{m}(j,k|N)(\langle i|\langle k|T|i\rangle|k\rangle)\\
&\overset{*}{=}\sum_{i,j}\mathbf{m}(i|N)\sum_k \mathbf{m}(k|N)\mathbf{m}(j|k,\mathbf{K}(k),N)(\langle i|\langle k|T|i\rangle|k\rangle)\\
&\overset{*}{=}\sum_{i,k}\mathbf{m}(i|N)\mathbf{m}(k|N)\sum_j \mathbf{m}(j|k,\mathbf{K}(k),N)\langle i|\langle k|T|i\rangle|k\rangle\\
&\overset{*}{<}\sum_{i,k}\mathbf{m}(i|N)\mathbf{m}(k|N)(\langle i|\langle k|T|i\rangle|k\rangle)\\
&\overset{*}{<}\mathrm{Tr}T\boldsymbol{\mu}\otimes\boldsymbol{\mu}\\
&< O(1).
\end{aligned}$$

So

$$\begin{aligned}
\mathbf{I}(|i\rangle:|j\rangle) &> \log\mathrm{Tr}\sum_T \mathbf{m}(O(1)T'|N^2)O(1)T'|i\rangle|j\rangle\langle i|\langle j|\\
&>^+ \log\mathrm{Tr}\sum_T \mathbf{m}(T|N^2)T'|i\rangle|j\rangle\langle i|\langle j|\\
&>^+ \log\sum_T \mathbf{m}(T|N^2)\sum_k \mathbf{m}(k|j,N)\langle k|\langle i|T|k\rangle|i\rangle\\
&>^+ \log\sum_T \mathbf{m}(T|N^2)\mathbf{m}(k|j,N)\langle k|\langle i|T|k\rangle|i\rangle\\
&=^+ \mathbf{I}(|k\rangle:|i\rangle)-\mathbf{K}(k|j).
\end{aligned}$$

(4) This follow as a special case of Theorem 31.

(5) Let $s(i,j)=\mathbf{m}(i|N)\mathbf{m}(j|N)2^{\mathbf{I}(|i\rangle:|j\rangle)}$. The function s is lower semicomputable relative to $\mathcal{H}$ because $\mathbf{m}$ and $\mathfrak{T}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}$ are lower computable relative to $\mathcal{H}$. Furthermore we have that

$$\begin{aligned}
\sum_{i,j}s(i,j) &= \sum_{i,j}\mathbf{m}(i|N)\mathbf{m}(j|N)\mathrm{Tr}\mathfrak{T}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}|i\rangle\langle i|\otimes|j\rangle\langle j|\\
&=\mathrm{Tr}\mathfrak{T}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}\sum_{i,j}\mathbf{m}(i|N))|i\rangle\langle i|\otimes\mathbf{m}(j|N)|j\rangle\langle j|\\
&<O(1)\mathrm{Tr}\mathfrak{T}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}\boldsymbol{\mu}\otimes\boldsymbol{\mu}<O(1).
\end{aligned}$$

Therefore $s(i,j)\overset{*}{<}\mathbf{m}(i,j|N,\mathcal{H})$ and so $\mathbf{I}(|i\rangle:|j\rangle) <^+ \log\mathbf{m}(i,j|N,\mathcal{H})|(\mathbf{m}(i|N)\mathbf{m}(j|N)) =^+ \mathbf{I}(i:j|N)+\mathbf{I}(i,j:\mathcal{H}|N)$.

(6) For $\mathbf{K}(i|N) <^+ \mathbf{I}(|i\rangle:|i\rangle)$, we prove the stronger statement: for elementary ρ, $2\mathbf{Hg}(\rho)-\mathbf{K}(\rho,\mathbf{Hg}(\rho))-2S(\rho) <^+ \mathbf{I}(\rho:\rho)$. Let $\nu = 2^{2\mathbf{Hg}(\rho)-2}(\rho\otimes\rho)$. The matrix $\nu\in\mathcal{T}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}$ because

$\mathrm{Tr}(\boldsymbol{\mu}\otimes\boldsymbol{\mu})\nu \leq 1$. Therefore

$$\begin{aligned}\mathbf{I}(\rho:\rho) =& \log \mathrm{Tr}\mathfrak{T}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}(\rho\otimes\rho)\\ &\geq \log \mathbf{m}(\nu)\mathrm{Tr}\nu(\rho\otimes\rho)\\ &>^+ \log \mathbf{m}(\nu)2^{2\mathbf{Hg}(\rho)}\mathrm{Tr}(\rho\rho\otimes\rho\rho)\\ &>^+ 2\mathbf{Hg}(\rho) - \mathbf{K}(\rho,\mathbf{Hg}(\rho)) + 2\log\sum_i \lambda_i^2,\end{aligned}$$

where λ_i are the eigenvalues of ρ. Due to concavity $-2S(\rho) \leq 2\log\sum_i \lambda_i^2$. So $\mathbf{I}(\rho:\rho) >^+ 2\mathbf{Hg}(\rho) - \mathbf{K}(\rho,\mathbf{Hg}(\rho)) - 2S(\rho)$. The inequality follows from $\mathbf{Hg}(|i\rangle) =^+ \mathbf{K}(i)$ and $S(|i\rangle\langle i|) = 0$. For $\mathbf{I}(|i\rangle : |i\rangle <^+ \mathbf{K}(i) + \mathbf{I}(i:\mathcal{H}|n)$, we note that it is a special case of (5).

(7) If $T \in \mathcal{T}_{\boldsymbol{\mu}_n\otimes\boldsymbol{\mu}_n}$, then $\mathrm{Tr}T \overset{*}{<} 2^{2n}$. This is because $1 \geq \mathrm{Tr}T(\boldsymbol{\mu}_n\otimes\boldsymbol{\mu}_n) \overset{*}{>} \mathrm{Tr}T(2^{-n}I\otimes 2^{-n}I)$. Since the set of lower computable matrices of trace not more than 2^{2n} is enumerable, $2^{-2n}\mathfrak{T}_{\boldsymbol{\mu}_n\otimes\boldsymbol{\mu}_n} \overset{*}{<} \boldsymbol{\mu}_{2n}$. Assume $\mathbf{I}(|i\rangle : |i\rangle) = \log \mathrm{Tr}\mathfrak{T}_{\boldsymbol{\mu}_n\otimes\boldsymbol{\mu}_n} |ii\rangle\langle ii| = 2n - c$. Then $-\log \mathrm{Tr}u_{2n}|ii\rangle\langle ii| <^+ c$. This means that $\mathbf{K}(ii|2^{2n}) <^+ c$. So $\mathbf{K}(i|2^n) <^+ c$. So $1 \geq \mathrm{Tr}\mathfrak{T}_{\boldsymbol{\mu}_n\otimes\boldsymbol{\mu}_n}\boldsymbol{\mu}\otimes\boldsymbol{\mu} \overset{*}{>} \mathbf{m}(i|2^n)^2\mathrm{Tr}\mathfrak{T}_{\boldsymbol{\mu}_n\otimes\boldsymbol{\mu}_n}|ii\rangle\langle ii| \overset{*}{>} 2^{2n-3c}$. Thus $c >^+ 2n/3$. This implies that $\mathbf{I}(|i\rangle : |i\rangle) = 2n - c <^+ 4n/3$.

(8) This follows from (5) and (6).

Corollary 11 *For elementary ρ,* $2\mathbf{Hg}(\rho) - \mathbf{K}(\rho,\mathbf{Hg}(\rho)) - 2S(\rho) <^+ \mathbf{I}(\rho:\rho)$.

Bibliography

[BDK+05] I Bjelaković, J Deuschel, T Krüger, R Seiler, R Siegmund-Schultze, and A Szkola. A quantum version of sanov's theorem. *Communications in Mathematical Physics*, 260(3):659–671, 2005.

[BH96] V. Buzek and M. Hillery. Quantum copying: Beyond the no-cloning theorem. *Phys. Rev. A*, 54:1844–1852, 1996.

[Bho21] T. Bhojraj. Algorithmic randomness and kolmogorov complexity for qubits. *CoRR*, abs/2106.14280, 2021.

[BKM+06] F. Benatti, T. Krüger, M. Müller, R. Siegmund-Schultze, and A. Szko—a. Entropy and Quantum Kolmogorov Complexity: A Quantum Brudno's Theorem. *Communications in Mathematical Physics*, 265(2), 2006.

[BOD14] F. Benatti, S. K. Oskouei, and A. S. Deh Abad. Gacs Quantum Algorithmic Entropy in Infinite Dimensional Hilbert Spaces. *Journal of Mathematical Physics*, 55(8), 2014.

[Bru78] A. Brudno. The Complexity of the trajectories of a Dynamical System. *Russian Mathematical Surveys*, 33(1):197, 1978.

[BvL01] A. Berthiaume, W. van Dam, and S. Laplante. Quantum Kolmogorov Complexity. *Journal of Computer and System Sciences*, 63(2), 2001.

[Deu99] David Deutsch. Quantum theory of probability and decisions. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 455(1988), 1999.

[Eps19a] S. Epstein. Algorithmic no-cloning theorem. *IEEE Transactions on Information Theory*, 65(9):5925–5930, 2019.

[Eps19b] S. Epstein. Algorithmic no-cloning theorem. *IEEE Transactions on Information Theory*, 65(9), 2019.

[Eps19c] S. Epstein. On the algorithmic probability of sets. *CoRR*, abs/1907.04776, 2019.

[Eps19d] S. Epstein. On the algorithmic probability of sets. *CoRR*, abs/1907.04776, 2019.

[Eps20] Samuel Epstein. An extended coding theorem with application to quantum complexities. *Information and Computation*, 275, 2020.

[Eps21a] S. Epstein. On the algorithmic content of quantum measurements. *CoRR*, abs/2102.03905, 2021.

[Eps21b] Samuel Epstein. All sampling methods produce outliers. *IEEE Transactions on Information Theory*, 67(11):7568–7578, 2021.

[Eps22] S. Epstein. The outlier theorem revisited. *CoRR*, abs/2203.08733, 2022.

[Eps23a] S. Epstein. On the algorithmic information between probabilities. *CoRR*, abs/2303.07296, 2023.

[Eps23b] S. Epstein. A quantum el theorem. *CoRR*, abs/2301.08348, 2023.

[Eps23c] S. Epstein. A quantum outlier theorem. *CoRR*, abs/2303.06256, 2023.

[Eve57] Hugh Everett. "relative state" formulation of quantum mechanics. *Rev. Mod. Phys.*, 29, 1957.

[G6̈1] Kurt Gödel. The modern development of the foundations of mathematics in the light of philosophy. *In: Kurt Gödel. Collected Works. Volume III. Oxford University Press.*, 1961.

[G0́1] P. Gács. Quantum Algorithmic Entropy. *Journal of Physics A Mathematical General*, 34(35), 2001.

[G2́1] Peter Gács. Lecture notes on descriptional complexity and randomness. *CoRR*, abs/2105.04704, 2021.

[Lev74] L. A. Levin. Laws of Information Conservation (Non-growth) and Aspects of the Foundations of Probability Theory. *Problemy Peredachi Informatsii*, 10(3):206–210, 1974.

[Lev84] L. A. Levin. Randomness conservation inequalities; information and independence in mathematical theories. *Information and Control*, 61(1):15–37, 1984.

[Lev13] L. A. Levin. Forbidden information. *J. ACM*, 60(2), 2013.

[Lev16] L. A. Levin. Occam bound on lowest complexity of elements. *Annals of Pure and Applied Logic*, 167(10):897–900, 2016. And also: S. Epstein and L.A. Levin, Sets have simple members, arXiv preprint arXiv:1107.1458, 2011.

[Mul08] M. Muller. Strongly Universal Quantum Turing Machines and Invariance of Kolmogorov Complexity. *IEEE Transactions on Information Theory*, 54(2), 2008.

[NS19] A. Nies and V. Scholz. Martin-Löf Random Quantum States. *Journal of Mathematical Physics*, 60(9), 2019.

[SBKW10] S. Saunders, J. Barrett, A. Kent, and D. Wallace. *Many Worlds?: Everett, Quantum Theory, & Reality*. OUP Oxford, 2010.

[Sch71] C. Schnorr. A unified approach to the definition of random sequences. *Mathematical systems theory*, 5(3), 1971.

[Sch95] B. Schumacher. Quantum coding. *Phys. Rev. A*, 51, 1995.

[Vai98] L Vaidman. On schizophrenic experiences of the neutron or why we should believe in the many-worlds interpretation of quantum theory. *International Studies in the Philosophy of Science*, 12(3):245–261, 1998.

[Vit01] P.M.B. Vitanyi. Quantum kolmogorov complexity based on classical descriptions. *IEEE Transactions on Information Theory*, 47(6):2464–2479, 2001.

[WZ82] W. Wootters and W. Zurek. A single quantum cannot be cloned. *Nature*, 299(5886):802–803, 1982.