ON THE NATURE AND DECAY OF QUANTUM RELATIVE ENTROPY

BY

NICHOLAS LARACUENTE

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Physics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2020

Urbana, Illinois

Doctoral Committee:

Professor Paul Kwiat, Chair
Professor Marius Junge, Director of Research
Professor Anthony Leggett
Associate Professor Eric Chitambar

Abstract

Historically at the core of thermodynamics and information theory, entropy's use in quantum information extends to diverse topics including high-energy physics and operator algebras. Entropy can gauge the extent to which a quantum system departs from classicality, including by measuring entanglement and coherence, and in the form of entropic uncertainty relations between incompatible measurements. The theme of this dissertation is the quantum nature of entropy, and how exposure to a noisy environment limits and degrades non-classical features.

An especially useful and general form of entropy is the quantum relative entropy, of which special cases include the von Neumann and Shannon entropies, coherent and mutual information, and a broad range of resource-theoretic measures. We use mathematical results on relative entropy to connect and unify features that distinguish quantum from classical information. We present generalizations of the strong subadditivity inequality and uncertainty-like entropy inequalities to subalgebras of operators on quantum systems for which usual independence assumptions fail. We construct new measures of non-classicality that simultaneously quantify entanglement and uncertainty, leading to a new resource theory of operations under which these forms of non-classicalty become interchangeable. Physically, our results deepen our understanding of how quantum entanglement relates to quantum uncertainty.

We show how properties of entanglement limit the advantages of quantum superadditivity for information transmission through channels with high but detectable loss. Our method, based on the monogamy and faithfulness of the squashed entanglement, suggests a broader paradigm for bounding non-classical effects in lossy processes. We also propose an experiment to demonstrate superadditivity.

Finally, we estimate decay rates in the form of modified logarithmic Sobolev inequalities for a variety of quantum channels, and in many cases we obtain the stronger, tensor-stable form known as a complete logarithmic Sobolev inequality. We compare these with our earlier results that bound relative entropy of the outputs of a particular class of quantum channels.

Acknowledgments

First, I thank the members of my thesis committee. Marius Junge, my research advisor, introduced me to the mathematics of quantum information, and to many of its other practitioners. Paul Kwiat was my original advisor at UIUC, and I have continued to seek his input and assistance in maintaining connection to the experimental community. Eric Chitambar has also given much guidance and support, and with him I have been discussing further joint research projects. Anthony J. Leggett helped me keep sight of fundamental physics between the extremes of pure mathematics and technological applications.

As well as my thesis committee, I acknowledge Mark Wilde for his advice and assistance. I thank my peer and most frequent co-author Li Gao, from whom I learned much of the math I had missed as a physics student. I thank the other co-authors who contributed to projects mentioned in this thesis.

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DMS-1144245. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation. I was partially supported by National Science Foundation grants DMS-1700168, DMS-1501103, and DMS-1800872. I was also supported by the Graduate College Distinguished Fellowship from the University of Illinois at Urbana-Champaign, by the Summer Predoctoral Institute in summer 2013, and as a teaching assistant for the department of physics.

Many of the important ideas in this thesis, especially in Chapters 3, 4, and 5, are rooted in discussions at the Institut Henri Poincaré in Paris. These include a conversation with Aurelian Gheondea about the differences in quantifying asymmetry between the school that followed Noether's theorem and that following the resource theory. They include conversations between Marius Junge and Reinhard F. Werner about entanglement. I believe this is also where Li Gao and Marius Junge first met Ivan Bardet, who inspired many of the studies in Chapter 7. I thank John Stack for inspiring conversations early in the program.

I would also like to specifically acknowledge the *Boulder School for Condensed Matter and Materials Physics* in the Summer of 2018, and the preceding *Rocky Mountain Summit on Quantum Information.* Conversations with Felix Leditzky and Graeme Smith in Boulder began efforts to experimentally demonstrate superadditivity of quantum rates as described in section 6.4. While having an experimentalist and a mathematician as research advisors had many benefits in terms of linking ideas from these extremes, it was nonetheless informative and transformative to receive instruction from the many quantum information theorists who lectured in Boulder that summer. Boulder is also where I met Kiel Howe, who suggested subsequent contact with Fermilab and that I attend *Next Steps in Quantum Science for High Energy Physics*.

In other important visits, I mention both the 2018 *Spring School in Quantum Computing* and the weeks I spent at Caltech's IQIM during Autumn of 2018, for both of which I can thank Thomas Vidick. I also acknowledge the 2019 *Thematic Research Program: Operator Algebras, Groups and Applications to Quantum Information* at ICMAT (Institute of Mathematical Sciences) in Madrid, the 2018 *Princeton Summer School on Condensed Matter Physics*, and UIUC's *Institute for Condensed Matter Theory Summer School 2016: Introduction to Topological Phases of Matter*. I thank Thomas Sinclair and Roy Arazia at Purdue for inviting me to *Quantitative Linear Algebra Meets Quantum Information Theory* at Purdue in 2019. I gained some particularly relevant insights at the workshop on *Algebraic and Statistical ways into Quantum Resource Theories* at Banff International Research Station. Finally, I should mention an enlightening visit with Nima Lashkari at Purdue university in the winter of 2019, where we discussed in detail the opportunities and challenges in new techniques based on operator algebras for high-energy physics.

Daniel T. Jensen was my mentor within UIUC's Leadership Center and helped me earn a *Leadership Certificate* along with the physics PhD. I also would like to thank May Berenbaum, Eric Larson, Daniel Miller, James O'Dwyer, and Andy Suarez, who showed me a different range of fascinating questions in mathematical biology, though ultimately I chose to focus on quantum information. Our director of graduate studies, Lance Cooper, helped me navigate the logistics of the program, which ultimately had me complete a master's in mathematics concurrently with this PhD. I thank Morten Lundsgaard for his advice on teaching.

For LaTex formatting, this dissertation uses the UIUCTHESIS2009 Package/Class by Charles Kiyanda, edited from the the UIUCTHESIS07 Package/Class by Tim Head based on the Peter Czoschke version, based on the original version by David Hull.

I would like to thank the many friends I've made on this journey, across academic divisions and in the Champaign-Urbana community. I thank my mother and father, possibly most of all for supporting me in the years preceding graduate school, in which I wandered through the technology startup world and ultimately found my way back to theoretical science. I thank Katie Bolan, who transformed the later years of graduate school from a task I needed to complete to a time I will cherish.

Table of Contents

List of Acronyms/Abbreviations

ASA	Adjusted Subadditivity
CHSH	Clauser, Horne, Shimony, Holt (usually referring to the entanglement test or quantum game)
CLSI	Complete (modified) Logarithmic Sobolev Inequality
CMI	Conditional Mutual Information
CNOT	The quantum “conditional not” gate
EPR	Einstein-Podolsky-Rosen (sometimes referring to a maximally entangled pair of qubits)
GNS	Gelfand-Naimark-Segal, usually referring to the GNS construction
GCMI	Generalized Conditional Mutual Information
IBM QX	The “Quantum Experience,” IBM’s publicly accessible test quantum computer
LO	Local Operations
QI	Quantum Information
MI	Mutual Information
MLSI	Modified Logarithmic Sobolev Inequality
UCR	Uncertainty Relation (sometimes referring to a single qubit split by bases)
SASA	Strong Adjusted Subadditivity
SSA	Strong Subadditivity
TRO	Ternary Ring of Operators
UIUC	University of Illinois at Urbana-Champaign
vNa	von Neumann algebra

List of Commonly Used Symbols

Basic Mathematics

$\mathbb{N}, \mathbb{Z}, \mathbb{R}^+, \mathbb{R}, \mathbb{C}$ - natural numbers, integers, non-negative real numbers, real numbers, and complex numbers

$\log_\alpha, \log$ - the logarithm with base α, and the natural logarithm, respectively

$\exp(\cdot)$ - the natural exponential

$\equiv$ - equivalence presented as a definition

$\cong$ - isometry for normed spaces, and isomorphism for algebraic sets

Vector & Matrix Spaces

$\mathcal{H}$ - a Hilbert space

$S^\alpha(\mathcal{H})$ - the α-normed operators on Hilbert space $\mathcal{H}$

$\mathbb{B}(\mathcal{H})$ - the (∞-normed) bounded operators on Hilbert space $\mathcal{H}$

$S(\mathcal{H})$ - the densities defined on the Hilbert space $\mathcal{H}$, given by 1-normed functionals or GNS construction

$U(\mathcal{H})$ - the unitaries on $\mathcal{H}$

$\mathbb{M}_d$ - the space of d-dimensional matrices

$l^d_\alpha(\mathbb{R}), l^d_\alpha(\mathbb{C})$ - the α-normed d-dimensional Banach space over real or complex numbers

$S^d_\alpha(\mathbb{R}), S^d_\alpha(\mathbb{C})$ - the α-normed d-dimensional spaces of matrices, equivalent to $S_\alpha(l^d_2(\mathbb{R}))$ and $S_\alpha(l^d_2(\mathbb{C}))$

$L^\alpha(S, B, \mu)$ - the α-normed space of continuous functions, where S is the space on which these functions are defined, B is the space to which they map, and μ is the integration measure with respect to which the norm is defined

$(l^d_\alpha)^*, (S^d_\alpha)^*, S(\mathcal{H})^*, ...$ - the Banach space duals of normed spaces

$A; \mathcal{H}_A, S_1(A), S(A), \mathbb{B}(A), U(A)$ - system A; the Hilbert space, 1-normed operators, states, bounded operators, and unitaries on A

A^c - the complement of subsystem A

$|A|, |\mathcal{H}_A|, |S(A)|, ...$ - the dimension of system A, equal to that of its Hilbert space, states, etc.

$A \times B$ - the setwise tensor product of pairs $(a, b) : a \in A, b \in B$, or occasionally denotes multiplication

$\otimes$ - the Cartesian tensor product

$\otimes_{i=1}^{n} A_i, \otimes_{i=1}^{n} a_i$ - the tensor product of spaces $A_1, ..., A_n$ or matrices/operators $a_1, ..., a_n$

$A^{\otimes n}, a^{\otimes n}$ - n copies of space A or matrix/operator a in tensor product

$\oplus$ - the direct sum

Vectors & Matrices

$\vec{\psi} = (\psi_1, ..., \psi_d)$ - a vector with entries $\psi_1, ..., \psi_d$ in dimension d

ρ^A, ρ_A - equivalent notations for restriction of a matrix or operator ρ to subsystem A

$\hat{1}$ - the identity matrix or operator, where sometimes a subscript denotes the system or dimension

$\hat{i}$ - the ith basis element of a vector space, where $i \in 1...d$ in dimension d

$\|\rho\|_\alpha$ - the α-norm of vector or matrix ρ

tr - the matrix or operator trace

tr_A - the partial trace over subsystem A, for a tensor space $A \otimes B$

$a^\dagger$ - the adjoint of matrix or operator a, equivalent to its Hermitian conjugate in finite dimension

a^* - the complex conjugate of a matrix, operator or number

h.c. - the Hermitian conjugate or adjoint of the previous expression

$[a, b]$ - the commutator of matrices or operators a and b

$\{a, b\}$ - the anticommutator of matrices or operators a and b, though the $\{...\}$ notation can also be used to define a set of elements, e.g. $\{f(i) : i \in 1...n\}$ for some function f

Quantum States & Operations

$|\psi\rangle, |\phi\rangle, ...$ - vectors in Hilbert space, via braket notation

$\langle\psi|, \langle\phi|, ...$ - dual vectors in Hilbert space, via braket notation

$\langle\psi|\phi\rangle$ - the Hilbert space antilinear inner product of $|\psi\rangle$ and $|\phi\rangle$

$|\psi\rangle\langle\phi|$ - the matrix given by the outer product of vectors $|\psi\rangle$ and $\langle\phi|$

$M_{|i\rangle}$ - measurement in the basis given by $\{|i\rangle\}$

$\langle\mathcal{O}\rangle_\rho$ - the expectation value of operator $\mathcal{O}$ with state or density ρ

Algebras

$\mathcal{N} \subseteq \mathcal{M}$ - $\mathcal{N}$ is a subset of $\mathcal{M}$, or a subalgebra if both are algebras

$\mathcal{N} \subset \mathcal{M}$ - $\mathcal{N}$ is a strict subset or subalgebra of $\mathcal{M}$

$\mathcal{N} \cap \mathcal{M}$ - the intersection of sets or algebras $\mathcal{M}$ and $\mathcal{N}$

$\mathcal{N} \vee \mathcal{M}$ - the algebra generated by the union of $\mathcal{M}$ and $\mathcal{N}$

$\mathcal{E}_{\mathcal{N}}$ - the conditional expectation to subalgebra $\mathcal{N}$

$\mathcal{N}', \mathcal{N}'_{\mathcal{M}}$ - the commutant of algebra $\mathcal{N}$, in the latter case explicitly specifying the larger algebra $\mathcal{M}$ (such that $\mathcal{N} \subseteq \mathcal{M}$) in which the commutant is taken

$\mathcal{E}_{G'}$ - the conditional expectation to subalgebra that commutes with group G of unitaries

$\mathcal{E}_{\rho'}$ - the conditional expectation to subalgebra that commutes with matrix ρ

$\mathbb{C}, \mathbb{C}\hat{1}$ - the algebra of complex numbers, where $\mathbb{C}$ may implicitly refer to complex multiples of the identity, while $\mathbb{C}\hat{1}$ does so explicitly

$M_{\mathcal{N}}$ - the conditional expectation to $\mathcal{N}$ with measurement of $\mathcal{N}$'s center

$\langle a, b, c, ... \rangle$ - the algebra generated by matrices or operators $a, b, c, ...$ (order does not matter)

$\mathcal{A}, \mathcal{B}, ...$ - script capital letters near the beginning of the alphabet usually denote the algebras of operators on systems $A, B, ...$, or in infinite dimension highlight that these algebras are factors in the von Neumann algebra sense, which are analogous to quantum subsystems

X, Y, Z - the Pauli matrices in dimension 2, unless noted to denote more general matrices or operators

$\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ - equivalent respectively to $\langle X \rangle, \langle Y \rangle, \langle Z \rangle$, shorthand for algebras generated by these operators

Entropy

Note: I may omit the subscript density matrix in denoting information quantities, especially if it is either obvious from context or unnecessary. By default, entropies are with the natural logarithm.

$D(\rho \| \sigma)$ - relative entropy between ρ and σ

$D_\alpha(\rho \| \sigma)$ - sandwiched Rényi relative α-entropy between ρ and σ

$D^{\mathcal{N}}(\rho)$ - relative entropy with respect to subalgebra $\mathcal{N}$, equivalent to $D(\rho \| \mathcal{E}_{\mathcal{N}}(\rho))$

$H(\rho)$ - von Neumann entropy of density ρ

$H(A)_\rho$ - entropy ρ's restriction to subsystem A

$H_\alpha(\rho), H_\alpha(A)_\rho$ - α-Rényi entropy of ρ and its restriction to subsystem A

$H(A|B)_\rho$ - conditional entropy of ρ's restriction to subsystem A conditioned on subsystem B

$H_\alpha(A|B)_\rho$ - sandwiched conditional Rényi α-entropy

$H(\mathcal{N})_\rho$ - entropy of ρ's restriction to subalgebra $\mathcal{N}$, equivalent to $H(\mathcal{E}_{\mathcal{N}}(\rho))$

$I(A : B)_\rho$ - mutual information between subsystems A and B for density ρ

$I(A : B|C)_\rho$ - conditional mutual information on C

$I(\mathcal{S} : \mathcal{T} \subseteq \mathcal{M})_\rho$ - generalized conditional mutual information for algebras $\mathcal{S}, \mathcal{T} \subseteq \mathcal{M}$

$I(\mathcal{S} : \mathcal{T})_\rho$ - equivalent to $I(\mathcal{S} : \mathcal{T} \subseteq \mathcal{M})_\rho$

$I_{\mathcal{L}}(\rho)$ - Fisher information of matrix or operator ρ for Lindbladian generator $\mathcal{L}$

Time-Evolution

$\mathfrak{H}$ - a Hamiltonian generator

$\mathcal{L}$ - a Lindbladian generator (not a Lagrangian)

$ad_{\mathcal{L}}(\rho), \mathcal{L}(\rho)$ - Lindbladian $\mathcal{L}$ applied to density ρ

Φ, Ψ - generally used to denote quantum channels

$\{\Phi^t\}$ - the family of quantum channels parameterized by t, usually assumed to form a semigroup

Φ^c - the complementary channel of Φ

A, A', B, E - in the context of a channel, usually A denotes a (possibly held-back) input system, A' the input sent into the channel, B the channel's output, and E the environment

$\Phi^\dagger$ - the adjoint of channel Φ with respect to the matrix or Hilbert space inner product

$R_{\sigma,\Phi}$ - the Petz recovery map parameterized by density σ for channel Φ

Z_λ, Z_k (specifically in Chapter 6) - erasure channel with success rate λ and k-success heralded channel

Shannon & Resource-Theoretic Measures

E_{sq} - squashed entanglement

I_{sq} - squashed mutual information for subalgebras

I_{conv} - convex-minimized mutual information

$\chi(\Phi)$ - the Holevo information of channel Φ

$I_c(A\rangle B)_\rho$ - the coherent information of a bipartite density ρ^{AB}

$C(\Phi)$ - the classical capacity of channel Φ

$Q(\Phi)$ - the quantum capacity of channel Φ

Chapter 1

Introduction

I thought of calling it 'information,' but the word was overly used, so I decided to call it 'uncertainty.' When I discussed it with John von Neumann, he had a better idea. Von Neumann told me, 'You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, no one really knows what entropy really is, so in a debate you will always have the advantage.' - Claude E. Shannon quoted in Scientific American (1971), v225, p180, accessed via `https://en.wikiquote.org/wiki/Claude_Elwood_Shannon` on 3/12/20.

Quantum information (QI) fuses math, physics, computer science, electrical engineering, a bit of chemistry, and probably more research areas. Ideas from coding theory help us understand fundamental theories of the universe [1], paradigms inspired by thermodynamic conversions model resources for quantum computing [2], and deep questions about operator algebras manifest in the theory of quantum correlations [3].

Traditional theory of computation separates the computational process from its underlying physics, hiding the mechanical, thermal, electronic and other processes underlying the computer from the abstractions of circuits, complexity, programming, etc. This separation of scales has been immensely successful, and many believe that error reduction and correction will make quantum computing follow the same trajectory. An alternate perspective views the intertwining of physical and informatic concepts as one of the most promising features of quantum information theory. Wheeler's "It from Bit" hypothesis [4, 5] posits that information fundamentally underlies physics, and some thermodynamic ideas [6, 7] suggest that physical consequences are unavoidable for information processing. The culmination of "It from Bit" is however not essential to the value of the connections that inspire it. Whether or not information is at the root of physics, or physics at the root of information, theoretical connections and parallels present an immense opportunity. By studying information and computation, we discover results in physics and mathematics.

Claude Shannon's choice to call his primary quantity of study "entropy" rather than "information" seems prescient in hindsight. As noted by Shannon, connotations of the word "information" may have interfered with the purpose of the theory. Colloquially, information often necessarily involves a person learning or transmitting, automatically assuming some anthropic component. Even in taking humans out of the story, one often cannot escape putting in computers or media. See for instance the following abridged definition:

Definition of ***information***

1. *(a) knowledge obtained from investigation, study, or instruction (b) intelligence, news (c) facts, data*
2. *the attribute inherent in and communicated by one of two or more alternative sequences or arrangements of something (such as nucleotides in DNA or binary digits in a computer program) that produce specific effects*

...

- Merriam-Webster Online, `https://www.merriam-webster.com/dictionary/information`, accessed on 3/18/20.

The intuition from computer bits was undoubtedly valuable, and early concepts of binary encoding probably helped inspire the idea that one could meaningfully separate the statistics of information from its semantics. Shannon's introduction of the concept was rejected from the math journal to which he had originally submitted largely because of its close ties to digital computing, which the reviewer (incorrectly in hindsight) believed to be a field with limited potential [8] *. Information theory instead first appeared in *Bell Systems Technical Journal* [9], and the new field found its first home in electrical engineering. Entropy nonetheless exists in natural systems and manifests in physical processes, independently from human engineering.

With its many applications and manifestations, entropy carries a wide variety of continuing mysteries. Quantum Shannon theory is notorious for intractable calculations [10, 11], and even apparently simple processes can show exotic effects that confound traditional methods of calculation [12]. The same properties that make quantum computing so potentially powerful also confound extrapolations from microscopic, qubit-by-qubit descriptions to their collective dynamics or time-evolution when coupled with unknown environments. The most computationally useful properties of quantum systems are almost by definition hard to simulate classically, and often arise from strong, non-perturbative interactions. Hence there is strong motivation to develop new analytical techniques.

The rest of this introduction describes some of the more specific themes of this thesis, and then summarizes the main chapters within.

1.1 Relative Entropy, Asymmetry and Subalgebras

The primary mathematical objects of study in this thesis are forms of relative entropy. Its finite dimensional form for a pair of density matrices ρ, σ is given by

$$D(\rho\|\sigma) = \mathrm{tr}(\rho \log \rho - \rho \log \sigma) \ .$$

*This story is bizarrely and literally close to the origins of this thesis. The great mathematician Joseph L. Doob is said to have made the decision to reject Shannon's paper. Doob's office was in Altgeld Hall at the University of Illinois, and it later became the office of my research advisor, Marius Junge.

As described in more detail in Chapter 2, relative entropy is mathematically central to information theory (also see [13]). Many other forms of entropy are special cases of relative entropy, and it bounds norm distances between quantum densities. Hence results for relative entropy often apply to many circumstances. Conceptually, relative entropy is a good example of a crossing point between disciplines. Key features of relative entropy include:

1. Relative entropy is extensive in the number of independent copies of a system. In conventional thermodynamics, quantities such as volume, total energy, particle number, and entropy are extensive, as they scale with the total amount of matter under consideration. This contrasts them with intensive quantities, such as temperature and density, which are constant in the amount of matter. In mathematical form,

$$D(\rho \otimes \omega \| \sigma \otimes \eta) = D(\rho \| \sigma) + D(\omega \| \eta) \ .$$

   The extensiveness of relative entropy first implies that resources it quantifies scale with the number of copies, which is often physically reasonable. Second, it often allows us to use information-theoretic techniques that repeat random processes many times and quantify typical outcomes, spreading the effects of rare outliers such that they become ignoreable.

2. Data processing: physical operations satisfying certain basic independence notions do not increase relative entropy. Mathematically, when Φ is a completely positive, trace-preserving map on matrices (in physical terms a quantum channel or open process),

$$D(\rho \| \sigma) \geq D(\Phi(\rho) \| \Phi(\sigma))$$

   for all densities ρ and σ. Variants of data processing lead to its use in quantifying resources, such as correlations, entanglement, asymmetry, and quantum coherence, under sets of operations that use but do not create these resources.

3. While the conventional von Neumann entropy diverges in infinite-dimensional quantum field theories, relative entropy may remain finite and maintain useful properties even where basic ideas such as subsystems no longer have meaning. Though relative entropy can be infinite even in finite dimension, many of the cases we study automatically enforce finiteness. In particular, we often study relative entropy with respect to a von Neumann subalgebra, which seems to be a fundamentally useful form.

4. The quantum form of relative entropy reduces gracefully to its classical cases. Meanwhile, measures of non-classicality based on relative entropy have proven faithful indicators of quantum properties.

In Section 2.4.1, I review several (often equivalent) formulations, from coding theory, from modular theory (a branch of operator algebra theory that has become popular in high-energy physics), and as a derivative of Banach space norms. The key ideas of this thesis arise largely from the interplay between different perspectives on relative entropy.

The relative entropy with respect to a subalgebra is given by

$$D^{\mathcal{N}}(\rho) \equiv D(\rho \| \mathcal{E}_{\mathcal{N}}(\rho)) = H(\mathcal{E}_{\mathcal{N}}(\rho)) - H(\rho) \ ,$$

where $H(\rho)$ is the von Neumann entropy as described in Section 2.4.2. In discussed Section 4.2, this is a common form in constructing quantum resource theories, and we may generally interpret this form as *relative entropy of asymmetry*. $D^{\mathcal{N}}$ may have a complete interpretation as an asymmetry measure, though this might be a stretch from the original intention of the resource theory of asymmetry. Nonetheless, asymmetry is a broad concept that can subsume ideas from coherence, conditional entropy, and many other areas of information theory. It also links to concepts that do not require any quantum underpinning, such as more traditional formulations as in Noether's theorem, and ideas on the tradeoffs between generality and precision in modeling. The aim of Chapter 5 is to show how these ideas are compatible and connected.

Nearly equivalent to the data processing inequality is the subadditivity of relative entropy for subalgebras. For this, we must introduce the idea of a von Neumann subalgebra, which is a particular kind of subset of matrices or operators I describe in detail in Section 2.3. Examples of subalgebras include bases of quantum systems, subsystems (up to an often trivial tensoring with complete mixture), and the invariant algebras of symmetry groups. Every von Neumann subalgebra has associated to it a map restricting to that subalgebra, called a *conditional expectation*. Given a pair of subalgebras $\mathcal{M}$ and $\mathcal{N}$ with *commuting* conditional expectations $\mathcal{E}_{\mathcal{N}}$ and $\mathcal{E}_{\mathcal{M}}$, the well-known subadditivity of relative entropy for subalgebras states that

$$D(\rho \| \mathcal{E}_{\mathcal{N}}(\rho)) + D(\rho \| \mathcal{E}_{\mathcal{M}}(\rho)) \geq D(\rho \| \mathcal{E}_{\mathcal{N} \cap \mathcal{M}}(\rho)) \ , \tag{1.1}$$

where $\mathcal{N} \cap \mathcal{M}$ is the intersection of subalgebras. This inequality also underlies forms of the uncertainty principle, as discussed in Chapter 3, and is a form of the strong subadditivity inequality for quantum entropy. In Section 3.1, I show that when $\mathcal{E}_{\mathcal{N}}$ and $\mathcal{E}_{\mathcal{M}}$ do not commute, but $\mathcal{N} \cap \mathcal{M}$ is a physically trivial algebra, we still obtain an inequality of the form

$$D(\rho \| \mathcal{E}_{\mathcal{N}}(\rho)) + D(\rho \| \mathcal{E}_{\mathcal{M}}(\rho)) \geq \alpha D(\rho \| \hat{1}/d)$$

in dimension d for some $\alpha > 0$ (see Theorem 3.2). This inequality has applications described in Section 7.2.1 to estimate the decay of relative entropy under continuous-time Markov processes. It further adds a general form of entropy inequality to the repertoire of quantum information, which is likely to have future applications. I also note some extensions of the conditional mutual information for commuting $\mathcal{E}_{\mathcal{N}}$ and $\mathcal{E}_{\mathcal{M}}$, as well as to pairs of compatible quantum channels, largely derived with Li Gao and Marius Junge.

1.2 Quantum vs. Classical Entropy

As this thesis focuses on *quantum* entropy, a key theme is how entropy reveals the difference between a quantum and classical universe. Since Einstein and his contemporaries [14, 15], physicists have wondered at results in quantum theory that seem to clash with classical understanding. The nature of this discourse has shifted from primarily challenging the validity of quantum theory to seeking applications that would be impossible in a classical world [16]. Entanglement and the uncertainty principle are two of the quintessential cases of quantum phenomena that one can formulate in terms of entropy [17, 18, 19], and ultimately in terms of relative entropy.

One basic notion of quantumness is *coherence*, which refers to the extent that a quantum system can have more entropy in a given basis than it does in its basis of minimal entropy. The concept is rooted in ideas like wave-particle duality: when a particle exhibits wavelike properties, it may overall be in a state of very low entropy due to its characterization in terms of a single frequency, but a measurement in the position basis shows a random outcome. Coherence is fundamentally linked to the uncertainty principle, as its existence depends on having multiple bases. It further distinguishes between quantum superpositions, which have definite phases between possible states, and probabilistic mixtures, which lose this phase information. When the subalgebra $\mathcal{N}$ corresponds to a measurement basis, $D^{\mathcal{N}}$ becomes the relative entropy of coherence, a resource-theoretic measure of quantum coherence (see Section 4.2).

When $\mathcal{N}$ and $\mathcal{M}$ are two bases, subadditivity and adjusted subadditivity become entropic uncertainty relations that are equivalent to subadditivity inequalities for coherence. When $\mathcal{N}$ and $\mathcal{M}$ are subsystems, $D^{\mathcal{N}}(\rho) - D^{\mathcal{N}}(\mathcal{E}_{\mathcal{M}}(\rho))$ becomes a form of generalized mutual information as described in detail in Chapter 3.

Several entanglement measures, including the squashed entanglement and entanglement of formation, have mathematical forms in terms of conditional mutual information and ultimately differences of relative entropies. Hence the subalgebra generalizations from Chapter 3 extend to these entanglement measures, and we can for instance quantify the entropic non-classicality present between two bases in an uncertainty principle. Studying these measures leads to a new resource theory, joining entanglement with uncertainty-based non-classicality, in which the two are interconvertible under free operations. We obtain the conversion:

$$2 \text{ UCR} \leftrightarrow \text{EPR} , \tag{1.2}$$

which exchanges two copies of a non-classical, single-qubit configuration for one maximally entangled pair of qubits. The ultimate result of this work, described in Chapter 4 and originally based in a joint project with Li Gao and Marius Junge, is a conceptual and operational unification of entanglement with uncertainty.

Non-classical aspects of quantum information show potential to enhance processing and communication, even when the inputs and outputs are entirely classical. A dramatic example is the superadditivity of transmission rates as discussed in depth in Chapter 6. Such effects are also evident in quantum games, scenarios in which several parties must generate correlated responses to inputs, usually with no or limited communication [20]. Most schemes for quantum computing also involve mapping classical problems to classical solutions using an internally quantum process. Superadditivity, quantum games, and quantum computers all contain indirect signatures of quantum processes. Were we to treat each such process as a black box, we would not directly receive any external quantum objects, but we would observe statistics or outcomes that could not plausibly arise from classical internals.

Studying what features of systems and entropies signal underlying quantum mechanisms also helps us understand cases that do not require quantum mechanics. Most of the results in Chapters 3 and 7 hold for classical stochastic systems. For instance, Theorem 3.2 gives an uncertainty-like bound on relative entropies with respect to classical subsystems when conditional expectations fail to commute. These results are usually less striking for classical relative entropy, in which calculations generally do not require the same level of non-commutative mathematical tools. Nonetheless, as noted in Chapter 5, the Rényi relative α-entropy of asymmetry as discussed in Chapter 3 probably has some interpretations in classical statistics and modeling.

Quantumness of information appears fundamentally connected to secrecy. Classically, secrecy of information fundamentally involves references to competing agents. Private classical transmissions almost necessarily involve a sender, a receiver, and some number of eavesdroppers. In contrast, information displaying fundamentally quantum aspects often requires privacy, related to the no cloning theorem for quantum states. An operation that perfectly copies a quantum state in one basis completely destroys coherence with respect to that basis - applying concepts discussed in Section 2.6, one can model a coherence-destroying operation as a copying channel. Similarly, the monogamy of quantum entanglement (used heavily in Chapter 6) implies some mutual exclusivity, implying that for a state to be highly entangled between two systems, it must not be highly entangled or even highly correlated between either and any other system. The inherent privacy required by quantum information underpins quantum cryptography, as any eavesdropper making extra copies of quantum data ultimately changes the statistics of the original. Importantly, these notions of quantum secrecy apply between systems and subalgebras, not requiring any anthropic agent as an eavesdropper. Quantum processes mapping an input to an output state generally include a complementary process that replaces output by the system's environment. In particular, Theorem 3.4 in this thesis shows a strong correspondence between correlations in quantum subalgebras and those of certain complementary algebras. Chapter 6 uses entanglement monogamy to bound quantum advantages when a larger number of randomly-selected subsystems are lost to the environment.

1.3 Decay and Decoherence

The downside of quantum information's obligatory secrecy is its apparent fragility. It is well-understood that a measurement converting quantum information to classical form eliminates or consumes the quantum power of resources. Unfortunately, the fact that copying need not involve an intentional agent means that most quantum systems constantly lose resources such as entanglement and coherence over time. Interactions that leak information to any part of the environment, such as the circuits controlling an apparatus, the walls of a sealed chamber, the dust in open air, or even other qubits in a quantum computer reduce the entanglement a quantum system maintains with any particular other. In quantum computing, this manifests in various forms of decoherence, which include passive decay of qubits toward equilibrium and errors introduced in quantum operations. Decoherence is one of the main challenges to scalable quantum technology. Though quantum error correction and encoding/decoding may ultimately enable quantum algorithms and communications that ignore decoherence, the reality of quantum experiments is even leading some theorists to consider what quantum advantages may appear in the presence of uncorrected noise [21].

The primary information-theoretic model of time-evolution, the quantum channel, is actually a general model of physical processes involving interactions with an initially uncorrelated environment. When the backaction of the system on the environment is sufficiently small or dissipates sufficiently quickly, we may use the more specialized model known as the quantum Markov semigroup. A quantum Markov semigroup is a family of quantum channels parameterized by time, modeling a continuous process. Quantum Markov semigroups usually appear to induce exponential decay toward a subspace of invariant *fixed point* states. For example, random transformations from a group appear to ultimately cause decay of a quantum state toward states that are invariant under the group's actions, even when the probability distribution on transitions is biased. Exponential decay of relative entropy with respect to a fixed point state or subspace is known as a modified logarithmic Sobolev inequality (MLSI), given mathematically as

$$D(\Phi^t(\rho)\|\Phi^\infty(\rho)) \leq e^{-\alpha t} D(\rho\|\Phi^\infty(\rho)) \ ,$$

where Φ^t is the time-parameterized quantum channel, Φ^∞ is the infinite-time limit of that process, and $\alpha > 0$ controls the decay rate. Related problems include estimating decoherence times and rates of quantum resource preservation or information transmission, including quantum analogs of Shannon's communication capacity. Relative entropy's centrality in quantum information make it a powerful target for decay bounds, which then usually apply to a wide variety of tasks and resources.

Exponential decay of relative entropy need not pose any fundamental barrier to quantum computing. The same forms of bound technically apply to classical systems. A single bit, stored in conventional electronic memory, technically has some probability to randomly flip states. While this implies exponential decay, the rate constant is also heavily suppressed by the large number of physical constituents maintaining that state. The advantage of classical information is that many classical systems almost trivially implement large-scale repetition codes, using an enormous number of physical degrees of freedom to encode small computational degrees of freedom. This is an advantage of separating physical from computational scales. Quantum error correction promises to do the same in the quantum realm, though it seems much harder for current technology.

Beyond its practical value to quantum technology, studying decoherence may help reveal how quantum information differs fundamentally from its classical analogs. The structure of quantum error correcting codes has inspired hypotheses on possible quantum structures underlying spacetime [1]. Structures that allow quantum properties to persist over long times and distances at high state complexity are a new frontier in physical sciences [22]. Chapter 7, which primarily describes the results of collaborations to theoretically understand relative entropy decay, also summarizes applications of complex interpolation, transference, and measure change techniques to quantify effects of collective decoherence and to analyze systems that can be arbitrarily entangled with an external auxiliary. How systems with strong quantum properties interact differently with their environments suggests ways in which being quantum changes the external nature of a system, and it may reveal natural constraints on the power of non-classicality.

1.4 Summary Outline of the Main Sections and Results

Most of the theorems, lemmas, corollaries, proofs, and examples in this thesis appear in publications or preprints. I have marked in each such result the original source, and whether the result has been substantially modified or updated in its thesis version. Results from co-authored papers may come from joint work, and often involved discussions with those co-authors. The emphasis of this thesis has however been selected to highlight the contributions and research interests of its author. Theorems that are not cited from a specific source may not appear in any current publication but could appear in future papers or upcoming versions of existing preprints. When quoting theorems from papers on which I was not a co-author, the authors are specifically identified.

- **Chapter 2** is a review of mathematical concepts used throughout the paper, such as von Neumann algebras, quantum entropies, and quantum channels. There is no new research in Chapter 2. In addition to recalling some of the definitions and classic results needed to understand this thesis, Chapter 2 is intended to convey some of the intuition underlying results therein. A particular theme is to connect approaches to similar ideas originating in different fields, such as operator algebras, physics of open systems, and quantum computer science.

- **Chapter 3** collects results primarily from [23, 24, 25, 26] on generalizations of the previously known and ubiquitous strong subadditivity inequality from subsystems to subalgebras, and ultimately to quantum channels. This chapter contains the central mathematical ideas of the thesis, which lead to further results described in Chapters 4, 5, and 7. The first major result in this chapter, appearing in Section 3.1 is the "adjusted subadditivity" (Theorem 3.2) that more fully extends the subadditivity of relative entropy with respect to subsystems to subalgebras, going beyond the known cases in which subalgebraic restrictions commute. This is based on a new technical inequality that multiplicatively compares relative entropies. The second big result of this chapter is the commutant duality (Theorem (3.4) appearing in Section 3.2. This chapter also includes an extension of generalized mutual information to more than two constituent algebras, and from subalgebras to some quantum channels.

- **Chapter 4** describes results primarily from [23], connecting quantum entanglement with the non-classicality found in quantum uncertainty relations. This gives physical interpretation to mathematical ideas of Chapter 3. We show that replacing subsystems by subalgebras in bipartite system decompositions yields a resource theory of bipartite non-classicality, in which non-classical entropy found in entangled states converts freely to and from that in qubits split between mutually unbiased bases. We also note some links to resource-theoretic coherence and asymmetry monotones.

- **Chapter 5** records an answer to a question I had posed when planning this thesis: how does the resource theory of asymmetry relate to the notion one obtains by extending Noether's theorem, and to symmetry in dynamical modeling? While not based on any current paper or preprint, this chapter fills in gaps in intuiting the relative entropy of asymmetry $D^{\mathcal{N}}$ as described in Section 4.2.

- **Chapter 6** describes results from [27] and from an ongoing collaboration [28] on the superadditivity of information transmission rates and non-classicality of quantum games, both entanglement-based advantages of quantum over classical methods. Superadditivity is a classic example of how quantum effects change information theory. The main theme of [27] is on how monogamy of entanglement limits quantum advantages under random subsystem loss. The thesis version of these results leverages a more recent continuity bound on a key entanglement measure to strengthen bounds from the original paper. The main bound is on superadditivity of classical bit transmission. Similar techniques bound the quantum advantages in multiplayer games, another well-known way in which the quantum nature of the universe can manifest even in the performance of an ultimately classical task. The goal of [28] is an experimental demonstration of superadditivity.

- **Chapter 7** summarizes and combines a mix of results from [29, 30, 31, 32, 25, 33, 26] and an upcoming paper with Marius Junge and Haojian Li on decay of quantum states toward equilibrium. The focus of this chapter is on how entropy measures the decay of quantum systems exposed to their environments. This is a slew of papers with several technical ideas and many collaborators. Hence the chapter does not go to into full technical depth on every mentioned result, seeking instead to collect and link the results of different projects as they relate to main themes of this thesis. The first set of results, in Section 7.1, additively (in some cases perturbatively) bound relative entropy of a particular class of quantum channels with respect to the output of conditional expectations, which are maximally mixing channels in this class. These are primarily based on joint work with Marius Junge and Li Gao. The second main class of results, appearing in Section 7.2 shows exponential decay of relative entropy in continuously time-evolving quantum systems with respect to their fixed points. These results are pieces in a set of projects driven largely by Marius Junge to show general decay estimates for quantum Markov processes.

Chapter 2

Background & Review of Mathematics for Quantum Information

The purpose of this section is to convey some intuition at the interface between the mathematical and physical sides of quantum information theory. None of the ideas in this section are new, and most would be common knowledge in their respective communities. Over the years in which this thesis formed, I saw many examples of how approaches to quantum information consider the same range of basic phenomena in wildly different ways.

I assume familiarity with basic linear algebra and calculus, including complex numbers, vectors, matrices, Cartesian tensor products, the trace and partial trace functions, diagonalization, integrals, etc. I also assume that the reader has seen undergraduate level quantum mechanics, including braket notation. Much of the formalism I use is based on that of Strocchi's "An Introduction to the Mathematical Structure of Quantum Mechanics [34]," which I found immensely useful when shifting to a more mathematical form of quantum information.

2.1 Banach Spaces, Hilbert Spaces, Probabilities & Amplitudes

In finite dimension d, a p-normed Banach space $l_p^d(\mathbb{C})$is a space of vectors $\vec{\psi} = (\psi_1, ..., \psi_d)$ such that $\psi_i \in \mathbb{C}$ for each $i \in 1...d$. $l_p^d(\mathbb{R})$ is defined correspondingly over the real numbers. The p-norm is the norm for l_p^d and given by

$$\|\vec{\psi}\|_p = \left(\sum_{i=1}^{d} |\psi_i|^p \right)^{1/p} .$$

If $p = \infty$, then the norm is given by the maximum vector entry, as $\lim_{p\to\infty} \sqrt[p]{(\psi_1^p, ..., \psi_d^p)} = \max_i \psi_i$.

The p-norm extends to spaces of countably infinite dimension, with the caveat that states are not assured to be normalizable in l_p^∞ unless $p = \infty$ or the vector's infinite tail is a convergent and finite series. In uncountably infinite dimension, it sometimes suffices to replace the sum by an integral. For example, given a set or interval S and measure μ on it, we may denote by $L^p(S, \mathbb{C}, \mu)$ the functions $\psi : S \to \mathbb{C}$ with norm given by

$$\|\psi\|_p = \left(\int_{x \in S} |\psi(x)|^p d\mu(x) \right)^{1/p} .$$

As a canonical example, we may consider $S = [0, 1]$, the unit interval on the reals, with uniform measure.

Each Banach space l_p^n or L^p has a dual space given by linear functionals on it and denoted $(l_p^d)^*$ or $(L^p)^*$. For $1 < p < \infty$, the dual space to the p-normed Banach space is a q-normed Banach space, and the space is reflexive in that the dual of the dual is the original space.

We may interpret the spaces l_1^d and L^1 as probability spaces, where a probability is a non-negative, real vector or function with 1-norm equal to 1. For example, the probability space for a (potentially biased) coin would be $l_1^2(\mathbb{R})$, which is the space of normalized vectors $\vec{\psi} = (\psi_1, \psi_2)$ over $\mathbb{R}$. The space $L^1([0, 1], \mathbb{R}, \mu)$ may have the interpretation of the space of probability distributions over position on a finite line segment of unit length. Technically, l_1^d is not just a space of probabilities, because it contains unnormalized vectors and vectors with negative entries. We will often ignore this distinction, and when we refer to probability spaces, assume we restrict our attention to probabilities.

Usual models of quantum mechanics start in Hilbert space, which in the settings considered so far is a Banach space such that $p = 2$. A Hilbert space $\mathcal{H}$ has an inner product $\langle \cdot | \cdot \rangle : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$ denoted $\langle \phi | \psi \rangle$ for elements ϕ and ψ. In the finite-dimensional Hilbert space l_2^d, $\langle \phi | \psi \rangle = \sum_{i=1}^d \phi_i^* \psi_i$, where "$*$" denotes the complex conjugate. In the space $L^2(S, \mathbb{C}, \mu)$,

$$\langle \phi | \psi \rangle = \int_{x \in S} \phi^*(x) \psi(x) d\mu(x) .$$

For the Hilbert space norm, $\|\psi\|_2 = \langle \psi | \psi \rangle$. We may denote a vector $\vec{\psi} \in l_2^d$ or function $\psi \in L^2(S, \mathbb{C}, \mu)$ by the "ket" notation $|\psi\rangle$ or a vector in the dual space by $\langle \psi |$ as in standard bra-ket notation. Any Hilbert space $\mathcal{H}$ (which might be l_2^d, L^2, or any of many other infinite-dimensional Hilbert spaces) is self-dual in that $\mathcal{H}^* = \mathcal{H}$.

In l_2^d, we interpret normalized vectors as containing quantum amplitudes for particular configurations. For example, we may consider normalized vectors in $l_2^2(\mathbb{C})$ given by $|\psi\rangle = (\psi_1, \psi_2)$ to represent the amplitudes of a spin-1/2 particle's "up" and "down" states along a given axis. In analogy to $L^1([0, 1], \mathbb{R}, \mu)$ as a space of probability distributions on positions, we may consider $L^2([0, 1], \mathbb{C}, \mu)$ to be the space of quantum amplitudes over positions on a line segment.

l_2^d and L^2 are not the only Hilbert spaces of physical relevance, but we defer construction of other examples until the Subsection 2.3.3 addressing infinite-dimensional von Neumann algebras.

2.2 Densities & Observables

We denote by S_p^d the Schatten class of p-normed, $d \times d$ matrices. The pth Schatten norm of a $d \times d$ matrix ρ is given by

$$\|\rho\|_p = \left(\sum_{i=1}^{d} |\rho_{ii}|^p \right)^{1/p} = \mathrm{tr}(|\rho|^p)^{1/p} \ .$$

By $S^p(\mathcal{H})$ we denote the class of infinite-dimensional operators on Hilbert space $\mathcal{H}$ with norm given by the analogous integral in functional calculus.

A *density matrix* or *density operator* combines classical probability with quantum amplitudes. A density matrix is a normalized element of S_1^d with real, non-negative eigenvalues. These are by their nature Hermitian. We may interpret the eigenvalues of a density matrix as probabilities of observing pure quantum states corresponding to its eigenvectors. In the usual methods of linear algebra, every density matrix has such a diagonalization via unitary matrices. By diagonalization, a density matrix can always be written in the form

$$\rho = \sum_i \rho_{ii} |i\rangle\langle i| \ ,$$

where $|i\rangle\langle i|$ denotes the ith eigenvector in ket-bra form. Hence we may think of a density matrix/operator as a probability density over a space of quantum amplitude vectors. For any quantum state $|\psi\rangle$, we may write $|\psi\rangle\langle\psi|$ to denote a density matrix that assigns probability 1 to eigenvector $|\psi\rangle$ and probability 0 to all other vectors. In general, we could imagine a density as modeling a quantum state that has been selected according to some random process. We will hence denote $S_1^d = S(l_2^d)$. More broadly, we will denote the densities over an arbitrary Hilbert space $\mathcal{H}$ by $S(\mathcal{H})$. When $\mathcal{H}$ is infinite-dimensional, we refer to density operators rather than density matrices.

Dual to S_1^d is S_∞^d, the space of matrices with norm given by maximum eigenvalue (in analogy to the ∞-normed Banach spaces). We may alternatively denote this space $\mathbb{B}(l_2^d)$, emphasizing that it is the space of bounded operators on d-dimensional Hilbert space. In dimension d, these are just matrices with finite eigenvalues. In infinite dimension, the boundedness of operators is often more subtle, but $\mathbb{B}(\mathcal{H})$ will be a useful construction in many contexts.

We formalize a complete, finite-dimensional von Neumann measurement by a basis $\{|i\rangle\}_{i=1}^d$ and a map $M_{\{|i\rangle\}} : \mathcal{H} \to S(\mathcal{H}) \otimes l_1^d$ such that

$$M_{\{|i\rangle\}}(|\psi\rangle) = M_{\{|i\rangle\}}\left(\sum_{i=1}^{d} \psi_i \, |i\rangle \right) = \sum_{i=1}^{d} |\psi_i|^2 |i\rangle\langle i| \otimes \hat{i} \ , \tag{2.1}$$

where the final $\hat{i}$ is a single-entry unit vector in the probability space l_1^d labeling the ith coordinate. By linearity, we may extend $M_{\{|i\rangle\}}$ to densities in S_1^d. The measurement loses any phase information contained in the amplitudes (ψ_i), which some interpretations describe as a collapse from quantum amplitude to classical probability. Replacing the sums in Equation (2.1) by integrals for continuous spaces, we construct an analogous operation.

Every operator X on Hilbert space $\mathcal{H}$ has an adjoint defined by $\langle\psi|X\phi\rangle = \langle X^\dagger\psi|\phi\rangle$ for all $\phi, \psi \in \mathcal{H}$. For finite-dimensional operators, this is equivalent to the Hermitian adjoint given by the complex conjugate of the matrix transpose. We may interpret self-adjoint (Hermitian) operators as observables, in that we may construct a measurement for a basis of eigenvectors. For observables with degenerate eigenvalues, we must broaden the notion of measurement to include subspaces in which distinct eigenvectors would not be distinguished. This will be easier to formalize in Section 2.3.1.

The expectation value of an observable $\mathcal{O}$ on a given density ρ usually has the form

$$\langle\mathcal{O}\rangle_\rho \equiv \mathrm{tr}(\mathcal{O}\rho)\ .$$

We construct higher-order moments, such as variance, from expectations of powers and powers of expectations. For a pure state $|\psi\rangle\langle\psi|$,

$$\langle\mathcal{O}\rangle_{|\psi\rangle\langle\psi|} = \mathrm{tr}(\mathcal{O}|\psi\rangle\langle\psi|) = \langle\psi|\mathcal{O}|\psi\rangle\ ,$$

recovering the usual formula from introductory-level quantum mechanics. The expectation is equivalent to the value obtained by averaging measurement outcomes over an asymptotically large ensemble of identical trials. The set of expectations over all observables completely fixes a density matrix.

2.3 von Neumann and C* Algebras

Readers with less mathematical background may wish to first read Subsection 2.3.2 to gain some intuition for the simplest cases.

An associative algebra has the structure of an algebraic ring, including addition and multiplication operations, and is also a vector space. A C^*-algebra $\mathcal{M}$ is an associative algebra over the real or complex numbers that is also a Banach space with a norm $\|\cdot\| : \mathcal{M} \to \mathbb{R}^+$ such that $\|XY\| \le \|X\|\|Y\|$ for any $X, Y \in \mathcal{M}$. A C*-algebra is closed under:

(1) Linear combinations: $\alpha a + \beta b \in \mathcal{M}$ if $a, b \in \mathcal{M}$ and $\alpha, \beta \in \mathbb{C}$

(2) Composition: $ab \in \mathcal{M}$ if $a, b \in \mathcal{M}$

(3) Hermitian conjugation: $a^\dagger \in \mathcal{M}$ if $a \in \mathcal{M}$

If we take a set of matrices with any Schatten norm, we may generate its C^*-algebraic closure under the above operations. Such algebras of matrices are relatively intuitive and familiar examples of finite-dimensional C^* algebras. In general, infinite-dimensional C^* algebras may support exotic phenomena [35] and remain an area of active research.

A linear functional on a C^*-algebra $\psi : \mathcal{M} \to \mathbb{C}$ with norm 1 is called a state. The Gelfand-Naimark-Segal (GNS) construction allows us to take a C^* algebra $\mathcal{M}$ and state ρ, and from these construct a $*$-representation π on a Hilbert space $\mathcal{H}$ with a cyclic, unit vector $\xi \in \mathcal{H}$ such that $\rho(a) = \langle \pi(a)\xi, \xi \rangle$ for any $a \in \mathcal{M}$. While traditional formulations of quantum mechanics usually start in Hilbert space, the GNS construction allows us to go the other way, obtaining a Hilbert space given an algebra and state. The word "state" is sometimes analogous to what we would normally think of as a density, though in other cases mathematical physicists refer to a vector in Hilbert space as a state. We will generally refer to potentially mixed states as densities in this dissertation.

A von Neumann algebra (vNa) is a C^*-algebra that is also closed in the weak operator topology. Equivalently, a vNa is a C^*-algebra that is the dual of some Banach space. One may also define a vNa as a subalgebra of the bounded operators on a Hilbert space that is closed under Hermitian conjugation. The additional properties of von Neumann algebras make them more common in physical models.

A finite dimensional von Neumann algebra is a direct sum of matrix algebras with multiplicities. Namely, any finite-dimensional von Neumann algebra $\mathcal{N}$ has the form

$$\mathcal{N} = \oplus_i (\mathbb{M}_{n_i} \otimes \mathbb{C}1_{m_i}), \tag{2.2}$$

where i indexes the diagonal blocks, $\mathbb{M}_{n_i}$ is the algebra of n_i-dimensional matrices, and m_i is the multiplicity of $\mathbb{M}_{n_i}$ within $\mathcal{N}$. Here $\mathbb{C}1_{m_i}$ is the physically trivial algebra of scalar multiples of an m_i-dimensional identity matrix. We note that this form is always block diagonal. We may begin to interpret it physically. The outer direct sum acts like a classical random variable, splitting the algebra into diagonal blocks. Within each of these blocks, we have a copy of the matrix algebra $\mathbb{M}_{n_i}$ that acts like it's tensored with a trivial algebra containing scalar multiples of the m_i-dimensional identity matrix.

2.3.1 Subalgebras, Subsystems & Measurements

A key aspect of von Neumann algebras is that they may contain subalgebras. Associated to any von Neumann subalgebra $\mathcal{N} \subseteq \mathcal{M}$ is a unique conditional expectation $\mathcal{E}_\mathcal{N} : \mathcal{M} \to \mathcal{N}$. In tracial algebras, the conditional expectation is formally defined as the operator with the property that

$$\mathrm{tr}(\mathcal{E}_{\mathcal{N}}(a)b) = \mathrm{tr}(\mathcal{E}_{\mathcal{N}}(a)\mathcal{E}_{\mathcal{N}}(b)) \; \forall a, b \in \mathcal{M} \; .$$

This property follows immediately from the fact that a conditional expectation is both self-adjoint and idempotent. One can easily derive that in tracial settings, any self-adjoint and idempotent operator is conversely a conditional expectation. Intuitively, the conditional expectation is the restriction from $\mathcal{M}$ to $\mathcal{N}$. Explicitly, we may write finite-dimensional conditional expectations in the form

$$\mathcal{E}_{\mathcal{N}}(\rho) = \oplus_i (\mathrm{tr}_{m_i}(P_i \rho P_i) \otimes \frac{\hat{1}_{m_i}}{m_i}) \; , \tag{2.3}$$

where ρ is an input density. Here P_i is the projection on to the ith block with $P_i \rho P_i \in M_{n_i} \otimes M_{m_i}$, tr_{m_i} is the partial trace on M_{m_i} and $\hat{1}_{m_i}$ is the identity matrix in M_{m_i}. We denote by $\mathcal{N}'$ the commutant of the subalgebra, which is another von Neumann algebra consisting of all $a \in \mathcal{M}$ such that $[a, b] = 0$ for all $b \in \mathcal{N}$. A finite-dimensional commutant has explicit form

$$\mathcal{N}' = \oplus_i (\mathbb{C}1_{n_i} \otimes \mathbb{M}_{m_i})$$

with conditional expectation given by

$$\mathcal{E}_{\mathcal{N}'}(\rho) = \oplus_i (\frac{\hat{1}_{n_i}}{n_i} \otimes \mathrm{tr}_{n_i}(P_i \rho P_i)) \; .$$

The center of a von Neumann algebra is the subalgebra that commutes with all elements in the original algebra. This corresponds to the block sum in i in Equation (2.2). When a von Neumann algebra has a trivial center such that there is only one value of i in Equation (2.2), it is called a *factor*. Physically, we will often associate a factor with the operators on a quantum system, and a commutative algebra as a classical system modeled as a probability space.

Given two von Neumann subalgebras $\mathcal{S}, \mathcal{T} \subseteq \mathcal{M}$, we denote by $\mathcal{S} \vee \mathcal{T}$ the joint algebra generated by the union of their elements, which is itself a von Neumann algebra. We denote by $\mathcal{S} \cap \mathcal{T}$ the intersection of elements, again a von Neumann algebra. It is generally true that $\mathcal{S} \cap \mathcal{T} \subseteq \mathcal{S}, \mathcal{T} \subseteq \mathcal{S} \vee \mathcal{T} \subseteq \mathcal{M}$. In Chapter 3, we will explore the consequences of whether $[\mathcal{E}_{\mathcal{S}}, \mathcal{E}_{\mathcal{T}}] = 0$, a condition known as a commuting square.

A simple and canonical example of subalgebras is that of tensor products between two systems. Let A, B denote a pair of quantum systems, associated respectively with the Hilbert spaces $\mathcal{H}_A, \mathcal{H}_B$, Schatten

1-norm classes $S(A) \equiv S(\mathcal{H}_B), S(B) \equiv S(\mathcal{H}_B)$, and bounded operator sets $\mathbb{B}(A) \equiv \mathbb{B}(\mathcal{H}_A), \mathbb{B}(B) \equiv \mathbb{B}(\mathcal{H}_B)$. We denote by $\mathcal{A} = \mathbb{B}(A)$ the algebra of operators on A, and by $\mathcal{B} = \mathbb{B}(B)$ the corresponding algebra on B. Here the joint algebra $\mathcal{A} \vee \mathcal{B}$ corresponds to $\mathbb{B}(A \otimes B) \equiv \mathbb{B}(A) \otimes \mathbb{B}(B)$. The conditional expectation $\mathcal{E}_\mathcal{A}$ is such that

$$\mathcal{E}_\mathcal{A}(\rho) = \mathrm{tr}_B(\rho) \otimes \frac{\hat{1}^B}{|B|} ,$$

assuming that the algebra supports a partial trace. Similarly, $\mathcal{E}_\mathcal{B}$ traces out A and replaces it by complete mixture. Even in infinite-dimensional settings that lack a finite trace (or in some cases any trace), conditional expectations to subfactors may have the interpretation of restriction to spatially or otherwise separated subsystems.

The other main example of a conditional expectation is given by a possibly incomplete measurement. First, we recall the form of von Neumann measurement $M_{\{|i\rangle\}}$ as defined in Equation (2.1) and note that this is a conditional expectation followed by a classical copying operation. It restricts to the commutative algebra of diagonal matrices in the given $\{|i\rangle\}$ basis. A more general notion of measurement first applies an arbitrary conditional expectation, and then copies the diagonal part indexed by "i" in Equation (2.2) to a classically probabalistic bit as stored in a classical register. Formally, we define $M_\mathcal{N} : \mathcal{M} \to \mathcal{N} \otimes l_1^c$ for a subalgebra $\mathcal{N} \subseteq \mathcal{M}$ by

$$M_\mathcal{N}(\rho) \equiv \oplus_i (\mathrm{tr}_{m_i}(P_i \rho P_i) \otimes \frac{\hat{1}_{m_i}}{m_i} \otimes |i\rangle\langle i|) ,$$

where $i = 1...c$, and c is the number of possible outcomes of the classical register. When $\mathcal{E}_\mathcal{N}$ is direct sum of projectors to the eigenspaces of a given observable operator, we refer to it as a *pinching* map. Equivalently, pinching removes the off-diagonal elements of a density matrix, and it is an example of a complete dephasing process. Unlike measuring in a complete basis, this form of measurement allows us to leave some subspaces untouched, as though we measure a subset of the system's aspects, and then decide based on those outcomes whether to measure more. It also allows for conditional partial traces, which go beyond the usual concept of a "measurement." $\mathcal{M}_\mathcal{N}$ is still not the most general form of measurement considered in the mainstream quantum information literature, which is the positive operator-value measurement (POVM). $M_\mathcal{N}$ will however suffice for almost all measurements considered in this thesis, and it fits the subalgebra theme.

Conditional expectations commonly arise as uniform averages over groups. Let μ be the *Haar measure* for a unitary subgroup $G \subseteq U(\mathcal{H})$. The Haar measure is the unique measure that is invariant under any group action. In many settings, it is a uniform measure that assigns equal weight to each group element.

By G' we denote the matrices that commute with all $u \in G$, which form a von Neumann algebra. Then

$$\mathcal{E}_{G'}(\rho) = \int_{g \in G} g \rho g^{\dagger} \mu(g) \ . \tag{2.4}$$

The prior example of the partial trace is a uniform average over the unitary group on the traced subsystem. Likewise, the pinching map is a uniform average over the unitaries applying a state-dependent phase in one basis. As discussed further in Section 4.2, when G is a symmetry group (or a representation of one), we may interpret $\mathcal{E}_{G'}$ as the projection to the symmetric subalgebra.

2.3.2 von Neumann Algebras on Zero, One or Two Bits or Qubits

To gain some intuition for von Neumann algebras, we will examine some of the simplest cases.

The complex numbers, denoted $\mathbb{C}$, form a one-dimensional von Neumann algebra, which is equivalent to the algebra of complex, 1×1 matrices, $\mathbb{M}_1(\mathbb{C})$. We may think of this as the algebra of operators on a one-dimensional Hilbert space l_2^1, containing a single state. Physically, it represents the usually trivial transformations of applying a global phase to states, and scaling the eigenvalues of observables (e.g. changing units). We may also think of $\mathbb{C}$ as the algebra generated by the identity matrix, $\hat{1}$, in any dimension. As all finite-dimensional von Neumann algebras contain the identity, $\mathbb{C}$ often appears when we take an operation or intersection that reduces all observables to the trivial case of full degeneracy, and reduces all states to the trivial case of complete mixture. We may think of $\mathbb{C}$ as the algebra of operators on zero qubits. While we might try to imagine $\mathbb{R}$ as the algebra on zero bits, this analogy is not useful enough to carry on further. We will generally use $\mathbb{C}$ to denote an algebra that is physically trivial.

We next consider an algebra of *commuting* 2×2 matrices. We may denote

$$\hat{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \ Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \ ,$$

where $\hat{1}$ is the identity matrix, and Z is the Pauli Z matrix. This algebra contains densities of the form

$$\rho = \begin{pmatrix} \rho_{11} & 0 \\ 0 & \rho_{22} \end{pmatrix} \cong \begin{pmatrix} \rho_{11} \\ \rho_{22} \end{pmatrix} \ ,$$

where $\rho_{11}, \rho_{22} \geq 0$, and $\rho_{11} + \rho_{22} = 1$. As shown, we may identify densities in this algebra with probabilities in l_1^2. This is the mathematical representation of one stochastic bit. The algebras generated by the different Pauli matrices are isometric, so $\langle X \rangle \cong \langle Y \rangle \cong \langle Z \rangle$. By convention, the Z eigenbasis is usually considered the *computational basis*, or the default basis for writing qubit states.

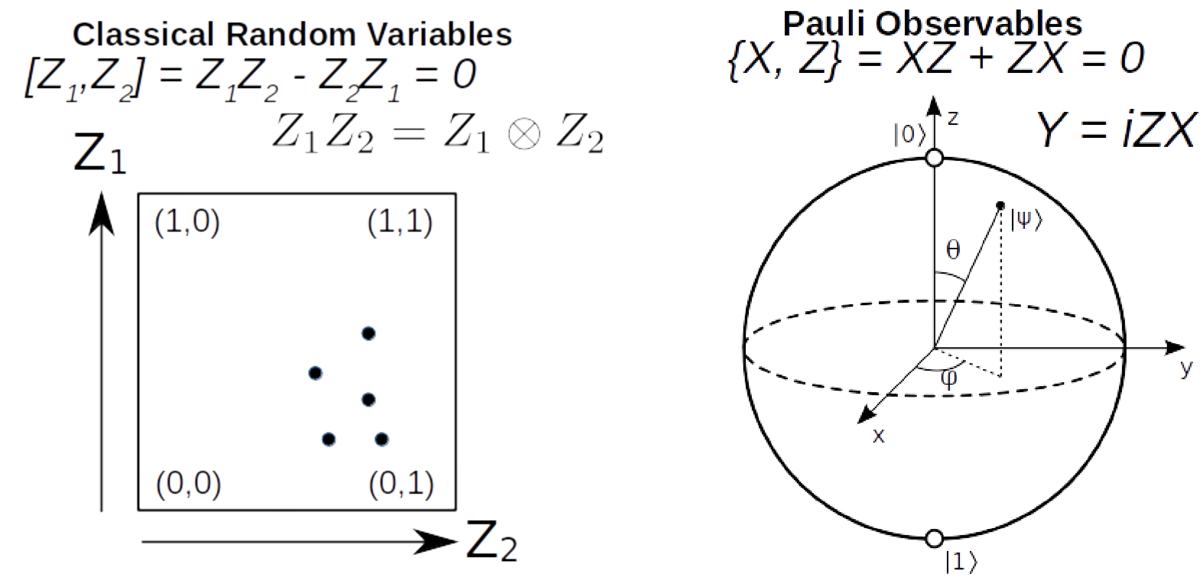

Figure 2.1: **Left:** In classical systems, a pair of observables Z_1, Z_2 will always commute. We can visualize the state of a single, probabalistic, binary variable as a point on the line $[0, 1]$, its position corresponding to the probability of observing each outcome. An uncorrelated state of two such observables is a point in the square formed by two such lines. We may weight multiple points to express correlated states, which are convex combinations of product states. **Right:** the Bloch sphere visualizes the space of quantum states. Here we take two observables that anticommute. The spherical geometry extends into three dimensions but removes the corners, which would represent points of impossibly high certainty in both incompatible observables. We are restricted to single points - a convex combination of points merely reduces to their midpoint. This reflects the fact that there is no classical notion of a correlated state between bases of the same qubit. Bloch sphere image by Smite-Meister - Own work, CC BY-SA 3.0, `https://commons.wikimedia.org/w/index.php?curid=5829358`

There are two basic ways to extend from $\langle Z\rangle$ to an algebra with two generators. The first of these is to take a pair of commuting generators Z_1, Z_2, which we may represent on $\mathbb{M}_4$ by

$$\hat{1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \; Z_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \; Z_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

We may associate densities in $\langle Z_1, Z_2\rangle$ with probabilities in l_1^4, corresponding to two stochastic bits. These algebras naturally support correlated states. For example, we may interpret the probability vector $(1/2, 0, 0, 1/2)$ as representing a random bit pair that is either in state 00 or 11, each with 50% probability, but definitely not in state 01 or 10. We have that $\langle X_1, X_2\rangle \cong \langle Y_1, Y_2\rangle \cong \langle Z_1, Z_2\rangle$.

Alternatively, we may take the algebra $\langle X, Z\rangle$, where X is the Pauli matrix given by

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} .$$

$\langle X, Z\rangle$ is $\mathbb{M}_2(\mathbb{C})$, the algebra of 2×2 complex matrices. It is equivalent to $\mathbb{B}(l_2^2)$, the algebra of bounded operators on 2-dimensional Hilbert space. $\langle X, Z\rangle$ is a factor in the von Neumann algebra sense. Compared with the classical case, we replace the commuting operators $[Z_1, Z_2] = 0$ by the anticommuting operators $\{X, Z\} = 0$. Instead of adding another stochastic bit, adding X turns our original stochastic bit into one qubit. In Figure 2.1, we visualize the algebras $\langle Z_1, Z_2\rangle$ and $\langle X, Z\rangle$. A key feature of the quantum algebra $\langle X, Z\rangle$ is the uncertainty principle: because X and Z neither commute nor contain any non-trivial, non-commuting sub-blocks, they cannot simultaneously have definite values when measured. This restricts the space of densities: it is for instance impossible to encode a state like $(1, 0, 0, 0) \in l_1^4(\mathbb{R})$ within a qubit density $\rho \in S_1^2$. On the other hand, XZ generates the Pauli Y. As visualized, $\langle X, Z\rangle$ therefore also contains states that would not appear in $\langle Z_1, Z_2\rangle$. In general, we may neither faithfully re-encode one qubit in two classical bits, nor compress two classical bits to one qubit.

A pinching in the eigenbasis of the Pauli Z operator projects $\langle X, Z\rangle$ to $\langle Z\rangle$. In this eigenbasis, it removes the off-diagonal elements of density matrices, transforming

$$\rho = \begin{pmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{pmatrix} \to \begin{pmatrix} \rho_{11} & 0 \\ 0 & \rho_{22} \end{pmatrix} .$$

Similarly, pinchings in the Pauli X and Y eigenbases project to corresponding stochastic bit algebras.

We may combine commuting with non-commuting operators, such as in the algebra $\langle X_1, X_2, Z_1, Z_2\rangle = \mathbb{M}_4(\mathbb{C})$, where $[X_1, X_2] = [Z_1, Z_2] = [X_1, Z_2] = [Z_1, X_2] = 0$, and $\{X_1, Z_1\} = \{X_2, Z_2\} = 0$. $\langle X_1, X_2, Z_1, Z_2\rangle$ is the algebra $\mathbb{B}(l_2^4)$ of bounded operators on two qubits, and it is a factor in the von Neumann algebra sense. In tensor product notation, $\langle X_1, X_2, Z_1, Z_2\rangle = \langle X_1, Z_1\rangle \otimes \langle X_2, Z_2\rangle = \mathbb{B}(l_2^2) \otimes \mathbb{B}(l_2^2)$. $\langle X_1, X_2, Z_1, Z_2\rangle$ is the minimal algebra capable of supporting quantum entanglement between subsystems.

A partial trace may project the algebra $\langle X_1, X_2, Z_1, Z_2\rangle$ back to $\langle X_1, Z_1\rangle$ or $\langle X_2, Z_2\rangle$. A pinching in both coordinates may yield $\langle Z_1, Z_2\rangle$, $\langle X_1, X_2\rangle$, or $\langle Y_1, Y_2\rangle$. We may also consider a Z-basis pinching in the 1st coordinate, yielding the algebra $\langle Z_1, X_2, Z_2\rangle = \langle Z_1\rangle \otimes \langle X_2, Z_2\rangle$, which contains a stochastic bit and a qubit. $\langle Z_1, X_2, Z_2\rangle$ supports correlated states as does $\langle Z_1, Z_2\rangle$, but not entanglement. Densities in $\langle Z_1, X_2, Z_2\rangle$

have the form

$$\rho = \begin{pmatrix} \rho_{11} & \rho_{12} & 0 & 0 \\ \rho_{21} & \rho_{22} & 0 & 0 \\ 0 & 0 & \rho_{33} & \rho_{34} \\ 0 & 0 & \rho_{43} & \rho_{44} \end{pmatrix},$$

which is a non-trivial example of the block diagonal form of von Neumann algebras.

Following [36], we may define an n-qubit algebra as any generated by operators $X_1, ..., X_n, Z_1, ..., Z_n$ such that $\{X_i, Z_i\} = 0$ for any $i \in 1...n$, and such that all other pairs commute. We may exchange any X_i or Z_i for Y_i while generating the same algebra. This algebraic construction of an n-qubit system is equivalent to the usual n-fold tensor product $\otimes_{i=1}^{n} \mathbb{B}(l_2^2(\mathbb{C})) \cong \mathbb{B}(l_2^{2^n}(\mathbb{C}))$. Finite-dimensional factors are always algebras of bounded operators on finite-dimensional Hilbert space. The idea of constructing algebras first, however, turns out to be important for infinite-dimensional systems.

2.3.3 A Brief (P)review of Types of von Neumann Algebras

So far, we have mostly considered finite-dimensional algebras of bounded operators on Hilbert space, $\mathbb{B}(l_2^d)$. Many basic examples of quantum systems are however infinite-dimensional, including the particle in potential well, and the harmonic oscillator. I will not directly address the case of the free particle, but for this I refer the reader to Strocchi's text [34]. Infinite-dimensional quantum systems are also at the heart of quantum field theory, which connects information theory to the physics of spacetime and matter. Different ways to construct infinite-dimensional von Neumann algebras result in distinct algebras and physical predictions. Classifying infinite-dimensional von Neumann algebras remains an area of active research. In this section, I briefly review a few of the constructions, highlighting cases in which unexpected physics may emerge. I roughly follow the treatment of Witten [37].

All finite-dimensional von Neumann algebras are type I. We refer to von Neumann algebras containing a trace (generalizing the matrix trace from linear algebra) as *tracial*, all of which are of type I or II. Type III von Neumann algebras are necessarily infinite-dimensional and lack a trace. While algebras of type III may appear bizarre from a traditional information theory perspective, they appear relevant and physically motivated in quantum field theory.

Type I

The von Neumann algebra community refers to finite-dimensional factors as type I_d, where d is the dimension of the underlying Hilbert space. These are always of the form $\mathbb{B}(l_2^d)$. Algebras of type I_∞ arise by taking

$d \to \infty$. In type I_∞, the identity operator has infinite trace. The eigenvalues of an operator should form a convergent series for that operator to have finite trace, and similarly, elements of l_2^∞ must have a convergent series of entries to be normalizable. As they have discrete energy spectra, the particle in a potential well and quantum harmonic oscillator yield algebras of type I_∞.

Type II

VNas of type II may have a finite trace but are not limit points of type I algebras. Formally, they are distinguished from type I by the lack of minimal projections. In type I, we may project to subspaces (e.g. of dimension 1) having no smaller, non-zero subspaces. This fails in type II. Unlike their type I_∞ counterparts, a type II vNa may have a trace that is finitely-valued on the identity element. Type II algebras can be more specifically of type II_1 or of type II_∞, where the latter is a type II_1 algebra tensored with the algebra of bounded operators on another infinitely-dimensional Hilbert space. Quantum systems described by type II algebras generally don't support decompositions into tensored subsystems, Schmidt decompositions of pure states, or other common constructions from finite-dimensional quantum information.

Instead of taking $d \to \infty$ for a d-dimensional Hilbert space, one could take a tensor product of Hilbert spaces $\mathcal{H}_1 \otimes ... \otimes \mathcal{H}_k$ and the limit $k \to \infty$. Naively, this produces a vector space of uncountably infinite dimension containing non-normalizable vectors, and it will not yield a von Neumann algebra.

A modification of the infinite tensor product yields the algebra known as the *hyperfinite II_1 factor*, as shown by Araki [38] and later summarized by Witten [37]. The matrix algebra $\mathbb{M}_2$ is a Hilbert space $\mathcal{H}$ with $\langle a|b\rangle = \mathrm{tr}(a^\dagger b)$ for any $a, b \in \mathbb{M}_2$. Let $\mathcal{M} \cong \mathbb{B}(l_2^2)$ be the algebra of left matrix multiplications, and $\mathcal{M}' \cong \mathbb{B}(l_2^2)$ be that of right multiplications, noting that $\mathcal{M}$ and $\mathcal{M}'$ indeed commute. $\mathcal{H} \cong l_2^2 \otimes l_2^2$, where $\mathcal{M}$ and $\mathcal{M}'$ act respectively on each qubit. The identity $\hat{1} \in \mathbb{M}_2$ plays the role of the completely entangled state between qubit Hilbert spaces. We then take vectors of the form

$$v_1 \otimes v_2 \otimes ... \otimes v_k \otimes ... \in \mathcal{H} \otimes \mathcal{H} \otimes ..., \tag{2.5}$$

such that $v_i = \hat{1}$ for all $i > k$. Given two such vectors v, w, there will be some k satisfying the aforementioned condition for both. This allows us to define the inner product via truncation, as $\langle v|w\rangle = \mathrm{tr}((v_1 \otimes ...v_k)^\dagger w_1 \otimes ...w_k)$ in $\mathcal{H}^{\otimes k}$, which is unaffected if k is taken larger than needed. With the inner product defined, we may complete the Hilbert space formed by limit points of the linear span of these vectors. Construction of operators in the corresponding algebra follows a similar pattern. We start with operators of the form

$$a_1 \otimes ... \otimes a_k \otimes ... \in \mathcal{M} \otimes \mathcal{M} \otimes ... \ ,$$

such that $a_i = \hat{1}$ for all $i > k$. We then take combinations and limits of such operators, which form a von Neumann algebra $\mathcal{A}$. There is a corresponding algebra $\mathcal{A}'$ formed analogously from tensor products of operators in copies of $\mathcal{M}'$. The trace of an element $a \in \mathcal{A}\mathcal{A}'$ is given by $\mathrm{tr}(a) = \langle \hat{1} | a | \hat{1} \rangle$, where $a = \hat{1} \otimes \hat{1} \otimes \ldots$. Here we see that the identity has trace 1. Hence we have constructed the hyperfinite II_1 factor.

The word "hyperfinite" comes from the fact that the aforementioned II_1 factor arises as an infinite limit of finite-dimensional constructions. There are non-hyperfinite II_1 factors. The Connes Embedding Problem asked how II_1 factors in general would relate to the hyperfinite II_1 factor, and turned out to be equivalent to whether correlations that might exist between sets of observables in (infinite-dimensional) commuting algebras would all be limit points of finite-dimensional correlation sets [3]. A group recently claimed to resolve the Connes Embedding problem in the negative [39] using complexity-theoretic methods.

Type III

Type III vNas lack a (non-trivial) trace, and in this sense they are the furthest from matrix algebras. One may construct a hyperfinite type III_λ algebra for $\lambda \in (0,1), \lambda \neq 1/2$ by a simple modification of the hyperfinite type II_1 algebra: instead of requiring all v_i in Equation (2.5) for $i > k$ to be $\hat{1}$, we require them to approach some normalized, 2×2, diagonal matrix with eigenvalues proportional to λ and $1 - \lambda$. In fact, we may think of the hyperfinite type II_1 factor as what we would obtain if we took a hyperfinite III_λ construction with $\lambda = 1/2$. Hyperfinite type III_1 algebras arise from cycling between several values of λ with several distinct limit points.

One way to intuit type III_λ algebras is that the object replacing a trace has some exponentially decaying weighting. We may relate the parameter λ to the temperature of a thermal state. The absence of a trace blocks many standard, linear-algebraic techniques for proving inequalities. Classifying type III algebras remains open and complicated, and these algebras have many properties that would seem exotic to quantum information theory. Like algebras of type II, algebras of type III generally do not decompose into tensored subsystems, and correlation measures such as entanglement entropy diverge. Nonetheless, quantum field theory appears to take place in type III_1. For more on the field theory applications of type III, we refer the curious reader to Witten's review [37].

2.4 Entropy

Entropy is the building block of information theory. The thermodynamic origin of entropy first connected information with physics, even before Shannon's work on communication [9] created the field of information theory as it is known today. Traditional treatments of entropy often assume a large number of events with identical statistical properties. This assumption will turn out to be of great value in what I call the Shannon-theoretic methods, which use the assumption of many trials to reduce deviations. Thermodynamically, this has an analog in the ergodic principle, which assumes that a thermal system will sample all or most possible configurations. Sometimes, the practical implications of Shannon-theoretic methods become questionable, as one cannot always rely on an asymptotically large number of instances without accepting the costs of scaling. Other ways of arriving at entropy appear in the norms of Banach spaces and in operator algebra theory. I try to keep these approaches on equal footing, seeing the connections between formulations. Often one approach reveals strategies of proof that would be extremely subtle or technical in others.

The convention for most of this thesis is to take entropy with respect to the natural logarithm, as this leads to major simplifications in a multitude of calculations, in return for a few formulas needing adjustments before comparison with results from computer science. When denoting entropies of e.g. single qubits, I will sometimes explicitly specific units of bits for entropies or ebits for entanglement measures, for which a single bit is equivalent to $\log 2 \approx 0.69314718056$ nats of unitless entropy. I also use the convention that H is the symbol for von Neumann or Shannon entropy, which is common in mathematical papers, rather than S as is common in information theory. Both letters H and S are unfortunately used for many other quantities, though it should usually be clear from context what the letter denotes.

2.4.1 Relative Entropy

Mark Wilde, author of the keystone text "From Classical to Quantum Shannon Theory" (published under the title "Quantum Information Theory [40]), describes the *relative entropy* as the "mother of all entropies" (see a presentation on recoverability, [41]). Vlatko Vedral's review of relative entropy [13] also discusses in detail its many uses in quantum information. Special cases of relative entropy include the von Neumann and Shannon entropies, conditional entropy, and mutual information (all of which I will recall subsequently). Hence I start with relative entropy as the root. The mathematical form of relative entropy, $D : S(\mathcal{H}) \times S(\mathcal{H}) \to \mathbb{R}^+$, is given in finite dimension by

$$D(\rho \| \phi) = \mathrm{tr}(\rho \log \rho - \rho \log \phi) \ . \tag{2.6}$$

Relative entropy is $+\infty$ when the support of ρ is not contained within that of ϕ, in which case there is no 0 to cancel the divergence in the 2nd logarithm. Relative entropy was originally introduced by Umegaki in 1962 [42]. While the form of Equation (2.6) is relatively simple in tracial von Neumann algebras, it is neither the most general nor always the most intuitive or useful way to express this quantity. Here I briefly review some forms of this entropy.

Relative Entropy as an Expectation in Shannon Theory

Given two densities ρ and ϕ, one might wonder if the expression $-\mathrm{tr}(\rho \log \phi)$ has any intuitive meaning. We recall as reviewed in Section 2.2 that $\langle \mathcal{O} \rangle_\rho = \mathrm{tr}(\mathcal{O}\rho)$. As $-\log \phi$ is Hermitian, we may interpret it as an observable, and $-\mathrm{tr}(\rho \log \phi) = \langle -\log \phi \rangle_\rho$. Since this section is intended to be more of an intuitive summary than a result in itself, we will assume for now that ρ and ϕ have non-degenerate eigenvalues, and we will focus on finite-dimensional densities.

Recall that ϕ generally has some eigenbases in which it is diagonal, and denote by $\mathcal{E}_{\phi'}$ the associated pinching map (in particular, we interpret ϕ' as the algebra of matrices that commute with ϕ). Obviously, $E_{\phi'}(\phi) = \phi$. By the properties of conditional expectations, we then have that $-tr(\rho \log \phi) = -\mathrm{tr}(\mathcal{E}_{\phi'}(\rho) \log \phi)$, allowing us to consider the entire expression in the diagonal basis of ϕ. Hence we may consider this term as though it were in terms of two classical random variables. Now we interpret $\log \phi$ in ϕ's diagonal basis. In dimension d, we may consider ϕ as though it describes a random process outputting eigenstates with probability given by its eigenvalues.

Since eigenstates of ϕ are orthogonal, we may consider each to be a distinct "message" in the sense of classical information theory. Formally, a message is indexed by the numbers $1...d$. As information theory addresses statistics rather than content of messages, the index is enough description. A message could be the state of a quantum system, a command transmitted by phone, the outcome of a die roll, etc - information theory considers the probability of each outcome, rather than its manifestation. We may then consider an alphabet of symbols, such as binary bits or English letters, and want to assign a string of said symbols (known as a *codeword*) to encode each message. A common goal in Shannon theory is to determine the assignment that minimizes the average codeword length for a given probability distribution on messages. The optimal average codeword length turns out to be proportional to the negative logarithm of the probability of each message (see Thomas & Cover for a review of this subject [43], or Shannon's paper [9] for the original *noiseless coding theorem*). Let $\mathcal{O}_{\mathrm{code}\ \phi} = -\log \phi$ denote the observable that is the codeword length in an optimal code for each ϕ eigenstate. Hence:

Remark 2.1. *Let ϕ and ρ be densities. Were one to know the probability distribution given by ϕ in advance and prepare an optimal encoding according to it, but then get messages distributed according to $\mathcal{E}_{\phi'}(\rho)$, one achieves the average codeword length $-tr(\rho \log \phi) = \langle \mathcal{O}_{code\ \phi} \rangle_\rho$. Hence*

$$D(\rho \| \phi) = \langle \mathcal{O}_{code\ \phi} \rangle_\rho - \langle \mathcal{O}_{code\ \rho} \rangle_\rho \ .$$

It is somewhat intuitive why $D(\rho\|\phi)$ is positive, as $\mathcal{O}_{code\ \rho}$ minimizes the average codeword length for ρ.

This intuition hides many important and often vexing details when ρ and ϕ don't commute. The "optimal codeword length" operator makes sense in the diagonal basis of ρ or ϕ, but its interpretation becomes fuzzier when we take its expectation with a density that is diagonal in a different basis. In deriving inequalities, the matrix logarithm is much harder to work with than the logarithm of a classical vector. Simple ideas from classical Shannon theory often require deep mathematics in the quantum setting.

Tomita-Takesaki Relative Entropy

Let $\mathcal{M}$ be a von Neumann algebra on Hilbert spaces $\mathcal{H}$ and $\tilde{\mathcal{H}}$ (taking $\mathcal{H} = \tilde{\mathcal{H}}$ is common but not strictly necessary). We note here that for a normalized density $\rho \in S_1(\mathcal{H})$, $\rho^{1/2}$ is a normalized vector in the matrix Hilbert space with inner product $\langle \rho | \phi \rangle = \mathrm{tr}(\rho^\dagger \phi)$, and we may identify it with a vector in $\mathcal{H}$ or $\tilde{\mathcal{H}}$. In this sense we may write $|\phi^{1/2}\rangle$ and use the usual braket notation and inner product. Let ϕ, ρ be densities or states on $\mathcal{M}$ such that $|\phi^{1/2}\rangle \in \mathcal{H}$, $|\rho^{1/2}\rangle \in \mathcal{H}'$, and $|\phi^{1/2}\rangle$ is:

1. Cyclic, in that $\{a \, |\phi^{1/2}\rangle : a \in \mathcal{M}\}$ is dense in $\mathcal{H}$.
2. Separating, in that if $a \in \mathcal{M}$ and $a \, |\phi^{1/2}\rangle = 0$, then $a = 0$.

The *Tomita-Takesaki operator* $S_{\phi,\rho}$, given by $S_{\phi,\rho} a \, |\phi^{1/2}\rangle = a^\dagger \, |\rho^{1/2}\rangle$ for any $a \in \mathcal{M}$, has polar decomposition

$$S_{\phi,\rho} = J_{\phi,\rho} \Delta_{\phi,\rho}^{1/2} \; .$$

where $J_{\phi,\rho}$ is the *relative modular conjugation* and $\Delta_{\phi,\rho}$ the *relative modular operator*. It's the latter we use here, writing

$$D(\rho \| \phi) = - \langle \rho^{1/2} | \log(\Delta_{\phi,\rho}) \rho^{1/2} \rangle \; .$$

This form of relative entropy does not rely on the presence of a trace, so it exists in field theories. In finite dimension, $\Delta_{\phi,\rho} = LR$, where $L(\eta) = \phi\eta$, and $R(\eta) = \eta\rho^{-1}$ for any input matrix η. High-energy physics often uses Tomita-Takesaki theory. Furthermore, the connection between modular theory and relative entropy was of great use in proving the data processing inequality that we study extensively in Chapter 3. In cases where the form in Equation (2.6) makes sense, it is equivalent to the Tomita-Takesaki form of relative entropy. For deeper treatments, see reviews by Summers [44], by Zhang and Wu [45], and by Witten [37].

Rényi Relative Entropy from Norms

Letting $\alpha \in [1, \infty]$, the sandwiched Rényi relative α-entropy is given by

$$D_\alpha(\rho\|\phi) = \alpha' \log \|\phi^{-1/2\alpha'} \rho \phi^{-1/2\alpha'}\|_\alpha, \tag{2.7}$$

or ∞ if $\mathrm{supp}(\rho) \not\subseteq \mathrm{supp}(\sigma)$, where $1/\alpha + 1/\alpha' = 1$. $D_\alpha(\rho\|\phi)$ was independently defined in [46] and [47]. It holds that $\lim_{\alpha \to 1} D_\alpha(\rho\|\phi) \to D(\rho\|\phi)$. The α-parameterized version appears in results summarized in Section 7.1.

There are several forms of quantum relative Rényi entropy [48]. Some forms of Rényi entropy avoid the requirements of asymptotically many copies [49], or characterize situations where it is not enough to succeed on average [50]. Sometimes it is easier to compute for $\alpha \neq 1$, which can yield estimates for or conjectures regarding the $\alpha = 1$ case. I focus on the sandwiched Rényi entropy here, as this is the form used in work related to this thesis.

2.4.2 von Neumann and Rényi Entropy

For a density ρ in a tracial algebra, the Rényi α-entropy is given by

$$H_\alpha(\rho) \equiv -D_\alpha(\rho\|\hat{1}) \ . \tag{2.8}$$

Since $\hat{1}$ is not a normalized density matrix, $D_\alpha(\rho\|\hat{1}) < 0$, so the entropy on the left positive. As $\alpha \to 1$, $H_\alpha(\rho) \to H(\rho) = -\mathrm{tr}(\rho \log \rho) = \langle \mathcal{O}_{\text{code } \rho} \rangle_\rho$, the von Neumann entropy. The Shannon entropy of a classical probability distribution is the von Neumann entropy of a diagonal density with entries according to that distribution. For a bipartite system $A \otimes B$, we recall the sandwiched conditional α-entropy of A conditioned on B as

$$H_\alpha(A|B)_\rho \equiv \inf_{\sigma_B} D_\alpha(\rho^{AB} \| 1_A \otimes \sigma_B). \tag{2.9}$$

As $\alpha \to 1$, $H_\alpha(A|B)_\rho \to H(A|B)_\rho = H(\rho^{AB}) - H(\rho^B)$. Following the notation for conditional entropy, we denote by $H(A)_\rho \equiv H(\rho^A)$ the entropy of ρ's restriction to subsystem A. By $H(\mathcal{A})_\rho \equiv H(\mathcal{E}_\mathcal{A}(\rho))$, we denote the entropy of ρ's conditional expectation to the $\mathcal{A}$ subalgebra. As the partial trace to the subsystem A effectively erases other systems so that they contribute no entropy, while the conditional expectation leaves them in complete mixture,

$$H(\mathcal{A})_\rho = H(A)_\rho + \log |A^c| \ ,$$

where A^c is the complement of the A system. In general, when $\mathcal{N} \subseteq \mathcal{M}$ is a von Neumann subalgebra, and $\rho \in \mathcal{M}_*$, we denote $H(\mathcal{N})_\rho \equiv H(\mathcal{E}_\mathcal{N}(\rho))$. We recall the mutual information

$$I(A:B)_\rho \equiv H(A) + H(B) - H(AB) = D(\rho \| \rho^A \otimes \rho^B) \ , \tag{2.10}$$

and with an extra system C, the conditional mutual information given by

$$I(A:B|C)_\rho \equiv H(AC) + H(BC) - H(ABC) - H(C) = D(\rho \| \exp(\log \rho^{AC} + \log \rho^{BC} - \log \rho^C)) \ . \tag{2.11}$$

In Chapter 3, we generalize the conditional mutual information to subalgebras.

2.5 Distance Between Densities

The main distance measures used in this thesis are p-norm distances, given for normalized densities ρ and ϕ as $\|\rho - \phi\|_p / \sqrt[p]{2}$ (up to normalization conventions). The 2-norm distance is the usual Euclidean distance often used in rotation-invariant vector spaces. More commonly applied to densities is the 1-norm distance (also known as *trace distance*), which is equal to the probability that an optimal set of measurements distinguishes the two densities.

Entropy is an extensive property on independent systems (on correlated systems this is more complicated, see Figure 4.2) with range in the positive reals. The relative entropy is not symmetric in its arguments, so it is not truly a notion of distance. The p-distance ranges from 0 to 1. Relative entropy is nonetheless often taken to characterize an entropic distance-like notion between densities. The relative entropy is related to the 1-distance by Pinsker's inequality, which states

$$D(\rho \| \phi) \geq \frac{1}{2} \|\rho - \phi\|_1^2 \ . \tag{2.12}$$

Inequalities in the other direction are more complicated (largely due to the unboundedness of relative entropy) but exist [51, 52]. Another extremely useful class of inequalities are known as the Fannes-Audenart or Alicki-Fannes inequalities. The form we will use is a recent refinement by Winter [53] for a bipartite system with two densities ρ and ϕ:

$$\frac{1}{2} \|\rho^{AB} - \sigma^{AB}\|_1 \leq \epsilon \implies |H(A|B)_\rho - H(A|B)_\sigma| \leq 2\epsilon \log |A| + (1+\epsilon) h(\frac{\epsilon}{1+\epsilon}) \ , \tag{2.13}$$

where $h(p) = -p \log_2 p - (1-p) \log_2 (1-p)$ is the binary entropy function.

2.6 Open & Time-Evolving Quantum Systems

The common description of time-evolution in a quantum system is the Schrödinger equation, given as

$$i\hbar\frac{\partial}{\partial t}|\psi(t)\rangle = \mathfrak{H}|\psi(t)\rangle \tag{2.14}$$

for a quantum pure state $|\psi(t)\rangle$. Here $\mathfrak{H}$ is the Hamiltonian, the observable operator corresponding to the system's energy. The effect of this time-evolution for a fixed amount of time t is to apply the unitary matrix $U_t = \exp(-iHt)$ to the state by left multiplication $|\psi\rangle \to U|\psi\rangle$. For a density ρ, $\rho \to U\rho U^\dagger$. One may equivalently apply the adjoint matrix to observables, $\mathcal{O} \to U^\dagger \mathcal{O} U$ in the Heisenberg picture. By the Stone von Neumann theorem, every finite-dimensional unitary is generated by some Hamiltonian in finite time.

The Schrödinger equation describes the evolution of closed quantum systems, so pure states say pure, and all unitaries are invertible. In fields like condensed matter and high-energy, this is often a reasonable assumption. In quantum information, however, exposure to environment is arguably the most important challenge. The mathematical formulation of an open quantum process (when there is no initial system-environment correlation) is the *quantum channel*: a completely positive, trace-preserving map $\Phi : S(\mathcal{H}_A) \to S(\mathcal{H}_B)$, for input system A and output B. The physical intuition for quantum channels as open processes comes from the *Stinespring dilation*: any channel Φ has the form

$$\Phi(\rho) = \mathrm{tr}_E(U\rho U^\dagger) ,$$

where $U : A \to BE$ is an isometry. We may think of a channel as attaching an initial environment, time-evolving unitarily, and then tracing out the final environment (which need not always be the same system as the initial environment). In finite-dimension, it is always sufficient to take a pure initial environment, and the size of environment needed to implement any quantum channel is at most $|A||B|$. Furthermore, all finite-dimensional Stinespring dilations are equivalent up to partial isometries, which are norm-preserving maps, and all minimal Stinespring dilations are equivalent up to unitaries on the environment. I illustrate the Stinespring dilation in Figure 2.2. Infinite-dimensional analogs exist but may involve different constructions.

Analogous to the Stinespring dilation, any finite-dimensional density is the marginal of a pure state on a larger system. For a density ρ^A on system A, $\rho = \mathrm{tr}_{A'}(|\rho^{1/2}\rangle\langle\rho^{1/2}|^{AA'})$, where $|A'| = |A|$. Writing ρ

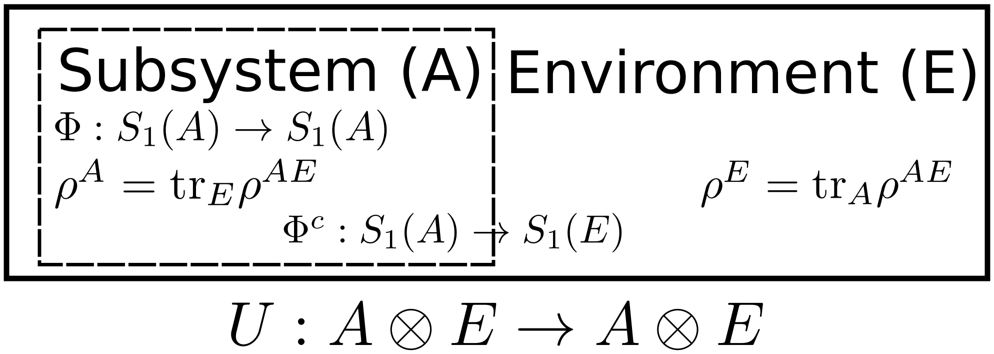

Figure 2.2: A diagram of the Stinespring dilation of finite-dimensional quantum systems. For simplicity, we assume that the channel, Φ, maps densities in system A to other densities in A.

explicitly as a convex combination of pure states,

$$\sum_i \rho_i |i\rangle\langle i|^A = \mathrm{tr}_{A'} \Big(\Big(\sum_i \sqrt{\rho_i}\, |i\rangle^A \otimes |i\rangle^{A'} \Big) \times h.c. \Big) .$$

Purification and Stinespring dilation allow formulations of open processes on mixed states in terms of unitary evolution on pure states, if desired.

2.6.1 Quantum Markov Semigroups

Though the quantum channel often describes a process occurring over time, it hides the passing of time in its internals. Given a quantum channel Φ occurring over a time $t \in \mathbb{R}^+$, one does not necessarily know the channel for another time length $s \neq t$. Physically, this is because the system may have some backaction on its environment. The process that occurs in time $2t$, for instance, is not assured to be equivalent to $\Phi \cdot \Phi$, since the environment in the later half may store pre-existing correlations with the input system. In many cases, however, physicists assume that the environment immediately dissipates such correlations and resets itself to an initial state. Mathematically, this allows us to more explicitly model the channel in terms of time, parameterizing it as Φ^t. Furthermore, we obtain the crucial semigroup property, $\Phi^t \cdot \Phi^s = \Phi^{t+s}$ for all $t, s \in \mathbb{R}^+$. A family of channels Φ^t for $t \in \mathbb{R}^+$ with the semigroup property is called a *quantum Markov semigroup* (QMS).

A QMS will have a *Lindbladian generator* $\mathcal{L}$ [54], so that $\Phi^t(\rho) = \partial\rho/\partial t = -\mathrm{ad}_{\exp(-t\mathcal{L})}(\rho)$ is the solution to a differential equation given by

$$\frac{\partial}{\partial t}(\rho) = -\mathrm{ad}_{\mathcal{L}}(\rho) .$$

This structure mirrors that of the Schrödinger equation, but instead of generating a group of unitaries, $\mathcal{L}$

generates a semigroup of usually non-invertible quantum channels. A usual physical scenario is coupling to a large bath of interacting particles. I describe results on quantum Markov semigroups in Chapter 7. To simplify the notation, I will often write $\mathcal{L}(\rho) \equiv \mathrm{ad}_{\mathcal{L}}(\rho)$.

2.6.2 Adjoints and Recovery

In Hilbert space, the adjoint of a unitary U is given by the inner product formula $\langle \phi | U \psi \rangle = \langle \phi U^\dagger | \psi \rangle$ and in finite dimension is equal to its Hermitian conjugate. Because matrices are a Hilbert space under $\langle a | b \rangle = \mathrm{tr}(a^\dagger b)$, we can similarly define the adjoint $\Phi^\dagger$ of a quantum channel Φ by $\langle a | \Phi(b) \rangle = \langle \Phi^\dagger(a) | b \rangle$. In general, the adjoint of a quantum channel $\Phi : S(\mathcal{H}_A) \to S(\mathcal{H}_B)$ is a unital (maps the identity to the identity), completely positive map $\Phi^\dagger : \mathbb{B}(\mathcal{H}_B) \to \mathbb{B}(\mathcal{H}_A)$, reversing the direction.

The adjoint of a unitary matrix is its inverse. In contrast, quantum channels generally lack inverses. While a self-adjoint unitary is the identity, there are many non-trivial examples of self-adjoint channels. Conditional expectations are a special case of self-adjoint quantum channels. Self-adjointness is a mathematical property that we might deem analogous to having no coherent or rotational part. Other examples of self-adjoint channels include depolarization, which replaces a state by complete mixture, and dephasing, which reduces the off-diagonal components of a density (a completely dephasing channel is a pinching conditional expectation). Nonetheless, the fact that the adjoint of a quantum channel switches the input with output space, as well as its role on unitaries, suggests that it has some markings of an inverse-like operation.

While non-unitary channels need not have inverses, it is still possible and fruitful to study operations that partially reconstruct the input state from the output, known as *recovery maps*. Probably the most canonical recovery map is the Petz recovery map $R_{\omega,\Phi} : S(\mathcal{H}_B) \to S(\mathcal{H}_A)$, given by

$$R_{\omega,\Phi}(\rho) = \omega^{1/2} \Phi^\dagger \big((\Phi(\omega)^{-1/2}) \rho (\Phi(\omega)^{-1/2}) \big) \omega^{1/2} \ .$$

In addition to the original channel, the Petz recovery map contains the extra parameter ω, a density in $S(\mathcal{H}_A)$. $R_{\omega,\Phi}$ perfectly recover ω. Furthermore,

$$D(\rho \| \omega) = D(\Phi(\rho) \| \Phi(\omega)) \iff R_{\omega,\Phi} \cdot \Phi(\rho) = \rho \ . \tag{2.15}$$

More recently, approximate versions of Equation (2.15) have appeared for modifications of the Petz map [55, 56, 57].

Chapter 3

Generalizing Strong Subadditivity

Strong subadditivity, arguably the most central inequality in quantum information, states in its usual form that

$$H(AC)_\rho + H(BC)_\rho \geq H(C)_\rho + H(ABC)_\rho \tag{3.1}$$

for any tripartite density ρ^{ABC}, as proven by Lieb and Ruskai in 1973 [58]. In words, strong subadditivity states that the sum of entropies of AC and BC is at least as large as the sum of the entropy of the joint system ABC and intersection system C. Strong subadditivity is closely related and nearly equivalent to the *data processing inequality of relative entropy*,

$$D(\rho\|\sigma) \geq D(\Phi(\rho)\|\Phi(\sigma)) \tag{3.2}$$

for any densities ρ, σ and quantum channel Φ, as proven by Lindblad in 1975 [59]. A later form of strong subadditivity comes from Petz in 1991, stating

Theorem 3.1 (Petz SSA, theorem 12 from [60] by Dénes Petz)**.** *Let $\mathcal{M}$ be a C^*-algebra, $\mathcal{E}_\mathcal{S}, \mathcal{E}_\mathcal{T}$ be conditional expectations to subalgebras $\mathcal{S}, \mathcal{T} \subseteq \mathcal{M}$ in commuting square, ρ be any density on $\mathcal{M}$, and σ be a density on $\mathcal{M}$ such that $\mathcal{E}_\mathcal{T}(\sigma) = \sigma$. Then*

$$D(\rho\|\sigma) + D(\mathcal{E}_{\mathcal{S}\cap\mathcal{T}}(\rho)\|\mathcal{E}_{\mathcal{S}\cap\mathcal{T}}(\sigma)) \geq D(\mathcal{E}_\mathcal{S}(\rho)\|\mathcal{E}_\mathcal{S}(\sigma)) + D(\mathcal{E}_\mathcal{T}(\rho)\|\mathcal{E}_\mathcal{T}(\sigma)) \ .$$

(re-proof in finite dimension, from [23]). This follows almost immediately from the usual data processing inequality. First, since $\sigma = \mathcal{E}_\mathcal{T}(\sigma)$, $\rho(\log \mathcal{E}_\mathcal{T}(\sigma) - \log \sigma) = 0$, and similarly for $\mathcal{E}_\mathcal{S}(\sigma) = \mathcal{E}_\mathcal{S}\mathcal{E}_\mathcal{T}(\sigma)$. This allows us to reduce the form above to the entropy difference,

$$H(\rho^\mathcal{S}) + H(\rho^\mathcal{T}) - H(\rho) - H(\rho^{\mathcal{S}\cap\mathcal{T}}) \ .$$

This chapter includes results appearing in [23, 24, 25], co-authored with Li Gao and Marius Junge, and also includes results appearing in [26].

We use the form of relative entropy with respect to a conditional expectation to rewrite the expression as

$$D(\rho \| \mathcal{E}_{\mathcal{S}}(\rho)) - D(\mathcal{E}_{\mathcal{T}}(\rho) \| \mathcal{E}_{\mathcal{S} \cap \mathcal{T}}(\rho)) \, .$$

Then positivity follows from the usual data processing inequality on relative entropy. ■

Petz gives a more general proof based on modular theory, which we do not repeat here. The key concept within Petz's SSA if that of the *commuting square*. A pair of von Neumann subalgebras $\mathcal{S}, \mathcal{T} \subseteq \mathcal{M}$ are in commuting square iff $[\mathcal{E}_{\mathcal{S}}, \mathcal{E}_{\mathcal{T}}] = 0$, in which case $\mathcal{E}_{\mathcal{S}} \mathcal{E}_{\mathcal{T}} = \mathcal{E}_{\mathcal{T}} \mathcal{E}_{\mathcal{S}} = \mathcal{E}_{\mathcal{S} \cap \mathcal{T}}$. Another form, as shown in [23] with Gao and Junge, states that

$$D(\rho \| \mathcal{E}_{\mathcal{S}}(\rho)) + D(\rho \| \mathcal{E}_{\mathcal{T}}(\rho)) \geq D(\rho \| \mathcal{E}_{\mathcal{S} \cap \mathcal{T}}(\rho)) \tag{3.3}$$

for any $\mathcal{S}, \mathcal{T} \subseteq \mathcal{M}$ in commuting square. We also show that the commuting square condition is necessary as well as sufficient for SSA to hold for all ρ. An important conceptual special case of SSA for subalgebras is the uncertainty principle with quantum memory [61] for mutually-unbiased bases:

$$H(\mathcal{X}|C) + H(\mathcal{Z}|C) \geq \log |\mathcal{M}| + H(\mathcal{M}|\mathcal{C}) \tag{3.4}$$

for $\mathcal{X}, \mathcal{Z} \subseteq \mathcal{M}$ corresponding to a pair of bases such that $[\mathcal{E}_{\mathcal{X}}, \mathcal{E}_{\mathcal{Z}}] = 0$. As studied in [62], there is a closely related inequality for the relative entropy of coherence in maximally incompatible bases. The connection between Equations (3.4), and (3.1) suggests a correspondence between quantum correlations and quantum uncertainty, which we explore further in Chapter 4. The connection between strong subadditivity and entropic uncertainty relations might be relatively underappreciated in physics. There is an intuitive parallel between the classical entropic subadditivity of a pair of commuting variables and the quantum uncertainty relation for a pair of anticommuting variables. Petz's SSA shows how strong subadditivity and an uncertainty principle share not just a common form, but sometimes a common mathematical derivation. If SSA and data processing for relative entropy are two closely related forms of what is essentially the same idea, then the entropic uncertainty relation appears to be a third.

Applications of strong subadditivity and data processing range from quantum Shannon theory [40] to quantum field theory [37] and holographic spacetime [63]. A common challenge in quantum information theory is to find entropy sums for which positivity is generally true *without* following from SSA [64], though sometimes these are more forthcoming in more specific contexts such as when holographic correspondences hold [65]. There are several simplifications and reviews of SSA's proof, including by Nielsen and Petz in

2001 [66], and by Ruskai in 2007 [67]. Carlen and Lieb show a tightening in [68]. As shown by Isaac Kim in 2012, SSA extends to an operator inequality [69]:

$$\mathrm{tr}_{AC}(\rho(\log \mathcal{E}_{\mathcal{AC}} + \log \mathcal{E}_{\mathcal{BC}} - \log \mathcal{E}_{\mathcal{C}} - \log)(\rho)) \geq 0 \ ,$$

for any tripartite ρ^{ABC}, and where "≥ 0" denotes operator non-negativity.

Petz's SSA is equivalent to positivity of the quantity defined as follows:

Definition 3.1 (GCMI, as in [23]). *Let ρ be a density on C^*-algebra $\mathcal{M}$, and $\mathcal{S}, \mathcal{T} \subseteq \mathcal{M}$ be subalgebras in commuting square. Then we define the generalized conditional mutual information (GCMI) as*

$$\begin{aligned} I(\mathcal{S} : \mathcal{T} \subseteq \mathcal{M})_\rho &\equiv D(\rho \| \mathcal{E}_{\mathcal{T}}(\rho)) + D(\rho \| \mathcal{E}_{\mathcal{S}}(\rho)) - D(\rho \| \mathcal{E}_{\mathcal{S} \cap \mathcal{T}}(\rho)) \\ &= D(\rho \| \mathcal{E}_{\mathcal{T}}(\rho)) - D(\mathcal{E}_{\mathcal{S}}(\rho) \| \mathcal{E}_{\mathcal{S} \cap \mathcal{T}}(\rho)) \\ &= D(\rho \| \sigma) + D(\mathcal{E}_{\mathcal{S} \cap \mathcal{T}}(\rho) \| \mathcal{E}_{\mathcal{S} \cap \mathcal{T}}(\sigma)) - D(\mathcal{E}_{\mathcal{S}}(\rho) \| \mathcal{E}_{\mathcal{S}}(\sigma)) - D(\mathcal{E}_{\mathcal{T}}(\rho) \| \mathcal{E}_{\mathcal{T}}(\sigma)) \\ &= H(\mathcal{S})_\rho + H(\mathcal{T})_\rho - H(\mathcal{S} \cap \mathcal{T})_\rho - H(\mathcal{M})_\rho \end{aligned}$$

for any density σ such that $\mathcal{E}_{\mathcal{T}}(\sigma) = \sigma$.

Technically, the entropy (as opposed to relative entropy) expression for $I(\mathcal{S} : \mathcal{T} \subseteq \mathcal{M})_\rho$ is only defined when the entropy is well-defined and finite, whereas the relative entropy expressions are more general. We will simply ignore the non-relative entropy form when entropies become undefined or divergent. We also use the shorthand notation:

$$I(\mathcal{S} : \mathcal{T})_\rho \equiv I(\mathcal{S} : \mathcal{T} \subseteq \mathcal{S} \vee \mathcal{T})_\rho \ . \tag{3.5}$$

In later sections, we will see some operational properties of and inequalities on $I(\mathcal{S} : \mathcal{T})_\rho$ that suggest it is a natural and meaningful restriction.

Within this Chapter, Section 3.1 shows that while SSA fails for any pair of subalgebras not in commuting square, a version adjusted by a multiplicative constant holds. Section 3.2 contains proof of a duality between the GCMI on a pair of algebras within their joint, and the commutants of those algebras within a purifying larger system. Section 3.3 defines a potential higher-order, multipartite entropy on subalgebras that subsumes some historical generalizations of mutual information. Section 3.4 discusses a generalization of the GCMI to particular quantum channels.

3.1 Adjusted Subadditivity

An immediate question following Petz's SSA (theorem 3.1) is what happens when von Neumann subalgebras $\mathcal{S}, \mathcal{T} \subseteq \mathcal{M}$ are not in commuting square. A classical example (see Example 7.2) is a pair of bases that generate the entire algebra $\mathcal{M}$ but which are not unbiased. With Gao and Junge in [23], we show that when $[\mathcal{E}_{\mathcal{S}}, \mathcal{E}_{\mathcal{T}}] \neq 0$ in finite dimension, SSA must fail for some ρ. In another joint paper submitted as a conference proceeding [24], we derive additive correction terms for the more general case of two quantum channels Φ and Ψ replacing $\mathcal{E}_{\mathcal{S}}, \mathcal{E}_{\mathcal{T}}$, which subsumes the non-commuting square case. When a density is sufficiently close to being in $\mathcal{S} \cap \mathcal{T}$, additive corrections depending only on $\mathcal{E}_{\mathcal{S}}, \mathcal{E}_{\mathcal{T}}$ become trivial. To see this clearly, consider the form

$$H(\mathcal{S})_\rho + H(\mathcal{T})_\rho - H(\mathcal{S} \cap \mathcal{T})_\rho - H(\mathcal{M})_\rho \geq -c \tag{3.6}$$

for some $c \geq 0$, and note that all terms on the left hand side become arbitrarily small as $\rho \to \sigma$ for any $\sigma \in \mathcal{S} \cap \mathcal{T}$. Hence this approaches the inequality $0 \geq -c$.

A different generalization of SSA beyond the commuting square case is to apply a multiplicative constant to the relative entropy form of Equation (3.3), obtaining an inequality that reads

$$D(\rho \| \mathcal{E}_{\mathcal{S}}(\rho)) + D(\rho \| \mathcal{E}_{\mathcal{T}}(\rho)) \geq \alpha D(\rho \| \mathcal{E}_{\mathcal{S} \cap \mathcal{T}}(\rho)) \ . \tag{3.7}$$

This form already appears in the context of "quasi-factorization" as studied by Ceci in [70]. Several works have noted the value of such inequalities in quantum decay estimates [71, 72, 73]. Some of these studies use a different notion of conditional expectation, so the sought inequalities are not always directly comparable to Equation (3.7). Furthermore, in this thesis I consider algebras such as those corresponding to bases of the same system, for which the notion of approximate tensorization is not truly applicable.

Unlike Equation (3.6), Equation (3.7) does not become trivial even when the left-hand side is small. In fact, one can see that if $D(\rho \| \mathcal{E}_{\mathcal{S} \cap \mathcal{T}}(\rho)) > 0$, then so is the left hand side, since ρ cannot be simultaneously in both subalgebras when not in their intersection. In the context of quasi-factorization, this inequality appears when $\mathcal{S}, \mathcal{T}$ are approximately in commuting square, or more specifically, approximately in tensor product position. We find however that this condition is not necessary. In [26], I define:

Definition 3.2 (Adjusted Subadditivity (ASA), definition 1.2 from [26]). *Let $\{\mathcal{E}_j : j \in 1...J \in \mathbb{N}\}$ be a set of conditional expectations, and $\mathcal{E}$ the conditional expectation onto their intersection algebra. We call the set α-subadditve if*

$$\sum_j D(\rho \| \mathcal{E}_j(\rho)) \geq \alpha D(\rho \| \mathcal{E}(\rho))$$

for some $\alpha > 0$. We call it completely α-subadditive if $\{\hat{1}^C \otimes \mathcal{E}_j\}$ is α-subadditive for any finite-dimensional extension C.

I then show:

Theorem 3.2 (theorem 1.4 from [26]). *Let $\{(\mathcal{N}_j, \mathcal{E}_j) : j = 1...J \in \mathbb{N}\}$ be a set of J von Neumann algebras of dimension d and associated conditional expectations. Assume that $\cap_j \mathcal{N}_j = \mathbb{C}1$, the physically trivial algebra of phase and normalization. Then $\{\mathcal{E}_j\}$ is α-subadditive for some $\alpha > 0$.*

In particular, let $S = \cup_{m \in \mathbb{N}} \{1...J\}^{\otimes m}$ be the set of sequences of indices $1...J$. For any $s \in S$, let $\mathcal{E}^s$ denote the composition of conditional expectations $\mathcal{E}_{j_1}...\mathcal{E}_{j_{|s|}}$ for $s = (j_1, ..., j_{|s|})$, where $|s|$ denotes the sequence length. Let $\mu : S \to [0,1]$ be a probability measure on S that is non-zero only on finite sequences, k_s be the maximum number of repeats of any index in s, and $k = \sum_s \mu(s) k_s$. If for all d-dimensional densities ρ,

$$\sum_{s \in S} \mu(s) \mathcal{E}^s(\rho) = (1 - \zeta)\hat{1}/d + \zeta \Phi(\rho)$$

for some unital channel Φ such that $\mathcal{E} \circ \Phi = \mathcal{E}$, then $\{\mathcal{E}_j\}$ is $1/(2k(\lceil \log_\zeta(2/(3d+5)) \rceil))$-subadditive, where $\lceil \cdot \rceil$ is the ceiling function, and assuming $\zeta \geq 2/(3d+5)$.

Finally, we conjecture:

Conjecture 3.1 (Strong Adjusted Subadditivity (SASA), conjecture 5.1 from [26]). *Let $\{(\mathcal{N}_j, \mathcal{E}_j) : j = 1...J \in \mathbb{N}\}$ be a set of J von Neumann algebras within a larger algebra $\mathcal{M}$ and associated conditional expectations. Then $\{\mathcal{E}_j\}$ is α-subadditive with an α that depends on the index $D(\mathcal{M} \| \mathcal{N})$, but not explicitly on the dimension of $\mathcal{M}$.*

Here $D(\mathcal{M} \| \mathcal{N}) = \sup_{\rho \in \mathcal{M}_*} D(\rho \| \mathcal{E}_\mathcal{N}(\rho))$ as defined in [25]. The key difference between adjusted subadditivity and quasi-factorization is that we do not require $\mathcal{S}, \mathcal{T}$ to be close to a commuting square. Rather, the important condition is a slight generalization of the idea that a sequence of conditional expectations $\mathcal{E}_\mathcal{S} \mathcal{E}_\mathcal{T} \mathcal{E}_\mathcal{S} \mathcal{E}_\mathcal{T}$... converges toward $\mathcal{E}_{\mathcal{S} \cap \mathcal{T}}$. We may for example choose a pair of arbitrary qubit bases, and as long as they are not the same basis, they will satisfy a form of adjusted subadditivity (see Example 7.2).

3.1.1 Relative Entropy with Respect to Near-Mixture

The main result underpinning adjusted subadditivity is a bound for a particular case of relative entropy:

Theorem 3.3 (theorem 1.3 from [26]). *Given $a \in [0,1]$ and two densities ρ, σ in dimension d such that $\rho \succ \sigma$ (ρ majorizes σ),*

$$D(\rho \| (1 - \zeta)\hat{1}/d + \zeta\sigma) - (1 - a) D(\rho \| \hat{1}/d) \geq 0$$

for any $\beta \in (0,1)$ and

$$\zeta \leq a \min\left\{\frac{1-\beta}{d+a(1-\beta)+1}, \frac{\beta}{(1-a\beta)d+a\beta+1}\right\} .$$

Theorem 3.3 is a *multiplicative* bound on the relative entropy of state to a copy that has been nearly replaced by the identity, in terms of $\log d - H(\rho)$. This particular form is related to the telescopic relative entropy [74]. It is easy to derive additive bounds of similar form. For example, we have:

Proposition 3.1. *Let ρ, σ, ω be three finite-dimensional densities, and $\zeta \in [0,1]$. Then*

$$D(\rho \| (1-\zeta)\omega + \zeta\sigma) \leq D(\rho\|\omega) - \log(1-\zeta)$$

Proof. We will use the form $D(\rho\|\omega) = \mathrm{tr}(\rho(\log\rho - \log\omega))$. Here

$$\begin{aligned}
&D(\rho\|(1-\zeta)\omega + \zeta\sigma) - D(\rho\|\omega) + \log(1-\zeta) \\
&= \mathrm{tr}(\rho(\log\omega - \log((1-\zeta)\omega + \zeta\sigma))) \\
&\leq \mathrm{tr}(\rho(\log\omega - \log((1-\zeta)\omega))) \\
&= -\log(1-\zeta) \ .
\end{aligned}$$

■

While Proposition 3.1 would initially appear to point the opposite way from Theorem 3.3, we will often find that when $\Phi(\rho) = (1-\zeta)\Psi(\rho) + \zeta\Theta(\rho)$ for quantum channels Φ, Ψ, Θ, it also holds that $\Psi(\rho) = (1-\epsilon)\Phi(\rho) + \epsilon\Theta'(\rho)$ for some channel Θ'. Unfortunately, bounds with additive corrections are insufficient to gain the full power of Theorem 3.2. On a technical level, if we have a bound of the type

$$\sum_j D(\rho\|\mathcal{E}_j(\rho)) \geq D(\rho\|\mathcal{E}(\rho)) - \delta$$

for some $\delta > 0$, then for ρ very close to the fixed point such that $\mathcal{E}(\rho) \approx \rho$, the right hand side actually becomes negative, and as such, the inequality trivial. Furthermore, multiplicative bounds will be essential to our primary application in merging decay inequalities as studied in Section 7.2.1. In that context, any additive correction results in a qualitatively weaker kind of bound.

The rest of this section will prove Theorem 3.2

Lemma 3.1 (lemma 2.1 from [26])**.** *Let ρ, ω be simultaneously diagonal densities of dimension d. Let $\delta > 0$. Let $i \neq j \in 1...d$ such that $\rho_i \geq \rho_j$, and $\rho_i\omega_j \geq \rho_j\omega_i$. Let $\omega \to \tilde{\omega}$ under the replacement $\omega_i \to \tilde{\omega_i} =$*

$\omega_i - \delta, \omega_j \to \tilde{\omega}_j = \omega_j + \delta$. *Then* $D(\rho\|\omega) \le D(\rho\|\tilde{\omega})$.

Proof. For any $a > b \in \mathbb{R}^+$, $(a-b)/a \le \log a - \log b \le (a-b)/b$, as one can verify from $(d/dx)(\log x) = 1/x$. Hence

$$
\begin{aligned}
D(\rho\|\tilde{\omega}) - D(\rho\|\omega) &= \rho_i(\log \omega_i - \log(\omega_i - \delta)) + \rho_j(\log \omega_j - \log(\omega_j + \delta)) \\
&\ge (\rho_i/\omega_i - \rho_j/\omega_j)\delta \ge 0
\end{aligned}
$$

■

Lemma 3.2 (lemma 2.2 from [26])**.** *Let* ρ *and* σ *be two densities such that* $\rho \succ \sigma$*, and* $\zeta \in [0,1]$*. Then*

$$
D(\rho\|(1-\zeta)\hat{1}/d + \zeta\sigma) \ge D(\rho\|(1-\zeta)\hat{1}/d + \zeta\rho) \; .
$$

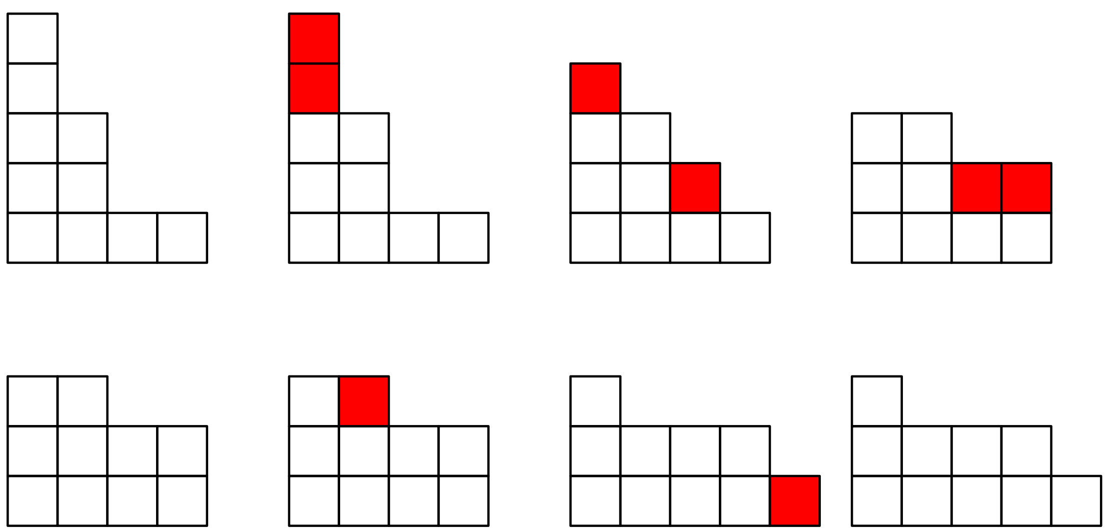

Figure 3.1: Visualization of the cascading redistribution algorithm, which converts the distribution on the top-left to that on the bottom-right. Duplicates figure 1 from [26].

Proof. The main idea of this proof is that if $\rho \succ \sigma$, then flattening ρ until it becomes σ only increases the value of $D(\rho\|(1-\zeta)\hat{1}/d + \zeta\rho)$. First,

$$
D(\rho\|(1-\zeta)\hat{1}/d + \zeta\sigma) \ge D(\rho\|(1-\zeta)\hat{1}/d + \zeta\mathcal{E}_{\rho'}(\sigma)) \; ,
$$

by data processing under $\mathcal{E}_{\rho'}$, the conditional expectation onto the subalgebra that commutes with ρ. We hence assume that $[\rho,\sigma] = 0$, and they are simultaneously diagonal. Let $\rho^\zeta = (1-\zeta)\hat{1}/d + \zeta\rho$, and define σ^ζ accordingly. Let $\vec{\rho}^\zeta$ and $\vec{\sigma}^\zeta$ be d-dimensional vectors of the eigenvalues of ρ^ζ and σ^ζ respectively, each in non-increasing order. Let $\omega = \rho^\zeta$. We alter $\vec{\omega}$ via a cascading probability redistribution procedure consisting of the following steps, which transform it into a copy of $\vec{\sigma}^\zeta$:

1. Start with the index i set to 1.
2. Let $\Delta = \vec{\omega}_i - \vec{\sigma}^\zeta_i$. If $\Delta_i > 0$, then
   (a) Subtract Δ from $\vec{\omega}_i$. Let $j = i + 1$.
   (b) If $\vec{\omega}_j < \vec{\sigma}^\zeta_j$, then let $\delta = \min\{\Delta, \vec{\sigma}^\zeta_j - \vec{\omega}_j\}$. Add δ to $\vec{\omega}_j$ and subtract it from Δ.
   (c) If $\Delta = 0$ or $j = d$, go on to step (3). Otherwise, increment $j \to j + 1$, and return to the previous substep (2b).
3. If $i < d - 1$, increment $i \to i + 1$ and return to step (2). Otherwise, the procedure is done.

See Figure 3.1. Since this procedure only subtracts from larger eigenvalues and adds to smaller ones, we apply Lemma 3.1 at each step that transfers probability mass from one index to another. If $\vec{\rho}_i \geq \vec{\rho}_j$, then $\vec{\rho}_i/\vec{\rho}^\zeta_i \geq \vec{\rho}_j/\vec{\rho}^\zeta_j$. Furthermore, it is always the case that $\vec{\omega}_i \leq \vec{\rho}^\zeta_i$ if we are moving probability mass out of $\vec{\omega}_i$, and always that $\vec{\omega}_j \geq \vec{\rho}^\zeta_j$ if we are moving probability into j. Hence $\vec{\rho}_i/\vec{\rho}_j \geq \vec{\omega}_i/\vec{\omega}_j$. This shows that $D(\rho\|\rho^\zeta) = D(\vec{\rho}\|\vec{\rho}^\zeta) \leq D(\vec{\rho}\|\vec{\sigma}^\zeta)$. Finally, using the simultaneous diagonality of ρ and σ, it is easy to see that $D(\vec{\rho}\|\vec{\sigma}^\zeta) \leq D(\rho\|\sigma^\zeta)$. ■

Lemma 3.3 (lemma 2.3 from [26]). *Let $b \geq 0$, and $a \in [0, 1]$. Then for $\zeta \leq a/(1 + b)$,*

$$a \log(1 + b) \geq \log(1 + \zeta b) .$$

Proof. We solve

$$(1 + b)^a \geq 1 + \zeta b$$

yielding

$$\zeta \leq \frac{(1 + b)^a - 1}{b} .$$

We then estimate

$$\frac{(1 + b)^a - 1}{b} = \frac{(1 + b)^{1+a} - 1 - b}{b(1 + b)} \geq \frac{1 + (1 + a)b - 1 - b}{b(1 + b)} = \frac{a}{1 + b}$$

by Bernoulli's inequality. ■

Lemma 3.4 (lemma 2.4 from [26].). *Let ρ be given in its diagonal basis by $(\rho_i)_{i=1}^n$, where n is the dimension of the system. Let $a, \beta \in (0,1)$, $\rho_i \geq 1/n \geq \rho_j$, and*

$$\zeta \leq a \min\left\{\frac{1-\beta}{n + a(1-\beta)+1}, \frac{\beta}{(1-a\beta)n + a\beta + 1}\right\} .$$

Then

$$\Big(\frac{\partial}{\partial \rho_i} - \frac{\partial}{\partial \rho_j}\Big) tr(\rho(a\log(n\rho) - \log((1-\zeta)\hat{1} + \zeta n\rho))) \geq 0 .$$

Note: in this Lemma and its proof, and thereafter within this section, we use the letter n rather than d for dimension to avoid possible confusion with the derivative.

Proof. Define $\delta \equiv \rho_i - \rho_j \geq 0$. For the purposes of the derivative, $\mathrm{tr}(\log(n\rho))$ is equivalent to $\mathrm{tr}(\log \rho)$, since they differ only by a constant $\log n$. We directly calculate,

$$\frac{\partial}{\partial \rho_k}(\rho_k \log((1-\zeta) + \zeta n \rho_k)) = \log((1-\zeta) + \zeta n\rho_k) + \frac{\zeta n \rho_k}{(1-\zeta) + \zeta n \rho_k} . \tag{3.8}$$

for any $k \in 1...n$. By setting $\zeta = 1$,

$$\Big(\frac{\partial}{\partial \rho_i} - \frac{\partial}{\partial \rho_j}\Big)\mathrm{tr}(\rho \log \rho) = \log(\rho_i/\rho_j) = \log(1 + \delta/\rho_j) .$$

Because

$$\frac{x}{1+x} \leq \log(1+x) \leq x$$

for all $x \geq 0$, we have for any $\beta \in [0,1]$ that

$$a\log(1+\delta/\rho_j) \geq a(1-\beta)\log(1+\delta/\rho_j) + a\beta\frac{\delta/\rho_j}{1+\delta/\rho_j} .$$

This will allow us to deal with the two terms in Equation (3.8) individually.

First, we handle the logarithm term, $\log((1-\zeta)+\zeta n\rho_k)$, by finding ζ such that

$$a(1-\beta)\log(1+\delta/\rho_j) \geq \log\Big(\frac{(1-\zeta) + \zeta n \rho_j + \zeta n \delta}{(1-\zeta)+\zeta n\rho_j}\Big) .$$

We rewrite the right hand side as

$$\log\Big(1 + \frac{\zeta n \delta}{(1-\zeta) + \zeta n \rho_j}\Big) .$$

On the left hand side, since $\rho_j \leq 1/n$, $\delta/\rho_j \geq n\delta$. We aim to show that

$$a(1-\beta)\log(1+n\delta) \geq \log\left(1 + \frac{\zeta}{(1-\zeta)+\zeta n\rho_j} n\delta\right) ,$$

which by Lemma 3.3 is achieved when

$$\frac{\zeta}{(1-\zeta)+\zeta n\rho_j} \leq \frac{a(1-\beta)}{1+n\delta} .$$

We know $\zeta/((1-\zeta)+\zeta n\rho_j) \leq \zeta/(1-\zeta)$ for any $n\rho_j \geq 0$. Hence any

$$\zeta \leq \frac{a(1-\beta)}{1+n\delta+a(1-\beta)} \tag{3.9}$$

is sufficiently small.

Next, we handle the fraction terms by finding ζ such that

$$a\beta\frac{\delta/\rho_j}{1+\delta/\rho_j} \geq \frac{\zeta n(\rho_j+\delta)}{(1-\zeta)+\zeta n(\rho_j+\delta)} - \frac{\zeta n\rho_j}{(1-\zeta)+\zeta n\rho_j} .$$

By Taylor expansion,

$$\begin{aligned}
&\frac{\zeta n(\rho_j+\delta)}{(1-\zeta)+\zeta n(\rho_j+\delta)} - \frac{\zeta n\rho_j}{(1-\zeta)+\zeta n\rho_j} \\
=&\frac{\zeta n(\rho_j+\delta)}{(1-\zeta)+\zeta n\rho_j}\sum_{k=0}^{\infty}\left(\frac{-\zeta n\delta}{(1-\zeta)+\zeta n\rho_j}\right)^k - \frac{\zeta n\rho_j}{(1-\zeta)+\zeta n\rho_j} .
\end{aligned}$$

Canceling the 0th order term,

$$\begin{aligned}
\ldots =&\frac{\zeta n\rho_j}{(1-\zeta)+\zeta n\rho_j}\sum_{k=1}^{\infty}\left(\frac{-\zeta n\delta}{(1-\zeta)+\zeta n\rho_j}\right)^k + \frac{\zeta n\delta}{(1-\zeta)+\zeta n\delta}\sum_{k=0}^{\infty}\left(\frac{-\zeta n\delta}{(1-\zeta)+\zeta n\rho_j}\right)^k \\
=&\frac{\zeta n\rho_j}{(1-\zeta)+\zeta n\rho_j}\sum_{k=1}^{\infty}\left(\frac{-\zeta n\delta}{(1-\zeta)+\zeta n\rho_j}\right)^k - \sum_{k=1}^{\infty}\left(\frac{-\zeta n\delta}{(1-\zeta)+\zeta n\rho_j}\right)^k \\
=&\left(\frac{\zeta n\rho_j}{(1-\zeta)+\zeta n\rho_j} - 1\right)\sum_{k=1}^{\infty}\left(\frac{-\zeta n\delta}{(1-\zeta)+\zeta n\rho_j}\right)^k .
\end{aligned}$$

Now to turn this back into the form of a fraction,

$$\begin{aligned}
\ldots =& \frac{-(1-\zeta)}{(1-\zeta)+\zeta n\rho_j}\sum_{k=1}^{\infty}\Big(\frac{-\zeta n\delta}{(1-\zeta)+\zeta n\rho_j}\Big)^k \\
=& \frac{(1-\zeta)\zeta n\delta}{((1-\zeta)+\zeta n\rho_j)^2}\sum_{k=0}^{\infty}\Big(\frac{-\zeta n\delta}{(1-\zeta)+\zeta n\rho_j}\Big)^k \\
=& \frac{(1-\zeta)\zeta n\delta}{((1-\zeta)+\zeta n\rho_j)^2}\frac{1}{1+\frac{\zeta n\delta}{(1-\zeta)+\zeta n\rho_j}} \\
=& \frac{1}{(1-\zeta)+\zeta n\rho_j}\frac{\zeta n\delta}{1+\frac{\zeta}{1-\zeta}(\rho_j+\delta)n}\ .
\end{aligned}$$

Since $0 \leq \rho_j \leq 1/n$,

$$\ldots \leq \frac{\frac{\zeta}{1-\zeta}n\delta}{1+\frac{\zeta}{1-\zeta}n\delta}\ .$$

We must find a ζ for which

$$a\beta\log(1+\delta/\rho_j) \geq a\beta\frac{\delta/\rho_j}{1+\delta/\rho_j} \geq a\beta\frac{n\delta}{1+n\delta} \geq \frac{\frac{\zeta}{1-\zeta}n\delta}{1+\frac{\zeta}{1-\zeta}n\delta}\ .$$

One can easily check that this is satisfied by

$$\frac{\zeta}{1+\zeta} \leq \frac{a\beta}{(1-a\beta)n\delta+1}\ ,$$

which follows from

$$\zeta \leq \frac{a\beta}{(1-a\beta)n\delta+a\beta+1}\ . \tag{3.10}$$

Finally, any

$$\zeta \leq \min\Big\{\frac{a(1-\beta)}{1+n+a(1-\beta)}, \frac{a\beta}{(1-a\beta)n+a\beta+1}\Big\}$$

satisfies both Inequalities (3.9) and (3.10). ■

Proof of Theorem 3.3. (Duplicates proof of theorem 1.3 from [26].) Let n be the dimension of ρ, since d may be confused with a derivative. The goal is to show that given some $a \in [0,1]$ and densities ρ, σ such that $\rho \succ \sigma$,

$$D(\rho\|(1-\zeta)\hat{1}/n+\zeta\sigma) - (1-a)D(\rho\|\hat{1}/n) \geq 0\ ,$$

for an appropriate value of $\zeta \in [0,1]$. We apply Lemma 3.2 to replace σ by ρ. For any states ρ and ω,

$$D(n\rho \| n\omega) = \mathrm{tr}((n\rho)(\log \rho + \log n - \log \omega - \log n)) = nD(\rho\|\omega) \ .$$

Hence, it suffices to show that

$$D(n\rho \| (1-\zeta)\hat{1} + \zeta n\rho) - (1-a)D(n\rho\|\hat{1}) \geq 0 \ ,$$

which expands as

$$... = n\mathrm{tr}(\rho(a \log(n\rho) - \log((1-\zeta)\hat{1} + \zeta n\rho))) \geq 0 \ .$$

The main insight behind this proof is Lemma 3.4. If $\rho = \hat{1}/n$, then both terms are 0, and the proof is trivially complete. If $\rho \neq \hat{1}/n$, then the total probability mass above $\hat{1}/n$ must equal that below $\hat{1}/n$ to maintain normalization. Hence we may apply the variational argument in Lemma 3.4 to continuously redistribute probability from larger to smaller until $\rho \to \hat{1}/n$. ∎

3.1.2 Proof of ASA

Lemma 3.5 (Chain Expansion / lemma 3.1 from [26])**.** *Let Φ be a quantum channel and $\mathcal{E}$ be a conditional expectation such that $\mathcal{E}(\Phi(\rho)) = \Phi(\rho)$. Then for any state ρ,*

$$D(\rho\|\Phi(\rho)) = D(\rho\|\mathcal{E}(\rho)) + D(\mathcal{E}(\rho)\|\Phi(\rho))$$

Proof.

$$\begin{aligned} & D(\rho\|\mathcal{E}(\rho)) + D(\mathcal{E}(\rho)\|\Phi(\rho)) \\ & = \mathrm{tr}(\rho \log \rho - \rho \log \mathcal{E}(\rho) + \mathcal{E}(\rho) \log \mathcal{E}(\rho) - \mathcal{E}(\rho) \log \Phi(\rho)) \\ & = \mathrm{tr}(\rho \log \rho - \rho \log \mathcal{E}(\rho) + \rho \log \mathcal{E}(\rho) - \rho \log \Phi(\rho)) \\ & = \mathrm{tr}(\rho \log \rho - \rho \log \Phi(\rho)) \ , \end{aligned}$$

where the 1st equality follows from expanding the relative entropy, and the 2nd from the defining property of conditional expectations. ∎

Lemma 3.6 (lemma 3.2 from [26])**.** *Let $\mathcal{E}$ be a conditional expectation and Φ be a quantum channel. Then*

$$D(\rho\|\mathcal{E}(\rho)) + D(\rho\|\Phi(\rho)) \geq D(\rho\|\mathcal{E}(\Phi(\rho))$$

Proof. By data processing on the 2nd term,

$$D(\rho\|\mathcal{E}(\rho)) + D(\rho\|\Phi(\rho)) \geq D(\rho\|\mathcal{E}(\rho)) + D(\mathcal{E}(\rho)\|\mathcal{E}(\Phi(\rho))) \ .$$

By Lemma 3.5 and the idempotence of conditional expectations, we obtain the desired result. ■

Lemma 3.7 (lemma 3.3 from [26]). *Let $\{(\mathcal{N}_j, \mathcal{E}_j) : j = 1...J \in \mathbb{N}\}$ be a set of J von Neumann algebras and associated conditional expectations. Let $\mathcal{E}$ be the conditional expectation to the intersection algebra, $\mathcal{N} = \cap_j \mathcal{N}_j$. Let $S = \cup_{m\in\mathbb{N}}\{1...J\}^{\otimes m}$ be the set of sequences of indices $1...k$. For any $s \in S$, let $\mathcal{E}^s$ denote the composition of conditional expectations $\mathcal{E}_{j_1}...\mathcal{E}_{j_m}$ for $s = (j_1, ..., j_m)$. Let $\mu : S \to [0,1]$ be a probability measure on S that is non-zero only if no index appears more than k times. If for all input densities ρ,*

$$\sum_{s\in S} \mu(s)\mathcal{E}^s(\rho) = (1-\zeta)\mathcal{E}(\rho) + \zeta\Phi(\rho)$$

for some channel Φ, and

$$D(\rho\|(1-\zeta)\mathcal{E}(\rho) + \zeta\Phi(\rho)) \geq f(\{\mathcal{N}_j\})D(\rho\|\mathcal{E}(\rho))$$

for some $f(\{\mathcal{N}_j\}) > 0$, then $\{\mathcal{E}_j\}$ is $f(\{\mathcal{N}_j\})/k$-subadditive.

Proof. ■

Proof of Theorem 3.2. (Based on and partially duplicating the proof of theorem 1.4 from [26].) First, let us assume that there exists such a μ and ζ as described in the second part of the Theorem. We know by the definition of k that

$$\sum_j D(\rho\|\mathcal{E}_j(\rho)) = \frac{1}{k}\Big(\sum_{s\in S}\mu(s)k_s\Big)\sum_j D(\rho\|\mathcal{E}_j(\rho)) \ ,$$

For each s such that $\mu(s) > 0$, we apply Lemma 3.6 iteratively until we obtain $D(\rho\|\mathcal{E}^s(\rho))$. For a sequence of conditional expectations $\mathcal{E}^s = \mathcal{E}_{j_1}...\mathcal{E}_{j_{|s|}}$, we need at most k_s copies of each $D(\rho\|\mathcal{E}_j(\rho))$, where k_s is the maximum count of any constituent conditional expectation. Hence

$$\mu(s)k_s\sum_j D(\rho\|\mathcal{E}_j(\rho)) \geq \mu(s)D(\rho\|\mathcal{E}^s(\rho)) \ .$$

All together,

$$\sum_j D(\rho\|\mathcal{E}_j(\rho)) \geq \frac{1}{k}\sum_{s\in S}\mu(s)D(\rho\|\mathcal{E}^s(\rho)) \ .$$

By convexity of relative entropy and the assumptions of the Theorem,

$$\sum_j D(\rho\|\mathcal{E}_j(\rho)) \geq \frac{1}{k} D(\rho\|(1-\zeta)\mathcal{E}(\rho) + \zeta\Phi(\rho)).$$

By a result of Uhlmann [75], $\rho \succ \Phi(\rho)$ for any unital channel Φ. If ζ is not sufficiently small, we use the fact that for any $m \in \mathbb{N}$,

$$((1-\zeta)\mathcal{E} + \zeta\Phi)^m(\rho) = (1-\zeta^m)\mathcal{E}(\rho) + \zeta^m \Phi^m(\rho)$$

by the idempotence of conditional expectations, and by the assumption that $\mathcal{E} \circ \Phi = \mathcal{E}$. To construct m powers of a convex combination of sequences of conditional expectations, we assume that we have $((1-\zeta)\mathcal{E} + \zeta\Phi)^{m-1}(\rho)$, apply Lemma 3.6 multiple times to append another copy of the conditional expectation sequence $\mathcal{E}^s$, and then convexify to reach $((1-\zeta)\mathcal{E} + \zeta\Phi)^m(\rho)$. We thereby find that

$$\sum_j D(\rho\|\mathcal{E}_j(\rho)) \geq \frac{1}{km} D(\rho\|(1-\zeta)\mathcal{E}(\rho) + \zeta\Phi(\rho)) \ .$$

By Theorem 3.3, this leads to the conclusion that

$$\sum_j D(\rho\|\mathcal{E}_j(\rho)) \geq \frac{1-a}{km} D(\rho\|(1-\zeta)\mathcal{E}(\rho) + \zeta\Phi(\rho))$$

for any $\beta, a \in (0,1)$ such that

$$\zeta^m \leq \min\left\{\frac{1-\beta}{d + a(1-\beta) + 1}, \frac{\beta}{(1-a\beta)d + a\beta + 1}\right\} .$$

In full generality, we would optimize a, β, and m for a given ζ. Here however, we will simplify by choosing $a = \beta = 1/2$. Hence we may choose

$$\zeta^m \leq \frac{2}{3d+5} \leq \min\left\{\frac{2}{4d+5}, \frac{2}{3d+5}\right\} ,$$

which is satisfied by $m \geq \lceil \log_\zeta(2/(3d+5)) \rceil$. Then

$$D(\rho\|(1-\zeta^m)\hat{1}/d + \zeta^m \Phi^m(\rho)) \geq \frac{1}{2} D(\rho\|\hat{1}/d) \ .$$

We apply Lemma 3.7 with ζ^m replacing ζ, mk replacing k, and with the above bound.

Finally, we must show that some sufficiently long chain of conditional expectations eventually converges to the intersection algebra, so that such a μ and ζ as in the Theorem always exist. Let $s = (1, ..., J)$, so that $\mathcal{E}^s$ is a sequence on all constituent conditional expectations. As a convex combination of unitaries, s is a contraction. One can easily see that it satisfies properties $(0.I - 0.IV)$ in [76]. As a result, $\lim_{m \to \infty} \frac{1}{m} \sum_{k=0}^{m} (\mathcal{E}^s)^k \to \mathcal{E}$ in 2-norm distance. In finite dimension, this eventually implies a bound on the smallest element, so for some m, it must yield a convex combination of the complete mixture with some unital channel. ■

3.2 Commutant & Complement Duality

In a finite-dimensional, tripartite system ABC, we can easily dualize strong subadditivity on its purification. Let $\rho^{ABC} = \mathrm{tr}_R(|\psi\rangle\langle\psi|^{ABCD})$, where D is an extra system of dimension at most equal to that of ABC (which is always possible by Schmidt decomposition). For any pure, bipartite state $|\phi\rangle\langle\phi|^{AB}$,

$$H(A)_{|\phi\rangle\langle\phi|} = H(B)_{|\phi\rangle\langle\phi|} \ . \tag{3.11}$$

Applying this to $|\psi\rangle\langle\psi|^{ABCD}$,

$$\begin{aligned} I(A : B|C)_\rho &= H(A|C)_\rho + H(B|C)_\rho - H(AB|C)_\rho \\ &= H(B|D)_{|\psi\rangle\langle\psi|} + H(A|D)_{|\psi\rangle\langle\psi|} - H(AB|D)_{|\psi\rangle\langle\psi|} = I(A : B|D)_{|\psi\rangle\langle\psi|} \ . \end{aligned} \tag{3.12}$$

When D is trivial (hence $\rho^{ABC} = |\psi\rangle\langle\psi|^{ABC}$),

$$H(B|A)_{|\psi\rangle\langle\psi|} = -H(B|C)_{|\psi\rangle\langle\psi|} \ . \tag{3.13}$$

which also follows from Equation (3.12).

As found in the process of writing [23], this actually fails for $I(\mathcal{S} : \mathcal{T} \subseteq \mathcal{M})_\rho$ in general. In that same work, however, we prove:

Theorem 3.4 (Commutant Duality, theorem 2.9 from [23])**.** *Let $\mathcal{S}, \mathcal{T} \subseteq \mathcal{M}$ be von Neumann subalgebras in commuting square, for which $\mathcal{M}$ is a finite-dimensional factor. Given a density ρ on $\mathcal{M}$, let $\mathcal{M}_*'$ be a purifying system such that $\rho^{\mathcal{S} \vee \mathcal{T}} = \mathcal{E}_{\mathcal{S} \vee \mathcal{T}}(|\psi\rangle\langle\psi|^{\mathcal{M}_* \mathcal{M}_*'})$. Then*

$$I(\mathcal{S} : \mathcal{T})_\rho = I(\mathcal{S}' : \mathcal{T}')_{|\psi\rangle\langle\psi|} \ .$$

Theorem 3.4 gives some intuition for Equation (3.12), which in the form of subalgebras becomes

$$I(\mathcal{A} \vee \mathcal{C} : \mathcal{B} \vee \mathcal{C})_\rho = I((\mathcal{A} \vee \mathcal{C})' : (\mathcal{B} \vee \mathcal{C})' \subseteq \mathcal{A} \vee \mathcal{B} \vee \mathcal{C} \vee \mathcal{D})_{|\psi\rangle\langle\psi|} = I(\mathcal{B} \vee \mathcal{D} : \mathcal{A} \vee \mathcal{D})_{\mathrm{tr}_C(|\psi\rangle\langle\psi|)} \ .$$

Following some ideas of Crann, Kribs, Levene & Todorov [77], the commutant of a factor is analogous to its complement in the subsystem sense, and to its environment. We might think of Theorem 3.4 as showing that the conditional mutual information on a tripartite system is equal to that of those systems' environments under purification. This may initially seem surprising: instead of asking about correlations between two parties conditioned in a third, we might as well consider the correlations in systems to which those parties lack access. We note however that $\mathcal{A}$ and $\mathcal{B}$ are still present in the commutant form, just switched in position - the real change replaces $\mathcal{C}$ by the purifying system $\mathcal{D}$.

That Theorem 3.4 holds for $I(\mathcal{S} : \mathcal{T})$ while no obvious analog appears for $I(\mathcal{S} : \mathcal{T} \subseteq \mathcal{M})$ when $\mathcal{M} \neq \mathcal{S} \vee \mathcal{T}$ suggests that $I(\mathcal{S} : \mathcal{T})$ has some potentially useful properties not found in the more general form, and that we might consider $\mathcal{S} \vee \mathcal{T}$ a canonical joint system for two subalgebras. In Chapter 4, we will see that $I(\mathcal{S} : \mathcal{T})$ also has sensible operational interpretations that may hold only when $\mathcal{M} = \mathcal{S} \vee \mathcal{T}$.

The relationship between complements and commutants in von Neumann algebras also appears in the complementary channels of conditional expectations. In particular,

Lemma 3.8 (based on lemma A.4 from [23])**.** *Let $\mathcal{S} \subseteq \mathcal{M}$ be a von Neumann subalgebra with conditional expectation $\mathcal{E}_\mathcal{S}$ in finite dimension. Then*

$$R_{\hat{1}/d, \mathcal{E}^c_\mathcal{S}} \circ \mathcal{E}^c_\mathcal{S} = \mathcal{E}_{\mathcal{S}'} \ ,$$

$$\mathcal{E}^c_\mathcal{S} \circ \mathcal{E}_{\mathcal{S}'} = \mathcal{E}^c_\mathcal{S} \ ,$$

where $R_{\hat{1}/d, \mathcal{E}^c_\mathcal{S}}$ is the Petz recovery map.

Proof. First, we explicitly calculate the complement of a conditional expectation on the d-dimensional complete mixture, $\hat{1}/d$,

$$\mathcal{E}^c_\mathcal{S}(\hat{1}/d) = \oplus_i \frac{n_i \hat{1}_{m_i^2}}{d m_i} \ . \tag{3.14}$$

This is easy to invert, so

$$(\mathcal{E}^c_\mathcal{S}(\hat{1}/d))^{-1/2} = \oplus_i \sqrt{\frac{d m_i}{n_i}} \hat{1}_{m_i^2} \ . \tag{3.15}$$

The effect of the two factors of this is to change the normalization of each block by $d|m_i/n_i$. Similarly, the two factors of $(\hat{1}/d)^{1/2}$ adjust the overall normalization by $1/d$. The d powers cancel, and we're left with a blockwise adjustment of m_i/n_i. We again use the blockwise structure to calculate the adjoint. First, define $\Theta : S_1^{n_i m_i} \to S_1^{m_i m_i}$ by $\Theta_i(\rho) \equiv \hat{1}_{m_i}/m_i \otimes \mathrm{tr}_{n_i}(\rho)$. We calculate

$$\Theta_i^\dagger(\sigma) = \hat{1}_{n_i} \otimes \mathrm{tr}_{m_i}(\sigma), \tag{3.16}$$

where the tr_{m_i} is positioned to remove the $\hat{1}_{m_i}/m_i$ attached by Θ_i when composing $\Theta_i^\dagger \Theta_i$. We may also directly calculate

$$R_{\hat{1}/n_i m_i, \Theta_i} = \frac{m_i}{n_i} \Theta_i^\dagger \tag{3.17}$$

This yields the result that

$$\mathcal{E}_{\mathcal{S}}^{c\dagger}(\sigma) = \oplus_i (1_{n_i} \otimes \mathrm{tr}_{m_i}(P_i \sigma P_i)) \ . \tag{3.18}$$

Then

$$\mathcal{E}_{\mathcal{S}}^{c\dagger} \circ \mathcal{E}_{\mathcal{S}}^{c}(\rho) = \oplus_i (1_{n_i}/m_i \otimes \mathrm{tr}_{n_i}(P_i \rho P_i)) \ . \tag{3.19}$$

Thus we find

$$R_{\hat{1}/d, \mathcal{E}_{\mathcal{S}}^c} \circ \mathcal{E}_{\mathcal{S}}^{c}(\rho) = \oplus_i (1_{n_i}/n_i \otimes \mathrm{tr}_{n_i}(P_i \rho P_i)) = \mathcal{E}_{\mathcal{S}'}(\rho) \ . \tag{3.20}$$

■

Lemma 3.8 describes something we will call a *recovered complement*. For a general channel Φ and state σ, we define the *recovered complement* to be the channel given by $\mathcal{R}_{\sigma, \Phi^c} \circ \Phi^c$. The recovered complement is the environment's reconstruction of the original input state after applying Φ. For conditional expectations, Lemma 3.8 shows an intuitive equivalence between the complement and commutant up to recovery. In general, the recovery map is not necessarily itself perfectly recoverable. First, we show two Lemmas.

Lemma 3.9 (lemma A.13 from [23]). *Let $\mathcal{S}, \mathcal{T} \subset \mathcal{M}$ be subalgebras such that $[\mathcal{E}_{\mathcal{S}'}, \mathcal{E}_{\mathcal{T}}] = 0$. Then $\mathcal{E}_{\mathcal{T}}^c \mathcal{E}_{\mathcal{S}} \mathcal{R}_{\hat{1}/d, \mathcal{E}_{\mathcal{T}}^c}$ is idempotent and a Petz map for itself.*

Proof. For idempotence, we apply Lemma 3.8 and calculate,

$$\mathcal{E}_{\mathcal{S}}^c \mathcal{E}_{\mathcal{T}} R_{\hat{1}/d, \mathcal{E}_{\mathcal{S}}^c} \mathcal{E}_{\mathcal{S}}^c \mathcal{E}_{\mathcal{T}} R_{\hat{1}/d, \mathcal{E}_{\mathcal{S}}^c} = \mathcal{E}_{\mathcal{S}}^c \mathcal{E}_{\mathcal{T}} \mathcal{E}_{\mathcal{S}'} \mathcal{E}_{\mathcal{T}} R_{\hat{1}/d, \mathcal{E}_{\mathcal{S}}^c} = \mathcal{E}_{\mathcal{S}}^c \mathcal{E}_{\mathcal{T}} R_{\hat{1}/d, \mathcal{E}_{\mathcal{S}}^c} \ . \tag{3.21}$$

To show that this is its own Petz map, we use the decomposition of Petz maps for channels composed in series. $\mathcal{E}_{\mathcal{T}}$ is its own Petz map. $\mathcal{E}_{\mathcal{S}}^c$ has Petz map $R_{\hat{1}/d, \mathcal{E}_{\mathcal{S}}^c}$ by definition, and by its form, $R_{\hat{1}/d, \mathcal{E}_{\mathcal{S}}^c}$ is recovered by $\mathcal{E}_{\mathcal{S}}^c$. ■

Proof of Theorem 3.4 as essentially duplicated from [23]. To save space, we will use the notational convention that $\mathcal{ST} \equiv \mathcal{S} \vee \mathcal{T}$ within this proof.

$$\begin{aligned} I(\mathcal{S} : \mathcal{T} \subset \mathcal{ST})_\rho &= H(\mathcal{S}) + H(\mathcal{T}) - H(\mathcal{S} \cap \mathcal{T}) - H(\mathcal{ST}) \\ &= D(\mathcal{E}_{\mathcal{ST}}(\rho) \| \mathcal{E}_{\mathcal{S}}(\rho)) - D(\mathcal{E}_{\mathcal{T}}(\rho) \| \mathcal{E}_{\mathcal{S}\cap\mathcal{T}}(\rho)) \\ &= D(\mathcal{E}^c_{(\mathcal{ST})'}(\rho) \| \mathcal{E}^c_{(\mathcal{ST})'} \mathcal{E}_{\mathcal{S}}(\rho)) - D(\mathcal{E}^c_{\mathcal{T}'}(\rho) \| \mathcal{E}^c_{\mathcal{T}'} \mathcal{E}_{\mathcal{S}\cap\mathcal{T}}(\rho)). \end{aligned} \tag{3.22}$$

In the first term of the last line, we use that $\mathcal{R}_{\hat{1}/d, \mathcal{E}^c_{(\mathcal{ST})'}} \mathcal{E}^c_{(\mathcal{ST})'} \mathcal{E}_{\mathcal{S}} = \mathcal{E}_{\mathcal{S}}$, and $\mathcal{R}_{\hat{1}/d, \mathcal{E}^c_{(\mathcal{ST})'}} \mathcal{E}^c_{(\mathcal{ST})'} = \mathcal{E}_{\mathcal{ST}}$ by Lemma 3.8, so the application of $\mathcal{E}^c_{(\mathcal{ST})'}$ to both arguments of $D(\cdot\|\cdot)$ is reversible by its Petz recovery. A similar argument holds for $\mathcal{E}^c_{\mathcal{T}'}$ in the second term. Application of a fully recoverable channel leaves relative entropy invariant by data processing in both directions. Let $\tilde{\mathcal{E}}_{\mathcal{S}} = \mathcal{E}^c_{(\mathcal{ST})'} \mathcal{E}_{\mathcal{S}} \mathcal{R}_{\hat{1}/d, \mathcal{E}^c_{(\mathcal{ST})'}}$, and $\tilde{\mathcal{E}}_{\mathcal{S}\cap\mathcal{T}} = \mathcal{E}^c_{\mathcal{T}'} \mathcal{E}_{\mathcal{S}\cap\mathcal{T}} \mathcal{R}_{\hat{1}/d, \mathcal{E}^c_{\mathcal{T}'}}$. By Lemma 3.9, these are idempotent and self-recovering. We then have

$$\begin{aligned} I(\mathcal{S} : \mathcal{T} \subset \mathcal{ST}) &= D(\mathcal{E}^c_{(\mathcal{ST})'}(\rho) \| \tilde{\mathcal{E}}_{\mathcal{S}} \mathcal{E}^c_{(\mathcal{ST})'}(\rho)) - D(\mathcal{E}^c_{\mathcal{T}'}(\rho) \| \tilde{\mathcal{E}}_{\mathcal{S}\cap\mathcal{T}} \mathcal{E}^c_{\mathcal{T}'}(\rho)) \\ &= H(\tilde{\mathcal{E}}_{\mathcal{S}} \mathcal{E}^c_{(\mathcal{ST})'}(\rho)) - H(\mathcal{E}^c_{(\mathcal{ST})'}(\rho)) + H(\mathcal{E}^c_{\mathcal{T}'}(\rho)) - H(\tilde{\mathcal{E}}_{\mathcal{S}\cap\mathcal{T}} \mathcal{E}^c_{\mathcal{T}'}(\rho)) \\ &= H(\mathcal{E}^c_{(\mathcal{ST})'} \mathcal{E}_{\mathcal{S}}(\rho)) - H(\mathcal{E}^c_{(\mathcal{ST})'}(\rho)) + H(\mathcal{E}^c_{\mathcal{T}'}(\rho)) - H(\mathcal{E}^c_{\mathcal{T}'} \mathcal{E}_{\mathcal{S}\cap\mathcal{T}}(\rho)) \; . \end{aligned} \tag{3.23}$$

Since $|\psi\rangle\langle\psi|$ is pure, the middle two terms equal $H(\mathcal{T}')_{|\psi\rangle\langle\psi|} - H(\mathcal{S}' \cap \mathcal{T}')_{|\psi\rangle\langle\psi|}$. We are done with these terms.

We turn our attention to the outer terms. First, $\mathcal{E}^c_{\mathcal{T}'} \mathcal{E}_{\mathcal{S}\cap\mathcal{T}} = \mathcal{E}^c_{\mathcal{T}'} \mathcal{E}_{\mathcal{T}} \mathcal{E}_{\mathcal{S}} = \mathcal{E}^c_{\mathcal{T}'} \mathcal{E}_{\mathcal{S}}$. Thereby, these terms become

$$H(\mathcal{E}^c_{(\mathcal{ST})'} \mathcal{E}_{\mathcal{S}}(\rho)) - H(\mathcal{E}^c_{\mathcal{T}'} \mathcal{E}_{\mathcal{S}}(\rho)) \tag{3.24}$$

Since there exists an isometry from the complementary channel of the minimal Stinespring dilation to any other, we are free to Stinespring dilate non-minimally without changing these entropies. Here we will use E to denote an environment system (not a channel), while $\mathcal{E}$ generally denotes a conditional expectation. Let $E_{(\mathcal{ST})'}$ be the environment of $\mathcal{E}_{(\mathcal{ST})'}$. Since $\mathcal{E}^c_{(\mathcal{ST})'} = \mathcal{E}^c_{\mathcal{S}'\cap\mathcal{T}'} = (\mathcal{E}_{\mathcal{T}'} \mathcal{E}_{\mathcal{S}'})^c$, we may compose the environments as $E_{(\mathcal{ST})'} \cong E_{\mathcal{T}'} \tilde{E}_{\mathcal{S}'}$ (which is a tensor system, not a channel composition). We then use the fact that $\mathcal{E}^c_{\mathcal{T}'} = (\mathcal{E}_{\mathcal{T}'} \mathcal{E}_{\mathcal{T}'\mathcal{S}'})^c$ to rewrite $E_{\mathcal{T}'} = E_{\mathcal{T}'\mathcal{S}'} \tilde{E}_{\mathcal{T}'}$, as though the channel $\mathcal{E}_{\mathcal{T}'\mathcal{S}'}$ had been applied first, and the channel $\mathcal{E}_{\mathcal{T}'}$ to the output of that, keeping both environments. In the term containing $\mathcal{E}^c_{\mathcal{T}'}$, we expand the environment to $E_{\mathcal{T}'\mathcal{S}'}\tilde{E}_{\mathcal{T}'}$. We also further expand $E_{(\mathcal{S}\mathcal{T})'} \cong E_{\mathcal{T}'}\tilde{E}_{\mathcal{S}'} \cong E_{\mathcal{T}'\mathcal{S}'}\tilde{E}_{\mathcal{T}'}\tilde{E}_{\mathcal{S}'}$. This leaves us with

$$\begin{aligned}
&H(E_{\mathcal{T}'\mathcal{S}'}\tilde{E}_{\mathcal{T}'}\tilde{E}_{\mathcal{S}'})_{\mathcal{E}_{\mathcal{S}}(\rho)} - H(E_{\mathcal{T}'\mathcal{S}'}\tilde{E}_{\mathcal{T}'})_{\mathcal{E}_{\mathcal{S}}(\rho)} \\
&= -D(E_{\mathcal{T}'\mathcal{S}'}\tilde{E}_{\mathcal{T}'}\tilde{E}_{\mathcal{S}'} \| E_{\mathcal{T}'\mathcal{S}'}\tilde{E}_{\mathcal{T}'} \otimes 1)_{\mathcal{E}_{\mathcal{S}}(\rho)} \\
&\leq -D(E_{\mathcal{T}'\mathcal{S}'}\tilde{E}_{\mathcal{S}'} \| E_{\mathcal{T}'\mathcal{S}'} \otimes 1)_{\mathcal{E}_{\mathcal{S}}(\rho)} \\
&= H(E_{\mathcal{T}'\mathcal{S}'}\tilde{E}_{\mathcal{S}'})_{\mathcal{E}_{\mathcal{S}}(\rho)} - H(E_{\mathcal{T}'\mathcal{S}'})_{\mathcal{E}_{\mathcal{S}}(\rho)}
\end{aligned} \tag{3.25}$$

by data processing. $\tilde{E}_{\mathcal{T}'}$ was the same system in both terms, and $\tilde{E}_{\mathcal{S}'}$ was split off before it. This can be written as

$$H((\mathcal{E}_{\mathcal{S}'}\mathcal{E}_{\mathcal{T}'\mathcal{S}'})^c \mathcal{E}_{\mathcal{S}}(\rho)) - H(\mathcal{E}^c_{\mathcal{T}'\mathcal{S}'}\mathcal{E}_{\mathcal{S}}(\rho)) \; . \tag{3.26}$$

We have however that $\mathcal{E}_{\mathcal{S}'}\mathcal{E}_{\mathcal{T}'\mathcal{S}'} = \mathcal{E}_{\mathcal{S}'}$, that $\mathcal{E}^c_{\mathcal{S}'} = \mathcal{E}^c_{\mathcal{S}'}\mathcal{E}_{\mathcal{S}}$, and that $\mathcal{E}^c_{\mathcal{T}'\mathcal{S}'}\mathcal{E}_{\mathcal{S}} = \mathcal{E}^c_{\mathcal{T}'\mathcal{S}'}\mathcal{E}_{\mathcal{S}\cap\mathcal{T}}\mathcal{E}_{\mathcal{S}} = \mathcal{E}^c_{\mathcal{T}'\mathcal{S}'}$, so we are left with

$$H(\mathcal{E}^c_{\mathcal{S}'}(\rho)) - H(\mathcal{E}^c_{\mathcal{T}'\mathcal{S}'}(\rho)) \; . \tag{3.27}$$

Since $|\psi\rangle\langle\psi|$ is pure, this becomes $H(\mathcal{S}')_{|\psi\rangle\langle\psi|} - H(\mathcal{T}'\mathcal{S}')_{|\psi\rangle\langle\psi|}$. This implies $I(\mathcal{S} : \mathcal{T} \subset \mathcal{S}\mathcal{T})_\rho \leq I(\mathcal{S}' : \mathcal{T}' \subset \mathcal{S}'\mathcal{T}')_{|\psi\rangle\langle\psi|}$. $\mathcal{M}$ being a factor implies the Theorem via double-commutants. ■

In fact, the phrase "recovered complement" is ambiguous unless we specify how to construct the a recovery map from a given state and channel. While the Petz map is a concrete candidate with many nice properties, we may consider optimizing recoverability. Let $\nu : (S_1(A) \to S_1(B)) \to [0, 1]$ be a measure on quantum channels and $\mu : S_1(A) \times (S_1(A) \to S_1(B)) \to [0, 1]$ be a measure on states for any input channel. Then

$$D_{chan}(\Phi, \rho|\mu, \nu) \equiv \int_R D(\rho\|R \circ \Phi(\rho)) d\nu(R) : R \text{ minimizes } \int_\eta D(\eta\|R \circ \Phi(\eta)) d\mu(\eta|R) \; . \tag{3.28}$$

We may define μ and ν to be uniform measures, or we may define μ to be the worst case measure that for each R maximizes $D(\eta\|R \circ \Phi(\eta))$. In either of these cases, if $D_{chan}(\Phi, \rho|\mu, \nu) = 0$, then the channel is essentially perfect and gives no information about the input state. Conversely, if $D_{chan}(\Phi^c, \rho|\mu, \nu) = 0$ for either of these cases (or for any faithful μ), then the channel is trivial - in this case the environment perfectly reconstructs the input.

For a usual quantum Stinespring dilation, if a channel's environment contains a perfect (up to recovery) copy of the input, then the output contains no information about it. That a perfect environment leaves no information in the output follows from taking the input to be maximally entangled with a reference system, and noting that after applying complement and recovery, the environment remains maximally entangled with the output (which may contain additional correlations with the environment that are input-independent). In contrast, a classical channel (taking classical input states) has no such restriction - here the environment may keep a perfect copy regardless of the output. This is one manifestation of the connection between quantum information and privacy. Much of the intuition behind quantum cryptography is that if the environment has retained information about the input, then the channel must disturb output statistics.

3.3 Higher-Order Inclusion-Exclusion Entropy

Definition 3.3. *Let $\mathcal{N}_1, ..., \mathcal{N}_L \subseteq \mathcal{M}$ be L von Neumann subalgebras of $\mathcal{M}$ for $L \in \mathbb{Z}^+$. Let ρ, σ be densities such that $\mathcal{E}_j(\sigma) = \sigma$ for some j. Let $\mathcal{N} = \cap_{i=1}^{L} \mathcal{N}_i$. Then we define*

$$I(\mathcal{N}_1, ..., \mathcal{N}_L)_{\rho,\sigma} \equiv \sum_{l=1}^{L} (-1)^l \sum_{\mathcal{C}_{l,L}} D(\rho^{\cup_{k \in \mathcal{C}_{l,L}} \mathcal{N}_k} \| \sigma^{\cup_{k \in \mathcal{C}_{l,L}} \mathcal{N}_k}) + D(\rho^{\mathcal{N}} \| \sigma^{\mathcal{N}}) \ ,$$

where each $\mathcal{C}_{l,L}$ is a combination of l integers chosen from $\{1...L\}$, and the sum is over all possible such combinations.

We denote $I(\mathcal{N}_1, ..., \mathcal{N}_L)_\rho \equiv I(\mathcal{N}_1, ..., \mathcal{N}_L)_{\rho,\hat{1}/d} = I(\mathcal{N}_1, ..., \mathcal{N}_L)_{\rho,\hat{1}}$ when $\hat{1}/d$ exists. One can easily check that

$$I(\mathcal{N}_1, ..., \mathcal{N}_L)_{\rho,\hat{1}/d} = \sum_{l=1}^{L} (-1)^{l+1} \sum_{\mathcal{C}_{l,L}} H(\rho^{\cup_{k \in \mathcal{C}_{l,L}} \mathcal{N}_k})_\rho - H(\mathcal{N})_\rho \tag{3.29}$$

for the general $\mathcal{N}$-subalgebra case in finite dimension. In fact, due to cancellation of the $\log d$ terms, we can replace $\hat{1}/d$ by $\hat{1}$ and find the same result. This generalizes to infinite dimensional, tracial von Neumann algebras. We therefore denote

$$I(\mathcal{N}_1, ..., \mathcal{N}_L)_\rho \equiv I(\mathcal{N}_1, ..., \mathcal{N}_L)_{\rho,\hat{1}} \ . \tag{3.30}$$

Conditional mutual information is the $L = 2$ case. The quantity I_3 posited as an indicator of quantum chaos [78] is another special case, as $I_3(A : B : C) = I(\mathcal{A}, \mathcal{B}, \mathcal{C})_{\rho,\hat{1}}$. For arbitrarily many parties, the subsystem case of $I(N_1, ..., N_L)$ is a quantum version of the *multivariate mutual information* introduced by McGill [79] and later (presumed independently) by Ting [80]. The general form as given in Definition 3.3 also makes sense in type III. The key feature of these entropies is that

$$I(\mathcal{N}_1, ..., \mathcal{N}_L)_{\rho,\sigma} = I(\mathcal{N}_1, ..., \mathcal{N}_{L-1})_{\rho,\sigma} - I(\mathcal{N}_1 \cup \mathcal{N}_L, ..., \mathcal{N}_{L-1} \cup \mathcal{N}_L)_{\rho,\sigma} \ . \tag{3.31}$$

In particular,

$$I_3(A:B:C) = I(A:B) - I(A:B|C)\ .$$

Equation (3.31) implies that $I(\mathcal{N}_1, ..., \mathcal{N}_{L-1}) = 0$ if $\mathcal{N}_j$ is in tensor position and uncorrelated with $\cup_{k\neq j}\mathcal{N}_k$ for any $j \in 1...L$. In that sense $I(\mathcal{N}_1, ..., \mathcal{N}_L)$ measures genuine L-party entanglement, discounting correlations between smaller numbers of parties. Following Theorem 3.4:

Corollary 3.1. *Let $\mathcal{N}_1, ..., \mathcal{N}_L \subseteq \mathcal{M}$ be von Neumann subfactors in finite dimension, and let $\rho = tr_E(|\psi\rangle\langle\psi|^{\mathcal{N}_1 \dots \mathcal{N}_L E})$ for some purification. Then*

$$I(\mathcal{N}_1, ..., \mathcal{N}_L)_{\rho,\hat{1}} = I(\mathcal{N}_1', ..., \mathcal{N}_L')_{|\psi\rangle\langle\psi|,\hat{1}}\ .$$

If $\mathcal{N}_1, ..., \mathcal{N}_L$ are subfactors, then it also holds that for odd L,

$$I(\mathcal{N}_1, ..., \mathcal{N}_L)_{\rho,\hat{1}} = -I(\mathcal{N}_1', ..., \mathcal{N}_L')_{|\psi\rangle\langle\psi|,\hat{1}}\ .$$

Corollary 3.1 follows from Equation (3.11). I had conjectured that $I(\mathcal{N}_1, ..., \mathcal{N}_L)_{|\psi\rangle\langle\psi|}$ could have definite sign for all even values of L, but this was falsified by a counterexample from Joshua Levin and Graeme Smith [81]. The $I(\mathcal{N}_1, ..., \mathcal{N}_L)$ form nonetheless has properties suggesting future work on possible interpretations.

3.4 Channels as Views of Quantum Systems

We end this chapter by replacing the conditional expectations in GCMI by quantum channels in specific circumstances in which it remains well-defined, positive, and monotonic. Inspired by the duality between the Heisenberg and Schrödinger pictures of quantum mechanics, we may think of degradations to a quantum state instead as degradations to an observer's access to that state. For some classes of state-modifying channels, we can modify the conditional expectations instead of the state itself. This yields an SSA-type inequality and generalized mutual information for pairs of channels.

Theorem 3.5. *Let $\mathcal{S}, \mathcal{T} \subseteq \mathcal{S} \vee \mathcal{T} \subseteq$ form a commuting square so that $\mathcal{S} \vee \mathcal{T}$ is a factor. Let $\rho^{\mathcal{S}\vee\mathcal{T}}$ be a density, and $\Phi : S_1(\mathcal{S} \vee \mathcal{T}) \to S_1(\mathcal{S} \vee \mathcal{T})$ and $\Psi : S_1(\mathcal{S} \vee \mathcal{T}) \to S_1(\mathcal{S} \vee \mathcal{T})$ be quantum channels such that:*

1. $[\mathcal{E}_{\mathcal{S}}, \Phi] = [\mathcal{E}_{\mathcal{T}}, \Psi] = [\mathcal{E}_{\mathcal{S}\mathcal{T}}, \Phi] = [\mathcal{E}_{\mathcal{S}\vee\mathcal{T}}, \Psi] = 0$.

2. $\mathcal{E}_{\mathcal{T}}\Phi = \Phi\mathcal{E}_{\mathcal{T}} = \mathcal{E}_{\mathcal{T}}$, *and* $\mathcal{E}_{\mathcal{S}}\Psi = \Psi\mathcal{E}_{\mathcal{S}} = \mathcal{E}_{\mathcal{S}}$.

3. $[\Phi, \Psi] = 0$.

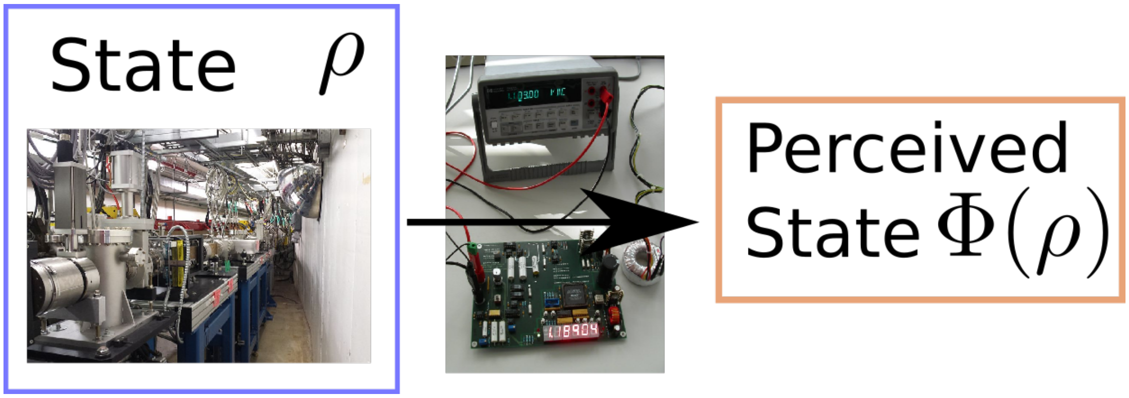

Figure 3.2: A physical system's mixed state corresponds to density ρ. A typical observer interacts with the system through layers of equipment and perception, which may impart locality, noise, basis restrictions, and other limitations. To account for imperfect access, we model the observer's perceived state by $\Phi(\rho)$. Voltmeter image CC BY-SA 3.0, `https://commons.wikimedia.org/w/index.php?curid=41945`. Photon Source Image by Alyssa Haroldsen - Own work, CC0, `https://commons.wikimedia.org/w/index.php?curid=32775941`.

Let $\Phi_S = \Phi \circ \mathcal{E}_{\mathcal{S}}$, and $\Psi_{\mathcal{T}} = \Psi \circ \mathcal{T}$. Then

$$\begin{aligned} I(\Phi_{\mathcal{S}} : \Psi_{\mathcal{T}})_\rho &\equiv H(\Phi_{\mathcal{S}}(\rho)) + H(\Psi_{\mathcal{T}}(\rho)) - H(\Phi \circ \Psi(\mathcal{E}_{\mathcal{S} \vee \mathcal{T}}(\rho))) - H(\mathcal{E}_{\mathcal{S} \cap \mathcal{T}}(\rho)) \\ &\geq -2 \log(F(\Phi \circ \Psi(\rho), R_{\Phi_{\mathcal{S}}(\rho), \mathcal{E}_{\mathcal{T}}} \circ \Psi_{\mathcal{T}}(\rho))) \geq 0. \end{aligned} \tag{3.32}$$

$I(\Phi_{\mathcal{S}} : \Psi_{\mathcal{T}})_\rho$ is monotonic under the transformations $\Phi \to \Theta \circ \Phi$ and $\Psi \to \Gamma \circ \Psi$ such that $\Theta \circ \Phi$ and $\Gamma \circ \Psi$ still obey conditions 1-3. In particular,

$$I(\Phi_{\mathcal{S}} : \Psi_{\mathcal{T}})_\rho - I(\Theta \circ \Phi_{\mathcal{S}} : \Psi_{\mathcal{T}})_\rho \geq -2 \log(F(\mathcal{E}_{\mathcal{S} \vee \mathcal{T}}(\Phi \circ \Psi(\rho)), R_{\mathcal{E}_{\mathcal{S}}(\Phi \circ \Psi(\rho)), \Theta} \circ \Theta(\mathcal{E}_{\mathcal{S} \vee \mathcal{T}}(\Phi \circ \Psi(\rho))))), \tag{3.33}$$

and similarly for $\Psi \to \Gamma \circ \Psi$, where R is a universal recovery map in the sense of [57, 56].

Remark 3.1. Let Φ and Ψ be any pair of quantum channels. We may write $\Phi = \Phi \circ \mathcal{E}_{\mathcal{S}}$, and $\Psi = \Psi \circ \mathcal{E}_{\mathcal{T}}$, where $\mathcal{E}_{\mathcal{S}}$ and $\mathcal{E}_{\mathcal{T}}$ are the conditional expectations onto the respective injective envelopes [83] of the output spaces of $\Phi^\dagger$ and $\Psi^\dagger$. The injective envelope gives the smallest such subalgebras. We may then check if $\Phi, \Psi, \mathcal{S}$, and $\mathcal{T}$ satisfy the conditions of Theorem 3.5. In general, we will get the strongest results from minimal $\mathcal{S}$ and $\mathcal{T}$.

Remark 3.2. We may extend Theorem 3.5 by applying another channel $\Omega : S_1(M) \to S_1(M)$ to ρ before all others. This simply replaces $\rho \to \Omega(\rho)$.

Example 3.1. ***(Uncertainty Relation with Memory and Noise)*** *Let $\mathcal{X}_A$ and $\mathcal{Z}_A$ be a pair of mutually unbiased bases with shift generators X_A and Z_A, such that $X_A Z_A = e^{i\theta} Z_A X_A$. Let $\Phi(\rho) = (1-p)\rho + pX_A \rho X_A$, and $\Psi = (1-q)\rho + qZ_A \rho Z_A$ be dephasing channels. Let B be an auxiliary memory system. We* *see that*

$$\begin{aligned}\Phi \circ \Psi(\rho) =&((1-p)(1-q)\rho + p(1-q)X_A\rho X_A \\ &+(1-p)qZ_A\rho Z_A + pqZ_AX_A\rho X_AZ_A\Big) = \Psi \circ \Phi(\rho).\end{aligned} \tag{3.34}$$

By substituting p or q with 1/2, $[\mathcal{E}_{\mathcal{Z}_A}, \Phi] = [\mathcal{E}_{\mathcal{X}_A}, \Psi] = 0$*. We also have that* $\mathcal{E}_{\mathcal{X}_A}\Phi = \Phi\mathcal{E}_{\mathcal{X}_A} = \mathcal{E}_{\mathcal{X}_A}$*, and* $\mathcal{E}_{\mathcal{Z}_A}\Psi = \Psi\mathcal{E}_{\mathcal{Z}_A} = \mathcal{E}_{\mathcal{Z}_A}$*. Hence* Φ *and* Ψ *satisfy the conditions of Theorem 3.5. Let* B *be an auxiliary or memory system, and* ρ^{AB} *be the full system-memory state. Then*

$$\begin{aligned}&H(\Phi_{\mathcal{Z}_A}|B)_\rho + H(\Psi_{\mathcal{X}_B}|B)_\rho - H(\Phi \circ \Psi(\rho)) \\ &\geq \log|A| - 2\log F(\Phi \circ \Psi(\rho), R_{\Phi\circ\mathcal{E}_{\mathcal{Z}_A}(\rho),\mathcal{E}_{\mathcal{X}_A}} \circ \Psi \circ \mathcal{E}_{\mathcal{X}_A}(\rho))\end{aligned} \tag{3.35}$$

via the Petz recovery map, where $H(\Phi_{\mathcal{Z}_A}|B)_\rho = H(\Phi_{\mathcal{Z}_A} \otimes \hat{1}^B(\rho)) - H(B)_\rho$*. Physically,* Φ *is a dephasing in the* $\mathcal{X}$ *basis, which appears as random shift noise in the* $\mathcal{Z}$ *basis but has no effect on an already fully* $\mathcal{X}$*-dephased state or* $\mathcal{X}$*-basis measurement. Hence* Φ *and* Ψ *each play the role of partially dephasing noise applied to the* $\mathcal{Z}$ *and* $\mathcal{X}$ *bases respectively.*

While $I(\Phi_\mathcal{S} : \Psi_\mathcal{T})_\rho$ is a very general positive and monotonic bipartite information, it requires many assumptions on Φ and Ψ, including that they factor through subalgebras and commute. In [24], Gao, Junge and I derive inequalities on channels without these assumptions. In general, the absence of a commuting square implies the existence of at least some cases with non-positive GCMI. While [24] is more geared toward proving the most general inequality possible, this section focuses on the cases for which we still have a potential operational interpretation in terms of bipartite, individual operations. It naturally leads to Chapter 4, which focuses on the operational interpretations of subalgebraic mutual information. It is through the operational picture that we will see why Theorem 3.5 must hold (see Remark 4.1), so we forgo the proof for now.

3.4.1 What Makes Conditional Expectations Special?

Ultimately, the proof that $I(\mathcal{S} : \mathcal{T})_\rho \geq 0$ and is monotonic under a reasonable set of operations (as discussed in Chapter 4) comes down to the fact that

$$I(\mathcal{S} : \mathcal{T})_\rho = D^\mathcal{S}(\rho) - D^\mathcal{S}(\rho^\mathcal{T}) = D^\mathcal{T}(\rho) - D^\mathcal{T}(\rho^\mathcal{S}) \ ,$$

which follows from the commuting square condition and Lemma 3.5. After that, we may apply data processing either to relate the two terms, or to the first term in either, such as when ρ undergoes a channel Φ that commutes with $\mathcal{E}_{\mathcal{S}}$ and is absorbed by $\mathcal{E}_{\mathcal{T}}$. A key consequence of Lemma 3.5 is that for any conditional expectation $\mathcal{E}$,

$$H(\mathcal{E}(\rho)) - H(\rho) = D(\rho \| \mathcal{E}(\rho)) \ .$$

In words, the entropy difference between ρ and $\mathcal{E}(\rho)$ is equal to the relative entropy. In general,

$$H(\Phi(\rho)) - H(\rho) \neq D(\rho \| \Phi(\rho))$$

for an arbitrary channel Φ, and we must require the input and output spaces of Φ to be the same for the latter to even make sense. Intuitively, a conditional expectation is a special degradation of a quantum state that introduces no rotation of bases, so it naturally guarantees that ρ and $\mathcal{E}(\rho)$ remain comparable. A mathematical notion of "rotation free" might be that $\Phi = \Phi^\dagger$ up to normalization under the inner product with respect to the partial trace, which we might interpret further as $\Phi = \mathcal{R}_{\hat{1}/d,\Phi}$, where $\mathcal{R}_{\hat{1}/d,\Phi}$ is the Petz map. A conditional expectation is always its own Petz map. While necessary, this condition is not entirely sufficient. The depolarizing channel (see for instance Equation (7.9)) is its own Petz recovery map, but it does not obey the chain rule of relative entropy. We require idempotence as well as self-adjointness, which imposes that the channel be a projector. To fully obtain the chain expansion as in Lemma 3.5, we actually need that the channel restrict to a subalgebra and be a conditional expectation.

In the course of this project, Junge and I spent substantial time trying to generalize to a meaningful notion of $I(\Phi : \Psi)$ for an arbitrary pair of channels Φ and Ψ. A major issue we encountered is the ambiguity in what would constitute a meaningful notion of an intersection or a union between the outputs of two channels. When the channels are conditional expectations to subalgebras given explicitly by generators in a larger joint algebra, it is clear at least in finite dimension how to compute the joint algebra as that generated by the union of generators, and an intersection algebra via a literal intersection of elements. A particular hint came from the equivalence to commutant algebras as in Theorem 3.4, in which it was necessary to use this minimal joint system (not a larger algebra) and maximal intersection (not a smaller algebra). Another hint appeared in deriving the results of Chapter 4, where choosing e.g. a larger-than-necessary joint algebra may break monotonicity of some natural operations. Hence we may guess that if there is a reasonable generalization to other channels, we would expect the joint system to be minimal in some sense, the intersection to be maximal, and that there is some notion of complementarity under which the complements of these would exchange roles.

One exception to the restriction that we consider conditional expectations is the mutual information, as

$$I(A : B)_\rho = D(\rho \| \rho^A \otimes \rho^B) = H(\rho^A \otimes \rho^B) - H(\rho^{AB}) \tag{3.36}$$

for a bipartite density ρ^{AB} on $A \otimes B$. While there is no conditional expectation, and in fact no channel mapping $\rho^{AB} \to \rho^A \otimes \rho^B$, the mutual information still reduces to a single entropy difference. The product states are not even a convex set. This may hint that there is a generalization beyond subalgebras that does not rely on quantum channels. This is a potential area of future research.

Chapter 4

Generalizing Non-Classicality

Several features of quantum physics distinguish it from that of classically stochastic systems. The uncertainty principle is one of these features, in which different potential measurements on a single system cannot simultaneously have definite outcomes. Another of these features is quantum entanglement, which guarantees that distant experiments will obtain correlated outcomes in each of several of these incompatible measurements. Entanglement and uncertainty are two fundamentally quantum phenomena that underpin research areas ranging from high-energy physics to quantum computing.

Quantum entanglement requires separated systems, so unlike the uncertainty principle, there is no way to observe it on quantum systems of dimension 2 or 3. The uncertainty principle has a simple form on qubits. Here we may construct 3 maximally incompatible measurements as Pauli observables X, Y, Z. Any state that would have a definite outcome of one such measurement is maximally random in the others. In contrast, to divide a system into subsystems and demonstrate entanglement requires that its dimension d (or a smaller positive integer, if one projects to a subspace) factor into a product $d_1 \cdot d_2$, where $d_1, d_2 > 1$. This is impossible when d is 2 or 3.

While we cannot factor a 2-dimensional quantum system into subsystems, we still find an analog of entanglement between pairs of incompatible bases. The underlying mathematical idea is that an observable such as X generates an algebra resembling a 2-dimensional, effectively classical probability space. When we compose algebras from two commuting observables of this type, we obtain a 4-dimensional classical space. Combining algebras from two anticommuting observables, such as X and Z, yields the matrix algebra associated with one qubit (see Section 2.3.2). Replacing subsystems by subalgebras, we generalize information measures used to quantify correlations, such as the conditional mutual information and squashed entanglement. First, we find that the generalized squashed entanglement can be non-zero between the X and Z bases of a single qubit. Second, we extend the theory of entanglement non-increasing operations (consisting of local operations and classical communication, as in [84]) to the subalgebra setting. This extended set of non-increasing operations allows one to convert two basis-split qubits into a pair of entangled qubits.

This chapter includes results appearing in [23], co-authored with Li Gao and Marius Junge.

In this chapter, we restrict to finite dimension, where Theorem 3.5 is the most general form of GCMI inequality for a pair of commuting, algebraically compatible channels. We saw in Chapter 3 the beginning of a link between inequalities on correlations and quantum uncertainty principles. In this chapter, we expand on this idea by examining the operations under which GCMI does not increase, in analogy to the local operations that never create correlations between subsystems. We will ultimately find operations that do not increase certain measures of non-classicality, but support interconversion between that in an uncertainty-based configuration and the usual notion of quantum entanglement.

In Section 4.1, I describe the interpretation of GCMI as a correlation-like resource. In Section 4.2, I recall resources measures of coherence and asymmetry and relate them to the relative entropy with respect to a subalgebra. In Section 4.3, I describe generalizations of entanglement measures based on GCMI to subalgebraic non-classicality. In Section 4.4, I describe a conversion protocol between subsystem entanglement and the non-classicality measured in an uncertainty relation.

4.1 Subalgebraic, Bipartite Correlations as a Resource

First, we construct a resource theory analogous to bipartite local operations (LO), under which CMI does not increase. The main ingredients to any resource theory are:

1. A set of *free operations* that may consume or convert, but not create the quantity defined as a resource.
2. A set of *free states* that have zero resource value.
3. One or more *monotones*, which are zero for free states, non-increasing under free operations, and interpreted as quantifying the resource value of a given state.

See [2] for background on the defining aspects of a resource theory. Resource theories abound in thermodynamics, where one converts and expends reserves of energy to perform heat or work. Another famous resource theory in quantum information is that of entanglement, which is non-increasing under local operations and classical communication. For a bipartite state ρ^{AB}, local operations would consist of all channels acting entirely on either A or B.

We generalize local operations to individual operations on a pair of algebras:

Definition 4.1 (Individual Operations, based on definitions from [23])**.** *Let $\mathcal{S}, \mathcal{T} \subseteq \mathcal{S} \vee \mathcal{T}$ and $\rho \in S_1(\mathcal{S} \vee \mathcal{T})$ be a density. We define an $\mathcal{S}$-operation ($\mathcal{S}$-op for short) as a sequence of the following:*

1. *We may extend $\mathcal{S} \to \mathcal{S} \otimes \mathcal{C}, \mathcal{T} \to \mathcal{T} \otimes 1, \rho \to \rho \otimes \sigma^C$, for any extra factor C with density σ.*

2. *We may transform $\rho \to \Phi(\rho)$ for any channel $\Phi : S_1(\mathcal{S} \vee \mathcal{T}) \to S_1(\mathcal{S} \vee \mathcal{T})$ that is $\mathcal{T}$-preserving up to isometry and satisfies $[\Phi, \mathcal{E}_{\mathcal{S}}] = 0$, without explicitly changing $\mathcal{S}$ or $\mathcal{T}$.*

3. *We may reduce $\mathcal{S} \to \tilde{\mathcal{S}}, \rho \to \mathcal{E}_{\tilde{\mathcal{S}} \vee \mathcal{T}}(\rho)$ without changing $\mathcal{T}$ for $\tilde{\mathcal{S}}$ satisfying $\tilde{\mathcal{S}} \subseteq \mathcal{S}$, $\mathcal{S} \cap \mathcal{T} = \tilde{\mathcal{S}} \cap \mathcal{T}$, and $\mathcal{E}_{\tilde{\mathcal{S}} \vee \mathcal{T}} \mathcal{E}_{\mathcal{S}} = \mathcal{E}_{\tilde{\mathcal{S}}}$.*

We define $\mathcal{T}$-operations analogously.

Directly following, we find:

Theorem 4.1 (based on theorem 2.12 in [23]). *Under an $\mathcal{S}$-operation for which $\mathcal{S} \to \tilde{\mathcal{S}}$ and $\rho \to \tilde{\rho}$,*

$$I(\mathcal{S} : \mathcal{T})_\rho \geq I(\tilde{\mathcal{S}} : \mathcal{T})_{\tilde{\rho}} \; . \tag{4.1}$$

As a consequence, I is non-increasing under any sequence of $\mathcal{S}$-ops and $\mathcal{T}$-ops.

proof of Theorem 4.1. We prove this by showing the monotonicity for each class of operations:

1. We use the additivity of entropy for product states. $H(\mathcal{T})$ and $H(\mathcal{S} \cap \mathcal{T})$ each gain contributions of $\log |C|$, which cancel. $H(\mathcal{S})$ and $H(\mathcal{S} \vee \mathcal{T})$ each gain contributions of $H(\sigma)$, which cancel between them.

2. Under the assumption that $\mathcal{E}_{\mathcal{T}}(\Phi(\rho)) = U \mathcal{E}_{\mathcal{T}}(\rho) U^\dagger$ for some unitary U, $H(\mathcal{T})_\rho = H(\mathcal{T})_{\Phi(\rho)}$ is unchanged. Moreover, by $[\Phi, \mathcal{E}_{\mathcal{S}}] = 0$,

$$\mathcal{E}_{\mathcal{S} \cap \mathcal{T}}(\Phi(\rho)) = \mathcal{E}_{\mathcal{T}} \mathcal{E}_{\mathcal{S}} \Phi(\rho) = \mathcal{E}_{\mathcal{T}} \Phi \mathcal{E}_{\mathcal{S}}(\rho) = U \mathcal{E}_{\mathcal{T} \cap \mathcal{S}}(\rho) U^\dagger \; , \tag{4.2}$$

and $H(\mathcal{S} \cap \mathcal{T})_\rho = H(\mathcal{S} \cap \mathcal{T})_{\Phi(\rho)}$. For the other two terms $H(\mathcal{S})_\rho - H(\mathcal{S} \vee \mathcal{T})_\rho = D^{\mathcal{S}}(\rho)$ is non-increasing under Φ that commutes with $\mathcal{E}_{\mathcal{S}}$.

3. $H(\mathcal{S} \cap \mathcal{T}) = H(\tilde{\mathcal{S}} \cap \mathcal{T})$ and $H(\mathcal{T})$ are unchanged. Again by the assumption $\mathcal{E}_{\tilde{\mathcal{S}} \vee \mathcal{T}} \mathcal{E}_{\mathcal{S}} = \mathcal{E}_{\tilde{\mathcal{S}}}$,

$$H(\mathcal{S})_\rho - H(\mathcal{S} \vee \mathcal{T})_\rho = D^{\mathcal{S}}(\rho) \geq D(\mathcal{E}_{\tilde{\mathcal{S}} \vee \mathcal{T}}(\rho) \| \mathcal{E}_{\tilde{\mathcal{S}} \vee \mathcal{T}}(\mathcal{E}_{\mathcal{S}}(\rho))) \geq D(\mathcal{E}_{\tilde{\mathcal{S}} \vee \mathcal{T}}(\rho) \| \mathcal{E}_{\tilde{\mathcal{S}}}(\rho)) \; .$$

■

As should be expected, individual operations restrict to local operations if we impose the constraint that $\mathcal{S}, \mathcal{T}$ must be factors. If they overlap, then their intersection becomes an auxiliary system. We thereby recover monotonicity of CMI under local operations. In general, $\mathcal{S}, \mathcal{T}$ need not be subsystems. A canonical such example is when we choose $\mathcal{S} = \mathcal{X}$, the algebra of observables in the Pauli $\mathcal{X}$ basis, and $\mathcal{T} = \mathcal{Z}$, that in the Pauli $\mathcal{Z}$ basis. In this case we still find that $\mathcal{X} \cap \mathcal{Z} = \mathbb{C}$, so there is no overlap between subsystems. Nonetheless, $\mathcal{X} \vee \mathcal{Z} = \mathcal{H}_2$, a qubit Hilbert space.

Remark 4.1. *Theorem 4.1 leads directly to Theorem 3.5, as we can rewrite $I(\Phi_{\mathcal{S}} : \Psi_{\mathcal{T}})$ as a $I(\mathcal{S} : \mathcal{T})_{\Phi \circ \Psi(\rho)}$, where Φ and Ψ are individual operations.*

We may also denote

$$I(\mathfrak{S} : \mathfrak{T} | \mathcal{S} \cap \mathcal{T}) \equiv I(\mathcal{S} : \mathcal{T}) \ . \tag{4.3}$$

This notation makes it clearer that $I(\mathfrak{S} : \mathfrak{T} | \mathcal{S} \cap \mathcal{T})$ does not allow the parties holding $\mathcal{S}$ and $\mathcal{T}$ to modify the intersection, and that $\mathcal{S} \cap \mathcal{T}$ should not be considered shared when studying correlations or operations.

Remark 4.2. *The individual operations defined in 4.1 essentially gives the intuition behind and proof for Theorem 3.5, and also suggest why it does not go much further than positivity of GCMI. In particular, we may rewrite*

$$I(\Phi_{\mathcal{S}} : \Psi_{\mathcal{T}})_\rho = I(\mathcal{S} : \mathcal{T})_{\Phi\Psi(\rho)} \ ,$$

recalling that $\Phi\Psi = \Psi\Phi$. Showing this inequality is a simple matter of applying the commutation and absorption relations assumed by Theorem 3.5. The main reason why this works is that these channels are individual operations.

(a) (b)

Figure 4.1: (a) Alice and Bob are two spatially-separated parties. No information written by Alice would be observable to Bob until sufficient time has passed for light to propagate. (b) Alice and Bob are not spatially separated, but their intersection is locked by imposed conditions. Any information written by Alice remains invisible to Bob. Images use LibreOffice clipart.

Definition 4.1 might not be the most general form of operations under which $I(\mathcal{S} : \mathcal{T})_\rho$ is monotonic. We could instead attempt to define a notion of individual operations in a game-like setting. Let us consider a set of experiments available to Alice, $\{a_1, ..., a_n\}$ and respectively to Bob, $\{b_1, ..., b_m\}$. These might index observables, or operators in a positive operator-valued measurement (POVM). We might then allow Alice and Bob to each perform any operations on the joint system in any order, as long as the other's experiments cannot tell whether those actions occurred. This guarantees that no communication occurs between Alice and Bob. Unfortunately, it may not entirely ensure monotonicity of GCMI - one could for example imagine that both players start with individual mixed states, but they replace those states by an entangled or correlated pair. What we will see shortly, however, is that we can add some free operations that change the algebras in more interesting ways. In this section, we also note:

Remark 4.3. Some I-non-increasing transformations are not really $\mathcal{S}$ and $\mathcal{T}$-operations but change the state and algebra in compensatory ways.

1. $I(\mathcal{S} : \mathcal{T})_\rho$ is unchanged under $\rho \to U\rho U^\dagger, \mathcal{S} \to \tilde{\mathcal{S}} = U\mathcal{S}U^\dagger$ and $\mathcal{T} \to \tilde{\mathcal{T}} = U\mathcal{T}U^\dagger$ for any isometry U. If U-conjugation commutes with $\mathcal{E}_\mathcal{S}, \mathcal{E}_\mathcal{T}$ and $\mathcal{E}_{\mathcal{S}\vee\mathcal{T}}$, then $\mathcal{S}, \mathcal{T} = \tilde{\mathcal{S}}, \tilde{\mathcal{T}}$, and we may change only the density. These are essentially coordinate changes.

2. $I(\mathcal{S} : \mathcal{T})$ is non-increasing under change of the algebra $\mathcal{T} \to \tilde{\mathcal{T}}, \rho \to \rho$ for which $\mathcal{T} \subseteq \tilde{\mathcal{T}}$, $\mathcal{S}$ and $\mathcal{S} \vee \mathcal{T} = \mathcal{S} \vee \tilde{\mathcal{T}}$ are unchanged, and both $\mathcal{T}, \mathcal{S}$ and $\tilde{\mathcal{T}}, \mathcal{S}$ form commuting squares. This locks elements of $\mathcal{S}$ in $\mathcal{S} \cap \mathcal{T}$.

3. Let $\mathcal{S}, \mathcal{T} \subseteq \mathcal{S} \vee \mathcal{T}$ be a commuting square. Let $\mathcal{E}_\mathcal{R}$ be a conditional expectation onto a subalgebra $\mathcal{R}$ and $\tilde{\mathcal{S}}$ be another algebra such that $\mathcal{T} \subset \mathcal{R}, \mathcal{R} \cap \mathcal{S} = \mathcal{R} \cap \tilde{\mathcal{S}}$, and $\mathcal{E}_\mathcal{R}$ commutes with $\mathcal{E}_\mathcal{S}$ and $\mathcal{E}_{\tilde{\mathcal{S}}}$. $I(\mathcal{S} : \mathcal{T})$ is non-increasing under the transformation $\mathcal{S} \to \tilde{\mathcal{S}}, \rho \to \mathcal{E}_\mathcal{R}(\rho)$ with $\mathcal{T}$ unchanged.

Since Alice and Bob are not in tensor position, the intuitive concept of spatial separation generally need not apply. For example, there is no way to have two parties share complementary bases of a single qubit without co-location. As illustrated in Figure 4.1, we replace spatial separation by secrecy. The individual channel Φ as in step 2 of Definition 4.1 have the interpretation of obeying the following constraints:

- $[\Phi, \mathcal{E}_\mathcal{S}] = 0$ implies all observables in $\mathcal{S}$ and $\mathcal{S} \vee \mathcal{T}$ perceive it as the same channel.

- $\Phi\mathcal{E}_\mathcal{T} = \mathcal{E}_\mathcal{T}\Phi = \mathcal{E}_\mathcal{T}$ implies invisibility to all experiments in $\mathcal{T}$.

The other steps in individual operations, as well as those from Remark 4.3 also forbid communication between an observer with access to $\mathcal{S}$ and one with access to $\mathcal{T}$. At times this may forbid some procedures that would intuitively be available. For example, a pair of observers holding the Pauli observables $\mathcal{X}$ and $\mathcal{Z}$ would each be forbidden from measuring their own observable, which would be sometimes detectable by an observer holding the other. They could however agree that one will drop access to observables in their subalgebra as in step 3 of an individual operation, and the other measures. Local operations are mutually commuting, so they do not require time-ordering, while free operations in more general commuting square settings may not commute. Even in the usual setting of local operations and classical communication, communication between the parties breaks the commutation of local operations and makes this class much more complicated [85]. In the rest of this chapter, we will see some applications of operations that may consume coordinated shared randomness or classical communication.

4.2 Coherence & Asymmetry

In this section, we return to the original theme of GCMI's positivity as unifying uncertainty-like and correlation-like information measures. First, we define the notation

$$D^{\mathcal{S}}(\rho) = D(\rho \| \mathcal{E}_{\mathcal{S}}(\rho)) \tag{4.4}$$

for an algebra $\mathcal{S}$. We recall the relative entropy of coherence for a density ρ as defined in [86] is actually given as $D^{\mathcal{X}}(\rho)$ for an algebra $\mathcal{X}$ that corresponds to a measurement basis. The relative entropy of coherence is a resource monotone under several sets of operations [87]. Similarly, when $\mathcal{S}$ is the fixed point algebra of a group action such that

$$\mathcal{E}_{\mathcal{S}} = \int g \rho g^{\dagger} dg$$

for some group elements $g \in G$ and for which dg is the Haar or effectively uniform measure, then $D^{\mathcal{S}}$ has the interpretation of the Holevo asymmetry measure, or relative entropy of asymmetry [88, 89, 90]. We also have that for a bipartite density ρ^{AB},

$$D^{\mathcal{A}}(\rho) = \log |B| - H(B|A) \; .$$

We may think of the relative entropy of coherence as corresponding to a states asymmetry with respect to a given basis. We may think of the conditional entropy as closely connected to a state's asymmetry with respect to observables on one subsystem. Hence we may again re-interpret algebraic SSA or (equivalently) positivity of GCMI as *subadditivity of asymmetry*. The form in Equation (3.3) makes this especially apparent.

Finally, when $\mathcal{S} \subseteq \mathcal{M}$, $D(\mathcal{E}_{\mathcal{M}}(\rho) \| \mathcal{E}_{\mathcal{S}}(\rho))$ has parallels with the relative entropy between channels, $D(\Phi(\rho) \| \Psi(\rho)) = \sup_{\rho} D(\Phi(\rho) \| \Psi(\rho))$ as in [91, 92, 93], which maximizes over all possible states and may add an arbitrary auxiliary system in the supremum. In [25] with Li Gao and Marius Junge, we study the special case $D(\mathcal{M} \| \mathcal{S}) = \sup_{\rho} D(\mathcal{E}_{\mathcal{M}}(\rho) \| \mathcal{E}_{\mathcal{S}}(\rho))$.

4.3 Subalgebraic, Bipartite Non-Classicality as a Resource

Many prior efforts have attempted to combine quantum coherence with quantum entanglement. Simultaneous work by Chitambar and Hsieh [94] and by Streltsov et. al. [95] considers a restrictive resource theory of local, incoherent operations and classical communication - in other words, at least one party in these scenarios can neither transmit nor receive quantum information, nor rotate local states out of a single preferred basis. Previous work has investigated the role of discord and entanglement measures in quantifying coherence [96, 97, 98, 99, 100, 101, 102, 103]. Luo and Sun have considered the relationship between coherence and uncertainty [104], and Hu and Fan further address the relationship between basis incompatibility, coherence and correlations [105]. Singh et. al. define an uncertainty relation for coherence [62].

A primary form of precedent for the results herein are the notions of "Generalized Local Operations" and related generalization of entanglement as studied in [106, 107, 108, 109]. These approaches focus on the problem of defining entanglement between indistinguishable particles having non-trivial spin statistics. If we for instance take a pair of qubit fermions in spin states $|0\rangle, |1\rangle$ and respective positions $|x_1\rangle, |x_2\rangle$, we would properly represent them in a Fock space with creation and annihilation operators. It is common however to treat each particle in a fixed-number state as though it were a subsystem, writing the state

$$\frac{1}{\sqrt{2}}(|0, x_1\rangle \otimes |1, x_2\rangle - |1, x_2\rangle \otimes |0, x_1\rangle) .$$

The antisymmetric state formally looks entangled between the two particles, but this is widely regarded as illusory and spurious. This form of "entanglement" formally appears between fermions that are spatially separated and do not appear to have interacted with each other, and it does not confer any measurable quantum effects between the regions containing these particles. The aforementioned works replace subsystems by projectors to subspaces, which allow them to consider the "states" of individual particles without artificial, unphysical entanglement appearing.

Our formalism is similar to those of [106] but with a few key distinctions. First, we restrict our attention to subalgebras, rather than subsystems, obtaining additional theorems at the price of specificity. Second, and more substantially, our goals differ. We do not focus on the formal challenges of modeling identical particles. Rather, our goal is to explain and interpret some apparent connections between entanglement and uncertainty that arise in the algebraic setting. These seem to fundamentally connect several quintessential phenomena in quantum mechanics.

The role of spatial separation in entanglement has been a matter of historical ambiguity. In the strictest sense, entanglement exists only between spatially separated systems. The relativistic lightspeed limit prevents instantaneous information exchange about local measurements on spatially separated systems. Hence non-classical correlations between spatially separated systems distinguish entanglement from alternative explanations [15]. Less strict interpretations of entanglement have considered distinct aspects of a single particle to be entangled with each other. The polarization of a single photon may, for instance, entangle with its location or frequency [110]. A further class of theories have extended more general definitions of entanglement [106, 108, 109, 37] to situations in which discernible subsystems cannot be identified. These varied definitions of entanglement all involve dividing a system into parts, but they differ on what constitutes a valid division.

Prior work like this and Example 4.1 shows that there is precedent for thinking of quantum entanglement in contexts other than spatially-separated systems. While traditional notions of entanglement have it closely connected to non-locality, neither spatial separation nor tensor products of Hilbert spaces are strictly necessary to observe entanglement-like phenomena.

Example 4.1 (Entanglement Between Degrees of Freedom)**.** *While many areas of quantum information by default assign a single aspect of each quantum particle a qubit or qudit degree of freedom (e.g. spin for matter, polarization for photons), quantum particles often have multiple aspects with states specified by quantum degrees of freedom with various dimensions. For example, a single photon carries not only a polarization qubit, but some continuous degrees of freedom corresponding to its position and energy, and a discrete but potentially infinite space of orbital angular momentum. Formally, we might denote the Hilbert space of 1-photon states as*

$$\mathcal{H}_{pol} \otimes \mathcal{H}_{orb} \otimes \mathcal{H}_{cts} \otimes ...$$

Due to this tensor product structure, states on $\mathcal{H}_{pol} \otimes \mathcal{H}_{orb}$ can appear formally entangled. The nature of the photon requires that these degrees of freedom be co-localized, and there is no way to create spatial separation. Nonetheless, "entanglement" between degrees of freedom of single photons appears frequently in optical experiments [111, 112]. The concept of hybrid entanglement combines inter-particle and intra-particle "entanglement." Some of the literature [113, 114] describes these aspects, particularly in optics, as manifesting a classical version of or analog to entanglement.

We define a generalized non-classicality analogous to the squashed entanglement as defined in [115],

$$E_{sq}(A:B)_\rho \equiv \inf_{\tilde{\rho}^{ABC}:\tilde{\rho}^{AB}=\rho^{AB}} I(A:B|C)_{\tilde{\rho}}. \tag{4.5}$$

The squashed entanglement quantifies non-classical correlation between subsystems as the minimal CMI over all possible environments. We may extend this to algebras, using GCMI:

Definition 4.2 (Squashed GCMI, definition 2.15 from [23]). *Let $\mathcal{S}, \mathcal{T}$ form a commuting square such that $\mathcal{S} \cap \mathcal{T} = \mathbb{C}1$. We define the squashed mutual information as*

$$I_{sq}(\mathcal{S} : \mathcal{T})_\rho \equiv \frac{1}{2} \inf_{\tilde{\mathcal{S}}, \tilde{\mathcal{T}}; \tilde{\rho} \in S_1(\tilde{\mathcal{S}} \vee \tilde{\mathcal{T}})} I(\tilde{\mathcal{S}} : \tilde{\mathcal{T}})_{\tilde{\rho}} , \tag{4.6}$$

where the minimization is constrained such that:

- $\tilde{\mathcal{S}}, \tilde{\mathcal{T}} \subseteq \tilde{\mathcal{S}} \vee \tilde{\mathcal{T}}$ *form a commuting square in finite dimension.*
- $\tilde{\mathfrak{S}} = \tilde{\mathcal{S}} \cap (\tilde{\mathcal{S}} \cap \tilde{\mathcal{T}})' = \mathcal{S}$, *and* $\tilde{\mathfrak{T}} = \tilde{\mathcal{T}} \cap (\tilde{\mathcal{S}} \cap \tilde{\mathcal{T}})' = \mathcal{T}$.
- $\tilde{\rho}$ *is a density on* $\tilde{\mathcal{S}} \vee \tilde{\mathcal{T}}$ *such that* $\mathcal{E}_{\tilde{\mathfrak{S}} \vee \tilde{\mathfrak{T}}}(\tilde{\rho}) = \rho$.

Remark 4.4. Were $\tilde{\mathcal{S}} \cap \tilde{\mathcal{T}}$ to contain a non-trivial commutative subalgebra, it would appear in $\tilde{\mathfrak{S}}$ and $\tilde{\mathfrak{T}}$, but this cannot happen when $\tilde{\mathfrak{S}} \cap \tilde{\mathfrak{T}} = \mathcal{S} \cap \mathcal{T} = \mathbb{C}$. Hence

$$I_{sq}(\mathcal{S} : \mathcal{T})_\rho = \frac{1}{2} \inf_{C; \tilde{\rho} \in S_1((\mathcal{S} \vee \mathcal{T}) \otimes \mathcal{C})} I(\mathcal{S} \otimes \mathcal{C} : \mathcal{T} \otimes \mathcal{C})_{\tilde{\rho}} \tag{4.7}$$

with the minimization constrained such that $\mathrm{tr}_C(\tilde{\rho}) = \rho$. It is not clear how $\tilde{\mathcal{S}}$ and $\tilde{\mathcal{T}}$ would interact with any non-trivial $\mathcal{S} \cap \mathcal{T}$, so we restrict to the case where there is no overlap.

Definition 4.3 (Convex Roof GCMI, definition 2.17 from [23]). *Let $\mathcal{S}, \mathcal{T}$ form a commuting square with density ρ such that $\mathcal{S} \cap \mathcal{T} = \mathbb{C}1$. We define the following convex roof measure,*

$$I_{conv}(\mathcal{S} : \mathcal{T})_\rho \equiv \frac{1}{2} \inf_{\{(p_x, \rho_x)\}} \sum_x p_x I(\mathcal{S} : \mathcal{T})_{\rho_x} = \inf_{\tilde{\rho} \in (\mathcal{S} \vee \mathcal{T}) \otimes \mathcal{X}} I(\mathcal{S} \otimes \mathcal{X} : \mathcal{T} \otimes \mathcal{X})_{\tilde{\rho}} , \tag{4.8}$$

where the infimum runs all finite family of densities $\rho_x \in S_1(\mathcal{S} \vee \mathcal{T})$ and probability distributions $\{p_x\}$ such that $\rho = \sum_x p_x \rho_x$, or equivalently the extension $\tilde{\rho} = \sum_x p_x \rho_x \otimes |x\rangle\langle x|$ over $(\mathcal{S} \vee \mathcal{T}) \otimes \mathcal{X}$.

We next show some properties of I_{sq} and I_{conv}:

Proposition 4.1 (Proposition 2.21 from [23]). Let $\mathcal{S}, \mathcal{T} \subset \mathcal{S} \vee \mathcal{T}$ form a commuting square with $\mathcal{S} \cap \mathcal{T} = \mathbb{C}1$. Then

i) I_{sq} and I_{conv} are convex in ρ. Hence they attain maximum values on pure states.

ii) If ρ is pure, then $I_{conv}(\mathcal{S} : \mathcal{T})_\rho = I_{sq}(\mathcal{S} : \mathcal{T})_\rho = \frac{1}{2} I(\mathcal{S} : \mathcal{T})_\rho$.

iii) I_{sq} and I_{conv} are continuous with respect to trace distance in ρ.

iv) $I_{sq}(\mathcal{S}:\mathcal{T})_\rho \leq I_{conv}(\mathcal{S}:\mathcal{T})_\rho$.

v) When $\mathcal{S}$,$\mathcal{T}$ are in tensor position $\mathcal{S}\otimes\mathcal{T}$, ρ is separable if $I_{sq}(\mathcal{S}:\mathcal{T})_\rho$ or $I_{conv}(\mathcal{S}:\mathcal{T})_\rho$ is 0.

I_{sq} has an additional property generalizing superadditivity of squashed entanglement:

vi) Let $\mathcal{S}_i, \mathcal{T}_i = \mathcal{S}_i \vee \mathcal{T}_i$ be commuting square for each $i \in 1,\cdots,n$. Then $\mathcal{S} = \otimes_{i=1}^n \mathcal{S}_i, \mathcal{T} = \otimes_{i=1}^n \mathcal{T}_i$ form a commuting square in $\mathcal{M} = \otimes_{i=1}^n(\mathcal{S}_i \vee \mathcal{T}_i)$, and

$$\sum_i I_{sq}(\mathcal{S}_i:\mathcal{T}_i)_{\mathcal{E}_{\mathcal{M}_i}(\rho)} \leq I_{sq}(\mathcal{S}:\mathcal{T})_\rho.$$

The equality is achieved for all product states $\rho = \otimes_{i=1}^n \rho_i$ such that $\rho_i \in S_1(\mathcal{M}_i)$. Consequently , if $\mathcal{S} = \mathcal{S}_1 \otimes \mathcal{S}_2$, and $\mathcal{T} = \mathcal{T}_1 \otimes \mathcal{T}_2$, then $I_{sq}(\mathcal{S}:\mathcal{T})_\rho \geq I_{sq}(\mathcal{S}_1:\mathcal{T}_1)_\rho$.

Proof of Proposition 4.1, based on proof from [23]. We assume that $\mathcal{M}$ is a minimal matrix algebra such that $\mathcal{S}\vee\mathcal{T} \subset \mathcal{M}$. We will write I_* when the given statement would be true for both I_{sq} and I_{conv}.
i) Let $\rho = \sum_x p_x \rho_x$ be a convex combination. Then

$$I_*(\mathcal{S}:\mathcal{T})_\rho \leq \frac{1}{2}\inf_{\{\tilde{\rho}_x\},\mathcal{C}} I(\mathcal{S}\otimes\mathcal{C}\otimes\mathcal{X}:\mathcal{T}\otimes\mathcal{C}\otimes\mathcal{X})_{\tilde{\rho}} = \sum_x p_x I_*(\mathcal{S}:\mathcal{T})_{\rho_x} \ , \tag{4.9}$$

where each $\tilde{\rho_x}$ is an extension of ρ_x, since the extensions considered herein are a subset of those considered by $I_*(\mathcal{S}:\mathcal{T})_\rho$.
ii) The extension of a pure state $\rho \in \mathcal{S}\vee\mathcal{T}$ can only be a tensor product $\rho\otimes\sigma$. By additivity of entropy of tensored densities,

$$\begin{aligned}
&I(\mathcal{S}\otimes\mathcal{C}:\mathcal{T}\otimes\mathcal{C})_{\rho\otimes\sigma}\\
&= H(\mathcal{S}\otimes\mathcal{C})_{\rho\otimes\sigma} + H(\mathcal{T}\otimes\mathcal{C})_{\rho\otimes\sigma} - H((\mathcal{S}\vee\mathcal{T})\otimes\mathcal{C})_{\rho\otimes\sigma} - H((\mathcal{S}\cap\mathcal{T})\otimes\mathcal{C})_{\rho\otimes\sigma}\\
&= H(\mathcal{S})_\rho + H(\mathcal{T})_\rho - H(\mathcal{S}\vee\mathcal{T})_\rho - H(\mathcal{S}\cap\mathcal{T})_\rho = I(\mathcal{S}:\mathcal{T})_\rho
\end{aligned}$$

iii) We have the following quantitative estimates: if $\|\rho-\eta\|_1 \leq \epsilon < 1$, then

$$|I_{sq}(\mathcal{S}:\mathcal{T})_\rho - I_{sq}(\mathcal{S}:\mathcal{T})_\eta| \leq 12\sqrt{\epsilon}\log|\mathcal{M}| + 3(1+2\sqrt{\epsilon})h\Big(\frac{1}{1+2\sqrt{\epsilon}}\Big) \ , \tag{4.10}$$

$$|I_{conv}(\mathcal{S}:\mathcal{T})_\rho - I_{conv}(\mathcal{S}:\mathcal{T})_\eta| \leq 6\sqrt{\epsilon}\log|\mathcal{M}| + 3(1+2\sqrt{\epsilon})h\Big(\frac{1}{1+2\sqrt{\epsilon}}\Big) \ , \tag{4.11}$$

where $h(p) = -p\log p - (1-p)\log(1-p)$ is the binary entropy function and $|\mathcal{M}|$ is the dimension of $\mathcal{M}$.

The proof follows the original argument in [115] for squashed entanglement. Note that

$$\begin{aligned} I_{sq}(\mathcal{S}:\mathcal{T})_\rho &= \inf_{\mathcal{C},\tilde{\rho}} \big(H(\mathcal{S}\mathcal{C})_{\tilde{\rho}} + H(\mathcal{T}\mathcal{C})_{\tilde{\rho}} - H((\mathcal{S}\vee\mathcal{T})\mathcal{C})_{\tilde{\rho}} - H(\mathcal{C})_{\tilde{\rho}}\big) \\ &= \inf_{\mathcal{C},\sigma} \big(H(\mathcal{M}|\mathcal{C})_{\mathcal{E}_{\mathcal{S}\mathcal{C}}(\tilde{\rho})} + H(\mathcal{M}|\mathcal{C})_{\mathcal{E}_{\mathcal{T}\mathcal{C}}(\tilde{\rho})} - H(\mathcal{M}|\mathcal{C})_{\mathcal{E}_{(\mathcal{S}\vee\mathcal{T})\mathcal{C}}(\tilde{\rho})}\big) \end{aligned} \tag{4.12}$$

Here $H(\mathcal{M}|\mathcal{C})$ is an ordinary conditional entropy. Without loss of generality, we assume that $I_{sq}(\mathcal{S}:\mathcal{T})_\rho \leq I_{sq}(\mathcal{S}:\mathcal{T})_\eta$. We first choose purifications $\rho^M = \mathrm{tr}_{M'}(|\psi\rangle\langle\psi|^{MM'})$ and $\eta^M = \mathrm{tr}_{M'}(|\phi\rangle\langle\phi|^{MM'})$, such that $\||\psi\rangle\langle\psi| - |\phi\rangle\langle\phi|\|_1 \leq 2\sqrt{\epsilon}$. For any extension $\tilde{\eta}^{MC}$, let $\Lambda : M' \to C$ be a quantum operation such that $(\hat{1}^M \otimes \Lambda)(|\psi\rangle\langle\psi|) = \tilde{\eta}^{MC}$. Let $\tilde{\rho} = (\hat{1}^M \otimes \Lambda)(|\phi\rangle\langle\phi|)$. Then $\|\tilde{\rho}^{MC} - \tilde{\eta}^{MC}\|_1 \leq 2\sqrt{\epsilon}$ by monotonicity of the trace distance under quantum operations. We then apply the Alicki-Fannes inequality (see Equation (2.13)) to each term in Equation (4.12), yielding

$$|I(\mathcal{S}\mathcal{C}:\mathcal{T}\mathcal{C})_{\tilde{\rho}} - I(\mathcal{S}'\mathcal{C}:\mathcal{T}'\mathcal{C})_{\tilde{\eta}}| \leq 12\sqrt{\epsilon}\log|M| + 3(1+2\sqrt{\epsilon})h\Big(\frac{1}{1+2\sqrt{\epsilon}}\Big) . \tag{4.13}$$

Then the (4.10) follows from that $I_{sq}(\mathcal{S}:\mathcal{T})_\rho \leq \frac{1}{2} I(\mathcal{S}\mathcal{C}:\mathcal{T}\mathcal{C} \subseteq \mathcal{S}\vee\mathcal{T}\mathcal{C})_{\tilde{\rho}}$ and $\tilde{\sigma}$ is arbitrary. I_{conv} admits similar extension definition of I_{sq}, where the extended system $\mathcal{C}$ is restricted to a classical system $\mathcal{X}$. We may thereby apply the same argument, but with the slightly better estimates on classically conditioned entropies.
iv) follows from the definition, as

$$I_{sq}(\mathcal{S}:\mathcal{T})_\rho = \frac{1}{2}\inf_{\tilde{\rho},\mathcal{C}} I(\mathcal{S}\otimes\mathcal{C}:\mathcal{T}\otimes\mathcal{C})_{\tilde{\rho}} \leq \frac{1}{2}\inf_{\tilde{\rho},\mathcal{X}} I(\mathcal{S}\otimes\mathcal{X}:\mathcal{T}\otimes\mathcal{X})_{\tilde{\rho}} = I_{conv}(\mathcal{S}:\mathcal{T})_\rho ,$$

where $\mathcal{X}$ is constrained to be classical, but $\mathcal{C}$ is a matrix algebra.
vii) Let $\mathcal{A},\mathcal{B}$ be matrix algebras such that $\mathcal{S}\subset\mathcal{A}$, $\mathcal{T}\subset\mathcal{B}$ and $\mathcal{S}\otimes\mathcal{T}\subset\mathcal{A}\otimes\mathcal{B}$. Suppose $I_{sq}(\mathcal{S}:\mathcal{T})_\rho = I_{sq}(A:B)_\rho \leq \epsilon$. Then by the faithfulness bounds in [116, 117], there exists a separable state σ^{AB} such that $\|\rho-\sigma\|1 \leq 3.1|B|\sqrt[4]{\epsilon}$, and (possibly different) separable η^{AB} such that $\|\rho-\eta\|_2 \leq 12\sqrt{\epsilon}$. Note that the conditional expectation is a contraction in trace norm. Then we have

$$\|\rho - \mathcal{E}_{\mathcal{S}\otimes\mathcal{T}}(\sigma)\|_1 \leq \|\mathcal{E}_{\mathcal{S}\otimes\mathcal{T}}(\rho) - \mathcal{E}_{\mathcal{S}\otimes\mathcal{T}}(\sigma)\|_1 \leq \min\{3.1|B|\sqrt[4]{\epsilon}, 12\sqrt{|A||B|\epsilon}\} ,$$

where $\mathcal{E}_{\mathcal{S}\otimes\mathcal{T}}(\sigma) = \mathcal{E}_\mathcal{S}\otimes\mathcal{E}_\mathcal{T}(\sigma)$ is separable. That $I_{conv} = 0$ implies separability follows from iv). Via the techniques updated in Chapter 6, we may obtain a bound of quadratic root order. By using the same argument with Proposition 1 in [117], we obtain a dimension-independent bound that when $I_{conv}(\mathcal{S}:\mathcal{T})_\rho \leq \epsilon$, $\|\rho-\sigma\|_1 \leq 2\sqrt{\log 2\epsilon}$ for some separable σ.

vi) follows from a chain rule of $I(\mathcal{S},\mathcal{T})$ proved in [23]. ■

I_{sq} and I_{conv} are non-increasing under an expanded set of operations, analogous to local operations and classical communication:

Theorem 4.2 (Individual Operations & Classical Communication, based on results in [23]). *Let $\mathcal{S},\mathcal{T}\subseteq\mathcal{S}\vee\mathcal{T}$ be a commuting square and $\rho\in S_1(\mathcal{S}\vee\mathcal{T})$ be a density. I_{sq} and I_{conv} are non-increasing under the following operations.*

1. *Individual $\mathcal{S}$- and $\mathcal{T}$-operations: more generally, if an operation $\mathcal{S}\to\tilde{\mathcal{S}},\mathcal{T}\to\tilde{\mathcal{T}},\rho\to\tilde{\rho}$ remains I-non-increasing for its all extensions $\mathcal{S}\otimes\mathcal{C}\to\tilde{\mathcal{S}}\otimes\mathcal{C},\mathcal{T}\otimes\mathcal{C}\to\tilde{\mathcal{T}}\otimes\mathcal{C},\sigma^{(S\vee T)\otimes\mathcal{C}}$ for any $\mathcal{C}$ and σ in $(S\vee T)\otimes\mathcal{C}$, then it does not increase I_{sq} or I_{conv}. Such operations include those in Remark 4.3 as well as individual operations.*

2. *Classical communication: Let $\mathcal{C}$ be a system such that $\mathcal{S}\otimes\mathcal{C},\mathcal{T}$ form a commuting square. I_{sq} and I_{conv} are non-increasing under the transformation $\mathcal{S}\otimes\mathcal{C}\to\mathcal{S},\mathcal{T}\to\mathcal{T}\otimes\mathcal{C},\rho\to\hat{1}\otimes\mathcal{E}_{\mathcal{X}}(\rho)$ for any basis $\mathcal{X}$ of C.*

They are also non-increasing under some other kinds of operations, which primarily affect algebras. That these operations have no obvious, non-trivial analogs in usual local operations and classical communication most likely follows from the fact that they make compensatory changes to the state and algebra. In the setting of subsystems, we rarely if ever change the algebras other than by appending or removing subfactors.

Theorem 4.3 (Operations Changing States with Algebras, based on results in [23]). *Let $\mathcal{S},\mathcal{T}\subseteq\mathcal{S}\vee\mathcal{T}$ be a commuting square and $\rho\in S_1(\mathcal{S}\vee\mathcal{T})$ be a density. I_{sq} and I_{conv} are non-increasing under the following operations.*

3. *Covariant Averaging: Let $\Psi(\rho)=\int_G g\rho g^\dagger d\mu(g)$ be a channel and G a unitary subgroup with probability measure μ. If $g\in G$ commutes with $\mathcal{E}_\mathcal{S},\mathcal{E}_\mathcal{T}$ and $\mathcal{E}_{\mathcal{S}\vee\mathcal{T}}$ μ-almost surely, then I_{sq} and I_{conv} are non-increasing under $\rho\to\Psi(\rho)$.*

4. *State-Compatible Algebra Changes: Let $\mathcal{E}_\mathcal{R}$ be a conditional expectation to a subalgebra $\mathcal{R}\subset\mathcal{S}\vee\mathcal{T}$ given by a covariant averaging defined above. Let $\tilde{\mathcal{S}}$ and $\tilde{\mathcal{T}}$ be two subalgebras forming a commuting square with $\tilde{\mathcal{S}}\vee\tilde{\mathcal{T}}=\mathcal{S}\vee\mathcal{T}$ and $\mathcal{R}$ form a commuting square for each of $\tilde{\mathcal{S}},\tilde{\mathcal{T}}$. Assume that $\mathcal{S}\cap\mathcal{R}=\tilde{\mathcal{S}}\cap\mathcal{R},\mathcal{T}\cap\mathcal{R}=\tilde{\mathcal{T}}\cap\mathcal{R}$ and $\tilde{\mathcal{S}}\cap\tilde{\mathcal{T}}\cap\mathcal{R}=\mathcal{S}\cap\mathcal{T}\cap\mathcal{R}$. Then I_{sq} and I_{conv} is non-increasing under the operation $\mathcal{S}\to\tilde{\mathcal{S}},\mathcal{T}\to\tilde{\mathcal{T}},\rho\to\mathcal{E}_\mathcal{R}(\rho)$.*

Proof of Theorems 4.2 & 4.3, from [23]. 1. Recall that

$$I_*(\mathcal{S}:\mathcal{T})_\rho = \inf_{\sigma,C} \frac{1}{2}\big(H(\mathcal{S}\otimes\mathcal{C})_\sigma + H(\mathcal{T}\otimes\mathcal{C})_\sigma - H((\mathcal{S}\vee\mathcal{T})\otimes\mathcal{C})_\sigma - H(\mathcal{C})_\sigma\big) \,, \tag{4.14}$$

where $I_* \in \{I_{sq}, I_{conv}\}$ with the infimum subject to corresponding constraints. Assume a transformation $\rho \to \tilde{\rho}$, $\mathcal{S} \to \tilde{\mathcal{S}}$ and $\mathcal{T} \to \tilde{\mathcal{T}}$ is I-non-increasing, and extends to an I-non-increasing transformation $\mathcal{S}\otimes\mathcal{C} \to \tilde{\mathcal{S}}\otimes\mathcal{C}$, $\mathcal{T}\otimes\mathcal{C} \to \tilde{\mathcal{T}}\otimes\mathcal{C}$, $\sigma^{(\mathcal{S}\vee\mathcal{T})\otimes\mathcal{C}} \to \tilde{\sigma}^{(\mathcal{S}\vee\mathcal{T})\otimes\mathcal{C}}$ that is also I-non-increasing for any extension $\sigma^{(\mathcal{S}\vee\mathcal{T})}$ of ρ, preserving the constraints of I_*. Then $\frac{1}{2}I(\mathcal{S}\otimes\mathcal{C}:\mathcal{T}\otimes\mathcal{C})_\sigma \geq \frac{1}{2}I(\tilde{\mathcal{S}}\otimes\mathcal{C}:\tilde{\mathcal{T}}\otimes\mathcal{C})_{\tilde{\sigma}}$ under the extended transformation. The latter is a candidate in the infimum within $I_*(\tilde{\mathcal{S}}:\tilde{\mathcal{T}})_{\tilde{\rho}}$. We may easily check that individual operations and those in Remark 4.3 extend this way for I_*.

2. For classical communication $\rho \to id\otimes\mathcal{E}_\mathcal{X}(\rho), \mathcal{S}\otimes\mathcal{C} \to \mathcal{S}, \mathcal{T} \to \mathcal{T}\otimes\mathcal{C}$, we proceed in the following steps:

   (a) By Remark 4.3 (3) and noting that $\mathcal{E}_{(\mathcal{S}\otimes\mathcal{X})\vee\mathcal{T}}$ is an $\mathcal{S}\otimes\mathcal{C}$-operation, we transform $\mathcal{S}\otimes\mathcal{C} \to \mathcal{S}\otimes\mathcal{X}, \rho \to \mathcal{E}_{(\mathcal{S}\otimes\mathcal{X})\vee\mathcal{T}}(\rho)$.

   (b) We transform $\mathcal{T}\otimes\mathcal{C} \to \mathcal{T}\otimes\mathcal{X}$ as per Remark 4.3, so that $\mathcal{X}$ appears in the intersection algebra. We have that

   $$I_*(\mathcal{S}\mathcal{X}:\mathcal{T})_\rho \leq \frac{1}{2}\inf_{\tilde{\rho},\mathcal{D}} I(\mathcal{S}\otimes\mathcal{X}\otimes\mathcal{D}, \mathcal{T}\otimes\mathcal{X}\otimes\mathcal{D})_\rho \,. \tag{4.15}$$

   (c) We transform $\mathcal{S}\otimes\mathcal{X} \to \mathcal{S}$, again candidate-wise in the infimum. This is allowed, because I_* may add a copy of X to the extending system. In doing so, the extra copy of $\mathcal{X}$ in $\mathcal{T}\otimes\mathcal{X}$ and $\mathcal{S}\vee(\mathcal{T}\otimes\mathcal{X})$ has no effect on entropies, since it's a completely correlated classical copy. Hence this enlarges the set of infimum candidates, lowering I_*.

   (d) We apply Remark 4.3 (3) again, this time noting that $\mathcal{E}_{\mathcal{S}\vee(\mathcal{T}\otimes\mathcal{X})}$ is a $\mathcal{T}\otimes\mathcal{X}$-operation, transforming $\mathcal{T}\otimes\mathcal{X} \to \mathcal{T}\otimes\mathcal{C}, \rho \to \mathcal{E}_{\mathcal{S}\vee(\mathcal{T}\otimes\mathcal{X})}(\rho)$.

   This sequence of free operations yields the desired configuration.

3. For covariant averaging, let C be the auxiliary system in the extension. For each U in the averaging form of Φ, we write $U\otimes 1_C$ as the unitary on the extended system. Since $\mathcal{E}_\mathcal{S}$ extends to $\mathcal{E}_{\mathcal{S}\otimes\mathcal{C}} = \mathcal{E}_\mathcal{S}\otimes id_\mathcal{C}$, it still commutes with conjugation by $U\otimes 1_C$ and similarly for $\mathcal{E}_{\mathcal{T}\otimes\mathcal{C}}$, $\mathcal{E}_{\mathcal{S}\vee\mathcal{T}\otimes\mathcal{C}}$ and $\mathcal{E}_{(\mathcal{S}\cap\mathcal{T})\otimes\mathcal{C}}$. Then $I(\mathcal{S}\mathcal{C}:\mathcal{T}\mathcal{C})_{(U\otimes 1)\tilde{\rho}(U^\dagger\otimes 1)} = I(\mathcal{S}\mathcal{C}:\mathcal{T}\mathcal{C})_{\tilde{\rho}}$ for each U. For each $\tilde{\rho}$ in the optimization of $I_*(\mathcal{S}:\mathcal{T})_\rho$, $I_*(\mathcal{S}:\mathcal{T})_{U\rho U^\dagger}$ achieves the same value with the extension $(U\otimes 1)\tilde{\rho}(U^\dagger\otimes 1)$. Hence we further have

$I_*(\mathcal{S}:\mathcal{T})_{U\rho U^\dagger} \leq I_*(\mathcal{S}:\mathcal{T})_\rho$ and they are equal by invertibility of U. By convexity of I_*,

$$I_*(\mathcal{S}:\mathcal{T})_{\Phi(\rho)} = I_*(\mathcal{S}:\mathcal{T})_{\int U\rho U^\dagger d\mu(U)} \leq \int I_*(\mathcal{S}:\mathcal{T})_{U\rho U^\dagger} d\mu(U) = I_*(\mathcal{S}:\mathcal{T})_\rho \,.$$

4. For state-compatible algebra changes, we return to the four-term entropy expression. For brevity, we may write $\mathcal{NC} = \mathcal{N} \otimes \mathcal{C}$ for a subalgebra $\mathcal{N} \subseteq \mathcal{S} \vee \mathcal{T}$, which is correct when $\mathcal{C}$ is in tensor position. Via the assumed commuting squares, $H(\mathcal{SC})_{\mathcal{E}_{\mathcal{RC}}(\tilde{\rho})} = H((\mathcal{S} \cap \mathcal{R})\mathcal{C})_{\tilde{\rho}}$ and similarly for $H(\mathcal{TC}), H((\mathcal{S} \vee \mathcal{T})\mathcal{C})$ and $H((\mathcal{S} \cap \mathcal{T})\mathcal{C})$. Then for any $\tilde{\rho}$ and $\mathcal{C}$,

$$\begin{aligned}
&H(\mathcal{SC})_{\mathcal{E}_{\mathcal{RC}}(\tilde{\rho})} + H(\mathcal{TC})_{\mathcal{E}_{\mathcal{RC}}(\tilde{\rho})} - H((\mathcal{S} \vee \mathcal{T})\mathcal{C})_{\mathcal{E}_{\mathcal{RC}}(\tilde{\rho})} - H((\mathcal{S} \cap \mathcal{T})\mathcal{C})_{\mathcal{E}_{\mathcal{RC}}(\tilde{\rho})} \\
=&H((\mathcal{S} \cap \mathcal{R})\mathcal{C})_{\tilde{\rho}} + H((\mathcal{T} \cap \mathcal{R})\mathcal{C})_{\tilde{\rho}} - H(((\mathcal{S} \vee \mathcal{T}) \cap \mathcal{R})\mathcal{C})_{\tilde{\rho}} - H((\mathcal{S} \cap \mathcal{T} \cap \mathcal{R})\mathcal{C})_{\tilde{\rho}} \\
=&H((\tilde{\mathcal{S}} \cap \mathcal{R})\mathcal{C})_{\tilde{\rho}} + H((\tilde{\mathcal{T}} \cap \mathcal{R})\mathcal{C})_{\tilde{\rho}} - H(((\tilde{\mathcal{S}} \vee \tilde{\mathcal{T}}) \cap \mathcal{R})\mathcal{C})_{\tilde{\rho}} - H((\tilde{\mathcal{S}} \cap \tilde{\mathcal{T}} \cap \mathcal{R})\mathcal{C})_{\tilde{\rho}} \\
=&H(\tilde{\mathcal{S}}\mathcal{C})_{\mathcal{E}_{\mathcal{RC}}(\tilde{\rho})} + H(\tilde{\mathcal{T}}\mathcal{C})_{\mathcal{E}_{\mathcal{RC}}(\tilde{\rho})} - H((\tilde{\mathcal{S}} \vee \tilde{\mathcal{T}})\mathcal{C})_{\mathcal{E}_{\mathcal{RC}}(\tilde{\rho})} - H((\tilde{\mathcal{S}} \cap \tilde{\mathcal{T}})\mathcal{C})_{\mathcal{E}_{\mathcal{RC}}(\tilde{\rho})}
\end{aligned} \tag{4.16}$$

We also have that $\mathcal{E}_{\mathcal{RC}}(\tilde{\rho}) = (\mathcal{E}_{\mathcal{R}} \otimes id_{\mathcal{C}})(\tilde{\rho})$ is an extension of $\mathcal{E}_{\mathcal{R}}(\rho)$ in $(\mathcal{S} \vee \mathcal{T})\mathcal{C}$. Hence $I_*(\mathcal{S}:\mathcal{T})_{\mathcal{E}_{\mathcal{R}}(\rho)}$ and $I_*(\tilde{\mathcal{S}}:\tilde{\mathcal{T}})_{\mathcal{E}_{\mathcal{R}}(\rho)}$ optimize over the same set of states, and they achieve the same values on each. Therefore, $I_*(\mathcal{S}:\mathcal{T})_\rho \geq I_*(\mathcal{S}:\mathcal{T})_{\mathcal{E}_{\mathcal{R}}(\rho)} = I_*(\tilde{\mathcal{S}}:\tilde{\mathcal{T}})_{\mathcal{E}_{\mathcal{R}}(\rho)}$ because $\mathcal{E}_{\mathcal{R}}$ is a covariant averaging for $\mathcal{S}, \mathcal{T} \subset \mathcal{S} \vee \mathcal{T}$.

■

The intuition for state-compatible algebra changes is to allow algebra changes that are undetectable by the two parties due to a particular form of state. For example, were ρ to be the complete mixture, we would have $\rho \in \mathbb{C}$, and we could arbitrarily assign any $\tilde{\mathcal{S}}, \tilde{\mathcal{T}} \subseteq \mathcal{S} \vee \mathcal{T}$ to the two parties without changing the resource value. The covariant averaging step first ensures that ρ is in such a particular algebra. Once we know this, we are free to replace the two parties' algebras as long as we know that the intersections with the state's algebra remain equivalent.

4.4 Converting Entanglement & Cross-Basis Non-Classicality

In this section, we derive

$$2 \text{ UCR} \leftrightarrow \text{EPR}\,, \tag{4.17}$$

that two copies of a basis-split qubit with maximum non-classicality convert under free operations to a maximally-entangled pair of qubit subsystems. As a first hint that this should be possible, consider the state

$$|\uparrow_Y\rangle = \frac{1}{\sqrt{2}}(|0\rangle + i\,|1\rangle) ,$$

an eigenstate of the Pauli Y operator, in the commuting square formed by $\mathcal{X}, \mathcal{Z} \subseteq \mathcal{H}_2$, two Pauli bases of a qubit Hilbert space. We see that

$$I_{sq}(|\uparrow_Y\rangle\langle\uparrow_Y|) = I_{conv}(|\uparrow_Y\rangle\langle\uparrow_Y|) = \frac{1}{2} \text{ ebits} .$$

Despite being a single qubit state, $|\uparrow_Y\rangle$ carries half an ebit of the algebraic generalization of squashed entanglement. As we prove, two copies will convert to a single EPR pair between tensor-separated subsystems under free operations for I_{sq} and I_{conv}.

Proof of equation (4.17)*, based on [23].* For qubit systems A and B, let $C_{\mathcal{O}_A \to V_B}$ denote the controlled gate that performs the unitary V_B if the binary observable $\mathcal{O}_A$ is in the "-1" eigenstate. For example, $C_{\mathcal{Z}_A \to X_B}$ is the standard controlled-NOT gate when $\mathcal{Z}$ eigenstates define the computational basis.

The main step is the state-compatible algebra replacement with $\mathcal{R} = \langle Y_A, Y_B \rangle$.

$$\mathcal{E}_{\langle Y_A, Y_B \rangle}(\rho) = \mathcal{E}_{\langle \mathcal{A}, Y_B \rangle} \mathcal{E}_{\langle Y_A, \mathcal{B} \rangle}(\rho) ,$$

because $\mathcal{E}_{\langle \mathcal{A}, Y_B \rangle} = \hat{1} \otimes \mathcal{E}_{\mathcal{Y}_B}$ and $\mathcal{E}_{\langle Y_A, \mathcal{A} \rangle} = \mathcal{E}_{\mathcal{Y}_A} \otimes \hat{1}$. Furthermore, $\mathcal{E}_{\mathcal{Y}_A}(\rho) = \frac{1}{2}(\rho + Y_A \rho Y_A)$. $Y_A \mathcal{S} Y_A = \mathcal{S}$. Similar arguments hold for Y_B with $\mathcal{S}$ and for $\mathcal{T}$. Hence $\mathcal{R}$ is generated by a convex combination of unitaries that commute with $\mathcal{E}_{\mathcal{S}}$ and $\mathcal{E}_{\mathcal{T}}$. It is easy to see that it commutes with $\mathcal{E}_{\mathcal{S} \vee \mathcal{T}}$, and $\mathcal{E}_{\mathcal{S} \cap \mathcal{T}}$, as the former is the whole $\mathcal{AB}$ algebra, and the latter is $\mathbb{C}1$. We then check that $\mathcal{R} \cap \langle Z_A, Z_B \rangle = \mathbb{C}1 = \mathcal{R} \cap \langle X_A, Z_A Z_B \rangle$. Similarly, $\mathcal{R} \cap \langle X_A, X_B \rangle = \mathbb{C}1 = \mathcal{R} \cap \langle X_A X_B, Z_B \rangle$. The latter algebras generate the same joint, and their intersection remains $\mathbb{C}1$. Hence this is a state-compatible algebra change.

Once we have $\mathcal{S} = \langle X_A, Z_A Z_B \rangle$, with $\mathcal{T} = \langle X_A X_B, Z_B \rangle$, we apply a controlled-NOT from B to A, $C_{Z_B \to X_A}$, as a coordinate change as per Remark 4.3. This changes the algebras to $\mathcal{S} = \mathcal{A}$, and $\mathcal{T} = \mathcal{B}$. For the state,

$$\begin{aligned} |\uparrow_Y \uparrow_Y\rangle =& \frac{1}{2}((|00\rangle - |11\rangle) + i(|01\rangle + |10\rangle)) \\ \to& \frac{1}{2}((|00\rangle - |01\rangle) + i(|11\rangle + |10\rangle)) = \frac{1}{\sqrt{2}}(|0-\rangle + i\,|1+\rangle). \end{aligned} \tag{4.18}$$

Now we show that the covariant averaging $\mathcal{E}_{\langle Y_A, Y_B\rangle} = \mathcal{E}_{\mathcal{Y}_A} \otimes \mathcal{E}_{\mathcal{Y}_B}$ is implementable with shared randomness and individual operations. For these purposes, let us consider the simplified commuting square $\mathcal{S} = \mathcal{X}_A, \mathcal{T} = \mathcal{Z}_A \subseteq \mathcal{A}$, and show how we can apply $\mathcal{E}_{\mathcal{Y}}$.

1. Extend $\mathcal{S}$ to $\mathcal{S} \to \mathcal{S} \otimes \mathcal{C}$ and $\mathcal{T}$ to $\mathcal{T} \otimes \mathcal{D}$, where $\mathcal{C}$ and $\mathcal{D}$ are qubit systems in state $\frac{1}{2}(|00\rangle\langle 00| + |11\rangle\langle 11|)$. This is the shared randomness, and one can easily generate this state with classical communication.

2. We apply the gates $C_{Z_C \to Z_A}$ and $C_{Z_D \to X_A}$ as $\mathcal{S}$ and $\mathcal{T}$-ops respectively. Hence we have

$$\rho \to \frac{1}{2}(\rho \otimes |00\rangle\langle 00| + Y\rho Y \otimes |11\rangle\langle 11|) \ . \tag{4.19}$$

   Importantly, the mixture removes order-dependence on which gate was applied first, which might otherwise appear as a phase difference.

3. As algebraic $\mathcal{S}$ and $\mathcal{T}$-ops, we remove $\mathcal{C}$ and $\mathcal{D}$. With these systems gone from $\mathcal{S} \vee \mathcal{T}$, we are free to trace them from the state as well. This transforms $\rho \to \mathcal{E}_{\mathcal{Y}}(\rho)$.

We then apply $\mathcal{E}_{\langle Y_A, Y_B\rangle}$ as $\mathcal{E}_{\mathcal{Y}_A} \otimes \mathcal{E}_{\mathcal{Y}_B}$. ■

In each of the individual $\mathcal{X}$ and $\mathcal{Z}$ bases, the $|{\uparrow_Y}\rangle$ state has one unit of relative entropy of coherence. The fundamentally bipartite GCMI would assign it one unit as well, though it's not immediately obvious whether this should be considered a correlated state. There is for instance no obvious way to extract shared randomness from this configuration. We would be able to extract shared randomness from two copies, in a configuration given by $\langle \mathcal{X}_A, \mathcal{X}_B\rangle$ and $\langle \mathcal{Z}_A, \mathcal{Z}_B\rangle$ within $\mathbb{B}(\mathcal{H}_2) \otimes \mathbb{B}(\mathcal{H}_2)$. First, each side would drop operations, so that the subalgebras become $\langle \mathcal{X}_A\mathcal{X}_B\rangle$ and $\langle \mathcal{Z}_A\mathcal{Z}_B\rangle$, where the absence of commas indicate direct multiplication (formally, $\mathcal{X}_A$ is shorthand for $\mathcal{X}_A \otimes \hat{1}_B$, $\mathcal{X}_B$ is shorthand for $\hat{1}_A \otimes \mathcal{X}_B$, and $\mathcal{X}_A\mathcal{X}_B$ is shorthand for $\mathcal{X}_A \otimes \mathcal{X}_B$). We would then see that measurements of $\mathcal{X}_A\mathcal{X}_B$ and $\mathcal{Z}_A\mathcal{Z}_B$ are each individual operations for each side, and that these are correlated in the state

$$|{\uparrow_Y}\rangle \otimes |{\uparrow_Y}\rangle = \frac{1}{\sqrt{2}}\Big(\frac{1}{\sqrt{2}}(|00\rangle - |11\rangle) + \frac{i}{\sqrt{2}}(|01\rangle + |10\rangle)\Big) \ .$$

For the Bell states

$$\frac{1}{\sqrt{2}}(|00\rangle \pm |11\rangle), \frac{1}{\sqrt{2}}(|01\rangle \pm |10\rangle) \ ,$$

$\mathcal{Z}_A\mathcal{Z}_B$ is the bit parity, and $\mathcal{X}_A\mathcal{X}_B$ observes the sign between terms. In fact, while the operator "$\mathcal{Z}_A\mathcal{Z}_B \otimes \mathcal{X}_A\mathcal{X}_B$" appears formally nonsensical, one can re-factor the Hilbert space in such a way that this makes sense

- these are potentially independent degrees of freedom. One can see that in the $|\uparrow_Y\rangle \otimes |\uparrow_Y\rangle$ state, $\mathcal{Z}_A\mathcal{Z}_B$ and $\mathcal{X}_A\mathcal{X}_B$ are perfectly anti-correlated.

The unifying principle behind the basis-split non-classical configuration and the usual notion of entanglement might be summarized as:

> Quantum states allow the entropy of the whole to be lower than each or any part.

We further illustrate this idea in Figure 4.2.

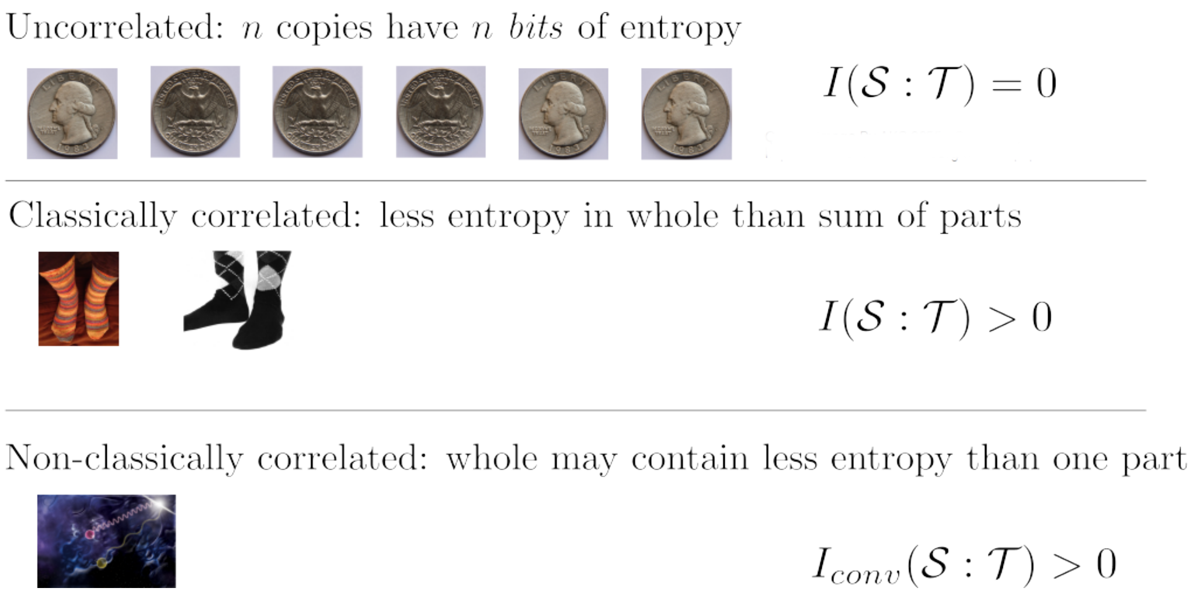

Figure 4.2: Flipping n fair coins results in n bits of entropy, an extensive property of uncorrelated subsystems. For an array of fully-correlated systems, entropy becomes an intensive property. Quantum entanglement allows the entropy of each or any subsystem to be higher than that of the whole. Most dramatically, a maximally entangled state between two parties has zero total entropy, but maximum subsystem entropy. Quarter Image By AKS.9955 - Own work, CC BY-SA 4.0, `https://commons.wikimedia.org/w/index.php?curid=53562236`. Colorful socks by Snapdragon66 - Own work, CC BY-SA 3.0, `https://commons.wikimedia.org/w/index.php?curid=31512542` B&W socks by Pearson Scott Foresman - Public Domain, `https://commons.wikimedia.org/w/index.php?curid=2572040`. Photon image from `https://commons.wikimedia.org/wiki/File:Quantum-Gravity-Photon-Race.jpg`, originally created by NASA .

Indeed, some of the intuition behind the "Einstein-Podolsky-Rosen paradox" is that local descriptions are inherently incomplete: there exist states for which adding more observables and even more regions of spacetime to a system's description seems to reduce the total amount of randomness or unpredictability. In classical information, adding more observables or variables can never result in lower entropy.

Chapter 5

Forms of (A)symmetry

A major and early theme of this thesis was to unify different notions of symmetry and relate them to quantum properties such as entanglement and coherence. On one hand, there is good case in [23] and [25] that entropic characterizations of entanglement and coherence relate closely to the Holevo asymmetry measure as recalled in Section 4.2. We recall the form of this measure for a von Neumann subalgebra $\mathcal{N} \subseteq \mathcal{M}$ and density matrix $\rho \in \mathcal{S}_1(\mathcal{M})$:

$$D^{\mathcal{N}}(\rho) = D(\rho \| \mathcal{E}_{\mathcal{N}}(\rho)) \ .$$

While the generality of this form suggests a common theme across many quantum resources and inequalities, it also might be so broad as to weaken the case that there is any specific, physical idea connecting them. Forms of entropy difference or relative entropy equivalent to cases of $D^{\mathcal{N}}$ are ubiquitous in information theory, where the conditional expectation in the second argument appears to confer essential properties. Ultimately, it might be that $D^{\mathcal{N}}$ is a unique information measure satisfying a set of axioms that include convexity, extensivity under tensor copying, monotonicity, generalization to infinite dimensions, a chain rule, and probably some other key properties. In this case it would tell us much about how to mathematically express resource theories, but less about the physical nature of different resources. Hence this thesis does not emphasize the interpretation in terms of asymmetry as much as originally expected.

I nonetheless briefly address the question of how notions of symmetry relate and the role of $D^{\mathcal{N}}$ as an asymmetry measure beyond its resource-theoretic interpretation. I start by recalling some prior notions:

1. Emmy Noether's original 1918 theorem on "Invariante Variationsprobleme [118]" states that every continuous symmetry of Lagrangian or Hamiltonian mechanics has associated with it a conserved "charge," a quantity unchanged under those dynamics. For example, momentum is the conserved Noether charge for spatially translation-invariant dynamics, and energy the charge for temporal translation-invariance. The classical electrical charge is also a Noether charge under the symmetries of electrodynamics [119].

2. Noether's theorem has a relatively simple quantum analog: for a Hamiltonian $\mathfrak{H}$, observable $\mathcal{O}$, and

Some of the results in this chapter may have been influenced by conversations with James O'Dwyer and Alice Doucet-Beaupré about projects with intended applications in biology, which are not included in this thesis.

time parameter t,

$$[\mathcal{O}, \mathfrak{H}] = 0 \iff \frac{d}{dt}\langle \mathcal{O} \rangle = 0, \tag{5.1}$$

as formulated in [120]. To deepen the analogy with the classical Noether's theorem, we note that $[\mathcal{O}, \mathfrak{H}] = 0$ implies that $\mathfrak{H}$ contains no dependence on operators that don't commute with $\mathcal{O}$. For example, if the momentum operator of a particle commutes with the Hamiltonian, then the Hamiltonian contains no powers of the position operator and is translation-invariant.

3. Baez and Fong [120] derive a stochastic Noether's theorem. Let $\mathcal{O}$ be an observable, and a density $\rho_t = \exp(t\mathcal{L})\rho_0$ for some stochastic (not quantum) Lindblad generator $\mathcal{L}$. Then

$$[\mathcal{O}, \mathcal{L}] = 0 \iff \frac{d}{dt}\langle \mathcal{O} \rangle = \frac{d}{dt}\langle \mathcal{O}^2 \rangle = 0. \tag{5.2}$$

   This formalism mirrors the quantum form of Noether's theorem. It differs from the quantum version in needing conservation of two moments to imply commutativity with the generator of time evolution. Gough et. al. and Gheondea extend this to quantum-stochastic Lindbladians [121, 122].

4. Marvian and Spekkens [123, 88, 124] develop an information-theoretic resource theory of asymmetry measures. A state has asymmetry with respect to some group of transformations, such as spatial translations, under which it is not invariant. The resource theory quantifies asymmetry of states and characterizes the possible conversions between these states under restricted operations. The resource theory of quantum coherence corresponds closely to a special case of asymmetry [87], though there are subtle differences in regularization for many copies of a system [89]. Similarly, conditional entropy resembles a form of asymmetry. We may consider a traced-out subsystem to be one to which all experiments are symmetric. This was the inspiration behind $D^{\mathcal{N}}$ and originally in [23]. As discussed in Chapter 4, we may interpret strong subadditivity as subadditivity of the asymmetry measure $D^{\mathcal{N}}$.

5. Symmetry often refers to invariance under subsystem or particle swaps, such as in bosonic or fermionic systems. The main argument of Section 6.2 is based on this type of symmetry, where we show that a large number of essentially interchangeable channel outputs cannot simultaneously and individually be entangled with an auxiliary system, or with each other. The entanglement monogamy argument derived by Ludovico Lami [125] is similar to that of the quantum de Finetti theorems [126], which bound non-classical correlations between symmetrically interchangeable systems.

6. Victor Albert's recent thesis [127] proves a partial version of Noether's theorem for Lindbladians, which describe continuous-time quantum Markov processes. When decoherence takes the form of a self-adjoint

Lindbladian, we may think of it as decaying asymmetry. The modified log-Sobolev inequalities studied in Section 7.2 show how self-adjoint Lindbladians generate exponential decay toward free states in the resource theory of asymmetry, and the entropy comparison Theorems of Section 7.1 compare decay in asymmetry given by a particular class of processes to the complete mixing out of asymmetry given by a conditional expectation.

7. In his work on the philosophy of science and modeling, Joe Rosen defines symmetry as "immunity to a possible change [128]." Rosen argues that the existence of repeatable laws or models relies on symmetry between different physical instances. While seeming at first philosophical, this notion is borne out in the idea of "sloppiness" in models [129], in notions of state or parameter space compression [130, 131], and in some basic approaches to forecasting [132]. Any sort of generalizing model relies on the expectation of symmetry between the system on which the model was trained or derived, and that to which it should generalize. We usually prefer that experimental outcomes be independent of the identity of the lab performing the experiment, the exterior environment of other experiments surrounding it, the absolute time of occurrence, etc. Less trivially, general physical laws predict aspects of wide varieties of systems, frequently ignoring most of physical details. Newton's universal law of gravitation, for instance, is a classic result of physics due to its independence from the particular material or absolute location of masses involved.

Traditional notions of symmetry, 1-3, refer to perfect invariance. In these settings, an observable that commutes with the Hamiltonian plays the role that the conserved charge would in the classical Noether's theorem. Conserved charges of continuous quantum symmetries become their generators. For example, invariance under spatial translation corresponds to momentum conservation, and the momentum operator generates spatial translations. Some studies [133] have considered perturbations from exact symmetry. In contrast, the resource theory of asymmetry is fundamentally a quantification, studying the extent of deviation even in states or processes that are far from symmetric. While the framing of [88] relates the monotonicity and conservation of asymmetry monotones to the conserved Noether charge, asymmetry resource monotones otherwise have little conceptual resemblance to Noether charges.

Consider a symmetry group $G \subseteq U(\mathcal{H})$ on Hilbert space $\mathcal{H}$. Associated with G is a fixed point algebra $\mathcal{N}_G \subseteq \mathbb{B}(\mathcal{H})$ and a conditional expectation $\mathcal{E}_G$ that projects densities onto the fully symmetric subspace. That a Hamiltonian $\mathfrak{H}$ is invariant under G is equivalent to $\mathfrak{H} \in \mathcal{N}_G$. Anything that commutes with the Hamiltonian is conserved under time-evolution. Hence

Remark 5.1. *Let $G \subseteq U(\mathcal{H})$ be a symmetry group. Then $\mathcal{N}_G'$, the commutant of the invariant subalgebra,* *is the algebra of operators that are conserved under time-evolution by any invariant Hamiltonian.* $\mathcal{N}_G'$ *is conceptually analogous to an algebra of Noether charges.*

We may consider more general classes of G that do not require full group structure. As discussed in [29], a transformation set that forms a quantum group or Hopf algebra has a fixed point algebra with commutant.

Given a quantum channel in finite dimension, we may Stinespring dilate, apply Stone's theorem, and ask if the associated generator commutes with a given symmetry group. Alternatively, if a quantum channel's Kraus operators are contained within the invariant subalgebra, then the channel commutes with (hence preserving) $\mathcal{N}_G'$. Hence the interpretation of $\mathcal{N}_G'$ as an algebra of conserved charges extends to open systems.

Remark 5.1 connects the traditional theory of symmetries and their corresponding charges, 1-3, to the resource-theoretic approach in quantum information, 4-6. A key observation is the appearance of the commutant, invoking a complementarity-like notion. The commutant of the invariant algebra contains the observables described in 2-3, which happen to commute with the Hamiltonian. There is more involved in this construction, as we have started with a description of invariances G rather directly with a commuting observable $\mathcal{O}$. Nonetheless, these definitions are often compatible. A system that is symmetric in the sense of 1-3 undergoes dynamics that would quantize to operators in the invariant subspace for a set of symmetries.

At first glance, it is not clear that the idea of symmetry in modeling and in the philosophy of science (point 7) relates to either. Qualitatively, the resource-theoretic approach favors states with high asymmetry, while the modeling approach benefits from symmetry between systems. They seem to disagree on whether mixture creates symmetry or asymmetry. To summarize the apparent difference:

Remark 5.2. *A completely mixed state or completely random process is uncorrelated between instances of an experiment, so it should be highly asymmetric from a classical modeling perspective. There are no generalizations between instances. In contrast, it is useless as an asymmetry resource, so it should be fully symmetric in this context.*

The resolution to this conundrum again recalls complementarity. A recurring concept in quantum information and particularly in Shannon theory is *decoupling*. In decoupling, given many copies of a quantum process with strongly-coupled inputs, one finds [134, 1] output subspaces that are independent of the environment. The nature of time-evolution for open quantum systems necessarily links noise with environmental backaction (see the Stinespring dilation in Section 2.6), so these subspaces are approximately noise-free.

For this illustration, consider determining the expectation value of an observable $\mathcal{O}$ by averaging trials from a repeated experiment with fixed initial configuration x. This experiment may involve any combination of quantum, stochastic, and classically deterministic processes. Inconsistencies in preparation and measurement between trials are implicitly included as part of these processes, as are non-deterministic quantum outcomes. Hence we model the desired expectation as $\langle\mathcal{O}\rangle_{\Phi(x)}$, where Φ is a channel with classical input x that outputs a density matrix. Despite potential non-determinism in $\Phi(x)$, $\langle\mathcal{O}\rangle_{\Phi(x)}$ is a fully-determined value in $\mathbb{R}$. Let $\vec{M}$ denote a vector of n measurement outcomes with initial configuration x, and $\langle\mathcal{O}\rangle_{\vec{M}}$ denote the average of $\mathcal{O}$ over those outcomes. In general, $\langle\mathcal{O}\rangle_{\Phi(x)} \neq \langle\mathcal{O}\rangle_{\vec{M}}$, since the particular content $\vec{M}$ can be probabalistic. We expect however that as $n\to\infty$, $\langle\mathcal{O}\rangle_{\vec{M}} \to \langle\mathcal{O}\rangle_{\Phi(x)}$ with probability approaching 1.

For any $i \in 1...n$, the ith measurement M_i is the outcome of a random variable, and $M_i - \langle\mathcal{O}\rangle_{\Phi(x)}$ is the outcome of a random variable with mean 0. Let $\vec{M} - \langle\mathcal{O}\rangle_{\Phi(x)}$ denote the vector $\vec{M}$ with each entry shifted by $-\langle\mathcal{O}\rangle_{\Phi(x)}$, so that its limit as $n\to\infty$ approaches zero. If (as is often the case) we can assume that $M_i - \mathcal{O}$ has random sign, then we may apply Khintchine's inequality (see [135]). In particular,

$$\frac{1}{\sqrt{2}n}\|\vec{M} - \langle\mathcal{O}\rangle_{\Phi(x)}\|_2 \leq \Big\langle\, \langle\mathcal{O}\rangle_{\vec{M}} - \langle\mathcal{O}\rangle_{\Phi(x)} \,\Big\rangle_{\{+,-\}} \leq \frac{1}{n}\|\vec{M} - \langle\mathcal{O}\rangle_{\Phi(x)}\|_2 \ ,$$

where $\langle ... \rangle_{\{+,-\}}$ denotes the expectation over random signs with fixed absolute values of entries ($|M_i|$). If $M_i - \langle\mathcal{O}\rangle_{\Phi(x)}$ is Gaussian-distributed with standard deviation β, then

$$\Big\langle\, \langle\mathcal{O}\rangle_{\vec{M}} - \langle\mathcal{O}\rangle_{\Phi(x)} \,\Big\rangle_{\{+,-\}} = \frac{\|\vec{M} - \langle\mathcal{O}\rangle_{\Phi(x)}\|_2}{n}\sqrt{\frac{2}{\pi}} \ .$$

Here the 2-norm, $\|\vec{M} - \langle\mathcal{O}\rangle_{\Phi(x)}\|_2$, acts like a variance as would be expected. This norm is however not a variance of the underlying process, but a function of the observed values - it might be different between different instances of $\vec{M}$. In any case, let us assume that

$$\Big\langle\, \langle\mathcal{O}\rangle_{\vec{M}} - \langle\mathcal{O}\rangle_{\Phi(x)} \,\Big\rangle_{\{+,-\}} = \frac{c\|\vec{M} - \langle\mathcal{O}\rangle_{\Phi(x)}\|_2}{n} \ ,$$

where $1/\sqrt{2} \leq c \leq 1$.

Let $\vec{p}_M = \|\vec{M} - \langle\mathcal{O}\rangle_{\Phi(x)}\|_2^2 / \|\vec{M} - \langle\mathcal{O}\rangle_{\Phi(x)}\|_1^2$, which is a probability vector. Note that $\|\vec{M} - \langle\mathcal{O}\rangle_{\Phi(x)}\|_1 = \langle|\mathcal{O} - \langle\mathcal{O}\rangle_{\Phi(x)}|\rangle_{\vec{M}}$, the average absolute difference between the observable and its expectation in $\vec{M}$. We may then interpret the Khintchine inequality as stating that

$$\langle\langle\mathcal{O}\rangle_{\vec{M}} - \langle\mathcal{O}\rangle_{\Phi(x)}\rangle_{\{+,-\}} = c\,\langle|\mathcal{O} - \langle\mathcal{O}\rangle_{\Phi(x)}|\rangle_{\vec{M}}\, e^{\frac{1}{2}(D_2(\vec{p}_M\|\hat{1}/n) - \log n)} \ . \tag{5.3}$$

As $\langle\langle\mathcal{O}\rangle_{\vec{M}} - \langle\mathcal{O}\rangle_{\Phi(x)}\rangle_{\{+,-\}}$ estimates the expected error induced in the expectation of $\mathcal{O}$ by a given vector of observations $\vec{M}$, minimizing the error corresponds to minimizing a Rényi relative 2-entropy.

We may interpret the asymmetry in Equation (5.3) as that with respect to hidden variables that take different values in each run of the experiment, but are irrelevant to the expectation value of $\mathcal{O}$. In classical, stochastic dynamics, all noise or randomness is indistinguishable from extra hidden variables. Quantum channels may include operations such as pinching, which according to quantum theory might not be determined by any pre-existing information in the environment. Nonetheless, via Stinespring dilation we may extract all randomness coming from a quantum channel to variables in the final environment, so that conditioned on the final environment's state, the physical process in each instance is a fully deterministic map from initial configuration to measured outcome. We then conclude:

Remark 5.3. *The squared ratio of the 2-norm to the 1-norm of $\vec{M} - \langle\mathcal{O}\rangle_{\Phi(x)}$, $\vec{p}_M$, reflects the extra, per-trial information contained in the vector of experimental trials $\vec{M}$ that differentiate each trial from the expectation. The Rényi relative 2-entropy of asymmetry of $\vec{p}_M$ with respect to the trivial algebra quantifies the expected decay of error with trial number n. Hence the decay of error when averaging repeated experiments exponentially follows the increase in symmetry measured by $\frac{1}{2}(\log n - D_2(\vec{p}_M \| \hat{1}/n))$.*

When $\vec{p}_M$ is maximally symmetric, we obtain the full, expected $1/\sqrt{n}$ convergence for the expectation of a binary random variable. More broadly, we still usually expect $O(1/\sqrt{n})$ convergence when the deviation of any trial from expectation is bounded, or when large deviations are rare. For processes with unbounded deviations that typically reach $O(n)$ or greater within n trials, there is no such convergence to expectation values.

The conceptual consequence of Remark 5.3 is that we can begin to view symmetry in the sense of 7 as linked with the Rényi relative entropy of asymmetry. Estimating the expectation of an observable by averaging trials is a relatively simplistic example, but a similar concept underlies many forms of predictive modeling. In assuming that there is any common structure between trials of an experiment, one assumes symmetry between them. The goal in modeling often involves highlighting some set of initial conditions as relevant, assuming irrelevance of other parameters. A maximally asymmetric model is a raw data table, conditioning all predictions on the label of the experiment, assuming no extrapolations whatsoever. A maximally symmetric model makes no distinctions, even between radically different systems. Useful models generally fall somewhere in between, predicting expectations conditioned on some initial conditions x (which by expectations of multiple observables also predicts higher moments). Key to extrapolation and generalization in models is symmetry as a form of decoupling from excess degrees of freedom.

Chapter 6

Entanglement, Rates & Superadditivity

The main result in this Chapter is a bound on the superadditivity of classical information rates for a particular class of channels that frequently destroy their inputs. This bound arises from the monogamy of quantum entanglement. Section 6.1 is a brief introduction to the classical capacity of a quantum channel. Section 6.2 contains the main result and its derivation. Section 6.3 applies a similar idea to bound the advantage from entanglement in a particular class of quantum games. Section 6.4 discusses a proposal to experimentally test superadditivity of the rate at which a quantum channel may transmit qubits.

6.1 Background on Holevo Information & Superadditivity

The Holevo information of a quantum channel Φ is formally for a classical reference system X, a quantum input system A', and a corresponding quantum output system B. It is given by

$$\chi(\Phi) = \sup_{\rho^{XA'}} I(X:B)_{(\hat{1}^X \otimes \Phi)(\rho)}$$

in analogy to Shannon's definition of the classical capacity [136, 137]. The Holevo information is an achievable rate of classical information transmission in the asymptotic limit of many channel copies and infinite encoding/decoding resources, though it is not a capacity. A channel's capacity for (not necessarily private) classical transmission is given by

$$C(\Phi) = \lim_{n\to\infty} \frac{1}{n} \chi(\Phi^{\otimes n}) \ .$$

The infinite limit is known as *regularization* of Holevo information, accounting for entangled encodings between inputs to different uses of the same channel. As shown by Hastings in 2009 [138], the Holevo information is superadditive in that $\chi(\Phi^{\otimes n}) > n\chi(\Phi)$ for some Φ. It remains an area of active research to find concrete, low-dimensional examples of superadditive Holevo information. While the effect appears small or absent in practical, low-dimensional settings, it nonetheless implies that calculating Holevo capacity is

This chapter includes results appearing in [27], co-authored with Li Gao and Marius Junge. It also references an ongoing project [28] with Spencer Johnson, Marius Junge, Eric Chitambar, and Paul Kwiat.

much harder than calculating classical capacity [10].

The *potential Holevo capacity* of a channel Φ is the maximum additional Holevo rate that Φ can contribute when paired with another channel [139]:

$$\chi^{(pot)}(\Phi) = \sup_{\Psi}(\chi(\Phi \otimes \Psi) - \chi(\Psi)) \ .$$

The potential capacity is trivially an upper bound on the regularized capacity, and by its nature additive. It is always true that $\chi(\Phi) \leq C(\Phi) \leq \chi^{(pot)}(\Phi)$, so one method to estimate $C(\Phi)$ is to upper bound $\chi^{(pot)}(\Phi) - \chi(\Phi)$. In finite dimension, $\chi(\Phi)$ is generally easier to compute than its regularization (though still non-trivial).

6.2 Heralded Channel Holevo Superadditivity Bounds from Entanglement Monogamy

The intuition for this section is that because entanglement is monogamous, losing a large number of channel instances to the environment effectively disentangles the remaining instances. This section describes the primary result in [27], but with an update that reflects some improvements to the faithfulness of squashed entanglement. Hence the enhancement of transmission rates due to superadditivity, an entanglement-dependent effect, is necessarily limited.

The original bounds were only useful in the limit of very high loss. The technique of using monogamy and faithfulness of a given entanglement measure appears much broader, however, and adds what we believe is a new approach to calculations that may become intractable due to superadditivity. Holevo capacity is a good example of a quantity that is clearly defined and well-studied, but for which there are few good analytical techniques, and which resists numerical calculation due to unbounded Hilbert space dimension. Hence this is a good problem on which to demonstrate the potential power of a general method. Furthermore, the method presented gains potency with further refinements to the faithfulness of monogamous entanglement measures. As a case in point, we review an updated version of the main Theorem that gives a stronger result than that in [27] simply by substituting a stronger and more recent faithfulness bound.

We denote for $\lambda \in [0, 1]$ the generalized quantum erasure channel with $1 - \lambda$ erasure probability,

$$Z_\lambda(\Phi)(\rho) \equiv \lambda \left|0\right\rangle \left\langle 0\right|^Y \otimes \Phi(\rho) + (1 - \lambda) \left|1\right\rangle \left\langle 1\right|^Y \otimes \sigma = \begin{bmatrix} \lambda\Phi(\rho) & \\ & (1-\lambda)\sigma \ . \end{bmatrix} \tag{6.1}$$

Here Y is an extra classical output system containing an erasure flag, and σ is any fixed density. A relatively simple consequence of our results is that

Theorem 6.1 (based on and updated from theorem 1.1 from [27])**.** *For any quantum channel Φ with output dimension at most d,*

$$C(Z_\lambda(\Phi)) \leq \lambda\chi(\Phi) + O(\lambda^{3/2}\log(1/\lambda)\cdot d\sqrt{\log d})\ . \tag{6.2}$$

This statement is stronger than that appearing in [27], because we use a stronger and more recent result on the faithfulness of squashed entanglement with trace distance. Our bound does not depend on the extent of Φ's superadditivity or on other details of Φ. It is relevant in the regime where $\lambda << 1/d$, which means that parallel copies of $Z_\lambda(\Phi)$ erase most of the inputs. The rest of this section shows the derivation of this and similar bounds.

6.2.1 Heralded Channels & Combinatoric Separability

Let $\{\Phi_1, \cdots, \Phi_n\}$ and $\{\Psi_1, ..., \Psi_n\}$ be two classes of quantum channels. For $k < n$, we define the *flagged switch channel* as follows,

$$Z_k(\Phi_1, \cdots, \Phi_n; \Psi_1, \cdots, \Psi_n)(\rho) = \frac{1}{\binom{n}{k}} \sum_{R\subset[n],|R|=k} \left(\Phi^R \otimes \Psi^{R^c}\right)(\rho) \otimes |R\rangle\langle R|^Y\ . \tag{6.3}$$

R labels which subset of the outputs were Φ channels. $\Phi^R \otimes \Psi^{R^c}$ denotes a tensor product containing Φ channels at the positions in R and Ψ channels at the others. The flagged switch channel also outputs a flag state $|R\rangle\langle R|^Y$, giving the output a classical description of which channels wound up as Φ vs. Ψ instances. It is key to our results that this is not known at the input. Related flagged channels appear in [124, 140, 141].

Let Θ be a trivial channel, such as a complete erasure, depolarization, or other state replacement. We define a *heralded channel* as

$$Z_k(\Phi_1, \cdots, \Phi_n) := Z_k(\Phi_1, \cdots, \Phi_n; \Theta, \cdots, \Theta)\ .$$

Heralded channels with differing Θ are interconvertible via local post-processing at the output as long as Θ's output has no input dependence. Hence the particular Θ used has no bearing on the Holevo information or capacity. The heralded channel is inspired by common situations in linear optics, where a gate, photon source or other element has some probability to perform the desired operation, but may fail in a way that is immediately apparent. Heralded operations are key to optical feed-forward schemes such as the Knill-

Laflamme-Milburn quantum computer [142]. For notational convenience, we may write

$$Z_k(\mathbf{\Phi};\mathbf{\Psi}) := Z_k(\Phi_1,\cdots,\Phi_n;\Psi_1,\cdots,\Psi_n)\ ,\ Z_k(\mathbf{\Phi}) := Z_k(\Phi_1,\cdots,\Phi_n)\ ,$$

where $\mathbf{\Phi} = \{\Phi_1,\cdots,\Phi_n\}$ and $\mathbf{\Psi} = \{\Psi_1,\cdots,\Psi_n\}$.

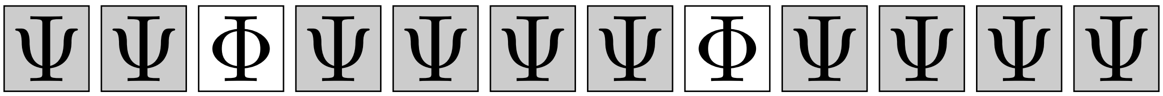

Figure 6.1: Diagram of possible cases of a flagged channel that usually applies Φ but sometimes applies Ψ. We don't know where the Φ instances will appear in advance but have an output flag distinguishing them.

We recall the squashed entanglement given by

$$E_{sq}(A,B)_\rho = \frac{1}{2}\inf\{I(A;B|C)_\rho \mid tr_C(\rho^{ABC}) = \rho^{AB}\}\ ,$$

where the infimum runs over all extensions ρ^{ABC} of ρ^{AB}.In Section 4.3, we constructed a generalization of squashed entanglement, I_{sq}. Here we need only consider tensor products of subsystems, not general subalgebras. We will use three properties of the squashed entanglement:

i) Convexity: let $\rho^{AB} = \sum_x p_x \rho_x^{AB}$ be a convex combination of states $\{\rho_x\}$. Then

$$E_{sq}(A,B)_\rho \le \sum_x p_x E_{sq}(A,B)_{\rho^x}\ . \tag{6.4}$$

ii) Monogamy: let $\rho^{AB_1\cdots B_k}$ be a $(k+1)$-partite state, then

$$\sum_{j=1}^{k} E_{sq}(A,B_j)_\rho \le E_{sq}(A,B_1...B_k)_\rho \le H(A)_\rho\ . \tag{6.5}$$

iii) 1-norm faithfulness:

$$\min\left\{\|\rho^{AB} - \sigma^{AB}\|_1 \mid \sigma = \sum_x p_x \sigma_x^A \otimes \sigma_x^B\right\} \le 4\sqrt{\log 2}(\min\{|A|,|B|\} - 1/2)\sqrt{E_{sq}(A,B)_\rho}\ . \tag{6.6}$$

Convexity and monogamy of squashed entanglement are basic properties known at its introduction in [115]. Its 1-norm faithfulness was proved by Brandão et al [116] and via a different method by Li and Winter [117], which we used in [27]. Here I use a further refinement that appeared in the thesis of Ludovico Lami [125] after we had submitted [27] for publication. Compared with the Brandão et al bound, Lami's squashed entanglement monogamy depends on the dimension of either A or B rather than both. Compared with Li and Winter's bound, Lami's is quadratic rather than quartic.

Lemma 6.1 (lemma 3.1 from [27]). *Let $Z_k(\mathbf{\Phi})$ be a heralded channel with fixed success number. For any state $\rho \in B_0 \otimes A_1 \cdots A_n$,*

$$E_{sq}(B_0, Z_k(\mathbf{\Phi}))_\rho \leq \frac{1}{L} H(B_0)_\rho \ ,$$

where $L = \lfloor n/k \rfloor$, the largest integer less or equal than n/k.

Proof. Let S_n be the symmetric group of the integer set $[n] = \{1, \cdots, n\}$. The heralded channel $Z_k(\mathbf{\Phi})$ can be rewritten as

$$Z_k(\mathbf{\Phi})(\rho) = \frac{1}{n!} \sum_{\sigma \in S_n} \left(\Phi^{\sigma(R)} \otimes \Theta^{\sigma(R)^c} \right)(\rho) \otimes |\sigma(R)\rangle\langle\sigma(R)|^Y \ ,$$

where R can be any k-subset of $[n]$ and the summation runs over all permutations σ. Then we find L disjoint subset $R_1, R_2, \cdots, R_L$ with each $|R_l| = k$, and have

$$Z_k(\mathbf{\Phi})(\rho) = \frac{1}{L} \sum_{l=1}^{L} \frac{1}{n!} \sum_{\sigma \in S_n} \left(\Phi^{\sigma(R_l)} \otimes \Theta^{\sigma(R_l)^c} \right)(\rho) \otimes |\sigma(R_l)\rangle\langle\sigma(R_l)| \ .$$

Exchanging the summation, we have

$$\begin{aligned}
E_{sq}(B_0, Z_k(\mathbf{\Phi}))_\rho =& E_{sq}\Big(B_0, \frac{1}{n!} \sum_{\sigma \in S_n} \frac{1}{L} \sum_{l=1}^{L} (\Phi^{\sigma(R_l)} \otimes \Theta^{\sigma(R_l)^c}) \otimes |\sigma(R_l)\rangle\langle\sigma(R_l)|^Y \Big)_\rho \\
\leq& \frac{1}{n!} \sum_{\sigma \in S_n} \frac{1}{L} \sum_{l=1}^{L} E_{sq}\Big(B_0, (\Phi^{\sigma(R_l)} \otimes \Theta^{\sigma(R_l)^c})\Big)_\rho \\
=& \frac{1}{n!} \sum_{\sigma \in S_n} \frac{1}{L} \sum_{l=1}^{L} E_{sq}\Big(B_0, \Phi^{\sigma(R_l)}\Big)_\rho \\
\leq& \frac{1}{n!} \sum_{\sigma \in S_n} \frac{1}{L} \sum_{l=1}^{L} E_{sq}\Big(B_0, A^{\sigma(R_l)}\Big)_\rho \ .
\end{aligned}$$

Here the first inequality follows from convexity and the fact $|\sigma(R_l)\rangle\langle\sigma(R_l)|^Y$ is a classical signal. The second equality is because Θ's are trivial channels. The last inequality is that squashed entanglement never increases under local operations. Note that for any permutation σ, $\sigma(R_1), \cdots, \sigma(R_L)$ are disjoint positions because $R_1, R_2, \cdots, R_L$ are. By monogamy of squashed entanglement, we know for any σ,

$$\sum_{l=1}^{L} E_{sq}(B_0, A^{\sigma(R_l)})_\rho \leq E_{sq}(B_0, A^{\sigma(R_1 \cdots R_l)})_\rho \leq H(B_0)_\rho \ ,$$

which completes the proof. ■

The above argument can be adapted to tensor products of heralded channels. Let $\mathbf{\Phi}_i = \{\Phi_{i,1}, \Phi_{i,2}, \cdots, \Phi_{i,n_i}\}, 1 \leq i \leq m$ be m classes of quantum channels such that each class consists of n_i quantum channels. Let $A = \otimes_{i=1}^{m}(\otimes_{j=1}^{n_i} A_{i,j})$ be the total input system and $B = \otimes_{i=1}^{m}(\otimes_{j=1}^{n_i} B_{i,j})$ be the quantum part of output system. Then the tensor product of heralded channels

$$\otimes_{i=1}^{m} Z_{k_i}(\mathbf{\Phi}_i)(\rho) = \frac{1}{\prod \binom{n_i}{k_i}} \sum_{R=R_1 \cdots R_m, |R_i|=k_i} (\Phi^R \otimes \Theta^{R^c})(\rho) \otimes |R\rangle\langle R|^Y$$

is from A to BY, where the heralding signal R now is an ensemble of the heralding signals $R_i \subset [n_i]$ for each $Z_{k_i}(\mathbf{\Phi}_i)$.

Lemma 6.2 (lemma 3.2 from [27])**.** *Let $Z_{k_i}(\mathbf{\Phi}_i), 1 \leq i \leq m$ be a family of heralded channels. For any state $\rho \in B_0 \otimes A$, where $A = \otimes_{i=1}^{m}(A_{1,i} \cdots A_{n_i,i})$, we have*

$$E_{sq}(B_0, \otimes_{i=1}^{m} Z_{k_i}(\mathbf{\Phi}_i))_\rho \leq \frac{1}{L} H(B_0)_\rho ,$$

where $L = \min_i \lfloor n_i/k_i \rfloor$.

Proof. Let $S_{n_1,\cdots,n_m} = S_{n_1} \times \cdots \times S_{n_m}$ be the direct product of the permutation groups $S_{n_1}, \cdots, S_{n_m}$. Denote $\sigma = (\sigma_1, \cdots, \sigma_m) \in S_{n_1,\cdots,n_m}$ where for each i such that $1 \leq i \leq m$, $\sigma_i \in S_{n_i}$. The tensor product of heralded channels can be rewritten as,

$$\otimes_{i=1}^{m} Z_{k_i}(\mathbf{\Phi}_i)(\rho) = \frac{1}{\prod n_i!} \sum_{\sigma} (\Phi^{\sigma(R)} \otimes \Theta^{\sigma(R)^c})(\rho) \otimes |\sigma(R)\rangle\langle\sigma(R)|^Y ,$$

where $\sigma(R) = \sigma_1(R_1) \cdots \sigma_m(\mathcal{R}_m)$ is the ensemble of shifted positions and the summation runs over all permutations $\sigma \in S_{n_1} \times \cdots \times S_{n_m}$. Given $L = \min_i \lfloor n_i/k_i \rfloor$, in each index set $[n_i]$ we can choose mutually disjoint subset $R_{1,i}, R_{2,i}, \cdots, R_{L,i}$ with $|R_{l,i}| = k$ for all l. Write $R(l) = R_{l,1} R_{l,2} \cdots R_{l,m}$. Following the same argument in Lemma 6.1, we have

$$
\begin{aligned}
&E_{sq}(B_0, \otimes_{i=1}^m Z_{k_i}(\mathbf{\Phi}_i))_\rho \\
=&E_{sq}\Big(B_0, \frac{1}{\prod n_i!}\sum_{\sigma\in S_{n_1,\cdots,n_m}}(\mathbf{\Phi}_1^{\sigma(R_1)}\otimes\cdots\otimes\mathbf{\Phi}_m^{\sigma(R_m)}\otimes\Theta^{\sigma(R)^c})(\rho)\otimes|\sigma(R)\rangle\langle\sigma(R)|^Y\Big)_\rho \\
\le&E_{sq}\Big(B_0, \frac{1}{\prod n_i!}\sum_{\sigma\in S_{n_1,\cdots,n_m}}\frac{1}{L}\sum_{1\le l\le L}(\mathbf{\Phi}_1^{\sigma(R_{l,1})}\otimes\cdots\otimes\mathbf{\Phi}_m^{\sigma(R_{l,m})}\otimes\Theta^{\sigma(R(l))^c})(\rho)\otimes|\sigma(R(l))\rangle\langle\sigma(R(l))|^Y\Big)_\rho \\
\le&\frac{1}{\Pi_i n_i!}\sum_{\sigma\in S_{n_1,\cdots,n_m}}\frac{1}{L}\sum_{l=1}^L E_{sq}\Big(B_0, \mathbf{\Phi}_1^{\sigma(R_{l,1})}\otimes\cdots\otimes\mathbf{\Phi}_m^{\sigma(R_{l,m})})\otimes\Theta^{\sigma(R(l))^c})\Big)_\rho \\
=&\frac{1}{\Pi_i n_i!}\sum_{\sigma\in S_{n_1,\cdots,n_m}}\frac{1}{L}\sum_{l=1}^L E_{sq}\Big(B_0, \mathbf{\Phi}_1^{\sigma(R_{l,1})}\otimes\cdots\otimes\mathbf{\Phi}_m^{\sigma(R_{l,m})})\Big)_\rho \\
\le&\frac{1}{\Pi_i n!}\sum_{\sigma\in S_{n_1,\cdots,n_m}}\frac{1}{L}\sum_{l=1}^L E_{sq}\Big(B_0, A^{\sigma(R(l))}\Big)_\rho .
\end{aligned}
$$

Note that for any permutation $\sigma \in S_{n_1,\cdots,n_m}$, $\sigma(R(1)),\cdots,\sigma(R(L))$ are disjoint index subsets because $R(1),R(2),\cdots,R(L)$ are. Thus by entanglement monogamy, for any σ,

$$\sum_{l=1}^L E_{sq}(B_0, A^{\sigma(R(l))})_\rho \le E_{sq}(B_0, A^{\sigma(R(1)R(2)\cdots R(L))})_\rho \le H(B_0)_\rho ,$$

which completes the proof. ■

Theorem 6.2 (based on theorem 3.4 from [27], updated via equation (6.6))**.** *Let* $\otimes_{i=1}^m Z_{k_i}(\mathbf{\Phi}_i) : A \to B$ *be the tensor of a family of heralded channels. Let* B_0 *be an extra system with dimension* $|B_0|$*. Suppose* $\underline{\lambda} = 1/\lfloor \min_i n_i/k_i \rfloor$ *is small enough that* $\delta = 4|B_0|\sqrt{\underline{\lambda} H(B_0)_\rho \log 2} \le 2$*. Then for any state* $\rho \in B_0 \otimes A$*, there exists a state* $\eta \in B_0 \otimes B$ *that is separable between* B_0 *and* B *such that*

$$\| id_{B_0} \otimes_{i=1}^m Z_{k_i}(\mathbf{\Phi}_i)(\rho) - \eta \|_1 \le \delta \le 2 ,$$

and

$$|H(B_0| \otimes_{i=1}^m Z_{k_i}(\mathbf{\Phi}_i))_\rho - H(B_0|B)_\eta| \le \delta \log|B_0| + \Big(1+\frac{\delta}{2}\Big)h\Big(\frac{\delta}{2+\delta}\Big) ,$$

where $h(\delta) = -\delta\log\delta - (1-\delta)\log(1-\delta)$*.*

Proof. By Lemma 6.2,

$$E_{sq}(B_0, \otimes_{i=1}^m Z_{k_i}(\mathbf{\Phi}_i))_\rho \le \underline{\lambda} H(B_0)_\rho .$$

Then we apply the faithfulness of squashed entanglement (Equation (6.6)) to show existence of a separable η such that

$$\| id_{B_0} \otimes_{i=1}^{m} Z_{k_i}(\mathbf{\Phi}_i)(\rho) - \eta \|_1 \leq \delta \leq 2 \ .$$

The Alicki-Fannes inequality (recalled in Equation (2.13)) completes the theorem. ■

6.2.2 Holevo Rate Bounds

Theorem 6.3 (based on and updated from theorem 4.1 from [27]). *Let $Z_{k_i}(\mathbf{\Phi}_i;\mathbf{\Psi}_i), 1 \leq i \leq m$ be a family of flagged switch channels and $\Phi_0 : A_0 \to B_0$ be an arbitrary channel. Suppose that $\underline{\lambda} = 1/\lfloor \min_i n_i/k_i \rfloor$ is small enough such that $\delta = 4|B_0|\sqrt{\underline{\lambda} \log |B_0| \log 2} \leq 2$. Then*

$$\chi(\Phi_0 \otimes_{i=1}^{m} Z_{k_i}(\mathbf{\Phi}_i;\mathbf{\Psi}_i)) \leq \chi(\Phi_0) + \chi(\otimes_{i=1}^{m} Z_{k_i}(\mathbf{\Phi}_i)) + \sum_{i=1}^{m} (1 - \frac{k_i}{n_i}) \sum_{j=1}^{n_i} \chi^{(pot)}(\Psi_{i,j}) + af(|B_0|, \delta) \ ,$$

where $h(\epsilon) = -\epsilon \log \epsilon - (1-\epsilon)\log(1-\epsilon)$, and

$$af(d, \delta) = 2\delta \log d + (2+\delta) h\Big(\frac{\delta}{2+\delta}\Big) \ .$$

Corollary 6.1 (based on and updated from corollary 4.2 from [27]). *Let the output system $B_{i,j}$ to each $\Phi_{i,j}$ be of dimension at most d and let $\underline{\lambda} = 1/\lfloor \min_i n_i/k_i \rfloor$ be small enough such that $\delta = 4|B_0|\sqrt{\underline{\lambda} \log |B_0| \log 2} \leq 2$. Then,*

$$\chi(\otimes_{i=1}^{m} Z_{k_i}(\mathbf{\Phi}_i;\mathbf{\Psi}_i)) \leq \sum_{i=1}^{m} \frac{k_i}{n_i} \sum_{j=1}^{n_i} \chi(\Phi_{i,j}) + \sum_{i=1}^{m} (1 - \frac{k_i}{n_i}) \sum_{j=1}^{n_i} \chi^{(pot)}(\Psi_{i,j}) + \Big(\sum_i k_i\Big) af(d, \delta) \ ,$$

with $af(d,\delta)$ defined as in Theorem 6.3. As a consequence, for a single flagged switch channel $Z_k(\mathbf{\Phi};\mathbf{\Psi})$,

$$C(Z_k(\mathbf{\Phi};\mathbf{\Psi})) \leq \frac{k}{n} \sum_{j=1}^{n} \chi(\Phi_j) + (1 - \frac{k}{n}) \sum_{j=1}^{n} \chi^{(pot)}(\Psi_j) + k \cdot af(d, \delta) \ .$$

Corollary 6.2 (based on and updated from corollary 4.3 from [27]). *Let the output system to each $\Phi_{i,j}$ be of dimension at most d, Φ_0 have output dimension d_0, and $\underline{\lambda} = 1/\lfloor \min_i n_i/k_i \rfloor$ be small enough such that*

$\delta = 4|B_0|\sqrt{\lambda \log d \log 2} \leq 2$ *and* $\delta_0 = 4|B_0|\sqrt{\lambda \log d_0 \log 2} \leq 2$. *Then,*

$$\begin{aligned}\chi(\Phi_0 \otimes \otimes_{i=1}^m Z_{k_i}(\mathbf{\Phi}_i;\mathbf{\Psi}_i)) \leq &\chi(\Phi_0) + \sum_{i=1}^m \frac{k_i}{n_i}\sum_{j=1}^{n_i}\chi(\Phi_{i,j}) + \sum_{i=1}^m (1-\frac{k_i}{n_i})\sum_{j=1}^{n_i}\chi^{(pot)}(\Psi_{i,j}) \\ &+ af(d_0,\delta_0) + \Big(\sum_i k_i\Big) af(d,\delta)\ ,\end{aligned}$$

where $af(d,\delta)$ *and* $af(d_0,\delta_0)$ *are defined as in Theorem 6.3.*

Lemma 6.3 (based on lemma 4.4 from [27])**.** *Let* $\Phi_0 : A_0 \to B_0$ *be a quantum channel. Let* B *be any quantum system, and* $\rho_x^{A_0B}$ *be a family of separable bipartite state. Then for any* $\eta = \sum_x p_x\eta_x$*, where* $\eta_x = id \otimes \Phi_0(\rho_x)$ *and* $\{p_x\}$ *probability distribution, we have*

$$H(B_0|B)_\eta - \sum_x p_x H(B_0|B)_{\eta_x} \leq \chi(\Phi_0)\ .$$

Proof. By separability, for each x, we may write

$$\eta_x = \sum_j p_{x,j}\eta_{x,j}^{B_0} \otimes \eta_{x,j}^B$$

as a convex combination of product states. Define a classical to quantum channel $\Phi^{cq} : X \to B$, where X is an extra classical system, by

$$\Phi^{cq}(|x,j\rangle\langle x,j|) = \eta_{x,j}^B\ .$$

From this, we define the classical-quantum states

$$\eta_x' = \sum_j p_{x,j}\eta_{x,j}^{B_0} \otimes |x,j\rangle\langle x,j|$$

and observe that $\eta_x = \Phi^{cq}(\eta_x')$ for each x. Applying the data processing inequality for conditional entropy,

$$\begin{aligned}H(B_0|B)_\eta - \sum_x p_x H(B_0|B)_{\eta_x} &\leq H(B_0)_\eta - \sum_x p_x H(B_0|B)_{\eta_x} \\ &\leq H(B_0)_\eta - \sum_x p_x H(B_0|X)_{\eta_x'} \\ &\leq H(B_0)_\eta - \sum_x p_x p_{x,j} H(B_0)_{\eta_{x,j}}\ .\end{aligned}$$

This has the form of the Holevo information for the state $\sum_{x,j} p_x p_{x,j}|x,j\rangle\langle x,j| \otimes \eta_{x,j}$. Therefore, it is less than the Holevo information of Φ_0. ■

Lemma 6.4 (based on lemma 4.5 from [27]). *Let $Z_{k_i}(\mathbf{\Phi}_i;\mathbf{\Psi}_i), 1 \leq i \leq m$ be a family of flagged switch channels and let Φ_0 be an arbitrary channel $\Phi_0 : A_0 \to B_0$. Then*

$$\chi(\Phi_0 \otimes_{i=1}^m Z_{k_i}(\mathbf{\Phi}_i;\mathbf{\Psi}_i)) \leq \chi(\Phi_0 \otimes_{i=1}^m Z_{k_i}(\mathbf{\Phi}_i)) + \sum_{i=1}^m (1-\frac{k_i}{n_i}) \sum_{j=1}^{k_i} \chi^{(pot)}(\Psi_{i,j}) \, .$$

Proof. Let $B = B_0 \otimes (\otimes_{i=1}^m \otimes_{j=1}^{n_i} B_{j,i})$ be the full quantum output system. For an classical quantum input state $\rho^{XA} = \sum_x p_x |x\rangle\langle x| \otimes \rho_x^A$, the output state is

$$\begin{aligned}
\omega^{XBY} &= \sum_x p_x |x\rangle\langle x| \otimes (\otimes_{i=1}^m Z_{k_i}(\mathbf{\Phi}_i;\mathbf{\Psi}_i)(\rho_x)) \\
&= \sum_x \sum_{R=R_1\cdots R_m, |R_i|=k_i} \frac{1}{\prod \binom{n_i}{k_i}} p_x |x\rangle\langle x|^X \otimes (\Phi^R \otimes \Psi^{R^c})(\rho_x) \otimes |R\rangle\langle R|^Y \, .
\end{aligned}$$

Note that the marginal distributions (reduced density) on the two classical system XY is independent. Thus we have

$$\chi(\Phi_0 \otimes_{i=1}^m Z_{k_i}(\mathbf{\Phi}_i;\mathbf{\Psi}_i)) = \sup_{\rho^{XA}} I(X:BY)_\omega = \sup_{\rho^{XA}} \frac{1}{\prod \binom{n_i}{k_i}} \sum_R I(X:B)_{\omega(R)} \, ,$$

where

$$\omega(R) = \sum_x p_x |x\rangle\langle x|^X \otimes (\Phi^R \otimes \Psi^{R^c})(\rho_x)$$

is the outcome of the heralding signal R. In each $I(X:B)_{\omega(R)}$, we could then replace R^c systems by the potential capacities of Ψ channels without decreasing the expression. That is

$$I(X:B)_{\omega(R)} \leq I(X:B)_{\omega(R)'} + \sum_{j \in R_i} \chi^{(pot)}(\Psi_{i,j}) \, ,$$

where

$$\omega(R)' = \sum_x p_x |x\rangle\langle x|^X \otimes (\Phi^R \otimes \Theta^{R^c})(\rho_x)$$

is the corresponding output of heralded channels. The result follows from summing over all R with uniform probabilities. ■

Lemma 6.5 (based on lemma 4.6 from [27]). *Let $\Phi_0 \otimes \Phi_1 : A_0 \otimes A \to B_0 \otimes B$ be a tensor product of channels. Let X be a classical system, with states $x \in X$. Let ρ_x be the input state to the channel for each x, and $\rho = \sum_x p_x \rho_x$. Let $\omega_x = \Phi_0 \otimes \Phi_1(\rho_x)$, and $\omega = \sum_x p_x \omega_x$. Assume $\exists$ a state $\eta = \sum_x p_x \eta_x$ that is* *separable between* B_0 *and* B *such that for each* x*,*

$$\|(id_{A_0} \otimes \Phi_1)(\rho_x) - \eta_x\|_1 \leq \delta \ .$$

Then

$$\begin{aligned} I(X : B_0 B)_\omega &\leq \chi(\Phi_0) + \chi(\Phi_1) + 2\delta \log |B_0| + (2+\delta) h\Big(\frac{\delta}{2+\delta}\Big) \\ &\leq \chi(\Phi_0) + \chi(\Phi_1) + 2\delta \log \Big(\frac{4|B_0|}{\delta}\Big) \ . \end{aligned}$$

Proof. For a classical-quantum state $\omega^{XB_0B} = \sum_x p_x |x\rangle\langle x| \otimes \omega_x^{B_0B}$,

$$\begin{aligned} I(X; B_0 B)_\omega &= H(B_0 B)_\omega - \sum_x p_x H(B_0 B)_{\omega_x} \\ &= \big(H(B_0|B)_\omega - \sum_x p_x H(B_0|B)_{\omega_x}\big) + \big(H(B)_\omega - \sum_x p_x H(B)_{\omega_x}\big) \ . \end{aligned} \tag{6.7}$$

The second half of this expression is upper bounded by the Holevo information of $\Phi_1(\rho^{XA})$ if

$$\omega^{XB_0B} = (id_X \otimes \Phi_0 \otimes \Phi_1)(\rho^{XA_0A}) \ .$$

For the first half, we use the factorization property,

$$\Phi_0 \otimes \Phi_1 = (\Phi_0 \otimes id_A) \circ (id_{A_0} \otimes \Phi_1) \ ,$$

where "*id*" represents the identity channel. Hence by assumption, $\exists \eta_x^{XA_0A}$ such that $\forall\ x \in X$,

$$\|(id_{A_0} \otimes \Phi_1)(\rho_x) - \eta_x\|_1 \leq \delta \ ,$$

and by the convexity of the 1-norm,

$$\|(id_{A_0} \otimes \Phi_1)(\rho) - \eta\|_1 \leq \delta \ .$$

The main subtlety of this part of the proof is that we need $\eta_x^{B_0}$ to be an output state of Φ_0, but the B part can be arbitrary. This is because we use this state to estimate only the $H(B_0|B)$ terms, and not the $H(B)$ terms. What we have just shown is that we can prepare a separable *input* state to the channel $\Phi_0 \otimes id_A$ that is close in the 1-norm sense to the output of $id_{A_0} \otimes \Phi_1$. By contractivity of the trace distance under quantum channels,

$$\|(\Phi_0 \otimes \Phi_1)(\rho_x) - \Phi_0 \otimes id_B(\eta_x)\|_1 \leq \delta, \text{ and } \|(\Phi_0 \otimes \Phi_1)(\rho) - \Phi_0 \otimes id_B(\eta)\|_1 \leq \delta \ .$$

By the Alicki-Fannes inequality 2.13, for all x,

$$|H(B_0|B)_{\omega_x} - H(B_0|B)_{\Phi_0 \otimes id_B(\eta_x)}| \leq \delta \log |B_0| + \Big(1 + \frac{\delta}{2}\Big) h\Big(\frac{\delta}{2+\delta}\Big) \ ,$$

and

$$|H(B_0|B)_{\omega} - H(B_0|B)_{\Phi_0 \otimes id_B(\eta)}| \leq \delta \log |B_0| + \Big(1 + \frac{\delta}{2}\Big) h\Big(\frac{\delta}{2+\delta}\Big) \ .$$

We now use the triangle inequality and Lemma 6.3.

$$\begin{aligned} &H(B_0|B)_\omega - \sum_x p_x H(B_0|B)_{\omega_x} \leq H(\Phi_0|B)_\eta - \sum_x p_x H(\Phi_0|B)_{\eta_x} + 2\delta \log |B_0| + (2+\delta) h\Big(\frac{\delta}{2+\delta}\Big) \\ &\leq \chi(\Phi_0) + 2\delta \log |B_0| + (2+\delta) h\Big(\frac{\delta}{2+\delta}\Big) \ . \end{aligned}$$

Returning to Equation (6.7), we obtain

$$I(X; B_0 B)_\omega \leq \chi(\Phi_0) + \chi(\Phi_1) + 2\delta \log |B_0| + (2+\delta) h\Big(\frac{\delta}{2+\delta}\Big) \ .$$

The final inequality in the Lemma follows from Equation (2.13). ■

Proof of Theorem 6.3, based on proof in [27]. First, by Lemma 6.4 it is sufficient to estimate $\chi(\Phi_0 \otimes_{i=1}^m Z_{k_i}(\mathbf{\Phi}_i))$. By Theorem 6.2, for each ρ_x there exists state $\eta_x^{A_0 AY}$ separable between A_0 and AY such that

$$\| id \otimes_{i=1}^m Z_{k_i}^{n_i}(id : \Theta)(\rho_x) - \eta_x \|_1 \leq 4|B_0| \sqrt{\underline{\lambda} H(B_0)_\rho \log 2} \ .$$

Let $\delta = 4|B_0|\sqrt{\underline{\lambda} \log |B_0| \log 2}$. Apply Lemma 6.5. ■

Proof of Corollary 6.1 based on proof in [27]. As before, Lemma 6.4 implies we need only estimate the Holevo information of $\otimes_{i=1}^m Z_{k_i}(\mathbf{\Phi}_i)$. We start by the averaging on the jth position being a position in the set R (of non-erased outputs) for a heralded channel $Z_k(\Phi_1, \cdots, \Phi_n)$. Since each R contains k positions, up to a re-ordering based on the classical signal,

$$
\begin{aligned}
Z_k(\Phi_1, ..., \Phi_n)(\rho) &= \frac{1}{\binom{n}{k}} \sum_R \Big(\Phi^R \otimes \Theta^{R^c} \Big)(\rho) \otimes |R\rangle\langle R| \\
&= \frac{1}{\binom{n}{k}} \sum_{|R|=k} \frac{1}{k} \sum_{j\in R} \Big(\Phi^R \otimes \Theta^{R^c} \Big)(\rho) \otimes |R\rangle\langle R| \\
&\to \frac{1}{n} \sum_{j=1}^{n} \Phi_j \otimes \frac{1}{\binom{n-1}{k-1}} \sum_{|R|=k-1, j\notin R} \Phi^R \otimes \Theta^{R^c}(\rho) \otimes |R\rangle\langle R| \qquad (6.8) \\
&= \frac{1}{n} \sum_{j=1}^{n} \Phi_j \otimes Z_{k-1}(\Phi_1, \cdots, \Phi_{j-1}, \Phi_{j+1}, \cdots, \Phi_n)(\rho) \ .
\end{aligned}
$$

Here in Equation (6.8), we rearrange the heralding sum by separating the j-th position in the index set R. Knowing this position in advance at the input does not lower the achievable rate, so we may estimate the channel capacity from above by moving this position to the front. Note that for the coefficients, $k\binom{n}{k} = n\binom{n-1}{k-1}$. Denote $\delta = 4|B_0|\sqrt{\underline{\lambda} \log |B_0| \log 2}$. We use the convexity of Holevo information over channels (reproved in [143] version 1, Proposition 4.3, though this had been previously known and is not hard to show) and Theorem 6.3,

$$
\begin{aligned}
\chi(\otimes_{i=1}^m Z_{n_i}(\mathbf{\Phi}_i)) \leq& \chi\bigg(\frac{1}{n_1} \sum_j \Phi_{1,j} \otimes Z_{k_1-1}(\Phi_{1,1}, \cdots, \Phi_{1,j-1}, \Phi_{1,j+1}, \cdots, \Phi_{n_1,1}) \otimes \otimes_{i=2}^m Z_{n_i}(\mathbf{\Phi}_i) \bigg) \\
\leq& \frac{1}{n_1} \sum_j \chi\bigg(\Phi_{1,j} \otimes Z_{k_1-1}(\Phi_{1,1}, \cdots, \Phi_{1,j-1}, \Phi_{1,j+1}, \cdots, \Phi_{n_1,1}) \otimes \otimes_{i=2}^m Z_{k_i}(\mathbf{\Phi}_i) \bigg) \\
\leq& \frac{1}{n_1} \sum_j \chi(\Phi_{1,j}) + 2\delta \log d + (2+\delta) h\Big(\frac{\delta}{2+\delta} \Big) \\
&+ \chi(Z_{n-1}(\Phi_{1,1}, ..., \Phi_{1,j-1}, \Phi_{1,j+1}, ..., \Phi_{1,k_1}) \otimes \otimes_{i=2}^m Z_{k_i}(\mathbf{\Phi}_i)) \Big) \ .
\end{aligned}
$$

We may repeat this procedure to separate out all Φ positions in each $Z_{k_i}(\mathbf{\Phi}_i)$. As we replace each $\Phi_{i,j}$ by its Holevo information plus the correction term, we are reducing n and k by the same amount, so $\underline{\lambda} = 1/\lfloor \min k_i/n_i \rfloor$ does not increase. Thus $2\delta \log d + (2+\delta)h(\delta/(2+\delta))$ is a uniform bound for the correction term at all steps. Therefore

$$
\chi(\otimes_{i=1}^m Z_{k_i}(\mathbf{\Phi}_i)) \leq \sum_{i=1}^m \frac{k_i}{n_i} \sum_{j=1}^{n_i} \chi(\Phi_{i,j}) + \Big(\sum_i k_i \Big) \Big(2\delta \log d + (2+\delta) h\big(\frac{\delta}{2+\delta} \big) \Big) \ .
$$

Regularizing the expression for $Z_k(\mathbf{\Phi})^{\otimes m}$ yields the bound for classical capacity. ■

Finally, Theorem 6.1 follows from the fact that in the limit of many copies of the channel, the number of successes of an erasure channel Z_λ concentrates around its average.

Theorem 6.4 (based on theorem 5.1 from [27])**.** *For any n and λ,*

$$|\chi(Z_\lambda(\Phi)^{\otimes n}) - \chi(Z^n_{\lfloor \lambda n \rfloor}(\Phi))| \leq \big(1 + \sqrt{n\lambda(1-\lambda)}\big)\chi^{(pot)}(\Phi) \ .$$

In particular, the classical capacity of the erasure channel $Z_\lambda(\Phi)$ can be rewritten by

$$C(Z_\lambda(\Phi)) = \lim_{n\to\infty} \frac{1}{n}\chi(Z^n_{\lfloor \lambda n \rfloor}(\Phi)) \ .$$

Even though the difference in Holevo information grows with the number of channel uses, the $O(\sqrt{n})$ growth is subleading in the asymptotic, many-copy limit. Under the $1/n$ regularization, it therefore goes to zero. To prove this, we rely on the following Lemma:

Lemma 6.6 (lemma 5.2 from [27])**.** *Let $k_1 \leq k_2$. For any classical-quantum input ρ,*

$$0 \leq I(X : Z^n_{k_2}(\Phi))_\rho - I(X : Z^n_{k_1}(\Phi))_\rho \leq (k_2 - k_1)\chi^{(pot)}(\Phi) \ .$$

Proof. By the zero contribution from erased outputs and classical conditionality of the mutual information,

$$I(X : Z^n_k(\Phi))_\rho = \frac{1}{\binom{n}{k}} \sum_{|R|=k} I(X : \Phi^R)_\rho \ .$$

Therefore, we have

$$\begin{aligned}
I(X : Z^n_{k_2}(\Phi))_\rho &= \frac{1}{\binom{n}{k_2}} \sum_{|R|=k_2} I(X : \Phi^R)_\rho \\
&= \frac{1}{\binom{n}{k_2}} \sum_{|R|=k_2} \frac{1}{\binom{k_2}{k_1}} \sum_{|P|=k_1, P\subset R} I(X : \Phi^P \otimes \Phi^{R\setminus P})_\rho \\
&\leq \frac{1}{\binom{n}{k_2}} \sum_{|R|=k_2} \frac{1}{\binom{k_2}{k_1}} \sum_{|P|=k_1, P\subset R} \big(I(X : \Phi^P)_\rho + (k_2 - k_1)\chi^{(pot)}(\Phi)\big) \\
&= \frac{1}{\binom{n}{k_1}} \sum_{|P|=k_1} I(X : \Phi^P)_\rho + (k_2 - k_1)\chi^{(pot)}(\Phi) \\
&= I(X : Z^n_{k_1}(\Phi))_\rho + (k_2 - k_1)\chi^{(pot)}(\Phi) \ .
\end{aligned}$$

The inequality follows from Lemma 6.3. The last step is because each k_1-subset P has been counted $\binom{n-k_1}{k_2-k_1}$ times as a subset of some k_2-set R. The inequality follows similarly. ■

Proof of Theorem 6.4, duplicating proof from [27]. Let A be the input system of Φ. Then the two channels

$Z_\lambda(\Phi)^{\otimes n}$ and $Z^n_{\lfloor \lambda n \rfloor}(\Phi)$ have the same input $A^{\otimes n}$. Denote ρ^{XA^n} by a classical-quantum state. By the triangle inequality and Lemma 6.6,

$$\begin{aligned}
|\chi(Z_\lambda(\Phi)^{\otimes n}) - \chi(Z^n_{\lfloor \lambda n \rfloor}(\Phi))| &= |\sup_\rho \sum_{k=0}^n \binom{n}{k} \lambda^k (1-\lambda)^{n-k} I(X : Z^n_k(\Phi)) - \sup_\rho I(X : Z^n_{\lfloor \lambda m \rfloor}(\Phi))_\rho| \\
&\leq \sup_\rho |\sum_{k=0}^n \binom{n}{k} \lambda^k (1-\lambda)^{n-k} I(X : Z^n_k(\Phi)) - I(X : Z^n_{\lfloor \lambda m \rfloor}(\Phi))_\rho| \\
&\leq \sum_{k=0}^n \binom{n}{k} \lambda^k (1-\lambda)^{n-k} |\sup_\rho I(X : Z^n_k(\Phi)) - I(X : Z^n_{\lfloor \lambda m \rfloor}(\Phi))_\rho| \\
&\leq \chi^{pot}(\Phi) \sum_{k=0}^n \binom{n}{k} \lambda^k (1-\lambda)^{n-k} |k - \lfloor \lambda n \rfloor| \\
&\leq \chi^{pot}(\Phi) \big(1 + \sum_{k=0}^n \binom{n}{k} \lambda^k (1-\lambda)^{n-k} |k - \lambda n|\big) \ .
\end{aligned}$$

Recall that the variance of binomial distribution is $n\lambda(1-\lambda)$. We obtain by Hölder's inequality that

$$\sum_{k=0}^n \binom{n}{k} \lambda^k (1-\lambda)^{n-k} |k - \lambda n| \leq (\sum_{k=0}^n \binom{n}{k} \lambda^k (1-\lambda)^{n-k} |k - \lambda n|^2)^{1/2} = \sqrt{n\lambda(1-\lambda)} \ .$$

■

Putting this together with the bounds on heralded channels completes Theorem 6.1.

6.2.3 Bounds for Rates with Finite Block Size

In practice, a common scenario in practice is frequent failure with small encoding block size, due to practical constraints on creating large, entangled states with high success probability. While the Shannon regime assumes local resources are effectively limitless and infallible, the reality of quantum technology is far from it. For short-distance quantum communication, such as between pieces of a distributed quantum computer, the encoding/decoding cost may be as larger or larger of a concern than the transmission channel. When block size is small and success rare, usually there will be only one success per block, eliminating superadditivity. Here we show that these assumptions result in superadditivity violations being at most of quadratic order in the transmission success probability. This does not rely on monogamy of entanglement, but reflects the quadratic-order probability that two successes appear in the same block. This result further gives some intuition for why in lossy channels, superadditive effects are often not large enough to justify the use of entangled blocks. There may yet be potential enhancements from encodings that use multiple degrees of freedom in the same information carrier (see Example 4.1), resulting in correlated loss.

One might wonder if we can apply the methods of this section or use entanglement monogamy to bound superadditivity of the quantum and private capacities. Technically, this might be the case, though the literal expressions are more complicated due to quantum subsystems potentially kept back at the input. Practically, however, both the (one-way) quantum and private capacities go to 0 when the failure probability is 1/2. The channel becomes *antidegradable*, meaning that an eavesdropper could reconstruct a copy of the output density from the channel's environment. It is intuitive why such a channel is useless for sending private information. Quantum information is inherently private due to effects related to entanglement monogamy and the no cloning theorem, so transmitting qubits is at least as hard as transmitting private bits. Even when some qubits may randomly reach the end, achieving a given quantum capacity requires the existence of an encoding/decoding scheme that reconstructs inputs with arbitrarily high success probability, and this is impossible for any block size given such a channel.

With feedback or two-way classical assistance, these channels often regain their ability to send and receive quantum or private data. This regime is however much more complicated to model, and even writing the quantum and private capacity with such assistance as a concise entropy expression remains an open question. Classically, capacity does not change with feedback, as one can always find a one-way encoding that reproduces whatever gains would appear from having the sender respond to information from the receiver. For quantum channels this is not the case, and the d-dimensional erasure channel with success probability λ is an especially dramatic example, in which the capacity for a channel with $\lambda = 1/2$ goes from zero with no extra assistance to $(\log d)/2$ with a polynomial amount of two-way classical resources [144]. Based on present-day technology, however, quantum data is likely to have a great deal of classical help, but also suffer from high loss and encoding overhead. It's plausible that classical assistance will play a much larger role in near-term quantum communication than superadditivity.

Proposition 6.1 (proposition 7.1 from [27]). *Let $\Phi_1, \cdots, \Phi_n$ be a family of quantum channels, and let $\lambda \ll 1/n$. Let $F^{(1)}$ be a function mapping densities to positive real numbers, and define its value for a quantum channel by $F^{(1)}(\Phi) = \max_\rho\{F^{(1)}(\Phi(\rho))\}$. Assume $F^{(1)}$ is superadditive on quantum channels, additive for separable input states, convex in the input state and channel, and admit a well-defined expression of the form $F^{(pot)}(\Phi) = \max_\Psi\{F^{(1)}(\Phi \otimes \Psi) - F^{(1)}(\Psi)\}$ such that $F^{(pot)}(\Theta) = 0$ if Θ is a trivial channel for which the output provides no information about the input. Then*

$$F^{(1)}(Z_\lambda(\Phi_1) \otimes ... \otimes Z_\lambda(\Phi_n)) - \lambda \sum_{i=1}^{n} F^{(1)}(\Phi_i) \leq O\left(\lambda^2 \sum_i \left(F^{(pot)}(\Phi_i) - F^{(1)}(\Phi_i)\right)\right) .$$

In words, Proposition 6.1 shows that a broad class of entropy functions depending on the output of a quantum channel Φ with fixed block size can be written as a first order additive term, plus a correction of quadratic order in the erasure probability, when applied to a frequently erasing channel. Theorem 6.1 bounds superadditivity in the Shannon regime with unbounded blocksize, while Proposition refprop:finiteblock gives a bound of potentially stronger order under the finite blocksize assumption. Furthermore, Proposition 6.1 may apply to a wider variety of information measures than just Holevo capacity.

Proof. If only one success has occurred at position i, then

$$F^{(1)}(\Theta_1 \otimes \cdots \otimes \Phi_i \otimes \cdots \otimes \Theta_n) = F^{(1)}(\Phi_i). \tag{6.9}$$

This follows from the fact that the trivial channels provide no information about the input, and being uncorrelated with it, are therefore not entangled either. Since F is additive on separable states and 0 on these trivial channels, they contribute nothing. For some tensor product $\Phi^R \otimes \Theta^{R^c}$, define $\Psi_1 \cdots \Psi_n$ such that $\Phi^R \otimes \Theta^{R^c} = \Psi_1 \otimes \cdots \otimes \Psi_n$.

$$\begin{aligned}
&F^{(1)}(\Psi_1 \otimes \cdots \otimes \Psi_n) \leq F^{(pot)}(\Psi_1) + F^{(1)}(\Psi_2 \otimes \cdots \otimes \Psi_n) \\
&= F^{(1)}(\Psi_1) + F^{(1)}(\Psi_2 \otimes \cdot \otimes \Psi_n) + F^{(pot)}(\Psi_1) - F^{(1)}(\Psi_1) \\
&\leq \sum_{i \in R} F^{(1)}(\Phi_i) + \sum_{i \in R : i \neq |R|} \Big(F^{(pot)}(\Phi_i) - F^{(1)}(\Phi_i)\Big),
\end{aligned} \tag{6.10}$$

where the last inequality is obtained by iterating and discarding the $F^{(pot)}(\Theta_i)$ terms that are zero anyway. Note that the final correction term has one fewer channel correction term than are channels, because Equation (6.9) shows that we do not need to add a correction term when only one success is involved. Using Equations

(6.10) and (6.9), and the convexity of $F^{(1)}$ in the channel,

$$
\begin{aligned}
&F^{(1)}(Z_\lambda(\Phi_1)\otimes\cdots\otimes Z_\lambda(\Phi_n)) \leq \sum_{k=0}^{n}\sum_{|R|=k}\lambda^k(1-\lambda)^{n-k}F^{(1)}(\Phi^R\otimes\Theta^{R^c})\\
&\leq \sum_{k=1}^{n}\sum_{|R|=k}\lambda^k(1-\lambda)^{n-k}\sum_{i\in R}F^{(1)}(\Phi_i)+\sum_{k=2}^{n}\sum_{|R|=k}\lambda^k(1-\lambda)^{n-k}\sum_{i\in R:i\neq|R|}(F^{(pot)}(\Phi_i)-F^{(1)}(\Phi_i))\\
&\leq \sum_{k=1}^{n}\binom{n}{k}\lambda^k(1-\lambda)^{n-k}\frac{k}{n}(\sum_{i=1}^{n}F^{(1)}(\Phi_i))+\sum_{k=2}^{n}\binom{n}{k}\lambda^k(1-\lambda)^{n-k}\frac{k-1}{n}\sum_{i=1}^{n}(F^{(pot)}(\Phi_i)-F^{(1)}(\Phi_i))\\
&\leq \lambda\sum_{i=1}^{n}F^{(1)}(\Phi_i)+\lambda\sum_{k=2}^{n}\binom{n-1}{k-1}\lambda^{k-1}(1-\lambda)^{n-k}(\sum_{i=1}^{n}F^{(pot)}(\Phi_i)-F^{(1)}(\Phi_i))\\
&\leq \lambda\sum_{i=1}^{n}F^{(1)}(\Phi_i)+\lambda(1-(1-\lambda)^{n-1})\sum_{i=1}^{n}(F^{(pot)}(\Phi_i)-F^{(1)}(\Phi_i))\\
&\leq \lambda\sum_{i=1}^{n}F^{(1)}(\Phi_i)+O\big(\lambda^2(\sum_{i=1}^{n}F^{(pot)}(\Phi_i)-F^{(1)}(\Phi_i))\big)\ .
\end{aligned}
$$

In the last inequality, we used that for small λ,

$$
\lambda(1-(1-\lambda)^{n-1}) \approx \lambda(1-(1-\frac{1}{n-1}\lambda)) = \frac{\lambda^2}{n-1}\ .
$$

■

6.3 Entanglement Monogamy Bounds Non-Classicality of Quantum Games

While the majority of [27] focuses on superadditivity of Holevo capacity, we also include a basic result on a special kind of quantum game. Consider a game with $n+1$ players: one Alice, and n "Bob" players. At the beginning of the game, all of the players share some quantum state $\rho \in S(AB_1...B_n)$, which might be entangled between any and all systems involved. A referee then chooses some $x \in X$, where X is the set of possible questions to Alice, and $y \in Y$, where Y is the set of possible questions to Bob. Furthermore, in our setting, the referee randomly selects some $i \in 1...n$ and sends only that "Bob" player y. Alice may then apply any local operations and measurements to A, and the chosen Bob to B_i. Alice and the ith Bob then return respective answers a and b to the referee. In a traditional, two-player game, $n = 1$.

There are many two-player games in which Alice and Bob gain some advantage by pre-sharing quantum entanglement. The general form of these games forces Alice and Bob to compute different parts of the solution to some problem that does not have a complete, universal solution. The referee checks that each piece is valid, and also that answers are consistent. In the CHSH game (named for Clauser, Horne, Shimony, & Holt, and based on the Bell test and inequality proposed by these authors [145]), Alice and Bob are sent random bits x and y, and they must return bits a and b such that $x \cdot y = a \oplus b$, where "$\cdot$" denotes multiplication in binary and "$\oplus$" denotes addition mod 2. One can easily write down a truth table of all possible questions and answers, and see that the best classical strategy is for both Alice and Bob to return 0, which succeeds in 75% of cases. Pre-shared classical randomness does not lead to a better strategy. With a shared Bell state, they can achieve an 85% success probability. Despite its simplicity, the CHSH game is deep and powerful, as it is *only* possible to win many copies of this game with near-optimal probability if Alice and Bob's pre-shared state nearly approximates that many copies of a Bell state [146]. The broader name for this property is *rigidity*: a game that is rigid has a specific pre-shared state (up to unitary conjugation) underpinning every successful strategy. Quantum games are one of the classic theoretical techniques to illuminate how quantum mechanics changes the rules of the universe.

For a two-player quantum game G, we denote by

$$val(G) = \sup_{a_x, b_y} \sum_{x,y,a,b} \pi(x,y)v(a,b,x,y) \int_\Omega a_x(\omega)b_y(\omega)d\mathbb{P}(\omega)$$

the winning probability of the best possible classical strategy for G, and by

$$val^*(G) = \sup_{\rho, E_x^a, F_y^b} \sum_{x,y,a,b} \pi(x,y)v(a,b,x,y)tr(\rho E_x^a \otimes F_y^b)$$

that of the best possible quantum strategy (which may include pre-shared entanglement).

There is some precedent for using monogamy to bound values of quantum games. Vidick et. al. used tripartite entanglement monogamy to show that quantum multi-prover games have nearly classical values, making their values similarly difficult to approximate [147]. Other previous authors noted the effects of monogamy of quantum correlations [148, 149] and the link between symmetric extendibility and hidden variable theories [150].

We modify the two-player quantum game to have n "Bob" players, only one of whom actually plays, such that neither Alice nor any "Bob" player know ahead of time which the referee will select. One way to model this is by requiring the players to pre-share a quantum state, and then play any of n quantum games $G_1, ..., G_n$, each differing only in its choice of "Bob" player. We then denote by $Aval(\{G_i\})$ the optimal winning probability for classical strategies averaged over $\{G_i\}$, and by $Aval^*(\{G_i\})$ that for quantum strategies.

Theorem 6.5 (based on and updated from theorem 7.4 from [27]). *Let $G_1, \cdots, G_n$ be a family of bipartite games that Alice play respectively with the players $B_1, \cdots, B_n$. Then*

$$Aval^*(\{G_i\}) - Aval(\{G_i\}) \leq 4\sqrt{\log 2}(\min\{|A|, |B|\} - 1/2)\frac{\log \min\{|A|, |B|\}}{\sqrt{n}} .$$

Proof. Let $G = (A, B, X, Y, \pi, v)$ be a bipartite game. For fixed systems $\mathcal{A}, \mathcal{B}$ and POVMs E_x^a, F_y^b, the value function can be viewed as a positive linear functional l_G on the trace-class $S_1(\mathcal{A} \otimes \mathcal{B})$,

$$l_G(\rho^{\mathcal{AB}}) = \sum_{x,y,a,b} \pi(x, y)v(a, b, x, y)tr(\rho E_x^a \otimes F_y^b) .$$

Note that l_G is of norm at most 1. Then for a separable σ and an arbitrary ρ,

$$l_G(\rho) \leq l_G(\rho - \sigma) + l_G(\sigma) \leq \|\rho - \sigma\|_1 + val(G) .$$

Now suppose that the quantum system $\mathcal{A}$ of Alice is of dimension at most d. We know by the monogamy of entanglement (Equation (6.5)) that for any $\rho^{AB_1 \cdots B_n}$,

$$\frac{1}{n}\sum_{i=1}^{n} E_{sq}(A, B_i)_\rho \leq \frac{1}{n}S(A)_\rho \leq \frac{\log d}{n} .$$

It follows from the faithfulness of squashed entanglement that there exists a state σ^{AB_i} separable on A and B_i such that

$$\|\rho^{AB_i} - \sigma^{AB_i}\|_1 \leq 4\sqrt{\log 2}(\min\{|A|, |B|\} - 1/2)\sqrt{E_{sq}(A, B_i)_\rho}$$

Thus,

$$l_{G_i}(\rho^{AB_i}) \leq 4\sqrt{\log 2}(\min\{|A|, |B|\} - 1/2)\sqrt{E_{sq}(A, B_i)_\rho} + val(G_i) ,$$

and then the average entangled value obeys

$$Aval_d^*(\{G_i\}) \leq Aval(\{G_i\}) + 4\sqrt{\log 2}(\min\{|A|, |B|\} - 1/2)\frac{1}{n}\sum_{i=1}^{n} \sqrt{E_{sq}(A, B_i)_\rho} .$$

By Hölder's inequality, we have

$$\sum_{i=1}^{n} \sqrt[4]{E_{sq}(A, B_i)_\rho} \leq \sqrt{n \sum_{i=1}^{n} E_{sq}(A, B_i)_\rho)} \leq \sqrt{n \log \min\{|A|, |B|\}} \ .$$

Therefore,

$$Aval_d^*(\{G_i\}) - Aval(\{G_i\}) \leq 4\sqrt{\log 2}(\min\{|A|, |B|\} - 1/2)\frac{\log \min\{|A|, |B|\}}{\sqrt{n}} \ .$$

■

6.4 Potential Experiment to Test Superadditivity of Coherent Information

In an ongoing effort, the Kwiat lab at UIUC seeks to make an experimental demonstration of the superadditivity of transmission rates through a quantum channel *. As noted, this is prohibitively difficult to show for Holevo information due to a lack of low-dimensional, easily preparable examples that are sufficiently robust to noise at the encoder and decoder. Instead, we consider a formula for the rate of one-way qubit transmission, the coherent information. For a state ρ^{AB}, the coherent information is simply the negative conditional entropy,

$$I_c(A\rangle B)_\Phi(\rho) = -H(B|A)_{\Phi(\rho)} = D^{\mathcal{A}}(\Phi(\rho)) - \log |B| \ . \tag{6.11}$$

For a quantum channel $\Phi : A' \to B$, we define

$$Q^{(1)}(\Phi) = \sup_{\rho^{AA'}} I_c(A\rangle B)_{(\hat{1}^A \otimes \Phi)(\rho)} \ .$$

$Q^{(1)}$ is an achievable rate of qubit transmission with asymptotically many channel copies and perfect encoding/decoding, though like the Holevo information, it requires regularization to become a capacity. The quantum capacity is given by

$$Q(\Phi) = \lim_{n\to\infty} \frac{1}{n} Q^{(1)}(\Phi^{\otimes n}) \ . \tag{6.12}$$

It is important to note that the quantum capacities with classical feedback, with two-way classical communication, and with quantum back-transmission are in general all different. Hence while many texts describe

*During the editing of this thesis, a new result appeared [151] claiming to experimentally demonstrate superadditivity of the same channel proposed in this section. Though the proposal here would not be the first demonstration of superadditive rates, efforts to improve upon these results continue.

the quantum capacity as the ultimate rate of quantum information transmission, this is under strong assumptions. As with the Holevo information, superadditivity of $Q^{(1)}$ demonstrates a clear distinction between the additive regime of classical rates and the quantum setting. Unlike Holevo information, $Q^{(1)}$ shows superadditivity on some known, low-dimensional, easily preparable input states. This makes it a much more practical candidate for an experimental demonstration.

The *dephrasure* channel proposed in [12] shows superadditivity of coherent information with few channel uses. Its original form is given as

$$\Phi^{dpr}_{p,q}(\rho) = (1-q)((1-p)\rho + pZ\rho Z) + q\mathrm{tr}(\rho)|e\rangle\langle e| \,, \tag{6.13}$$

where ρ is the input density, Z the Pauli matrix, q the probability of erasure, and p the probability of dephasing given that the channel did not erase. For the dephrasure channel, $Q^{(1)}(\Phi^{dpr}_{p,q} \otimes \Phi^{dpr}_{p,q}) > 2Q^{(1)}(\Phi^{dpr}_{p,q})$ in some parameter regimes. Superadditivity of coherent information appears with a simple repetition code input,

$$|\phi_2(\lambda)\rangle \equiv \sqrt{\lambda}^{AA'} |000\rangle + \sqrt{1-\lambda}\,|111\rangle^{AA'} \tag{6.14}$$

with a 1-qubit A and 2-qubit A' system (for the 3 total qubits shown). This puts the dephrasure channel potentially within the precision and dimension ranges achievable by quantum optics.

One of the major barriers to quantum optics is creating entanglement and applying multi-qubit gates between photons. While theoretical schemes including gate-based [142] and cluster-state optical quantum computers achieve scalability in principle, many of the usual methods for quantum computing [152, 153] impose high overhead in straightforward implementations. There is still ongoing research into large-scale optical quantum computing that shows potential promise [154].

One workaround for the difficulty of entangling photons exploits multiple degrees of freedom on each photon. Widely explored phenomena around this idea include hyperentanglement [155], in which multiple photons are entangled across multiple degrees of freedom simultaneously, and hybrid entanglement, which may refer to cross-photon or same-photon, cross-aspect entanglement [156, 157]. As an example, time bin encoding allows one photon to carry a large amount of information [158], which can be useful in brightness-limited quantum cryptography and tests of high-dimensional entanglement.

The effort to experimentally demonstrate superadditivity of the dephrasure channel continues [28]. This is an early step toward determining whether superadditivity is likely to be of practical value in quantum communication.

Chapter 7

Relative Entropy Decay

Previous chapters of this thesis have focused on the nature of quantum entropy and its non-classical properties. Chapter 3 essentially dealt with statics, deriving relations between different entropies on fixed states. Chapters 4 and 6 introduce dynamics in the use of entropy as a non-classical resource but still delegate transformations to discrete operations, encodings and transmissions. This chapter focuses on entropy as a dynamical quantity, subject to (often unwanted) interactions between system and environment. The decay of quantum states toward equilibrium is a primary driver of decoherence in quantum technologies, one of the main challenges to scalability of quantum information.

As briefly noted, the relative entropy can serve as a distance-like quantity between states. $D(\rho\|\omega) = 0 \iff \rho = \omega$, so it is faithful. Since $D(\rho\|\omega) \neq D(\omega\|\rho)$ in general, we cannot literally interpret it as a distance. It is however bi-convex in its two arguments, satisfies the data processing inequality as recalled in Chapter 3, and upper bounds the 1-norm distance. In many cases, we wish to compare a given state ρ to some state that plays the role of ultimate decay, such as a free state in a resource theory, a fixed point state as described in Section 7.2, or the output of a worst-case member of a family of quantum channels. Often we will find that such a comparison involves the expression $D(\rho\|\mathcal{E}(\rho))$ for some conditional expectation, which we saw in Chapters 3, 4, and 5 has good properties, such as being always finite and equaling $H(\mathcal{E}(\rho)) - H(\rho)$.

We can apply our results to estimate decay of asymmetry and coherence as resources (see Section 4.2), quantum capacity, and well as other resources characterized by relative entropy with respect to a subalgebraic projection of the input state. By Pinsker's inequality (Equation (2.12)), decay of relative entropy implies 1-norm distance decay. It therefore lets us estimate decoherence times, defined as

$$t_\epsilon^{dec}(\{\Phi^t\}) = \inf\{t \geq 0 : \|\Phi^t(\rho) - \Phi^\infty(\rho)\|_1 \leq \epsilon \forall \rho \in S(\mathcal{H})\} \tag{7.1}$$

for a family of channels $\{\Phi^t\}$ parameterized by $t \in \mathbb{R}^+$.

$D(\mathcal{E}(\rho)\|\rho)$ would not be as nice a form, as it is often infinite, does not equal an entropy difference, etc.

This chapter includes results appearing in [33], co-authored with Ivan Bardet, Marius Junge, Cambyse Rouzé, and Daniel Stilck França, results appearing in [29, 30, 31, 25], co-authored with Li Gao and Marius Junge, results appearing in [32], co-authored with Marius Junge and Cambyse Rouzé, and results appearing in [26].

This asymmetry in the relative entropy may not be a flaw. Decoherence seems to have a preferred direction - it would be a wonderful but unlikely phenomenon if densities were to decohere toward long-range Bell pairs, cluster states, or other states with useful, non-classical properties.

In this chapter more than previous ones, I rely on quantum channels and quantum Markov semigroups as recalled in Section 2.6. Like entropy, the quantum channel is historically analogous to the channel in classical information theory, but its physical interpretations go well beyond the context of communication. In Section 7.1, I recall some results with Gao and Junge on a method for estimating the change in relative entropy with respect to a subalgebra for a class of parameterized quantum channels. In Section 7.2, I describe and recall a class of exponential decay bounds for channels modeling continuous decay. In Section 7.3, I briefly discuss some reasons why decay toward states without accessible quantum properties might be naturally expected.

7.1 Interpolation to Estimate Relative Entropy

This section summarizes a method to estimate relative entropy at the output of channels from a particular class. The main mathematical base comes from the paper co-authored with Gao and Junge on "TRO channels [29]." Let B and E be some output and environment systems, denoting by $\mathbb{B}(B, E) \cong \mathbb{B}(B \otimes E)$ the bounded operators from B to E. A *ternary ring of operators* (TRO) is a subalgebra $X \subset \mathbb{B}(B, E)$ such that $\forall x, y, z \in X$, $xy^*z \in X$ [159].

We may construct a Hilbert space $X_2 \subseteq B \otimes E$ by applying the inner product $\langle x|y\rangle = \mathrm{tr}(x^* y)$ to elements of X. We define an observable algebra accessible to the environment $E(X) = \mathrm{span}\{x^* y | x, y \in X\}$, which we call the *right algebra of* X. Similarly, we define the *left algebra* $B(X) = \mathrm{span}\{xy^* | x, y \in X\}$, which is accessible at the output. Let $\mathcal{N} \subseteq \mathbb{B}(X_2)$ be a subalgebra such that $\mathcal{E}_{\mathcal{N}} \mathcal{E}_{E(X)} = \mathcal{E}_{E(X)} \mathcal{E}_{\mathcal{N}}$ (a commuting square). Then $\forall f \in \mathcal{N}$, we define a channel $\Phi_f : \mathbb{B}(X_2) \to \mathbb{B}(B)$ given by

$$\Phi_f(|x\rangle \langle y|) = x f y^* . \tag{7.2}$$

Φ_1 is what we call the *TRO channel* - this is the conditional expectation onto $E(X)$, or a full averaging over $\mathcal{N}$. f is the *symbol* for Φ_f. When $E(X)$ is a commutative group, we have a particularly simple form of channel,

$$\Phi_{G,f}(\rho) = \sum_{g \in G} f(g) U_g \rho U_g^\dagger , \tag{7.3}$$

where in this case $f(g)$ is equivalent to a probability distribution on G. Another of our papers [160] contains more complicated examples, such as depolarizing and quantum group channels. Once we have established a

TRO, we look to operator theory to estimate entropy:

Theorem 7.1 (Kosaki's Interpolation [161], modified as in [29])**.** *Let X be a TRO. For positive $\sigma \in B(X)$, $1 \leq p \leq \infty$ and $\theta \in [0,1]$, let $X_{p,\theta,\sigma,\rho}$ be X with the norm $\|x\|_{p,\theta,\sigma,\rho} = \|\sigma^{\theta/p} x \rho^{(1-\theta)/p}\|_p$. Then $X_{p,\theta,\sigma,\sigma} = [X_\infty, X_{1,\theta,\sigma,\rho}]_{1/p}$ is an interpolation space, where X_∞ is X with the $p = \infty$ norm $\|x\|_\infty$. In particular, this means that*

$$\|x\|_{p,\theta,\sigma,\rho} \leq \|x\|_\infty^{1-1/p} \|x\|_{1,\theta,\sigma,\rho}^{1/p} \tag{7.4}$$

The interpolation theorem allows us to estimate one hard-to-estimate norm by a product of easier norms. Our main consequence is the following *entropy comparison theorem*:

Theorem 7.2 (Comparison Theorem from [29])**.** *Let Φ_f be a TRO channel for TRO X with symbol f. Then for any $\sigma \in B(X)$, $1 \leq \alpha \leq \infty$ st. $1/\alpha + 1/\alpha' = 1$, and ρ,*

$$\|\sigma^{-1/2\alpha'} \Phi_1(\rho) \sigma^{-1/2\alpha'}\|_\alpha \leq \|\sigma^{-1/2\alpha'} \Phi_f(\rho) \sigma^{-1/2\alpha'}\|_\alpha \leq \|f\|_\alpha \|\sigma^{-1/2\alpha'} \Phi_1(\rho) \sigma^{-1/2\alpha'}\|_\alpha \tag{7.5}$$

We then make use of the fact that various entropy expressions are equivalent to matrix norms, in particular the sandwiched Rényi relative α-entropy as defined in Equation (2.7). Hence the comparison theorem, restated in terms of α-entropy, becomes (for normalized f),

$$D_\alpha(\Phi_1(\rho)\|\sigma) \leq D_\alpha(\Phi_f(\rho)\|\sigma) \leq D_\alpha(\Phi_1(\rho)\|\sigma) + D_\alpha(f\|\hat{1}^E/|E|) \tag{7.6}$$

We also derived a version for continuous environments in [33]:

Theorem 7.3 (Theorem 2.1 from [33])**.** *Let G be a compact group with Haar measure μ, and f a bounded, measurable function such that $\int f(g) d\mu(g) = 1$. Let Φ_f be a quantum channel given by*

$$\Phi_f(\rho) = \int f(g) u_g \rho u_g^\dagger d\mu(g) \ ,$$

where u_g is given by some projective representation of G. Then $\Phi_{\hat{1}}$ is a conditional expectation, and for any σ such that $\Phi_1(\sigma) = \sigma$ and $\alpha \in [1, \infty]$,

$$D_\alpha(\Phi_f(\rho)\|\sigma) \leq D_\alpha(\Phi_1(\rho)\|\sigma) + D_\alpha(f\mu\|\mu) \ ,$$

where $D_\alpha(f\mu\|\mu) \equiv \alpha' \log \int_G f^\alpha d\mu(g)$. Here $D_1(f\mu\|\mu) = \int f \log f d\mu$, and $D_\infty(f\mu\|\mu) = \log \|f\|_\infty$.

I leave the proofs of the comparison Theorem to our three respective papers on the topic, [29, 30, 33].

While the original method was developed for quantum channel capacity, the formalism should apply to relative entropy of asymmetry and coherence at the output of channels, and possibly inform general scenarios in which a subsystem of a larger quantum system acquires entropy during interaction.

One primary application of the comparison Theorem, and the focus of [29, 30] is to estimate quantum capacity. For a finite group G and channel Φ_f in the form of Equation (7.3),

$$Q(\Phi_1(\rho)) \leq Q(\Phi_f(\rho)) \leq Q(\Phi_1(\rho)) + \log |G| - H(f) \ . \tag{7.7}$$

The comparison Theorem bypasses the challenge of superadditivity (as discussed at length in Chapter 6), estimating capacities that would at first glance appear difficult. The bounds implied by the comparison theorem are tensor-stable, so they are stable under regularization.

The comparison Theorem is a perturbative bound, but rather than perturb around a ground state or weakly-interacting model, it perturbs around a conditional expectation. The success of this technique relies on the relative simplicity of computing $D^{\mathcal{N}}$ for a subalgebra $\mathcal{N}$. Fukuda and Wolf had previously studied additivity problems for classical capacity of direct sums of quantum channels [162], which are closely related and often equivalent to conditional expectations. In [25], we further address calculation of $D^{\mathcal{N}}$ and its relation to the mathematical questions of index theory. Conditional expectations also provide simple examples of channels that are not degradable, but have additive coherent information and thereby often admit simple, analytical formulae for several kinds of capacity. The broader philosophy of the comparison Theorem is that even though conditional expectations are often strongly-interacting processes, they present points at which entropies are easy to calculate, and easy to perturb around.

7.2 Modified Logarithmic Sobolev Inequalities (MLSIs) for Quantum Markov Semigroups

Classical Sobolev inequalities and logarithmic Sobolev inequalities have a long history, only a recent sliver of which is directly used in this thesis. Sobolev inequalities concern norms on Sobolev spaces, which combine L^p norms of a function with its derivatives. Logarithmic Sobolev inequalities relate notions of entropy to Dirichlet forms [163, 164]. Due to some difficulties in extending this formalism to the general quantum case, Kastoryano and Temme defined [164] (and Bardet later extended [165]) the modified logarithmic Sobolev inequalities. Let $\{\Phi^t : t \in \mathbb{R}^+\}$ be a semigroup of quantum channels. $\{\Phi^t\}$ satisfies an α-MLSI for $\alpha \in \mathbb{R}^+$ if

$$D(\Phi^t(\rho)\|\Phi^\infty(\rho)) \leq e^{-\alpha t} D(\rho\|\Phi^\infty(\rho)) \ , \tag{7.8}$$

where Φ^∞ is the infinite-time limit point of Φ^t, for all input densities ρ. We will often refer to Φ^∞ as projecting to a fixed point subspace. Many results are simpler if Φ^∞ is a single state. The canonical example of single-state noise is the depolarizing channel in dimension d:

$$\Phi^t_{dep(t_0)}(\rho) = e^{-t/t_0}\rho + (1 - e^{-t/t_0})\frac{\hat{1}}{d} \ . \tag{7.9}$$

This is known to have MLSI with $\alpha = 1/t_0$, using simple convexity arguments. We also obtain many simplifications when Φ^t is self-adjoint for all t, in which case the fixed point subspace is a subalgebra, and Φ^∞ becomes a conditional expectation. The depolarizing channel is self-adjoint, and its fixed point conditional expectation is the projection to $\mathbb{C}$, leaving any input state in complete mixture. A self-adjoint semigroup with a non-trivial fixed point algebra is that of continuous pinching, given by

$$\Phi^t_{deph(t_0)}(\rho) = e^{-t/t_0}\rho + (1 - e^{-t/t_0})\mathcal{E}_{\{|i\rangle\}}(\rho) \ , \tag{7.10}$$

where $\{|i\rangle\}$ is some measurement basis. Continuous dephasing also has $(1/t_0)$-MLSI. More broadly, semigroups that exponentially replace ρ by some conditional expectation of it have this form of MLSI [165].

A stronger form of MLSI is a complete logarithmic Sobolev inequality (CLSI):

$$D((\Phi^t \otimes \hat{1}^B)(\rho)\|(\Phi^\infty \otimes \hat{1}^B)(\rho)) \leq e^{-\alpha t} D(\rho\|(\Phi^\infty \otimes \hat{1}^B)(\rho)) \ , \tag{7.11}$$

where ρ is any state on the input space of $\{\Phi^t\}$ and a finite-dimensional extension B. Unlike MLSI, CLSI is automatically tensor-stable: if $\{\Phi^t\}$ and $\{\Psi^t\}$ are two semigroups with α-CLSI, then $\{\Phi^t \otimes \Psi^t\}$ also has α-CLSI. CLSI is usually harder to prove than MLSI for two reasons. First, the presence of the untouched B system immediately eliminates the possibility of a unique fixed point density. Second, most dimension-dependent bounds become trivial with an arbitrarily large extension.

MLSI implies norm decay. In contrast, some channels lack MLSI despite showing norm decay. For example, we may define a state replacement channel with pure fixed point as

$$\Phi^t_{rep(t_0,|\psi\rangle)}(\rho) = e^{-t/t_0}\rho + (1 - e^{-t/t_0})|\psi\rangle\langle\psi| \tag{7.12}$$

for some fixed $|\psi\rangle$. It is easy to see that $\|\Phi^t_{rep(t_0,|\psi\rangle)}(\rho) - \Phi^\infty_{rep(t_0),|\psi\rangle}(\rho)\|_1 \leq \exp(-t/t_0)\|\rho - \Phi^\infty_{rep(t_0),|\psi\rangle}(\rho)\|_1$. For the relative entropy, however,

$$D(\Phi^t_{rep(t_0,|\psi\rangle)}(\rho)\|\Phi^\infty_{rep(t_0,|\psi\rangle)}(\rho)) = \begin{cases} \infty & \rho \neq |\psi\rangle\langle\psi| \\ 0 & \rho = |\psi\rangle\langle\psi| \end{cases} \tag{7.13}$$

for all $t < \infty$. Self-adjoint channels are always unital, and as we saw in Section 3.4.1, relative entropy with respect to a conditional expectation has especially good properties. Hence for these examples, relative entropy is a good characterization of decay. In Section 7.2.4, I summarize joint work with Marius Junge and Cambyse Rouzé [32] on relative entropy decay for non-self-adjoint channels, for which the fixed point channel is generally not a conditional expectation. We obtain good results for semigroups that are nearly unital, but MLSI may not be a good way to characterize decay to pure states.

Bounds arising from Theorems 7.2 and 7.3 do not imply MLSI, since they contain additive constants. One may nonetheless apply such bounds to semigroups. When $D(\rho\|\Phi^\infty(\rho))$ is relatively large compared to the additive corrections, the additive bounds can be stronger. In contrast, for large enough t or $\rho \approx \Phi^\infty(\rho)$, there will be a point at which any fixed additive correction makes the bound trivial. MLSI does not become trivial unless ρ is already a fixed point.

The rest of this section and most of the chapter are about methods to prove MLSI or CLSI for a given Lindbladian. Directly expanding the Taylor series of an operator or super-operator exponential often leads to intractable calculations, so it can be hard to infer properties of the semigroup by inspecting the form of its generator. I have worked with many collaborators, including Ivan Bardet, Li Gao, Marius Junge, Haojian Li, Cambyse Rouzé, and Daniel Stilck França, on several approaches to estimate MLSI and CLSI from a known Lindblad generator.

7.2.1 MLSI Merging

In [26] and as described in section 3.1, I recall adjusted subadditivity (ASA, Definition 3.2 and Theorem 3.2) of relative entropy of non-commuting conditional expectations. The main application of adjusted subadditivity is as follows:

Theorem 7.4 (MLSI Merge, theorem 1.5 from [26]). *Let $\{\Phi^t_j : j \in 1...J \in \mathbb{N}\}$ be self-adjoint quantum Markov semigroups such that $\Phi^t_j = ad_{\exp(-\mathcal{L}_j t)}$ with fixed point conditional expectation $\Phi^\infty_j = \mathcal{E}_j$. Let Φ^t be the semigroup generated by $\mathcal{L} = \sum_j \mathcal{L}_j + \mathcal{L}_0$, where $\mathcal{L}_0$ generates Φ^t_0 with fixed point algebra containing the fixed point algebra of $\{\Phi_j\}$. Let $\{\mathcal{E}_j\}$ be α-subadditive, and Φ^t_j have λ_j-MLSI for each j. Then Φ^t has*

$\alpha \times \min_j\{\lambda_j\}$*-MLSI.*

There are similarities between Theorem 7.4 and the results and techniques of [70, 71, 72, 73].

Proof. We start by writing $\Phi^t(\rho) = \Phi^\tau(\rho')$, where $\tau < t$, and $\rho' = \Phi^{t-\tau}(\rho)$. Let

$$\mathrm{wsb}(\epsilon, \mathcal{N} \subseteq \mathcal{M}) = 2\epsilon D(\mathcal{M}\|\mathcal{N}) + (1+2\epsilon) h\Big(\frac{\epsilon}{1+2\epsilon}\Big) \ , \tag{7.14}$$

as in proposition 3.7 of [25], where h is the binary entropy, $\mathcal{N} \subset \mathcal{M}$ is a subalgebra, and $D(\mathcal{M}\|\mathcal{N}) = \sup_\rho D(\mathcal{E}_\mathcal{M}(\rho)\|\mathcal{E}_\mathcal{N}(\rho))$. Rather than depend on dimension as would the Fannes-Audenart bound, this bounds entropy in terms of trace distance and subalgebra index. Let $\mathcal{N}$ be the fixed point algebra of Φ^t with conditional expectation $\mathcal{E}$, and $\mathcal{N}_j$ the fixed point algebra of Φ^t_j. Let $\mathcal{M}$ be the original algebra containing ρ. By the Suzuki-Trotter expansion, the Taylor expansion of an exponential, or by direct commutation, $\Phi^\tau(\rho) = \Phi^\tau_1 ... \Phi^\tau_J \Phi^\tau_0(\rho) + O((J+1)\tau^2)$. Let $\hat{\rho} = \Phi^\tau_0(\rho')$. Hence

$$D(\Phi^\tau(\hat{\rho})\|\mathcal{E}(\hat{\rho})) \leq D(\Phi^\tau_1 ... \Phi^\tau_J(\hat{\rho})\|\mathcal{E}(\hat{\rho})) + \mathrm{wsb}(O(J\tau^2), \mathcal{N} \subseteq \mathcal{M}). \tag{7.15}$$

Define γ_j by

$$D(\Phi^\tau_{j+1} ... \Phi^\tau_J(\hat{\rho})\|\mathcal{E}_j(\Phi^\tau_{j+1} ... \Phi^\tau_J(\hat{\rho}))) = \gamma_j D(\Phi^\tau(\hat{\rho})\|\mathcal{E}(\hat{\rho})) \ . \tag{7.16}$$

Then

$$\begin{aligned}
& D(\Phi^\tau_j ... \Phi^\tau_J(\hat{\rho})\|\mathcal{E}(\hat{\rho})) \\
& = D(\Phi^\tau_j ... \Phi^\tau_J(\hat{\rho})\|\mathcal{E}_j(\Phi^\tau_j ... \Phi^\tau_J(\hat{\rho}))) + D(\mathcal{E}_j(\Phi^\tau_{j+1} ... \Phi^\tau_J(\hat{\rho}))\|\mathcal{E}(\hat{\rho})) \\
& \leq (1 - \lambda_j \tau + O(\lambda_j^2 \tau^2)) D(\Phi^\tau_{j+1} ... \Phi^\tau_J(\hat{\rho})\|\mathcal{E}_j(\Phi^\tau_{j+1} ... \Phi^\tau_J(\hat{\rho}))) + D(\mathcal{E}_j(\Phi^\tau_{j+1} ... \Phi^\tau_J(\hat{\rho}))\|\mathcal{E}(\hat{\rho})) \\
& \leq (1 - \lambda_j \gamma_j \tau + O(\lambda_j^2 \tau^2)) (D(\Phi^\tau_{j+1} ... \Phi^\tau_J(\hat{\rho})\|\mathcal{E}_j(\Phi^\tau_{j+1} ... \Phi^\tau_J(\hat{\rho}))) + D(\mathcal{E}_j(\Phi^\tau_{j+1} ... \Phi^\tau_J(\hat{\rho}))\|\mathcal{E}(\hat{\rho}))) \\
& = (1 - \lambda_j \gamma_j \tau + O(\lambda_j^2 \tau^2)) D(\Phi^\tau_{j+1} ... \Phi^\tau_J(\hat{\rho})\|\mathcal{E}(\hat{\rho})) \ ,
\end{aligned}$$

where the first equality follows from Lemma 3.5 and the first inequality from CLSI of Φ^τ_i. The second inequality follows from Equation (7.16), where we replace the $-\lambda_j \tau D(\Phi^\tau_{j+1} ... \Phi^\tau_J(\hat{\rho})\|\mathcal{E}_j(\Phi^\tau_{j+1} ... \Phi^\tau_J(\hat{\rho})))$ subtracted term by $-\lambda_j \gamma_j \tau D(\Phi^\tau(\hat{\rho})\|\mathcal{E}(\hat{\rho}))$ and then note that this is at least as large as the entire original expression. Iterating, we obtain

$$D(\Phi^\tau_i ... \Phi^\tau_J(\hat{\rho})\|\mathcal{E}(\hat{\rho})) \leq \Big(1 - \sum_{j=1}^J \lambda_j \gamma_j \tau + O\Big(\tau^2 \sum_j \lambda_j^2\Big) D(\mathcal{M}\|\mathcal{N}_j)\Big) D(\hat{\rho}\|\mathcal{E}(\rho)) \ . \tag{7.17}$$

Now we must show that the subtracted sum $\sum_j \lambda_j \gamma_j$ is bounded below. Here we invoke α-subadditivity. First, note that for any j,

$$D(\Phi^\tau_{j+1}...\Phi^\tau_J(\hat\rho) \| \mathcal{E}_j(\Phi^\tau_{j+1}...\Phi^\tau_J(\hat\rho))) \geq D(\hat\rho \| \mathcal{E}_j(\hat\rho)) + \text{wsb}(O((J-j)\tau^2), \mathcal{N}_j \subseteq \mathcal{M}) \ .$$

Hence

$$\sum_j D^{\mathcal{N}_j}(\Phi^\tau_{j+1}...\Phi^\tau_J(\hat\rho)) = \sum_j \gamma_j D(\hat\rho \| \mathcal{E}(\hat\rho)) \geq \alpha D(\hat\rho \| \mathcal{E}(\hat\rho)) + \sum_{j=1}^J \text{wsb}(O((J-j)\tau^2), \mathcal{N}_j \subseteq \mathcal{M}) \ .$$

This implies that $\sum_j \gamma_j \geq \alpha$. $D(\hat\rho \| \mathcal{E}(\rho)) \geq D(\Phi^t(\rho) \| \mathcal{E}(\rho))$ by data processing, so either the latter is zero, or the former is at least some ϵ that is independent of τ. If $D(\Phi^t(\rho)\|\mathcal{E}(\rho)) = 0$, then the Theorem is trivially complete. Otherwise, returning to Equations (7.17) and (7.15),

$$D(\Phi^\tau(\hat\rho) \| \mathcal{E}(\hat\rho)) \leq \Big(1 - \alpha\tau \min_j \{\lambda_j\} + \text{err}\Big) D(\hat\rho \| \mathcal{E}(\hat\rho)). \tag{7.18}$$

if we let

$$\begin{aligned} \text{err} = & \frac{\text{wsb}(O(J\tau^2), \mathcal{N} \subseteq \mathcal{M})}{\epsilon} \\ & + \sum_{j=1}^J \Big(\frac{\text{wsb}(O((J-j)\tau^2), \mathcal{N}_j \subseteq \mathcal{M})}{\epsilon} + O(\lambda_j^2 \tau^2) D(\mathcal{M} \| \mathcal{N}_j) \Big) \ . \end{aligned} \tag{7.19}$$

Furthermore,

$$D(\Phi^\tau_0(\rho') \| \mathcal{E}(\rho')) \leq D(\rho' \| \mathcal{E}(\rho')), \tag{7.20}$$

as the left hand side follows from data processing on the right hand side. Since $\Phi^t = (\Phi^\tau)^{t/\tau}$,

$$D(\Phi^t(\rho) \| \mathcal{E}(\rho)) \leq (1 - \alpha \min_j \{\lambda_j\}\tau + \text{err})^{t/\tau} D(\rho \| \mathcal{E}(\rho)) \ . \tag{7.21}$$

Since $\text{err} \leq O(J \log J \tau^2 \log \tau (D(\mathcal{M}\|\mathcal{N}) + \sum_j \lambda_j^2 \log \lambda_j D(\mathcal{M} \| \mathcal{N}_j)))$,

$$\lim_{\tau \to 0} (1 - \alpha \min_j \{\lambda_j\}\tau + \text{err})^{t/\tau} = e^{-\alpha \min_j \{\lambda_j\} t} \ .$$

This completes the Theorem. ■

Corollary 7.1 (CLSI Merge, corollary 4.1 from [26])**.** *Let $\{\Phi^t_j\}$ and Φ^t be as in Theorem 7.4, with $\{\mathcal{E}_j\}$*

being completely α-subadditive, and each Φ_j^t having λ_j-CLSI. Then Φ^t has $\alpha \times \min_j\{\lambda_j\}$-CLSI.

Proof. Corollary 7.1 motivates further research to prove Conjecture 3.1 Apply Theorem 7.4 for any extension $\{\hat{1}^C \otimes \Phi_j^t\}$ and $\{\hat{1}^C \otimes \mathcal{E}_j\}$. ∎

Merging MLSI or CLSI allows us to break down a complicated Lindbladian into simpler constituents. As examples, copied from the original paper:

Example 7.1 (Mutually Unbiased Basis Decay, example 4.4 in [26]). *Let X, Z be the Pauli matrices, and $\langle X \rangle, \langle Z \rangle$ their corresponding generated subalgebras. It is well-known that the subgroup $\Phi_{\langle X \rangle}^t$ generated by $\mathcal{L}_{\langle X \rangle}(\rho) = \rho - X\rho X$ has 1-CLSI, and so does $\Phi_{\langle Z \rangle}^t$. Here we have*

$$\Phi_{\langle X \rangle}^t(\rho) = e^{-t\mathcal{L}_{\langle X \rangle}}(\rho) = e^{-t}\rho + (1 - e^{-t})X\rho X \ .$$

Hence

$$\begin{aligned}
\Phi_{\langle X \rangle}^t(\mathcal{E}_{\langle Z \rangle}(\rho)) &= \frac{1}{2}\big(e^{-t}\rho + (1-e^{-t})X\rho X + e^{-t}Z\rho Z + (1-e^{-t})XZ\rho ZX\big) \\
&= \frac{1}{2}\big(e^{-t}\rho + (1-e^{-t})X\rho X + e^{-t}Z\rho Z + (1-e^{-t})ZX\rho XZ\big) \\
&= \mathcal{E}_{\langle Z \rangle}(\Phi_{\langle X \rangle}^t(\rho)) \ ,
\end{aligned}$$

and similarly, $[\Phi_{\langle Z \rangle}^t, \mathcal{E}_{\langle X \rangle}] = 0$. As these form a commuting square, Corollary 7.1 implies that the semigroup generated by $\mathcal{L} = \mathcal{L}_{\langle X \rangle} + \mathcal{L}_{\langle Z \rangle}$ has 1-CLSI.

This example generalizes to any pair of mutually unbiased bases $\mathcal{X}, \mathcal{Z}$ in finite dimension.

Remark 7.1. *Until recently, proving CLSI for the Lindbladian with two Pauli generators given by $\mathcal{L}(\rho) = 2\rho - X\rho X + Z\rho Z$ remained open. Because the conditional expectations $\mathcal{E}_{\langle X \rangle}$ and $\mathcal{E}_{\langle Z \rangle}$ commute, Petz's algebraic SSA shows that these conditional expectations are 1-subadditive, satisfying the assumptions of Corollary 7.1. Hence CLSI merging resolved an apparently simple open question that had eluded far more sophisticated techniques. It turns out that 1-CLSI for this Lindbladian is not hard to prove, and Junge showed around the same time that one can do so from the properties of conditional expectations and the Fisher information as discussed in Section 7.2.3. More broadly, any time we can write and Lindbladian as a sum of simpler Lindbladians for which the fixed point conditional expectations commute, we may apply Corollary 7.1 with constant factor 1.*

Example 7.2 (Tilted Incompatible Basis Decay, example 4.5 in [26]). *Let $\mathcal{X}$ and $\mathcal{Z}$ now denote a pair of bases in dimension d, not necessarily mutually unbiased. Let $\{|i_X\rangle : i = 1...d\}$ and $\{|i_Z\rangle : i = 1...d\}$*

index these bases, and $\inf_{i,j} |\langle i_X | j_Z \rangle|^2 = \epsilon$ *for some* $0 \leq \epsilon \leq 1/d$*. Because a pinching to either basis is completely dephasing in that basis, a subsequent pinching in the other leaves any input state in a classical mixture with minimal probability mass no less than* ϵ*. We may write each collapse as a convex combination of unitaries of the form* $|i\rangle \rightarrow \exp(2\pi i k/d)\,|i\rangle$ *for each* $k \in 1...d$*. Hence it depolarizes with probability* ϵ*, so* $\mathcal{E}_{\mathcal{X}} \circ \mathcal{E}_{\mathcal{Z}}(\rho) = \epsilon \hat{1}/d + (1-\epsilon)\Phi(\rho)$ *for some unital* Φ*. By Theorem 3.2,*

$$D(\rho \| \mathcal{E}_{\mathcal{X}}(\rho)) + D(\rho \| \mathcal{E}_{\mathcal{Z}}(\rho)) \geq \frac{D(\rho \| \hat{1}/d)}{4(\log_{1-\epsilon}(2/(3d+5)) + 1)} \; .$$

Expanding the relative entropies, we find this is equivalent to

$$H(\mathcal{X})_\rho + H(\mathcal{Z})_\rho \geq \Big(2 - \frac{1}{4(\lceil \log_{1-\epsilon}(2/(3d+5)) \rceil)}\Big) H(\rho) + \frac{\log d}{4(\lceil \log_{1-\epsilon}(2/(3d+5)) \rceil)} \; .$$

An uncertainty relation in the conventional form would be

$$H(\mathcal{X})_\rho + H(\mathcal{Z})_\rho \geq H(\rho) + \log d - c$$

for some constant $c > 0$*, as required for non-commuting conditional expectations as shown in [23]. When* $\rho \approx \hat{1}/d$*, the right hand side of the conventional uncertainty relation becomes negative, and the bound becomes trivial. In contrast, adjusted subadditivity still gives a positive, non-trivial bound on the sum of basis entropies.*

Furthermore, let $\Phi^t_{\mathcal{X}}$ *and* $\Phi^t_{\mathcal{Z}}$ *be quantum Markov semigroups with* β*-MLSI for some* $\beta > 0$ *and respective Lindbladian generators* $\mathcal{L}_{\mathcal{X}}, \mathcal{L}_{\mathcal{Z}}$*. Then* $\{\Phi^t = e^{-(\mathcal{L}_{\mathcal{X}} + \mathcal{L}_{\mathcal{Z}})t}\}$ *has* $1/(4(\lceil \log_{1-\epsilon}(2/(3d+5)) \rceil))$*-MLSI.*

7.2.2 MLSI Merging for Finite, Symmetric, Ergodic Graphs

As a broader class of examples, MLSI merging proves MLSI for decay models given by finite, ergodic graphs with fixed point algebra $\mathbb{C}$. Let $G = (V, E)$ be a finite graph with vertices V and edges $E \subseteq V \times V$. To each e one may assign an element of a group, represented on a Hilbert space $\mathcal{H}$ by unitaries $\{u_e\}$, Let each $\langle u_e \rangle$ generates a unitary subgroup and that the Lindbladian given on an input density ρ. Then consider

$$\mathcal{L}(\rho) = \sum_{e \in E} u_e \rho u_e - |E| \; . \tag{7.22}$$

First, we use MLSI merging to show MLSI for many Lindbladians of this form:

Corollary 7.2 (Graph Corollary to Theorem 7.4)**.** *Let* $\mathcal{L}$ *be defined as in Equation* (7.22) *for a graph*

$G = (V, E)$. Assume for each $e \in E$ that $\mathcal{L}_e(\rho) = u_e \rho u_e - 1$ has MLSI with constant α_e, and that $\mathcal{L}$ has fixed point algebra $\mathbb{C}$. Then $\mathcal{L}$ has MLSI.

To explicitly compute the MLSI constant, we must consider the fixed point algebras of each $\mathcal{L}_e$ and from them derive the subadditivity adjustment constant in Theorem 3.2. It is also possible to derive CLSI for decay given by finite graphs, which I describe in Section 7.2.5.

7.2.3 Geometry & Transference from Classical Markov Semigroups

The primary method developed in [31] and applied in [33] is a group transference technique. Given a Lindbladian $\mathcal{L}$ constructed from elements of a group G, we often may also construct an analogous classical Markov semigroup generator L on $L_\infty(G)$, the space of bounded, measurable functions from G to $\mathbb{C}$. Log-Sobolev inequalities and decay to equilibrium in classical systems is a well-studied topic with a long history of results. A relatively recent paper by Carlen and Maas [166] began to link quantum decay processes with classical return-to-equilibrium, expressing quantum Lindbladians as a heat semigroup and bringing a variety of geometric methods to the problem. In [31], we build on the work of Carlen and Maas by deriving the transference principle and making heavy use of information geometry.

A core concept in information geometry is the Fisher information, given for a semigroup generator $\mathcal{L}$ by

$$I_{\mathcal{L}}(\rho) = \mathrm{tr}(L(\rho) \log \rho) \ . \tag{7.23}$$

A self-adjoint generator has α-MLSI iff

$$\alpha D(\rho \| \mathcal{E}_{fix}(\rho)) \leq I_{\mathcal{L}}(\rho) \tag{7.24}$$

for all densities ρ. Given a function $f \in L_\infty(G)$, we may calculate not only its Fisher information, but the Fisher information for a matrix-valued version in $L_\infty(G) \times \mathbb{M}_d$ for $d \in \mathbb{N}$. In [31], we use a variety of geometric and analytic methods to show that for any self-adjoint generator $\mathcal{L}$ and $\beta \in (0, 1)$, $\mathcal{L}^\beta$ has α_β-CLSI for some $\alpha_\beta > 0$. This does not ensure that $\mathcal{L}$ has CLSI, as α_β may approach 0 as $\beta \to 1$. It does however show that there is a dense set of generators in $\mathbb{M}_d$ with CLSI for some constant.

The focus of [31] is largely mathematical, whereas [33] is mostly about applications. Let $k_t(g, h)$ be a kernel representation such that the stochastic Markov semigroup S_t on functions $f \in L_\infty(G)$ has the form

$$S_t(f)(g) = \int_G k_t(g, h) f(h) d\mu_G(h) \ , \tag{7.25}$$

where μ_G is the Haar measure on G. Then we construct the transferred QMS on $a \in \mathbb{M}_d$

$$T_t(a) = \int_G k_t(g^{-1}, \hat{1}) u_g^\dagger a u_g d\mu_G(g) \ . \tag{7.26}$$

We construct the co-representation

$$\pi : \mathbb{B}(\mathcal{H}) \to L_\infty(G, \mathbb{B}(\mathcal{H})), \pi(a)(g) = u_g^\dagger a u_g \tag{7.27}$$

and have that

$$\pi \circ T_t = (S_t \otimes \hat{1}_{\mathbb{B}(\mathcal{H})}) \circ \pi \ . \tag{7.28}$$

In diagrammatic form,

$$\begin{array}{ccc} \mathbb{B}(\mathcal{H}) & \to_{\mathcal{E}_{fix}} & \mathcal{N}_{fix} \\ \downarrow_\pi & & \downarrow_\pi \\ L_\infty(G, \mathbb{B}(\mathcal{H})) & \to_{\mathcal{E}_{\mu_G}} & \mathbb{B}(\mathcal{H}) \end{array}$$

in analogy to the commuting squares and diagrams of Chapter 3. The gist of the paper is then that estimates of norm bounds and CLSI constants for S_t should also hold for T_t. Importantly, the constants involved depend on the structure of the group rather than on the dimension of the Hilbert space on which it acts. This allows us to estimate decay times for *collective channels*, which apply highly correlated errors across multiple registers.

The first collective channel we consider is called "weak" collective decoherence, as it involves qubit dephasing by one Pauli generator. Without loss of generality, let this be the Pauli Z matrix, so that

$$\mathcal{L}_n^{wcd}(\rho) = a - Z^{\otimes n} \rho Z^{\otimes n} \tag{7.29}$$

acting on densities on the n-qubit Hilbert space $l_2^{2^n}(\mathbb{C})$. It was previously known that $t_\epsilon^{dec}(\{\exp(\mathcal{L}_n^{wcd})\}) \sim O(n)$ for this semigroup (see [167]). In [33], we show that

$$\| \exp(-t\mathcal{L}_n^{wcd}(\rho) - \mathcal{E}_{fix}(\rho) \|_1 \leq \sqrt{2 + \sqrt{\pi/t}} e^{-t/2} \ , \tag{7.30}$$

so that

$$t_\epsilon^{dec}(\{\exp(\mathcal{L}_n^{wcd})\}) \leq \max \left\{ 1, \log \left(\frac{\epsilon^2}{2 + \sqrt{\pi}} \right) \right\} \ . \tag{7.31}$$

We also consider the "strong" collective decoherence channel that involves all Pauli matrices,

$$\mathcal{L}_n^{scd}(\rho) = a - \sum_{W=X,Y,Z} W^{\otimes n} \rho W^{\otimes n} \tag{7.32}$$

with decoherence time

$$t_\epsilon^{dec}(\{\exp(\mathcal{L}_n^{scd}(\rho))\}) \leq \frac{34}{3} - 8 \log \epsilon + 2 \log \Big(1 + \frac{3}{2} \log \frac{3}{4} \Big) \, . \tag{7.33}$$

We further estimate the decoherence time of the random swap channel,

$$\mathcal{L}^{swap}(\rho) = \frac{1}{n} \sum_{i,j=1}^{n} (\rho - u_{ij}^{sw} \rho u_{ij}^{sw}) \, , \tag{7.34}$$

where u_{ij}^{sw} is the unitary matrix that swaps the ith with the jth qubit system. We find

$$t_\epsilon^{dec}(\{\exp(\mathcal{L}^{swap})\}) \leq \log n + \frac{1 - \log \epsilon}{2} \, . \tag{7.35}$$

In [33], we also find estimates for the quantum capacities of this and other channels.

7.2.4 Non-Self-Adjoint Semigroups & Measure Change

So far, all of the techniques discussed apply to self-adjoint semigroups, for which the fixed point states form a subalgebra. As mentioned near the beginning of this section, it is not possible to extend MLSI-type estimates to semigroups with pure fixed points, for which the relative entropy is neither finite nor continuous. Considering semigroups with unique fixed point densities, we might expect that when the fixed point is nearly complete mixture, we could perturbatively compare to a self-adjoint semigroup. We show this in [32], allowing us to estimate CLSI constants for non-self-adjoint semigroups with faithful fixed point densities. As expected, these estimates approach the self-adjoint estimates as the fixed point density approaches complete mixture, but yield extremely slow decay as the fixed point state approaches purity.

The gist of this argument comes from a Holley-Stroock change of measure argument. Given two continuous measures μ, ν on $\mathbb{R}^n$,

$$D(\nu \| \mu) = \mathrm{Ent}_\mu(f) \equiv \int f \log f d\mu - \Big(\int f d\mu \Big) \log \Big(\int f d\mu \Big) \, , \tag{7.36}$$

where f is the Radon-Nikodym derivative $d\nu / d\mu$, or the pointwise ratio of measures. As shown by Holley and Stroock [168],

$$\mathrm{Ent}_\mu(f) = \inf_{c>0} (\int f \log f - c \log c + c - f) d\mu \ . \tag{7.37}$$

First, we note the similarities between this form and that of the relative entropy of asymmetry $D^{\mathcal{N}}$ as studied throughout this thesis. Second, it is easy to see from this form that

$$\mathrm{Ent}_\nu(f) \leq \left\| \frac{d\nu}{d\mu} \right\|_\infty \mathrm{Ent}_\mu(f) \ . \tag{7.38}$$

Furthermore, we may use the form of the classical Fisher information,

$$I_L^\mu(f) = \int L(f) \log f d\mu = \int \frac{|\nabla f|^2}{f} d\mu \tag{7.39}$$

to see that

$$I_L^\nu(f) \leq \left\| \frac{d\nu}{d\mu} \right\|_\infty I_L^\mu(f) \tag{7.40}$$

We recall the Fisher version of α-MLSI, equivalent to

$$\alpha \mathrm{Ent}_\mu(f) \leq I_L^\mu(f) \ . \tag{7.41}$$

So far, all of these results are for classical functions f rather than quantum densities. In the paper, we find however that we can work around this. A common trick is to use the chain rule of relative entropy, so that

$$D(\rho \| \omega) = D(\rho \| \mathcal{E}_{\omega'}(\rho)) + D(\mathcal{E}_{\omega'}(\rho) \| \omega) \ , \tag{7.42}$$

where $\mathcal{E}_{\omega'}(\rho)$ projects to a diagonal subalgebra generated by ω. Since $\hat{1}/d$ is always simultaneously diagonal with all densities,

$$\frac{D(\rho \| \omega)}{D(\rho \| \hat{1}/d)} \leq \frac{D(\mathcal{E}_{\omega'}(\rho) \| \omega)}{D(\mathcal{E}_{\omega'}(\rho) \| \hat{1}/d)} \tag{7.43}$$

for all densities ω if $D(\rho \| \omega) \geq D(\rho \| \hat{1}/d)$. Hence we may use the classical formalism.

In [32], we study Lindbladians of the form

$$\mathcal{L}^\dagger(a) = \sum_j \left(e^{-\omega_j/2} A_j^* [a, A_j] + e^{\omega_j/2} [A_j, a] A_j^* \right) \tag{7.44}$$

as given by Carlen and Maas [166]. It holds that

$$\mathcal{L}_0^\dagger(a) = \sum_j \left(A_j^*[a, A_j] + [A_j, a]A_j^* \right) \tag{7.45}$$

is a self-adjoint Lindbladian, to which we compare. Ultimately, we find

Theorem 7.5 (theorem 1.2 from [32]). *Let $\mathcal{L}$ be a quantum Markov semigroup generator in the form of Equation* (7.44) *with a unique, faithful fixed point density $\eta = \sum_{k=1}^d \eta_k |k\rangle\langle k|$. Then*

$$\alpha(\mathcal{L}_0) \leq \frac{\max_k \eta_k}{\min_k \eta_k} \max_j e^{\omega_j/2} \alpha(\mathcal{L}) \ ,$$

where α denotes the CLSI constant.

As a particular case, let

$$\omega_\beta \equiv \frac{1}{Z_\beta} \sum_{k=1}^m e^{-\beta E_k} |k\rangle\langle k| = \frac{1}{Z_\beta} \sum_j e^{-\beta E_j} \sum_{k : E_k = E_j} |k\rangle\langle k| \ , \tag{7.46}$$

which has the form of a thermal state at temperature proportional to $1/\beta$. Then:

Corollary 7.3 (Corollary 7.1 in [32]). *Let ω_β be a thermal state as in Equation* (7.46) *and the fixed point of QMS $\{\Phi^t\}$ generated by Lindbladian $\mathcal{L}$ with m energy levels taking the form in equation* (7.46). *Let $\mathcal{L}_0$ be the corresponding self-adjoint Lindbladian. Then*

$$\alpha_{\text{CLSI}}(\mathcal{L}_0) \leq e^{\beta E_m} \max_j e^{\beta(E_{j+1} - E_j)/2} \, \alpha_{\text{CLSI}}(\mathcal{L}) \ . \tag{7.47}$$

We might think of decay to complete mixture as the infinite-temperature limit of decay to thermal mixture. Corollary 7.3 allows us to generalize to finite-temperature in finite dimension. This gives reasonably strong estimates at high temperature, but as the temperature falls, the fixed point approaches purity and estimate becomes loose. Again, this appears not to be a weakness in the change of measure, but reflects that relative entropy approaches discontinuity as the thermal state approaches a pure ground state.

The form of relative entropy used here is somewhat particular. In general, it is difficult to derive purely multiplicative bounds relating the relative entropies for different densities, even when these densities have the same support. The primary technical challenge of Section 3.1 was to find an inequality of this form, from which we obtain powerful entropy inequalities and almost immediate application to decay estimates. For the CLSI results of [32], however, Equation (7.38) suffices.

7.2.5 CLSI for Finite, Symmetric, Ergodic Graphs

In this subsection, I revisit the question of Section 7.2.2. Using the methods described in Sections 7.2.4 and 7.2.3, as well as new techniques developed by Marius Junge and Haojian Li, we show that any graph with n vertices has $O(1/n^2)$-CLSI. This is a much stronger result than that given by MSLI merging alone, as the CLSI result is tensor-stable and dimension-independent. MLSI merging may however prove a better MLSI constant for highly interconnected graphs, as the $1/n^2$ dependence comes from comparing to the CLSI constant of a cycle or a broken cycle, which have essentially the minimum number of transitions for the graph to be fully connected. At the other extreme, an example in [31] is the fully-connected graph, which has a CLSI constant that is $O(1)$ in the number of vertices. Still, the universal, polynomial decay constant shown in this section is a significant step toward showing CLSI for an extremely broad range of systems. These results will appear in an upcoming manuscript with Marius Junge and Haojian Li.

Our starting point is a result appearing in the aforementioned upcoming manuscript:

Theorem 7.6 (Cycle CLSI)**.** *The cyclic graph with n vertices and minimal weight w has cw/n^2-CLSI for some constant c.*

The key to showing that finite graphs have CLSI is analogous to the covering isomorphisms studied in geometry. We construct homomorphisms between graphs. These homomorphisms preserve Fisher information and relative entropy up to multiplicative factors. Some of these factors arise from multiple vertices in the input graph mapping to a single vertex in the target, which causes polynomial overcounting in the degree of the target graph. Others arise from change of measure arguments as discussed in Section 7.2.4.

Definition 7.1. *Let $G = (V, E)$ be a graph by its vertex set V and edges $E \subseteq \{(v_1, v_2) : v_1, v_2 \in V\}$. Let μ be a probability measure on V. A cover is a finite graph and measure (G', μ') with a map $\phi : G' \to G$ such that*

1. *ϕ is surjective.*
2. *$\phi(E') \subseteq E$.*
3. *$\mu = \mu' \circ \phi$.*

Definition 7.2. *For a graph G with cover (G', μ', ϕ), we define the multiplicity m_e of an edge $e \in E$ as the number of edges in E' that map to e, or $m_{E',\phi}(e) \equiv |\{e' \in E' | \phi \times \phi(e') = e\}|$. We similarly define the multiplicity of a vertex $v \in V$ to be $m_{V',\phi}(v) = |\{v' \in V' | \phi(v') = v\}|$.*

The Fisher information of a function $f \in l_\infty(G,\mu)$ decomposes into a sum over edges. First, let

$$I_{p,q}(f) = \operatorname{tr}(f(p) - f(q)) J_{f(p)}(f(p) - f(q)) \ ,$$

where $J_{f(p)}$ is an operator integral given by

$$J_{f(p)}(y) = \int_0^\infty (f(p) + r)^{-1} y (f(p) + r)^{-1} dr \ ,$$

and $p, q \in V$. Then the Fisher information over the graph is given by

$$I_{G,\mu}(f) = \sum_{p \in V} \mu(p) \sum_{q:(p,q) \in E} I_{p,q}(f) \ .$$

Lemma 7.1. *Let $G = (V,E)$ and $G' = (V',E')$ be a pair of symmetric graphs with a surjective map $\phi : V' \to V$ such that $e' \in E' \implies \phi(e') \in E$, where $\phi(e') = (\phi(v_1'), \phi(v_2'))$ for $e' = (v_1', v_2')$. Let $\mu' : V' \to [0,1]$ be a measure on the vertices of V', and $\mu_\phi : V \to [0,1]$ be the measure given by*

$$\mu_\phi(p) = \sum_{p' \in V' : \phi(p') = p} \mu'(p') \ .$$

If (G', E', μ') has λ'-CLSI, then (G, E, μ_ϕ) has $\lambda' / \sup_{(p,q) \in E} m_{E,E',\phi}(p,q)$-CLSI.

Proof. Let $\pi_\phi : l_\infty(G) \to l_\infty(G')$ be given by $\pi_\phi(f) = f \circ \phi$. The Fisher information,

$$\begin{aligned} I^F_{G',\mu'}(\pi_\phi(f)) &= \sum_{p' \in V'} \mu'(p') \sum_{q':(p',q') \in E'} \operatorname{tr}\Big((\pi_\phi(f)(q') - \pi_\phi(f)(p')) J_{f(p')} (\pi_\phi(f)(q') - \pi_\phi(f)(p')) \Big) \\ &= \sum_{p' \in V'} \mu'(p') \sum_{q':(p',q') \in E'} \operatorname{tr}\Big((f(\phi(q')) - f(\phi(p'))) J_{f(p')} (f(\phi(q')) - f(\phi(p'))) \Big) \ . \end{aligned}$$

The sum $\sum_{p' \in V'} \sum_{q':(p',q') \in E'}$ is equivalent to the sum $\sum_{p \in V} \sum_{q:(p,q) \in E} \sum_{p':\phi(p')=p} \sum_{q':\phi(q')=q}$, since each edge $(p', q') \in E'$ maps via ϕ to a single edge $(p,q) \in E$. Hence

$$I_{G',\mu'}(\pi_\phi(f)) = \sum_{p \in V} \sum_{q:(p,q) \in E} \sum_{p':\phi(p')=p} \sum_{q':\phi(q')=q} \mu'(p') I_{(p,q)}(f) \ .$$

Since $I_{(p,q)}(\pi_\phi(f))$ contains no explicit dependence on p' or q', we re-arrange this sum as

$$I(\pi_\phi(f)) = \sum_{p \in V} \sum_{q:(p,q) \in E} I_{(p,q)}(f) \Big(\sum_{p':\phi(p')=p} \mu'(p') \sum_{q':\phi(q')=q} \Big) \ .$$

The final sum over q' is no larger than $\beta(p,q)$. Hence

$$I_{G',\mu'}(\pi_\phi(f)) \leq \sum_{p\in V} \mu_\phi(p) \sum_{q:(p,q)\in E} m_{E',\phi}(p,q) I_{(p,q)}(\pi_\phi(f))$$
$$\leq \max_{p,q\in E} m_{E',\phi}(p,q) I_{G,\mu_\phi}(f) \ .$$

Let

$$D_\mu(\rho\|\sigma) = \mu(\rho\log\rho - \rho\log\sigma) - \rho + \sigma) \ ,$$

a weighted form of Lindblad's relative entropy for unnormalized but diagonal ρ, σ. We then find that

$$D_{\mu'}(\pi(\rho)\|\mathcal{E}_\mathbb{C}(\pi(\rho)) = D_{\mu_\phi}(\rho\|\mathcal{E}_\mathbb{C}(\sigma))$$

by the $*$-homomorphism property of π. ■

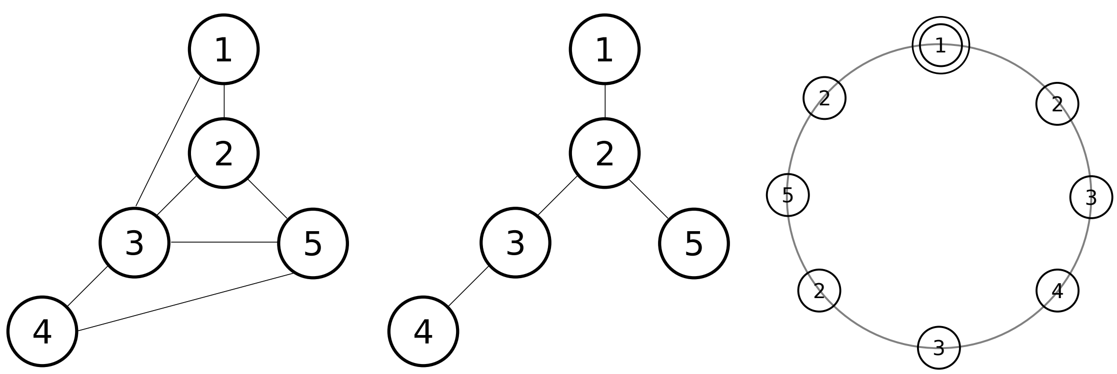

Figure 7.1: Left: a graph with labeled vertices. Center: a spanning tree. Right: a cycle traversing that tree, starting and ending at 1.

Lemma 7.2. *Let $G_T = (V_T, E_T)$ be a finite, undirected n-vertex tree (a graph with no cycles). Let d_v be the degree of vertex $v \in V$, and $d_{tot} = \sum_{v\in V_T} d_v$ be the sum of degrees of all nodes. Let μ_T be a probability measure on V_T given by*

$$\mu_T(v) = \begin{cases} d_v/(d_{tot}+1) & v \text{ is not root} \\ (d_v+1)/(d_{tot}+1) & v \text{ is the root .} \end{cases}$$

Then (G_T, μ_T) has $O(1/n^2)$-CLSI

Proof. Let us assume that each node's children in G_T are ordered by choosing an arbitrary order if necessary. We use the well-known pre-order traversal/labeling algorithm, given by the following recursive procedure and recording each edge traversed:

1. Begin by selecting the root of the tree as the current vertex. Set $j = 1$ and label the current/root vertex as $j = 1$.

2. If the current vertex is not a leaf, select the first child vertex not already visited, label it j, record the the edge $(j_{current}, j_{child})$ in the traversal list, increment $j \to j + 1$, and make the child the current vertex. If the current vertex has no more unvisited children, then return to the parent node, traversing and recording that edge.

3. Return to step 2 and repeat until returning to the parent node.

See Figure 7.1. The procedure traverses each edge twice, producing a list of the form $(j_1, j_2), ...(j_{2n-1}, j_{2n})$, where $j_1 = j_{2n} = 1$ labels the root. We construct a cyclic graph $G' = (V', E')$ with $2n - 1$ vertices. For each $k \in 1...2n-1$, we label the kth vertex in the cycle by the starting vertex in the kth entry of the traversal list. We now have the function $\phi : V' \to V_T$ that for each vertex in V' assigns the unique vertex in V having the same label. This allows us to apply Lemma 7.1. By Lemma 7.6, with measure μ_ϕ, G_T has $O(1/n^2)$-CLSI. We then note that each vertex's multiplicity is equal to its degree, except for the root node, which is visited one extra time. ■

Theorem 7.7. *Every symmetric, connected graph $G = (V, E)$ with n vertices of degree at most d and measure $\mu(v) = 1/n$ has $O(1/(d + 1)^2 n^2)$-CLSI.*

Proof. First, we recall that any connected, undirected graph contains at least one spanning tree. For a given graph G, let $G_T = (V_T, E_T)$ denote a chosen spanning tree. Let $\mu_T : V_T \to [0, 1]$ be the measure induced by projecting μ to G_T and normalizing. Then for any $f \in l_\infty(G)$,

$$I_{G_T,\mu_T}(f) \leq I_{G,\mu}(f)$$

by the form of Fisher information that sums over edges. Hence we study the Fisher information of the spanning tree G_T.

We must compare μ_T with μ. For a function $f \in l_\infty(G_T)$,

$$\begin{aligned} I_{G_T,\mu_T}(f) &= \sum_{p \in V_T} \mu_T(p) \sum_{q:(p,q) \in E_T} \operatorname{tr}\Big((f(q) - f(p))J_{f(p)}(f(q) - f(p)))\Big) \\ &\leq \Big(\max_p \frac{\mu_T(p)}{\mu(p)}\Big) \sum_{p \in V_T} \mu_T(p) \sum_{q:(p,q) \in E_T} \operatorname{tr}\Big((f(q) - f(p))J_{f(p)}(f(q) - f(p)))\Big) \\ &= \Big(\max_p \frac{\mu_T(p)}{\mu(p)}\Big) I^F_{G_T,\mu_T}(f) \ . \end{aligned}$$

We observe that $\max_p \mu_T(p)/\mu(p) \leq d+1$, since the root of the spanning tree is visited a number of times equal to its degree plus one, and all other vertices are visited a number of times equal to their degree.

Finally, we must compare the relative entropy with respect to measure μ_ϕ to that of a uniform measure on the original graph. The relative entropy with respect to fixed point algebras is the same for the spanning tree as for the original graph, as they have the same fixed point algebra and vertex set. As originally found in [169] and recalled in [32],

$$D_\mu(\rho \| \mathcal{E}_{\mathbb{C}}(\sigma)) \leq \left\| \frac{\mu}{\mu_\phi} \right\|_\infty D_{\mu_\phi}(\rho \| \mathcal{E}_{\mathbb{C}}(\sigma))$$

for discrete probability measures μ and μ_ϕ. In the worst case for uniform μ, this adds another factor of the largest multiplicity of a repeatedly visited label, which is $d+1$. ■

Remark 7.2. *We can refine Theorem 7.7 to use the minimal degree of any spanning tree in place of d. The depth-first and breadth-first search algorithms are both sufficient to find a spanning tree, but other algorithms may achieve better minimal degree, running time, etc.*

Remark 7.3. *We could replace the factors of $d+1$ by d if we replaced the comparison to a cycle by that with a broken cycle (one link missing). This may change the constant factor, so it is not clear whether it ultimately improves the estimate.*

7.3 Links to Computation, Complementarity & Asymmetry

In many experiments, a main barrier to scalable quantum technology is the error induced by cross-qubit gates. According to a recent paper by on circuit compilation for the IBM QX, "a program with more than 16 CNOT operations, has less than 50% chance of executing correctly," "it is most important to consider CNOT and readout error rates... Optimization based on qubit coherence time is also useful, but less critical here because gate errors severely limit useful computation time," and "for this machine, single-qubit error rates are considerably smaller so our formulation chooses to ignore them [170]." The IBM QX is a superconducting-based architecture, as is Google's primary test device [16]. Neutral atom quantum computing also shows promise but suffers from low fidelity of two-qubit gates. Trapped ion qubits have high-fidelity, entangling gates within arrays, but "it will likely become necessary to link separate registers of trapped ion chains with photonic interfaces [171]." Photonic quantum computing with traditional gate-based approaches has immense overhead to simulate photon-to-photon interactions from linear optical media [153], and obtaining high-fidelity, low-loss cross-photon gates from nonlinear optical media remains an open challenge. While these examples are far from the comprehensiveness needed to show a universal pattern, they are enough to arouse suspicion.

That most errors occur in interacting quantum systems and in measurements (classical-quantum interactions) shows a potential hole in the Markov chain approach to quantifying decoherence. The change in environment during execution of a gate is a controlled form of non-stationarity, a breaking of time translation symmetry. Nevertheless, we might consider many approaches to switch on an interaction, time-evolve with a new and short-time stationary Markov process, and then switch back to the original state.

Remark 7.4. *Let $\mathcal{L}_1, ..., \mathcal{L}_n$ be a family of n Lindbladians with the same fixed point algebra, $\mathcal{N}$, applied sequentially in time intervals of lengths $t_1, ..., t_n \in \mathbb{R}^+$ to construct the quantum channel*

$$\Phi(\rho) = \exp(-t_n \mathcal{L}_n) ... \exp(-t_1 \mathcal{L}_1)(\rho) \ .$$

Then if $\mathcal{L}_i$ has α_i-CLSI, we may combine decay estimates and find that

$$D(\Phi(\rho) \| \mathcal{E}_{\mathcal{N}}(\rho)) \leq e^{-\alpha_1 t_1 + ... \alpha_n t_n} D(\rho \| \mathcal{E}_{\mathcal{N}}(\rho)) \ . \tag{7.48}$$

Another possible gap in the approaches of Section 7.2 is that we have focused on drift-free Lindbladians, which have no coherent time-evolution (no Hamiltonian part). These are good models of stochastic processes that only induce decoherence, but it can be complicated to combine them with the Hamiltonian time-evolution occurring during gate execution. Indeed, Lindblad generators that generate some unitary rotation are almost automatically non-self-adjoint. The methods of Section 7.2.4 begin to address this problem. Here we include another short result that helps reconcile these problems:

Corollary 7.4. *Let $\mathcal{L}_0$ be a self-adjoint Lindbladian with α-(M/C)LSI and fixed point algebra $\mathcal{N}$. Let $\mathfrak{H}$ be a Hamiltonian such that $[\mathcal{E}_{\mathcal{N}}, \mathfrak{H}] = 0$, and such that $\mathcal{N}$ is also an invariant subalgebra of $\mathfrak{H}$. Then the combined time-evolution has α-(M/C)LSI.*

Proof. This proof closely follows that of MLSI merging in Subsection 7.2.1, via the Suzuki-Trotter expansion for Lindbladians. In particular, let us write

$$\mathcal{L}(\rho) = -i[\mathfrak{H}, \rho] + \mathcal{L}_0(\rho) \ .$$

Let τ be a short time interval, so that

$$\exp(-t\mathcal{L})(\rho) = \left(e^{-\mathcal{L}_0 t} e^{-it\mathfrak{H}} \right)^{t/\tau} + O(\tau^2) \ .$$

We then directly compute that

$$
\begin{aligned}
&D\Big(\Big(e^{-\mathcal{L}_0 t}e^{-it\mathfrak{H}}\Big)^{t/\tau}(\rho)\Big\|\mathcal{E}_{\mathcal{N}}(\rho)\Big) = D\Big(\Big(e^{-\mathcal{L}_0 t}e^{-it\mathfrak{H}}\Big)^{(t-\tau)/\tau}e^{-\tau\mathcal{L}_0}e^{-i\tau\mathfrak{H}}(\rho)\Big\|\mathcal{E}_{\mathcal{N}}(\rho)\Big) \\
&= D\Big(\Big(e^{-\mathcal{L}_0 t}e^{-it\mathfrak{H}}\Big)^{(t-\tau)/\tau}e^{-\tau\mathcal{L}_0}e^{-i\tau\mathfrak{H}}(\rho)\Big\|e^{-\tau\mathcal{L}_0}e^{-i\tau\mathfrak{H}}\mathcal{E}_{\mathcal{N}}(\rho)\Big) \\
&= D\Big(\Big(e^{-\mathcal{L}_0 t}e^{-it\mathfrak{H}}\Big)^{(t-\tau)/\tau}e^{-\tau\mathcal{L}_0}e^{-i\tau\mathfrak{H}}(\rho)\Big\|e^{-\tau\mathcal{L}_0}\mathcal{E}_{\mathcal{N}}(e^{-i\tau\mathfrak{H}}(\rho))\Big) \\
&= D\Big(\Big(e^{-\mathcal{L}_0 t}e^{-it\mathfrak{H}}\Big)^{(t-\tau)/\tau}(\hat{\rho})\Big\|\mathcal{E}_{\mathcal{N}}(\hat{\rho})\Big)\ ,
\end{aligned}
$$

where $\hat{\rho} = e^{-\tau\mathcal{L}_0}e^{-i\tau\mathfrak{H}}(\rho)$. Then we see that

$$
\begin{aligned}
&D\Big(e^{-\tau\mathcal{L}_0}e^{-i\tau\mathfrak{H}}(\rho)\Big\|\mathcal{E}_{\mathcal{N}}(\rho)\Big) = D\Big(e^{-\tau\mathcal{L}_0}e^{-i\tau\mathfrak{H}}(\rho)\Big\|\mathcal{E}_{\mathcal{N}}(e^{-i\tau\mathfrak{H}}(\rho))\Big) \\
&\leq e^{-\alpha\tau}D\Big(e^{-i\tau\mathfrak{H}}(\rho)\Big\|\mathcal{E}_{\mathcal{N}}(e^{-i\tau\mathfrak{H}}(\rho))\Big) \\
&= e^{-\alpha\tau}D(\rho\|\mathcal{E}_{\mathcal{N}}(\rho))\ ,
\end{aligned}
$$

where the last equality follows from data processing under the invertible unitary $\exp(i\tau\mathfrak{H})$. Hence we may iterate t/τ times and take the limit $\tau \to \infty$ to conclude that

$$
D\Big(e^{-it\mathcal{L}}(\rho)\|\mathcal{E}_{\mathcal{N}}(\rho)\Big) \leq e^{-\alpha t}D(\rho\|\mathcal{E}_{\mathcal{N}}(\rho))\ .
$$

This is the proof for MLSI. For CLSI, we use the same proof for all extensions B and $\mathcal{L} \otimes \hat{1}^B$. ■

When we can separate the coherent time-evolution from stochastic decay to equilibrium, often the aim is for the coherent evolution timescale to be much faster than the the stochastic decay timescale, such that meaningful computation or error correction happens before decoherence reduces the system to a computationally useless state or passes the error correction threshold.

The prominence of interacting qubits as an error source hints at a potential advantage of the CLSI approach over the more usual MLSI. When two systems have CLSI, we may often use MLSI or CLSI merging to obtain immediate estimates for decay of their simultaneous application. This is generally not the case for two systems having MLSI. While mathematically modeling real errors that occur in quantum technologies leaves many remaining open questions, we have taken steps toward adapting to the challenge.

One common explanation for the difficulty in protecting quantum information is the need to maintain purity in multiple bases, which prevents schemes including the classical repetition code from functioning for qubits. Quantum information is necessarily private due to its multi-basis nature: copying in any bases reduces it to classical mixture. Recalling Figure 2.1, collapsing the Bloch sphere to any plane puts it within a set of states that are representable within two classical bits, and full collapse to any axis reduces it to a single classical bit. Copying as the primary explanation for decoherence and quantum collapse is the primary idea behind the “einselection” explanation [172].

The tendency for errors to occur primarily in *interacting* gates rather than single-qubit gates is also suggestive. From the perspective of Hamiltonian engineering, the goal is to dynamically turn on an interaction term between two qubits without increasing interaction terms with other qubits, whether these qubits are other parts of the computation or come implicitly from the experimental environment (e.g. the walls of the chamber and any light in the room might be thought of as extra qubits, generally in thermal mixture). A useful gate-implementing Hamiltonian strongly breaks time symmetry (the process implementing a traditional gate should run for a precise duration, though paradigms such as adiabatic computation may only improve with longer runtime) and most forms of interchange symmetry (a 2-qubit gate should not have the same effect if we interchange an active qubit with another qubit in the system or environment). We return

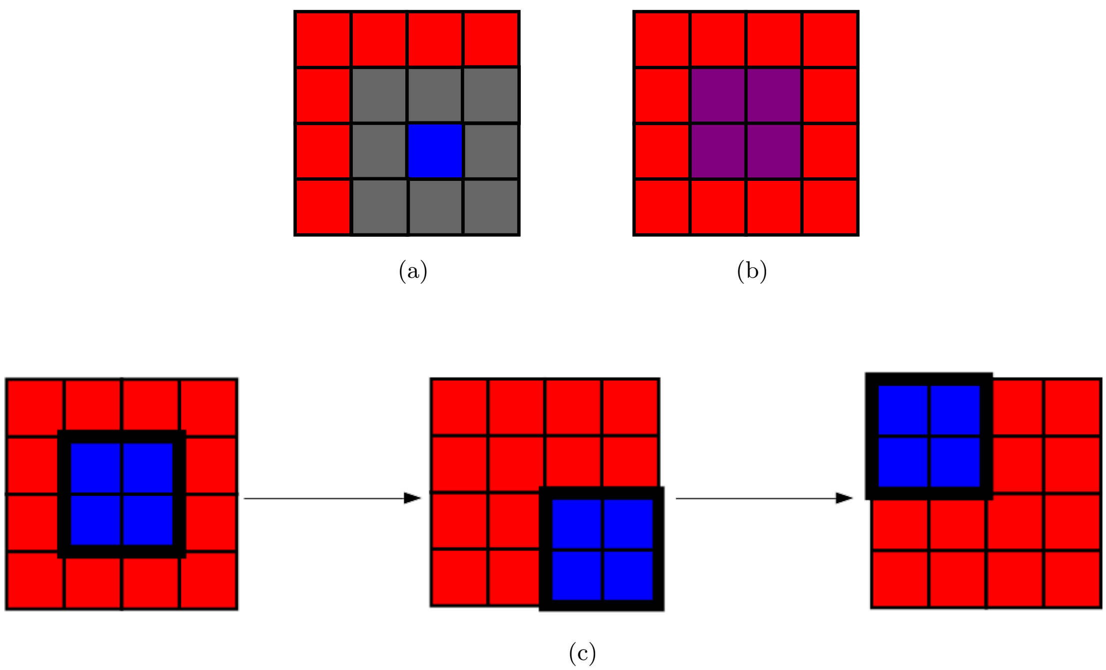

Figure 7.2: (a) Experimental isolation and control of a single qubit or a small number of qubits is possible with high-fidelity. (b) Strong interactions between qubits is possible as well. (c) The challenge is to obtain dynamically-controlled coupling between specific clusters of qubits while maintaining isolation from the environment, which may include qubits that will be active in other parts of the computation.

to the observation that $D(\rho\|\Phi^\infty(\rho)) \neq D(\Phi^\infty(\rho)\|\rho)$, the former appearing to be an extremely strong characterization of decay when Φ^∞ is a conditional expectation. As noted in Remark 7.5, the ergodic principle or principle of maximum entropy in thermodynamics is equivalent to maximizing the possible interchanges of particular states under which the system and environment are symmetric. $\mathcal{E}_\mathcal{N}(\rho)$, where $\mathcal{E}_\mathcal{N}$ is a conditional expectation, projects to a state of maximal asymmetry with respect to any set of transformations leaving $\mathcal{N}$ invariant. When $\mathcal{N} = \mathbb{C}$, this is the maximally symmetric state of the system. Decay to a thermal state (as discussed in Section 7.2.4) maximizes symmetry of the system-bath composite, as opposed to of the system alone. Hence the directionality of $\rho \to \Phi^\infty(\rho)$ is not coincidental. Physical instances of Φ^∞ tend to lose information or asymmetry, increasing entropy of the system-environment combination. Indeed, $D(\rho\|\Phi^\infty(\rho))$ is finite when Φ^∞ is guaranteed to increase support, so the form of the measure favors processes that solely spread rather than concentrate probability. A potential lingering question is whether some good properties of $D(\rho\|\Phi^t(\rho))$ or its extension to include part of an environment, such as finiteness or equivalence to a difference of entropies, might connect to the First and Second Laws of Thermodynamics.

Remark 7.5. *The favored macrostate of a thermal system maximizes the number of microscopic symmetries, defined as interchanges between microstates that preserve macroscopic observables.*

As an area of future inquiry, a primary drawback or limitation of the relative entropy of asymmetry $D^{\mathcal{N}}$ is that the comparison to a subalgebra disallows non-trivially weighted fixed point densities or distributions. It is plausible that thermal states, when expanded to include the bath as well as system, ultimately admit models in terms of unital comparison points. Marius Junge and I spent some time discussing extension of $D^{\mathcal{N}}$ to factorizable channels, which may arise from coupling to an environment that is initially completely mixed. This is unfortunately muddled by the fact that the unitary processes implementing channels with initially mixed environments are not unique.

As shown by Stinespring dilation, all processes are invertible unitaries to the combination of the system and its complete environment (unless decoherence results largely from some non-quantum effect, as has been proposed, though not necessarily at scales that would affect quantum computers [173]). Via purification, it is not hard to see that any state with no coherence in a given basis, or that has lost all entanglement with other systems, can be written as entirely copied in its environment. As briefly mentioned in Section 3.2, the quantity

$$D(\rho\|R \circ (\Phi^t)^c(\rho)) \ ,$$

where $(\Phi^t)^c$ is the complement of Φ^t, and where R is a sensible recovery map, approaches zero as $\Phi^t(\rho)$ loses all aspects that would distinguish it from classicality.

The idea of passive error correction [174] is to invert the favorability of globally high-entropy or high-symmetry states, constructing a system-environment coupling for which the overall state of maximum entropy is actually favorable to computation or memory. In Kitaev's passive anyon code [174], an active error-corrected subspace becomes the ground state subspace of a Hamiltonian, so that overall states of low

Figure 7.3: The loss of quantum information or of quantum aspects of information coincides with a copy appearing in the environment. Photon image from `https://commons.wikimedia.org/wiki/File:Quantum-Gravity-Photon-Race.jpg`, originally created by NASA

temperature favor the code space. Unfortunately, some active steps seem necessary for the scheme to be truly scalable [175].

Active error correction may also pay the thermodynamic price for maintaining computation against the pull of thermodynamics. Landauer's principle assigns a thermal cost to computations that erase information [6], which is usually small compared with the actual heat dissipated in computing. In principle a fully private classical computer should take great care to ensure decoupling between the heat, light, sounds, and other observable aspects of the computer's running from the running computation (see for instance "Van Eck Phreaking" [176], in which one uses electromagnetic fields near a computer monitor to eavesdrop on its contents). The required secrecy of quantum information transforms privacy from an option to a necessity, as any state-dependent leakage of electromagnetic or thermal energy induces decoherence. Refrigeration (and with it, passive error correction) is one way of expending work to maintain an internal state of lower entropy that isn't perfectly shielded from its environment. Active error correction may function analogously, allowing the system to constantly dump the entropy introduced by natural thermodynamics back to its environment without revealing the state of computation.

On the practical side, we are likely to deal with noise for long enough that many prominent theorists are trying to work with and around it [21]. A few theorists have long considered the possibility of decoherence as its own force, going beyond quantum mechanics [177, 178], but this is largely separate from quantum information considerations. MLSI-type bounds have found renewed interest in the mathematical community, including work by Beigi, Capel, Datta, Kastoryano, King, Lucia, Montanaro, Muller-Hermes, Olkiewicz, Peres-Garcia, Temme, Wolf & Zegarlinski to name a few. A potential angle of future study is how the decay of relative entropy may reveal fundamental aspects of relative entropy, quantum mechanics, and thermodynamics. Rather than view the asymmetry in its arguments and apparent favoring of conditional expectations as flaws, these aspects of relative entropy could hint at thermodynamic interpretations and help connect the decay of quantum information to its fundamental distinction from classical information.

Chapter 8

Conclusions and Outlook

Applications of information-theoretic techniques to physics and mathematics questions are recently growing, as are applications of operator algebra techniques to information theory. Often techniques with great mathematical sophistication enable progress on questions that resisted earlier approaches, such as when noncommutative analysis and geometry recovers results on quantum entropy that were relatively simple for classical information theory. In contrast, a theme emphasized in this thesis is how our understanding of even seemingly simple systems may improve with deeper mathematical structure. This is especially apparent in Chapter 4. Replacing subsystems by von Neumann subalgebras has well-known uses in field theory and in studying identical particles, for which the partial trace is insufficient or non-existent. We however applied this method to analyze non-classical correlations in 2×2 and 4×4 matrices representing qubits and pairs of qubits, where conventional notions of pinching and partial traces to subsystems nonetheless remain valid. In doing so, we better understand the connections between entanglement, coherence, and uncertainty.

A continuing open challenge is to reconcile many information-theoretic approaches with the main problems in scalable quantum technology. Near-term quantum test computers such as the IBM QX enable studies probing the realistic nature of quantum decoherence, which often differs from expectations [179]. Resource theories make an extremely sharp distinction between free and non-free operations, but true costs of quantum operations are typically far messier. The operational assumptions of Shannon-regime quantum communication, for instance, assume perfect encoding, decoding, and quantum memory, attributing all costs to noisy channels. At the other end, a typical gate-based computation model makes no special distinction for operations that must cross long distances, which frequently limit the meaningful size of quantum computers and thereby block effective use of many-qubit error-correcting codes. As alluded to in Chapter 5, the solution is almost certainly *not* to add immense technical detail and specificity to models, which by preventing generalization obliterates robustness to noise or to variation between experiments. Large-scale computer simulation is unlikely to solve these challenges in the long run, as the main promise of quantum technology is computation beyond the regimes accessible to conventional supercomputers. The underlying philosophy of resource theories is to distinguish between important and negligible effects, as does nearly any good physical model. There are still important questions about what effects we should consider important, which are now more tractable because of recent experimental progress.

Maintaining experimental grounding need not be mutually exclusive with studying fundamental theory. While the main themes of this thesis result from cross-pollination between different theoretical areas, several lines of inquiry came from directly observing the realities of experiments in Paul Kwiat's laboratory and others. In particular, the main themes of Chapter 6 come from the observation that real optics experiments often "pump" a photon source repeatedly, generating a huge number of experimental trials, and then discard most of them due to detectable photon loss or apparent misfiring of the source. Chapter 5 also references the practical need to infer expectations and moments from finitely many trials, which favors high symmetry of experimental outcomes with respect to per-trial details. As noted in Chapter 7, the apparent fragility of quantum information may further our understanding of its fundamental distinctions from classical information. Even in the absence of computing and telecommunication hardware that would support directly applying many of the quantum analogs of classic ideas from information theory, the methods of analysis learned in studying quantum information have already had a lasting impact on high-energy physics, condensed matter, and even operator algebras.

References

[1] Ahmed Almheiri, Xi Dong, and Daniel Harlow. Bulk locality and quantum error correction in AdS/CFT. *Journal of High Energy Physics*, 2015(4):163, April 2015.

[2] Eric Chitambar and Gilad Gour. Quantum resource theories. *Reviews of Modern Physics*, 91(2):025001, April 2019.

[3] Marius Junge, Miguel Navascues, Carlos Palazuelos, David Perez-Garcia, Volkher B. Scholz, and Reinhard F. Werner. Connes' embedding problem and Tsirelson's problem. *Journal of Mathematical Physics*, 52(1):012102, January 2011. Publisher: American Institute of Physics.

[4] John Archibald Wheeler. Information, Physics, Quantum: The Search for Links. In *Feynman and Computation*, pages 309–336. CRC Press, 1 edition, 1990.

[5] Rachel Thomas. It from bit? *Plus Magazine*, December 2015. Library Catalog: plus.maths.org.

[6] R. Landauer. Irreversibility and Heat Generation in the Computing Process. *IBM Journal of Research and Development*, 5(3):183–191, July 1961.

[7] Joshua A. Grochow and David H. Wolpert. Beyond Number of Bit Erasures: New Complexity Questions Raised by Recently Discovered Thermodynamic Costs of Computation. *SIGACT News*, 49(2):33–56, June 2018.

[8] Simone Santini. We Are Sorry to Inform You ... *Computer*, 38(12):128–127, December 2005. Conference Name: Computer.

[9] Claude E. Shannon. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3):379–423, July 1948.

[10] Peter W. Shor. Quantum information theory: The bits don't add up. *Nature Physics*, 5(4):247–248, April 2009.

[11] Felix Leditzky, Mohammad A. Alhejji, Joshua Levin, and Graeme Smith. Playing games with multiple access channels. *Nature Communications*, 11(1):1–5, March 2020. Number: 1 Publisher: Nature Publishing Group.

[12] Felix Leditzky, Debbie Leung, and Graeme Smith. Dephrasure Channel and Superadditivity of Coherent Information. *Physical Review Letters*, 121(16):160501, October 2018.

[13] Vlatko Vedral. The role of relative entropy in quantum information theory. *Reviews of Modern Physics*, 74(1):197–234, March 2002. Publisher: American Physical Society.

[14] Albert Einstein, Boris Podolsky, and Nathan Rosen. Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? *Physical Review*, 47(10):777–780, May 1935.

[15] B. Hensen, N. Kalb, M. S. Blok, A. E. Dréau, A. Reiserer, R. F. L. Vermeulen, R. N. Schouten, M. Markham, D. J. Twitchen, K. Goodenough, D. Elkouss, S. Wehner, T. H. Taminiau, and R. Hanson. Loophole-free Bell test using electron spins in diamond: second experiment and additional analysis. *Scientific Reports*, 6(1):1–11, August 2016. Number: 1 Publisher: Nature Publishing Group.

[16] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C. Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando G. S. L. Brandao, David A. Buell, Brian Burkett, Yu Chen, Zijun Chen, Ben Chiaro, Roberto Collins, William Courtney, Andrew Dunsworth, Edward Farhi, Brooks Foxen, Austin Fowler, Craig Gidney, Marissa Giustina, Rob Graff, Keith Guerin, Steve Habegger, Matthew P. Harrigan, Michael J. Hartmann, Alan Ho, Markus Hoffmann, Trent Huang, Travis S. Humble, Sergei V. Isakov, Evan Jeffrey, Zhang Jiang, Dvir Kafri, Kostyantyn Kechedzhi, Julian Kelly, Paul V. Klimov, Sergey Knysh, Alexander Korotkov, Fedor Kostritsa, David Landhuis, Mike Lindmark, Erik Lucero, Dmitry Lyakh, Salvatore Mandr, Jarrod R. McClean, Matthew McEwen, Anthony Megrant, Xiao Mi, Kristel Michielsen, Masoud Mohseni, Josh Mutus, Ofer Naaman, Matthew Neeley, Charles Neill, Murphy Yuezhen Niu, Eric Ostby, Andre Petukhov, John C. Platt, Chris Quintana, Eleanor G. Rieffel, Pedram Roushan, Nicholas C. Rubin, Daniel Sank, Kevin J. Satzinger, Vadim Smelyanskiy, Kevin J. Sung, Matthew D. Trevithick, Amit Vainsencher, Benjamin Villalonga, Theodore White, Z. Jamie Yao, Ping Yeh, Adam Zalcman, Hartmut Neven, and John M. Martinis. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, October 2019. Number: 7779 Publisher: Nature Publishing Group.

[17] Ryszard Horodecki, Paweł Horodecki, Michał Horodecki, and Karol Horodecki. Quantum entanglement. *Reviews of Modern Physics*, 81(2):865–942, June 2009.

[18] Stephanie Wehner and Andreas Winter. Entropic uncertainty relations - a survey. *New Journal of Physics*, 12(2):025009, 2010.

[19] Patrick J. Coles, Mario Berta, Marco Tomamichel, and Stephanie Wehner. Entropic uncertainty relations and their applications. *Reviews of Modern Physics*, 89(1):015002, February 2017.

[20] Richard Cleve, Peter Hoyer, Ben Toner, and John Watrous. Consequences and Limits of Nonlocal Strategies. *arXiv:quant-ph/0404076*, April 2004. arXiv: quant-ph/0404076.

[21] John Preskill. Quantum Computing in the NISQ era and beyond. *Quantum*, 2:79, August 2018.

[22] John Preskill. Quantum computing and the entanglement frontier. *arXiv:1203.5813 [cond-mat, physics:quant-ph]*, November 2012. arXiv: 1203.5813.

[23] Li Gao, Marius Junge, and Nicholas LaRacuente. Unifying Entanglement with Uncertainty via Symmetries of Observable Algebras. *arXiv:1710.10038 [quant-ph]*, October 2017. arXiv: 1710.10038.

[24] Li Gao, Marius Junge, and Nicholas LaRacuente. Uncertainty Principle for Quantum Channels. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pages 996–1000, June 2018.

[25] Li Gao, Marius Junge, and Nicholas LaRacuente. Relative entropy for von Neumann subalgebras. *arXiv:1909.01906 [math]*, September 2019. arXiv: 1909.01906.

[26] Nicholas LaRacuente. Adjusted Subadditivity of Relative Entropy for Non-commuting Conditional Expectations. *arXiv:1912.00983 [quant-ph]*, December 2019. arXiv: 1912.00983.

[27] Li Gao, Marius Junge, and Nicholas LaRacuente. Heralded channel Holevo superadditivity bounds from entanglement monogamy. *Journal of Mathematical Physics*, 59(6):062203, June 2018. Publisher: American Institute of Physics.

[28] Spencer Johnson, Nicholas LaRacuente, Marius Junge, Eric Chitambar, and Paul Kwiat. Towards an Experimental Demonstration of Superadditivity through the Dephrasure Channel. In *APS Division of Atomic, Molecular and Optical Physics APS Meeting Abstracts*, May 2019.

[29] Li Gao, Marius Junge, and Nicholas LaRacuente. Capacity bounds via operator space methods. *Journal of Mathematical Physics*, 59(12):122202, December 2018.

[30] Li Gao, Marius Junge, and Nicholas LaRacuente. Capacity Estimates via Comparison with TRO Channels. *Communications in Mathematical Physics*, 364(1):83–121, November 2018.

[31] Li Gao, Marius Junge, and Nicolas LaRacuente. Fisher Information and Logarithmic Sobolev Inequality for Matrix Valued Functions. *arXiv:1807.08838 [quant-ph]*, July 2018. arXiv: 1807.08838.

[32] Marius Junge, Nicholas LaRacuente, and Cambyse Rouzé. Stability of logarithmic Sobolev inequalities under a noncommutative change of measure. *arXiv:1911.08533 [math-ph, physics:quant-ph]*, November 2019. arXiv: 1911.08533.

[33] Ivan Bardet, Marius Junge, Nicholas LaRacuente, Cambyse Rouzé, and Daniel Stilck Frana. Group transference techniques for the estimation of the decoherence times and capacities of quantum Markov semigroups. *arXiv:1904.11043 [quant-ph]*, April 2019. arXiv: 1904.11043.

[34] Franco Strocchi. *An Introduction to the Mathematical Structure of Quantum Mechanics: A Short Course for Mathematicians*. World Scientific, 2008. Google-Books-ID: Bn7MaT3X8fkC.

[35] Richard Cleve, Benoit Collins, Li Liu, and Vern Paulsen. Constant gap between conventional strategies and those based on C*-dynamics for self-embezzlement. *arXiv:1811.12575 [quant-ph]*, April 2019. arXiv: 1811.12575.

[36] Rui Chao, Ben W. Reichardt, Chris Sutherland, and Thomas Vidick. Overlapping Qubits. In Christos H. Papadimitriou, editor, *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)*, volume 67 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 48:1–48:21, Dagstuhl, Germany, 2017. Schloss DagstuhlLeibniz-Zentrum fuer Informatik. ISSN: 1868-8969.

[37] Edward Witten. APS Medal for Exceptional Achievement in Research: Invited article on entanglement properties of quantum field theory. *Reviews of Modern Physics*, 90(4):045003, October 2018.

[38] Huzihiro Araki and E. J. Woods. A Classification of Factors. *Publications of the Research Institute for Mathematical Sciences*, 4(1):51–130, April 1968.

[39] Zhengfeng Ji, Anand Natarajan, Thomas Vidick, John Wright, and Henry Yuen. MIP*=RE. *arXiv:2001.04383 [quant-ph]*, January 2020. arXiv: 2001.04383.

[40] Mark Wilde. *Quantum Information Theory*. Cambridge University Press, April 2013. Google-Books-ID: T36v2Sp7DnIC.

[41] Mark Wilde. Recoverability in Quantum Information Theory, 2015. http://markwilde.com/publications/LSU-Phys-2015-colloquium.pdf.

[42] Hisaharu Umegaki. Conditional expectation in an operator algebra. IV. Entropy and information. *Kodai Mathematical Seminar Reports*, 14(2):59–85, 1962. Publisher: Tokyo Institute of Technology, Department of Mathematics.

[43] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 2006.

[44] Stephen J. Summers. Tomita-Takesaki Modular Theory. *arXiv:math-ph/0511034*, November 2005. arXiv: math-ph/0511034.

[45] Lin Zhang and Junde Wu. Tomita-Takesaki Modular Theory vs. Quantum Information Theory. *arXiv:1301.1836 [math-ph, physics:quant-ph]*, January 2013. arXiv: 1301.1836.

[46] Martin Müller-Lennert, Frédéric Dupuis, Oleg Szehr, Serge Fehr, and Marco Tomamichel. On quantum Renyi entropies: a new generalization and some properties. *Journal of Mathematical Physics*, 54(12):122203, December 2013. arXiv: 1306.3142.

[47] Mark M. Wilde, Andreas Winter, and Dong Yang. Strong Converse for the Classical Capacity of Entanglement-Breaking and Hadamard Channels via a Sandwiched Rényi Relative Entropy. *Communications in Mathematical Physics*, 331(2):593–622, October 2014.

[48] Koenraad M. R. Audenaert and Nilanjana Datta. -z-Rényi relative entropies. *Journal of Mathematical Physics*, 56(2):022202, February 2015. Publisher: American Institute of Physics.

[49] Marco Tomamichel. A Framework for Non-Asymptotic Quantum Information Theory. *arXiv:1203.2142 [math-ph, physics:quant-ph]*, July 2013. arXiv: 1203.2142.

[50] Giuseppe Vallone, Davide G. Marangon, Marco Tomasin, and Paolo Villoresi. Quantum randomness certified by the uncertainty principle. *Physical Review A*, 90(5):052327, November 2014. Publisher: American Physical Society.

[51] Koenraad M. R. Audenaert and Jens Eisert. Continuity bounds on the quantum relative entropy. *Journal of Mathematical Physics*, 46(10):102104, October 2005. Publisher: American Institute of Physics.

[52] Koenraad M. R. Audenaert and Jens Eisert. Continuity bounds on the quantum relative entropy II. *Journal of Mathematical Physics*, 52(11):112201, November 2011. Publisher: American Institute of Physics.

[53] Andreas Winter. Tight uniform continuity bounds for quantum entropies: conditional entropy, relative entropy distance and energy constraints. *Communications in Mathematical Physics*, 347(1):291–313, October 2016. arXiv: 1507.07775.

[54] Daniel Manzano. A short introduction to the Lindblad master equation. *AIP Advances*, 10(2):025106, February 2020. Publisher: American Institute of Physics.

[55] Omar Fawzi and Renato Renner. Quantum Conditional Mutual Information and Approximate Markov Chains. *Communications in Mathematical Physics*, 340(2):575–611, December 2015.

[56] Marius Junge, Renato Renner, David Sutter, Mark M. Wilde, and Andreas Winter. Universal Recovery Maps and Approximate Sufficiency of Quantum Relative Entropy. *Annales Henri Poincaré*, 19(10):2955–2978, October 2018.

[57] Mark M. Wilde. Recoverability in quantum information theory. *Proc. R. Soc. A*, 471(2182):20150338, October 2015.

[58] Elliott H. Lieb and Mary Beth Ruskai. Proof of the strong subadditivity of quantummechanical entropy. *Journal of Mathematical Physics*, 14(12):1938–1941, December 1973.

[59] Gran Lindblad. Completely positive maps and entropy inequalities. *Communications in Mathematical Physics*, 40(2):147–151, 1975. Publisher: Springer-Verlag.

[60] Dénes Petz. On certain properties of the relative entropy of states of operator algebras. *Mathematische Zeitschrift*, 206(1):351–361, January 1991.

[61] Mario Berta, Matthias Christandl, Roger Colbeck, Joseph M. Renes, and Renato Renner. The uncertainty principle in the presence of quantum memory. *Nature Physics*, 6(9):659, September 2010.

[62] Uttam Singh, Arun Kumar Pati, and Manabendra Nath Bera. Uncertainty Relations for Quantum Coherence. *Mathematics*, 4(3):47, July 2016.

[63] Elena Caceres, Arnab Kundu, Juan F. Pedraza, and Walter Tangarife. Strong subadditivity, null energy condition and charged black holes. *Journal of High Energy Physics*, 2014(1):84, January 2014.

[64] Mohammad Alhejji and Graeme Smith. Monotonicity Under Local Operations: Linear Entropic Formulas. *arXiv:1811.08000 [quant-ph]*, November 2018. arXiv: 1811.08000.

[65] Ning Bao, Sepehr Nezami, Hirosi Ooguri, Bogdan Stoica, James Sully, and Michael Walter. The Holographic Entropy Cone. *Journal of High Energy Physics*, 2015(9), September 2015. arXiv: 1505.07839.

[66] Michael A. Nielsen and Denes Petz. A simple proof of the strong subadditivity inequality. *arXiv:quant-ph/0408130*, August 2004. arXiv: quant-ph/0408130.

[67] Mary Beth Ruskai. Another Short and Elementary Proof of Strong Subadditivity of Quantum Entropy. *Reports on Mathematical Physics*, 60(1):1–12, August 2007. arXiv: quant-ph/0604206.

[68] Eric A. Carlen and Elliott H. Lieb. Bounds for Entanglement via an Extension of Strong Subadditivity of Entropy. *Letters in Mathematical Physics*, 101(1):1–11, July 2012.

[69] Isaac H. Kim. Operator extension of strong subadditivity of entropy. *Journal of Mathematical Physics*, 53(12):122204, November 2012.

[70] Filippo Cesi. Quasi-factorization of the entropy and logarithmic Sobolev inequalities for Gibbs random fields. *Probability Theory and Related Fields*, 120(4):569–584, August 2001.

[71] Angela Capel, Angelo Lucia, and David Pérez-Garca. Quantum conditional relative entropy and quasi-factorization of the relative entropy. *Journal of Physics A: Mathematical and Theoretical*, 51(48):484001, November 2018. arXiv: 1804.09525.

[72] Ivan Bardet, Angela Capel, Angelo Lucia, David Pérez-Garca, and Cambyse Rouzé. On the modified logarithmic Sobolev inequality for the heat-bath dynamics for 1D systems. *arXiv:1908.09004 [cond-mat, physics:math-ph, physics:quant-ph]*, August 2019. arXiv: 1908.09004.

[73] Ivan Bardet, Angela Capel, and Cambyse Rouzé. Approximate tensorization of the relative entropy for noncommuting conditional expectations. *arXiv:2001.07981 [math-ph, physics:quant-ph]*, January 2020. arXiv: 2001.07981.

[74] Koenraad M. R. Audenaert. Telescopic Relative Entropy. In Dave Bacon, Miguel Martin-Delgado, and Martin Roetteler, editors, *Theory of Quantum Computation, Communication, and Cryptography*, Lecture Notes in Computer Science, pages 39–52, Berlin, Heidelberg, 2014. Springer.

[75] Armin Uhlmann. Relative entropy and the Wigner-Yanase-Dyson-Lieb concavity in an interpolation theory. *Communications in Mathematical Physics*, 54(1):21–32, February 1977.

[76] Marius Junge and Quanhua Xu. Noncommutative Maximal Ergodic Theorems. *Journal of the American Mathematical Society*, 20(2):385–439, 2007.

[77] Jason Crann, David W Kribs, Rupert H Levene, and Ivan G Todorov. Private algebras in quantum information and infinite-dimensional complementarity. *Journal of Mathematical Physics*, 57(1):015208, 2016.

[78] Pavan Hosur, Xiao-Liang Qi, Daniel A. Roberts, and Beni Yoshida. Chaos in quantum channels. *Journal of High Energy Physics*, 2016(2):4, February 2016.

[79] William J. McGill. Multivariate information transmission. *Psychometrika*, 19(2):97–116, June 1954.

[80] Hu Kuo Ting. On the Amount of Information. *Theory of Probability & Its Applications*, 7(4):439–447, January 1962.

[81] Joshua Levin and Graeme Smith. Private Communications, December 2019.

[82] Don N. Page. Average Entropy of a Subsystem. *Physical Review Letters*, 71(9):1291–1294, August 1993. arXiv: gr-qc/9305007.

[83] Vern Paulsen. *Completely Bounded Maps and Operator Algebras*. Cambridge University Press, 2002. Google-Books-ID: VtSFHDABxMIC.

[84] Michael A. Nielsen. Conditions for a Class of Entanglement Transformations. *Physical Review Letters*, 83(2):436–439, July 1999.

[85] Eric Chitambar, Debbie Leung, Laura Mančinska, Maris Ozols, and Andreas Winter. Everything You Always Wanted to Know About LOCC (But Were Afraid to Ask). *Communications in Mathematical Physics*, 328(1):303–326, May 2014.

[86] Tillmann Baumgratz, Marcus Cramer, and Martin B. Plenio. Quantifying Coherence. *Physical Review Letters*, 113(14):140401, September 2014. Publisher: American Physical Society.

[87] Alexander Streltsov, Gerardo Adesso, and Martin B. Plenio. Colloquium: Quantum coherence as a resource. *Reviews of Modern Physics*, 89(4):041003, October 2017.

[88] Iman Marvian and Robert W. Spekkens. Extending Noethers theorem by quantifying the asymmetry of quantum states. *Nature Communications*, 5:3821, May 2014.

[89] Gilad Gour, Iman Marvian, and Robert W. Spekkens. Measuring the quality of a quantum reference frame: The relative entropy of frameness. *Physical Review A*, 80(1):012307, July 2009.

[90] Joan Alfina Vaccaro, Fabio Anselmi, Howard Mark Wiseman, and Kurt Jacobs. Tradeoff between extractable mechanical work, accessible entanglement, and ability to act as a reference system, under arbitrary superselection rules. *Physical Review A*, 77(3):032114, 2008.

[91] Tom Cooney, Milán Mosonyi, and Mark M. Wilde. Strong Converse Exponents for a Quantum Channel Discrimination Problem and Quantum-Feedback-Assisted Communication. *Communications in Mathematical Physics*, 344(3):797–829, June 2016.

[92] Gilad Gour and Mark M. Wilde. Entropy of a quantum channel. *arXiv:1808.06980 [cond-mat, physics:hep-th, physics:math-ph, physics:quant-ph]*, January 2020. arXiv: 1808.06980.

[93] Kun Fang, Omar Fawzi, Renato Renner, and David Sutter. Chain Rule for the Quantum Relative Entropy. *Physical Review Letters*, 124(10):100501, March 2020. Publisher: American Physical Society.

[94] Eric Chitambar and Min-Hsiu Hsieh. Relating the Resource Theories of Entanglement and Quantum Coherence. *Physical Review Letters*, 117(2):020402, July 2016.

[95] Alexander Streltsov, Swapan Rana, Manabendra Nath Bera, and Maciej Lewenstein. Towards Resource Theory of Coherence in Distributed Scenarios. *Physical Review X*, 7(1):011024, March 2017.

[96] Alexander Streltsov, Uttam Singh, Himadri Shekhar Dhar, Manabendra Nath Bera, and Gerardo Adesso. Measuring Quantum Coherence with Entanglement. *Physical Review Letters*, 115(2):020403, July 2015.

[97] Kok Chuan Tan, Hyukjoon Kwon, Chae-Yeun Park, and Hyunseok Jeong. Unified view of quantum correlations and quantum coherence. *Physical Review A*, 94(2):022329, August 2016.

[98] Xiao-Li Wang, Qiu-Ling Yue, Chao-Hua Yu, Fei Gao, and Su-Juan Qin. Relating quantum coherence and correlations with entropy-based measures. *arXiv:1703.00648 [quant-ph]*, March 2017. arXiv: 1703.00648.

[99] Yunchao Liu, Qi Zhao, and Xiao Yuan. Quantum coherence via conditional entropy. *arXiv:1712.02732 [quant-ph]*, December 2017. arXiv: 1712.02732.

[100] Xianfei Qi, Ting Gao, and Fengli Yan. Measuring coherence with entanglement concurrence. *Journal of Physics A: Mathematical and Theoretical*, 50(28):285301, 2017.

[101] Seungbeom Chin. Generalized coherence concurrence and path distinguishability. *Journal of Physics A: Mathematical and Theoretical*, 50(47):475302, 2017.

[102] Huangjun Zhu, Masahito Hayashi, and Lin Chen. Coherence and entanglement measures based on Rényi relative entropies. *Journal of Physics A: Mathematical and Theoretical*, 50(47):475303, 2017.

[103] Yuan Sun, Yuanyuan Mao, and Shunlong Luo. From quantum coherence to quantum correlations. *EPL (Europhysics Letters)*, 118(6):60007, 2017.

[104] Shunlong Luo and Yuan Sun. Quantum coherence versus quantum uncertainty. *Physical Review A*, 96(2):022130, August 2017.

[105] Ming-Liang Hu and Heng Fan. Relative quantum coherence, incompatibility, and quantum correlations of states. *Physical Review A*, 95(5):052106, May 2017.

[106] Howard Barnum, Emanuel Knill, Gerardo Ortiz, Rolando Somma, and Lorenza Viola. A Subsystem-Independent Generalization of Entanglement. *Physical Review Letters*, 92(10):107902, March 2004.

[107] Łukasz Derkacz, Marek Gwóźdź, and Lech Jakóbczyk. Entanglement beyond tensor product structure: algebraic aspects of quantum non-separability. *Journal of Physics A: Mathematical and Theoretical*, 45(2):025302, 2012.

[108] Aiyalam P. Balachandran, Thupil R. Govindarajan, Amilcar R. de Queiroz, and Andrés F. Reyes-Lega. Entanglement and Particle Identity: A Unifying Approach. *Physical Review Letters*, 110(8):080503, February 2013.

[109] Aiyalam P. Balachandran, Thupil R. Govindarajan, Amilcar R. de Queiroz, and Andrés F. Reyes-Lega. Algebraic approach to entanglement and entropy. *Physical Review A*, 88(2):022301, August 2013.

[110] Hai-Woong Lee and Jaewan Kim. Quantum teleportation and Bell's inequality using single-particle entanglement. *Physical Review A*, 63(1):012305, December 2000.

[111] D. Boschi, S. Branca, F. De Martini, L. Hardy, and S. Popescu. Experimental Realization of Teleporting an Unknown Pure Quantum State via Dual Classical and Einstein-Podolsky-Rosen Channels. *Physical Review Letters*, 80(6):1121–1125, February 1998.

[112] Markus Michler, Harald Weinfurter, and Marek Żukowski. Experiments towards Falsification of Noncontextual Hidden Variable Theories. *Physical Review Letters*, 84(24):5457–5461, June 2000.

[113] Robert J. C. Spreeuw. A Classical Analogy of Entanglement. *Foundations of Physics*, 28(3):361–374, March 1998.

[114] Xiao-Feng Qian and J. H. Eberly. Entanglement and classical polarization states. *Optics Letters*, 36(20):4110–4112, October 2011. Publisher: Optical Society of America.

[115] Matthias Christandl and Andreas Winter. Squashed entanglement: An additive entanglement measure. *Journal of Mathematical Physics*, 45(3):829–840, February 2004.

[116] Fernando G. S. L. Brando, Matthias Christandl, and Jon Yard. Faithful Squashed Entanglement. *Communications in Mathematical Physics*, 306(3):805, August 2011.

[117] Ke Li and Andreas Winter. Squashed Entanglement, k-Extendibility, Quantum Markov Chains, and Recovery Maps. *Foundations of Physics*, 48(8):910–924, August 2018.

[118] Emmy Noether. Invariant variation problems. *Transport Theory and Statistical Physics*, 1(3):186–207, January 1971. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/00411457108231446.

[119] Thomas B Mieling. Noethers theorem applied to classical electrodynamics. page 7, 2017. https://homepage.univie.ac.at/thomas.mieling/pdf/electrodynamics-noether.pdf.

[120] John C. Baez and Brendan Fong. A Noether theorem for Markov processes. *Journal of Mathematical Physics*, 54(1):013301, January 2013.

[121] John E. Gough, Tudor S. Ratiu, and Oleg G. Smolyanov. Noethers theorem for dissipative quantum dynamical semi-groups. *Journal of Mathematical Physics*, 56(2):022108, February 2015.

[122] Aurelian Gheondea. On Propagation of Fixed Points of Quantum Operations and Beyond. *arXiv:1611.04742 [math]*, November 2016. arXiv: 1611.04742.

[123] Iman Marvian and Robert W. Spekkens. The theory of manipulations of pure state asymmetry: I. Basic tools, equivalence classes and single copy transformations. *New Journal of Physics*, 15(3):033001, 2013.

[124] Iman Mashad Marvian. Symmetry, Asymmetry and Quantum Information. September 2012.

[125] Ludovico Lami. Non-classical correlations in quantum mechanics and beyond. *arXiv:1803.02902 [math-ph, physics:quant-ph]*, March 2018. arXiv: 1803.02902.

[126] Matthias Christandl, Robert Knig, Graeme Mitchison, and Renato Renner. One-and-a-Half Quantum de Finetti Theorems. *Communications in Mathematical Physics*, 273(2):473–498, July 2007.

[127] Victor V. Albert. Lindbladians with multiple steady states: theory and applications. *arXiv:1802.00010 [cond-mat, physics:math-ph, physics:quant-ph]*, January 2018. arXiv: 1802.00010.

[128] Joe Rosen. Symmetry at the Foundation of Science and Nature. *Symmetry*, 1(1):3–9, June 2009.

[129] Mark K. Transtrum, Benjamin B. Machta, Kevin S. Brown, Bryan C. Daniels, Christopher R. Myers, and James P. Sethna. Perspective: Sloppiness and emergent theories in physics, biology, and beyond. *The Journal of Chemical Physics*, 143(1):010901, July 2015.

[130] David H. Wolpert, Joshua A. Grochow, Eric Libby, and Simon DeDeo. Optimal high-level descriptions of dynamical systems. *arXiv:1409.7403 [cond-mat, q-bio]*, September 2014. arXiv: 1409.7403.

[131] Benjamin B. Machta, Ricky Chachra, Mark K. Transtrum, and James P. Sethna. Parameter Space Compression Underlies Emergent Theories and Predictive Models. *Science*, 342(6158):604–607, November 2013.

[132] Edward N. Lorenz. Atmospheric Predictability as Revealed by Naturally Occurring Analogues. *Journal of the Atmospheric Sciences*, 26(4):636–646, July 1969.

[133] Philippe Faist, Sepehr Nezami, Victor V. Albert, Grant Salton, Fernando Pastawski, Patrick Hayden, and John Preskill. Continuous symmetries and approximate quantum error correction. *arXiv:1902.07714 [cond-mat, physics:hep-th, physics:quant-ph]*, February 2019. arXiv: 1902.07714.

[134] Patrick Hayden, Michał Horodecki, Andreas Winter, and Jon Yard. A Decoupling Approach to the Quantum Capacity. *Open Systems & Information Dynamics*, 15(01):7–19, March 2008. Publisher: World Scientific Publishing Co.

[135] Jelani Nelson. Dimensionality Reduction Notes. page 9, August 2015. https://people.eecs.berkeley.edu/˜minilek/madalgo2015/notes1.pdf.

[136] Alexander S. Holevo. Some estimates for the amount of information transmittable by a quantum communications channel. *Akademiya Nauk SSSR. Institut Problem Peredachi Informatsii Akademii Nauk SSSR. Problemy Peredachi Informatsii*, 9(3):3–11, 1973.

[137] Benjamin Schumacher. Sending classical information via noisy quantum channels. *Physical Review A*, 56(1):131–138, 1997.

[138] Matthew B. Hastings. Superadditivity of communication capacity using entangled inputs. *Nature Physics*, 5(4):255–257, April 2009.

[139] A. Winter and D. Yang. Potential Capacities of Quantum Channels. *IEEE Transactions on Information Theory*, 62(3):1415–1424, March 2016.

[140] Fernando G. S. L. Brando, Jonathan Oppenheim, and Sergii Strelchuk. When Does Noise Increase the Quantum Capacity? *Physical Review Letters*, 108(4):040501, January 2012. Publisher: American Physical Society.

[141] Graeme Smith and Jon Yard. Quantum Communication with Zero-Capacity Channels. *Science*, 321(5897):1812–1815, September 2008.

[142] Emanuel Knill, Raymond Laflamme, and Gerard J. Milburn. A scheme for efficient quantum computation with linear optics. *Nature*, 409(6816):46–52, January 2001.

[143] Li Gao, Marius Junge, and Nicholas LaRacuente. Capacity Estimates via comparison with TRO channels. *arXiv:1609.08594 [quant-ph]*, September 2016. arXiv: 1609.08594.

[144] Debbie Leung, Joungkeun Lim, and Peter Shor. Capacity of Quantum Erasure Channel Assisted by Backwards Classical Communication. *Physical Review Letters*, 103(24):240505, December 2009.

[145] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt. Proposed Experiment to Test Local Hidden-Variable Theories. *Physical Review Letters*, 23(15):880–884, October 1969. Publisher: American Physical Society.

[146] Ben W. Reichardt, Falk Unger, and Umesh Vazirani. A classical leash for a quantum system: Command of quantum systems via rigidity of CHSH games. *arXiv:1209.0448 [quant-ph]*, September 2012. arXiv: 1209.0448.

[147] Julia Kempe, Hirotada Kobayashi, Keiji Matsumoto, Ben Toner, and Thomas Vidick. Entangled Games Are Hard to Approximate. *SIAM Journal on Computing*, 40(3):848–877, January 2011.

[148] Ben Toner. Monogamy of non-local quantum correlations. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 465(2101):59–69, January 2009.

[149] Michael P. Seevinck. Monogamy of correlations versus monogamy of entanglement. *Quantum Information Processing*, 9(2):273–294, April 2010.

[150] Barbara M. Terhal, Andrew C. Doherty, and David Schwab. Symmetric Extensions of Quantum States and Local Hidden Variable Theories. *Physical Review Letters*, 90(15):157903, April 2003.

[151] Shang Yu, Yu Meng, Raj B. Patel, Yi-Tao Wang, Zhi-Jin Ke, Wei Liu, Zhi-Peng Li, Yuan-Ze Yang, Wen-Hao Zhang, Jian-Shun Tang, Chuan-Feng Li, and Guang-Can Guo. Experimental observation of coherent-information superadditivity in a dephrasure channel. *arXiv:2003.13000 [quant-ph]*, April 2020. arXiv: 2003.13000.

[152] Ying Li, Peter C. Humphreys, Gabriel J. Mendoza, and Simon C. Benjamin. Resource Costs for Fault-Tolerant Linear Optical Quantum Computing. *Physical Review X*, 5(4):041007, October 2015. Publisher: American Physical Society.

[153] Pieter Kok, W. J. Munro, Kae Nemoto, T. C. Ralph, Jonathan P. Dowling, and G. J. Milburn. Linear optical quantum computing with photonic qubits. *Reviews of Modern Physics*, 79(1):135–174, January 2007.

[154] Terry Rudolph. Why I am optimistic about the silicon-photonic route to quantum computing. *APL Photonics*, 2(3):030901, March 2017. Publisher: American Institute of Physics.

[155] Paul G. Kwiat. Hyper-entangled states. *Journal of Modern Optics*, 44(11-12):2173–2184, November 1997. Publisher: Taylor & Francis _eprint: https://www.tandfonline.com/doi/pdf/10.1080/09500349708231877.

[156] Leonardo Neves, Gustavo Lima, Aldo Delgado, and Carlos Saavedra. Hybrid photonic entanglement: Realization, characterization, and applications. *Physical Review A*, 80(4):042322, October 2009. Publisher: American Physical Society.

[157] I. Nape, B. Ndagano, B. Perez-Garcia, R. I. Hernandez-Aranda, F. S. Roux, T. Konrad, and A. Forbes. Hybrid entanglement for quantum information and communication applications. In *Optical Trapping and Optical Micromanipulation XIV*, volume 10347, page 1034711. International Society for Optics and Photonics, August 2017.

[158] Todd Pittman. Viewpoint: Its a Good Time for Time-Bin Qubits. *Physics*, 6, October 2013. Publisher: American Physical Society.

[159] Manmohan Kaur and Zhong-Jin Ruan. Local Properties of Ternary Rings of Operators and Their Linking C *-Algebras. *Journal of Functional Analysis*, 195(2):262–305, November 2002.

[160] Li Gao, Marius Junge, and Nicholas LaRacuente. Capacity Bounds via Operator Space Methods. *arXiv:1509.07294 [quant-ph]*, September 2015. arXiv: 1509.07294.

[161] Hideki Kosaki. Applications of the complex interpolation method to a von Neumann algebra: Non-commutative Lp-spaces. *Journal of Functional Analysis*, 56(1):29–78, March 1984.

[162] Motohisa Fukuda and Michael M. Wolf. Simplifying additivity problems using direct sum constructions. *Journal of Mathematical Physics*, 48(7):072101, July 2007.

[163] Leonard Gross. Logarithmic Sobolev Inequalities. *American Journal of Mathematics*, 97(4):1061–1083, 1975. Publisher: Johns Hopkins University Press.

[164] Michael J. Kastoryano and Kristan Temme. Quantum logarithmic Sobolev inequalities and rapid mixing. *Journal of Mathematical Physics*, 54(5):052202, May 2013. Publisher: American Institute of Physics.

[165] Ivan Bardet. Estimating the decoherence time using non-commutative Functional Inequalities. *arXiv:1710.01039 [math-ph, physics:quant-ph]*, October 2017. arXiv: 1710.01039.

[166] Eric A. Carlen and Jan Maas. Gradient flow and entropy inequalities for quantum Markov semigroups with detailed balance. *Journal of Functional Analysis*, 273(5):1810–1869, September 2017.

[167] Ivan Bardet and Cambyse Rouzé. Hypercontractivity and logarithmic Sobolev Inequality for non-primitive quantum Markov semigroups and estimation of decoherence rates. *arXiv:1803.05379 [math-ph, physics:quant-ph]*, July 2018. arXiv: 1803.05379.

[168] Richard Holley and Daniel Stroock. Logarithmic Sobolev inequalities and stochastic Ising models. *Journal of Statistical Physics*, 46(5):1159–1194, March 1987.

[169] Michel Ledoux. *The Concentration of Measure Phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, Rhode Island, February 2005.

[170] Prakash Murali, Jonathan M. Baker, Ali Javadi-Abhari, Frederic T. Chong, and Margaret Martonosi. Noise-Adaptive Compiler Mappings for Noisy Intermediate-Scale Quantum Computers. In *Proceedings of the Twenty-Fourth International Conference on Architectural Support for Programming Languages and Operating Systems*, ASPLOS '19, pages 1015–1029, Providence, RI, USA, April 2019. Association for Computing Machinery.

[171] Kenneth R. Brown, Jungsang Kim, and Christopher Monroe. Co-designing a scalable quantum computer with trapped atomic ions. *npj Quantum Information*, 2:16034, November 2016.

[172] Wojciech Hubert Zurek. Decoherence, einselection, and the quantum origins of the classical. *Reviews of Modern Physics*, 75(3):715–775, May 2003.

[173] GianCarlo C. Ghirardi, Alberto Rimini, and Tullio Weber. Unified dynamics for microscopic and macroscopic systems. *Physical Review D*, 34(2):470–491, July 1986. Publisher: American Physical Society.

[174] Alexei Yu. Kitaev. Fault-tolerant quantum computation by anyons. *Annals of Physics*, 303(1):2–30, January 2003.

[175] Fernando Pastawski, Alastair Kay, Norbert Schuch, and Ignacio Cirac. Limitations of Passive Protection of Quantum Information. *Quantum Info. Comput.*, 10(7):580–618, July 2010.

[176] Van Eck phreaking (Wikipedia), November 2019. Page Version ID: 926374915.

[177] Matteo Carlesso, Angelo Bassi, Paolo Falferi, and Andrea Vinante. Experimental bounds on collapse models from gravitational wave detectors. *Physical Review D*, 94(12):124036, December 2016.

[178] Bassam Helou, B.J.J. Slagmolen, David E. McClelland, and Yanbei Chen. LISA pathfinder appreciably constrains collapse models. *Physical Review D*, 95(8):084054, April 2017.

[179] Joshua Morris, Felix A. Pollock, and Kavan Modi. Non-Markovian memory in IBMQX4. *arXiv:1902.07980 [quant-ph]*, February 2019. arXiv: 1902.07980.