

Additivity of the Quantum and Classical Capacities of Quantum Channels

by

Alexander Kazachek

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Applied Mathematics (Quantum Information)

Waterloo, Ontario, Canada, 2025

© Alexander Kazachek 2025

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Quantum channels enable communication through the transmission of quantum states. Quantum Shannon theory investigates these channels, aiming to characterize their capacity for information transmission under various conditions. While this characterization is well-established for classical communication channels, quantum channels exhibit significantly more complex and mathematically intricate behavior, making a complete understanding elusive. A key challenge is the phenomenon of non-additivity, where combining quantum channels can enhance information flow by leveraging quantum effects. In this work, we focus on two types of non-additivity: those of classical capacity and quantum capacity.

We present new constructive counterexamples demonstrating the non-additivity of the minimum output p -Rényi entropy for $p > 2$. These examples achieve non-additivity at lower values of p than previously known constructions of the same dimension. We also show that several plausible generalizations of antisymmetric spaces – such as through alternative symmetries or higher tensor powers – cannot produce non-additivity using current techniques.

Additionally, we advance the study of resonant multilevel amplitude damping channels. We analytically derive their degradability regions, previously inferred using a heuristic assumption supported by numerical evidence, and formulate conjectures on their capacity based on our own numerical evidence. Specifically, we conjecture that their coherent information is optimized on diagonal states and that they are always weakly additive. However, we find that coherent information activation is possible, as strong non-additivity arises in certain regions when combined with erasure channels.

Acknowledgements

My biggest thanks goes to my supervisor, Graeme Smith. Your flexibility throughout these past few years has been invaluable. I greatly attribute not only my academic but also professional success to your understanding and support. I consider myself very fortunate to have ended up in your group.

I am also indebted to Peixue Wu. Without your guidance and direction, I would not have been able to put this thesis together. Thank you for helping me navigate the literature, and for answering the countless trivial questions I asked.

My undergraduate supervisor, Tatyana Barron, has been a tremendous help as well. You played a pivotal role in the beginning of my research journey by taking me on for a summer project. I am grateful you continued to reach out even after I had graduated as well, not only to provide opportunities but also just to check in.

Of course, I extend my love to my family. To my mum and dad, your daily and unconditional support for the last three-and-twenty years has sculpted me into who I am now. Also, to my sister Katya, you have been a pillar for me in so many ways these past few years.

There are so many other people without whom I would not be here today. To name only but a few, I am also thankful for Luis Mazzali, Sina Delsooz, Madhu Gunasingam, Jacob Ender, Sophie Wu, Zachary Mann, Ben Liu, Hyder Butt, Chris Mayer, Helen He, and, of course, Alice Kuo.

Table of Contents

Author's Declaration	ii
Abstract	iii
Acknowledgements	iv
List of Figures	viii
List of Tables	ix
1 Motivation	1
2 Mathematical Background	3
2.1 Functional Analysis	3
2.2 Tensor Products	8
2.3 Quantum States and Channels	10
2.4 Classical Information Theory	17
2.5 From Classical to Quantum Information	18
2.6 Majorization and Convexity	21
2.7 Semidefinite Programming	22
2.8 Irreducible Representations of Symmetric Groups	23
3 Classical Capacity	29
3.1 Holevo Capacity and Minimum Output Entropy	29
3.2 Numerical Evaluation and Bounds	33
3.3 Non-Additivity via Dimensional Bounds on Schmidt Coefficients	36
3.4 Antisymmetric Spaces	40

4	Quantum Capacity	47
4.1	Coherent Information	47
4.2	Degradability	48
4.3	Numerical Evaluation and Bounds	52
4.4	The Platypus Family	56
4.5	Potential Theoretical Frameworks for Non-Additivity	59
4.5.1	Spin Alignment Conjecture	59
4.5.2	ε -Log Singularities	59
4.5.3	Information Leakage	61
4.5.4	Private Channels	61
5	Numerical Searches for Non-Additive Classical Capacity	63
5.1	Numerical Framework and Challenges	63
5.1.1	Optimization of Bounds on Schmidt Coefficients	63
5.1.2	Numerically Constructing Orthonormal Bases	64
5.1.3	Solving for Minimal Non-Additive p	64
5.2	Numerical Experiments	66
5.2.1	Irreducible Representations of Symmetric Groups	67
5.2.2	Maximal Completely Entangled Spaces	71
5.2.3	Extensions of Antisymmetric Spaces	74
6	ReMAD Channel Additivity and Degradability	77
6.1	Full Channel Analysis	78
6.1.1	Degradability	78
6.1.2	Coherent Information	83
6.2	Subchannel Degradability Analysis	85
6.3	Additivity of Coherent Information	89
	References	93

Appendices	101
A Symbolic Study of Subchannels and SDPs for Degradability	102
B Optimization of Coherent Information	107
C Optimization of Schmidt Coefficients and Numerical Procedures Involving Young Tableaux	112

List of Figures

5.1	Descriptions of the irreducible representation W_{nk}^λ of maximal dimension relative to W_{nk} for $n = 2, \dots, 50$. As λ always satisfies $ \lambda = 2$, the left plot shows the ratio λ_2/λ_1 of its two rows. The right plot is the relative dimension within the corresponding ambient space W_{n2}	69
6.1	Degradable and antidegradable regions of the ReMAD channel $\mathcal{N}_{r,s,t}$ respectively in blue and red.	83
6.2	Analysis of the coherent information of the ReMAD channel $\mathcal{N}_{r,s,t}$	85
6.3	Search space for non-additivity of coherent information for $\mathcal{N}_{r,s,t}$. Red indicates points where \mathcal{N}_{12} is degradable. Note that gaps of points are areas missing due to the constraint $1 - s - t \geq 0$. We use a grid resolution of 0.02.	90
6.4	Regions of the ReMAD channel $\mathcal{N}_{r,s,t}$ which exhibit strong non-additivity with the erasure channel $\mathcal{E}^{1/2}$. We define non-additivity as an activation greater than 10^{-3} . Optimization took approximately 3 weeks to perform.	91

List of Tables

5.1 Minimal Schmidt bound μ for which non-additive p -Rényi entropy may still be numerically computed, grouped in terms of relative size c of subspace W . Up to 64-bit precision, μ_{\min} is the smallest value of μ so $\mu^p > 0$ and non-additivity occurs for $p \in [1, 100]$. We also give $\mu_{\max} = 1 - \mu_{\min}$ to account for symmetry. Values of p which appear to violate monotonicity are in bold, indicating the challenges imposed by numerical precision in these edge cases. 66

5.2 Bounds of the Schmidt coefficients for subspaces and bipartitions of $W_{5,3}^\lambda$ with $\lambda = (1, 2, 5)$. For each dimension $3 \leq w \leq 6 = \dim W_{5,3}^\lambda$, we search across all subspaces $W \subseteq W_{5,3}^\lambda$ of dimension w , and across all bipartitions where Bob holds m components. The largest Schmidt coefficient found for each w and m pair is reported as μ_{\max} . We also include the relative dimension c and $N = \binom{6}{w}$, the number of subspaces W . This optimization took approximately 14 hours, with 25 trials of 1000 iterations each for every subspace W , and a learning rate of 10^{-4} 71

5.3 Schmidt bound μ computed for W_{nk}^\otimes . Performed 50 local optimizations running for 5000 maximum iterations for each with a learning rate of 10^{-4} , taking about 7 hours. Also given is the relative dimension $c = \dim W_{nk}^\otimes / k^n$ as well as the maximum Schmidt coefficient $\mu \in [1/2, 1]$ which could result in non-additivity $\mu_{\max} = \sqrt{c}$ for the particular relative dimension. In all cases, $\mu > \mu_{\max}$, meaning no non-additivity occurs. 73

5.4 Non-additive extensions of W_{2k}^\wedge via the inclusion of a single vector from $W_{2k}^\bar{\wedge}$. “Best Index” refers to which vector yielded the corresponding smallest values p_{\min} and μ_{\min} . The total number of non-additive examples found for each choice of k is reported in the last column, and p_{\max} is the largest value of non-additive p among all these examples. For comparison, $p_0(k)$ is given as the previously best known value of p_{\min} . Performed for 10000 iterations and 240 initial starts with a learning rate of 10^{-4} . Searching over a single index took approximately 15 minutes. 76

6.1 Degradable and antidegradable regions of the subchannel \mathcal{N}_{12} . The regions are verified using COROLLARY 2, in particular to ensure the specific edge cases examined in order to have tractable spectra are indeed the only relevant regions. 89

1 Motivation

Quantum channels enable communication between parties through the transmission of quantum states. A central goal of quantum information theory, and more specifically quantum Shannon theory, is to characterize this communication.

The foundational concepts of channels and information originated classically with the seminal work of Shannon [Sha48]. Core concepts in quantum information theory, such as entropy and information, were then developed by way of analogy to these classical definitions. However, a hurdle emerged when trying to port over the definition of channel capacity, a measurement of the maximum reliable transmission rate of a channel in the presence of its inherent noise.

Shannon proposed that capacity should be understood in terms of the asymptotic probability of error-free communication, given an agreed-upon encoding and decoding scheme between the sender and receiver. Formally, for a channel \mathcal{N} whose capacity we wish to define as $F(\mathcal{N})$, we consider messages of increasing length n and evaluate the asymptotic average behavior of an associated quantity $F^{(n)}(\mathcal{N})$:

$$F(\mathcal{N}) = \lim_{n \rightarrow \infty} \frac{1}{n} F^{(n)}(\mathcal{N}).$$

We refer to this as regularizing $F^{(n)}(\mathcal{N})$. For classical channels, the quantity $F^{(n)}(\mathcal{N})$ derived by Shannon is the maximal mutual information, which has the property that $F^{(n)}(\mathcal{N}) = nF^{(1)}(\mathcal{N})$, referred to as the additivity of $F^{(n)}(\mathcal{N})$. Therefore, this regularization is trivial and the capacity is given by a so-called single symbol formula: $F(\mathcal{N}) = F^{(1)}(\mathcal{N})$.

Regularization as a necessary procedure appeared all-but-forgotten until quantum channels came into the fray with the advent of quantum Shannon theory. Then, several different notions of quantum channel capacity were developed, with all of the major ones fitting the framework of typicality and regularization developed by Shannon. However, it was not clear that their associated quantities $F^{(n)}(\mathcal{N})$ were also additive.

Additivity appeared necessary for the capacities $F(\mathcal{N})$ to be tractable, not just analytically but even numerically. And, slowly, counterexamples arose showing that each $F^{(n)}(\mathcal{N})$ of interest was non-additive for some channel \mathcal{N} . As a consequence, knowing the capacity of a quantum channel is generically impossible.

Thus, Quantum Shannon theory is significantly more complicated and elusive than its classical ancestor. However, this also means it is a richer theory since quantum channels can exhibit behavior impossible for their classical counterparts. In this thesis, we will focus on two notions of quantum channel capacity – the classical and quantum capacities.

For the former, we present explicit constructions of channels which exhibit stronger non-additivity than previously known examples. Specifically, these channels have non-additive minimum output p -Rényi entropy for values of p smaller than any other existing examples of the same dimension. We also show that several other potential approaches to finding non-additive channels, namely bipartitions of antisymmetric spaces in large tensor powers as well as irreducible representations of symmetric groups, are unlikely to be fruitful. We do this by proving or providing strong numerical evidence for their inability to be non-additive within current frameworks.

In regards to the latter, we thoroughly study a recently introduced class of channels, the resonant multilevel amplitude damping channels. These channels unify several other channels with exotic information-theoretic properties. We provide a rigorous derivation of their degradability regions, which were known previously but hinged on a numerical heuristic. We also form some conjectures on their capacities due to our own numerical investigations, claiming that their coherent information is weakly additive and also attained on diagonal arguments. Moreover, we present numerical results showing they can exhibit strong non-additivity when combined with erasure channels.

Throughout, we also provide a thorough and self-contained review of quantum channel capacities, including the history of the field and its key results, as well as the current state of research into non-additivity and relevant conjectures. Any theorems and conjectures which are novel are explicitly stated throughout to be so, and are contained in [SECTION 3.4](#), [SECTION 5.2.1](#), [SECTION 5.2.3](#), [SECTION 6.1](#), and [SECTION 6.3](#).

2 Mathematical Background

2.1 Functional Analysis

We start by presenting some basic facts and definitions from functional analysis and linear algebra. Although we will generally be working only with finite-dimensional matrix spaces later, a great deal of the modern literature, particularly for classical capacity, is often presented in the full generality of an abstract Hilbert space. So, we opt for the same presentation here. Several great standard references exist, such as [RS81].

We will let V be a generic complex Hilbert space with norm $\|\cdot\|_V$ and inner product $\langle \cdot, \cdot \rangle_V$. Assume likewise for W, X , and Y . We use the convention that inner products are antilinear in their second component.

We will let $L(V, W)$ represent the set of (bounded!) linear operators $T: V \rightarrow W$. We use T as a generic example and employ the shorthand $L(V, V) = L(V)$. As a reminder, bounded means there is some constant $c \geq 0$ so $\|Tx\|_W \leq c\|x\|_V$ for all $x \in V$. The identity operator on V will be $\mathbb{1}_V$, and if V is finite-dimensional we might write $\mathbb{1}_{|V|}$. Throughout, we will drop subscripts if the context is clear.

Definition 1 (Orthonormal Basis). *We say a collection $e_i \in V$ forms an orthonormal basis if $\langle e_i, e_j \rangle = \delta_{ij}$ and any $x \in V$ can be written as*

$$x = \sum_i \langle x, e_i \rangle e_i.$$

Above, δ_{ij} is the Kronecker delta. The reason for separability is to ensure all our spaces have a countable orthonormal basis. Sometimes the equation above is called Parseval's identity. There is a helpful inequality when working with orthonormal bases.

Theorem 1 (Bessel). *If e_i is an orthonormal basis for V , then*

$$\sum_i |\langle x, e_i \rangle|^2 \leq \|x\|^2$$

for all $x \in V$.

Definition 2 (Operator/Infinity Norm). We define the operator norm of T by

$$\|T\|_\infty = \sup_{x \in V} \frac{\|Tx\|_W}{\|x\|_V}.$$

The operator norm goes by many names. In the event that $V = W$ as sets but $V \neq W$ as Hilbert spaces because they are equipped with different norms (as an example, we will call them the V -norm and W -norm), it sometimes goes by the name V -to- W norm, written $\|T\|_{V \rightarrow W}$. It is often also written $\|T\|_{\text{op}}$, however we opt for the symbol infinity as it may be thought as taking the p -norm (defined below) in the limiting case $p \rightarrow \infty$. We also note the following elementary equalities:

$$\|T\|_\infty = \sup_{\substack{x \in V \\ \|x\|_V=1}} \|Tx\|_W = \sup_{\substack{x \in V \\ \|x\|_V=1}} \frac{|\langle Tx, Tx \rangle_W|}{\|Tx\|_W} = \sup_{\substack{x \in V, y \in W \\ \|x\|_V=1=\|y\|_W}} \langle Tx, y \rangle_W.$$

Definition 3 (p -Norm). Over \mathbb{R}^n and \mathbb{C}^n and $p > 1$ we have the p -norm

$$\|x\|_p = \left(\sum_{i=1}^n x_i^p \right)^{\frac{1}{p}},$$

where x_i is the i -th entry of x . We may extend this definition to the complex or real ℓ^p -spaces for vectors of countable length.

We know the p -norm is only given by an inner product when $p = 2$, where it is the Euclidean inner product. There is a useful inequality for this family of norms.

Theorem 2 (Hölder). If p, q are such that $1/p + 1/q = 1$, then

$$\sum_{i=1}^n |x_i y_i| \leq \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} \left(\sum_{i=1}^n |y_i|^q \right)^{\frac{1}{q}}$$

for over \mathbb{R}^n and \mathbb{C}^n .

Now, we will define some particular classes of operators.

Definition 4 (Self-Adjoint). The adjoint of an operator $T \in L(V, W)$ is an operator $S \in L(W, V)$ satisfying $\langle Tx, y \rangle_W = \langle x, Sy \rangle_V$ for all $x \in V$ and $y \in W$. We say that $T \in L(V)$ is self-adjoint if $S = T$.

Theorem 3 (Riesz-Fréchet). Consider the dual of V , the set of bounded linear functionals: $V^* = L(V, \mathbb{C})$. Then, for every $\varphi \in V^*$, there exists a vector $x \in V$ such that $\varphi(y) = \langle y, x \rangle$ for all $y \in V$.

It is a fact, stemming from Riesz-Fréchet, that the adjoint of a map is not only always defined, but is also unique, and so we use the notation $S = T^*$. Also, since we are assuming our operators are bounded, the notions of symmetric, self-adjoint, and Hermitian all coincide. We will typically call them self-adjoint.

Definition 5 (Unitary). We say $T \in L(V)$ is unitary if $T^*T = \mathbb{1} = TT^*$. We say two operators $S, T \in L(V)$ are unitarily equivalent if $S = UTU^*$ for some unitary U .

Unitary operators are those which preserve angles, in the sense that if $U \in L(V)$ is unitary then $\langle Ux, Uy \rangle = \langle x, y \rangle$ for any $x, y \in V$. We may understand them therefore as not changing the underlying space, but merely rotating it. Several classes of operators we will encounter in quantum information theory are unique only up to unitaries.

Definition 6 (Compact). We say $T \in L(V, W)$ is compact if for every convergent sequence $x_n \in V$, there is a convergent subsequence of $Tx_n \in W$.

Compact operators are essentially as good as working with operators of finite-dimensional range, which are essentially as good as working with matrices. In fact, as mentioned, we will really only be working with matrices later.

Definition 7 (Spectrum). The spectrum of T is

$$\sigma(T) = \{\lambda \in \mathbb{C} : T - \lambda \mathbb{1}_W \notin L(V, W)\}.$$

Theorem 4 (Spectral Theorem). If $T \in L(V)$ is self-adjoint and compact, then $\sigma(T) = \{\lambda_n\} \subseteq \mathbb{R}$ is countable and consists entirely of eigenvalues, and we may write

$$T = \sum_n \lambda_n P_{\lambda_n}$$

where P_{λ_n} is a projection onto the eigenspace of eigenvalue λ_n . Moreover, either $\sigma(T)$ is finite, or $\lambda_n \rightarrow 0$ and the unit eigenvectors form an orthonormal basis of V .

This spectral theorem is crucial for understanding our later definition of quantum states, where these eigenvalues will be the probabilities of measurement outcomes. It is also a good place to motivate and introduce Dirac notation.

Dirac notation is where we represent a vector $x \in V$ in a Hilbert space with the notation of a “ket”: $|x\rangle \in V$. Associated to each vector x is its dual functional $\langle \cdot, x \rangle$ due to Riesz-Fréchet, which we represent with a “bra” $\langle x|$. We may write the inner product of two vectors x and y as

$$\langle y, x \rangle = (\langle \cdot, x \rangle)(y) = \langle x|y\rangle.$$

The projection onto x may be written as

$$P_x = \langle \cdot, x \rangle x = x \langle \cdot, x \rangle = |x\rangle \langle x|.$$

Algebraic manipulation of vectors is often simpler with Dirac notation, and it allows for the omission of indices. For example, we may rewrite the spectral theorem as

$$T = \sum_{\lambda \in \sigma(T)} \lambda |\lambda\rangle \langle \lambda|.$$

In finite-dimensional vector spaces, the kets $|x\rangle$ correspond to column vectors, while the bras $\langle x|$ correspond to the transposed row vectors.

The following result actually holds more generally for bounded normal operators (those which commute with their adjoints). However, we will not prove this theorem in full generality since dropping compactness means we no longer only have to consider eigenvalues, and would instead need to invoke more sophisticated tools (namely Gelfand’s formula). The extension to normal operators simply rephrases the result in terms of singular values, which we will discuss right after.

Proposition 1. *If T is compact and self-adjoint, then $\|T\|_\infty = \max |\sigma(T)|$.*

Proof. By hypothesis we may spectrally decompose T so

$$T = \sum_{\lambda \in \sigma(T)} \lambda |\lambda\rangle \langle \lambda|.$$

Essentially by definition, $\|T\|_\infty \geq |\lambda|$ for all $\lambda \in \sigma(T)$. To get the reverse inequality, for an arbitrary unit vector $x \in V$ we write $x = \sum_\lambda x_\lambda |\lambda\rangle$ so that

$$\|Tx\|^2 = \left\| \sum_\lambda x_\lambda \lambda |\lambda\rangle \right\|^2 = \sum_\lambda |x_\lambda \lambda|^2.$$

Let $\hat{\lambda} = \max_\lambda |\lambda|$. Thus,

$$\sum_\lambda |x_\lambda \lambda|^2 \leq \hat{\lambda}^2 \sum_\lambda |x_\lambda|^2 = \hat{\lambda}^2$$

since x is a unit vector. In all, $\|Tx\| \leq \hat{\lambda}$. However, we know that this inequality is saturated when $x = |\hat{\lambda}\rangle$. So, we have deduced that

$$\|T\|_{\infty} = \sup_{\|x\|=1} \|Tx\| = \hat{\lambda}.$$

as needed. □

We will need to define one more type of operator, both to work with quantum states and also to introduce a norm we will use throughout our study of classical capacity.

Definition 8. We say that $T \in L(V)$ is positive if $\langle Tx, x \rangle \geq 0$ for all $x \in V$.

It is not hard to see that T^*T is positive, since $\langle T^*Tx, x \rangle = \langle Tx, Tx \rangle = \|x\|^2$. When T is further compact, the spectral theorem lets us easily make sense of the operator $\sqrt{T^*T}$ by taking the square roots of its eigenvalues in its spectral decomposition.

Definition 9 (Schatten p -Norm). The Schatten p -norm of a compact T is

$$\|T\|_p = \left(\sum_{s \in \sigma(\sqrt{T^*T})} s^p \right)^{\frac{1}{p}} = \left(\sum_{\lambda \in \sigma(T)} \sqrt{|\lambda|}^p \right)^{\frac{1}{p}}.$$

These values s are often known as the singular values of T , and the notation $|T| = \sqrt{T^*T}$ is often employed as well. We make special note of them as singular values are often used directly in the numerical procedures we will employ. As implied, our T does not need to be compact for this definition to make sense. We can define it for bounded operators in general, if we were to introduce continuous functional calculus to define the square root of an operator.

The Schatten p -norm will be important in the study of classical capacity, as we will find out we can rephrase its non-additivity problem into an inequality involving the Schatten p -norms of some matrices. Accordingly, we will finish this section by introducing our notation for matrix spaces.

Definition 10 (Matrix Space). The space of n -by- m matrices with complex entries will be denoted $M_{nm}(\mathbb{C})$, or simply $M_n(\mathbb{C})$ if $n = m$. For self-adjoint matrices, we write $M_{nm}^*(\mathbb{C})$.

2.2 Tensor Products

We now move on to tensor products, which are the main structural tool used to build spaces in quantum information theory. They allow for us to combine Hilbert spaces with an intricate product-like structure. An excellent reference is [Rom05].

Definition 11 (Tensor Product). *Let e_i and f_j respectively be orthonormal bases of V and W . Then, the tensor product space $V \otimes W$ is given by formal linear combinations $e_i \otimes f_j$ completed under the inner product*

$$\langle e_i \otimes f_j, e_n \otimes f_m \rangle_{V \otimes W} = \langle e_i, e_n \rangle_V \langle f_j, f_m \rangle_W.$$

The tensor product of operators $S \in L(V, X)$ and $T \in L(W, Y)$ is defined by

$$(S \otimes T)(e_i \otimes f_j) = S e_i \otimes T f_j,$$

and $S \otimes T \in L(V \otimes W, X \otimes Y)$. These definitions extend by linearity to non-basis tensors.

When working in finite dimensions we can explicitly realize tensor products by working with Kronecker products, the block matrix

$$S \otimes T = \begin{pmatrix} s_{11}T & \cdots & s_{1n}T \\ \vdots & \ddots & \vdots \\ s_{m1}T & \cdots & s_{mn}T \end{pmatrix},$$

where $S \in M_{mn}(\mathbb{C})$. We would also like to mention that the tensor product can be constructed in a basis-free manner, although it would be inconvenient for us to detail.

Definition 12 (Decomposable). *A tensor $t \in V \otimes W$ is decomposable (or simple) if there exist $x \in V$ and $y \in W$ such that $t = x \otimes y$.*

Decomposability is the concept which gives rise to quantum entanglement, as it is the sense in which two spaces cannot be fully separated. The closest we can get to a canonical separation is in the following result, which defines the coefficients we will study in detail to understand classical capacity.

Theorem 5 (Schmidt Decomposition). *Given any $t \in V \otimes W$, there exist orthonormal bases $e_i \in V$ and $f_i \in W$ such that*

$$t = \sum_i \mu_i(t) e_i \otimes f_i,$$

where $\mu_i(t) \geq 0$. We call these $\mu_i(t)$ Schmidt coefficients and order them such that $\mu_i(t) \geq \mu_{i+1}(t)$.

Now, Dirac notation can also greatly clean up the notation used in tensor products. Above, we could write

$$e_i \otimes f_j = |e_i\rangle \otimes |f_j\rangle = |ij\rangle = |ij\rangle_{VW},$$

where in the last equality we add subscripts to keep track of the spaces involved. When it comes to operators, we can also write them in several suggestive ways:

$$|i\rangle \langle j|_V \otimes |k\rangle \langle \ell|_W = |ik\rangle \langle j\ell|_{VW}.$$

Now that we know how to combine two Hilbert spaces, we show how to remove one of them with a partial trace. We can also perform a transpose on an individual component of the product.

Definition 13 (Partial Trace). *We define the partial trace of system W of an operator $T \in L(V \otimes W, X)$ by*

$$\text{tr}_W T = \sum_i (\mathbb{1}_V \otimes \langle i|_W) T (\mathbb{1}_V \otimes |i\rangle_W)$$

for any orthonormal basis $|i\rangle_W$. Likewise, we may define $\text{tr}_V T$.

Definition 14 (Partial Transpose). *Let $T \in L(V \otimes W, X)$ and fix an orthonormal basis $|i\rangle_W$. Define the transpose operator*

$$\tau_W: L(W) \rightarrow L(W) \text{ by } |i\rangle \langle j|_W \mapsto |j\rangle \langle i|_W.$$

We then define the partial transpose of T on W by $T^{\tau_W} = (\mathbb{1}_V \otimes \tau_W)T$. We similarly define T^{τ_V} .

The partial transpose is the key ingredient in one of the most important results in the study of entanglement [Per96; HHH96]. The operation is subtle, however, since it is in general basis-dependent as implied in its definition.

We will close with a special norm associated with tensor products, which crops up specifically in the context of quantum information.

Definition 15 (Diamond Norm). *Let $T \in L(V, W)$ and define its diamond norm*

$$\|T\|_{\diamond} = \max_{\substack{S \in L(V \otimes V) \\ \|S\|_1 \leq 1}} \|(T \otimes \mathbb{1}_V)S\|_1.$$

2.3 Quantum States and Channels

Now, we begin our discussion of genuine quantum mechanical phenomenon. For references, we suggest [BZ17; NC10]. We will primarily make use of three Hilbert spaces: Alice A , Bob B , and an ambient environment E . We will continue to use W for a generic Hilbert space, typically as an ancilla. However, we will now assume all these spaces are finite-dimensional. Note though that quantum channels are a perfectly fine concept to study in infinite-dimensional Hilbert spaces, see for example [HS13; Haa+21].

Definition 16 (Quantum State). *A quantum state in A is an operator $\rho \in L(V)$ which is positive, self-adjoint, and has unit trace. The set of all states in A is denoted $S(A)$.*

Definition 17 (Pure). *We say a state $\rho \in S(A)$ is pure if it is given by a rank-one orthogonal projection, that is if $\rho = |x\rangle\langle x|$ for some $x \in A$. The set of all pure states in A is $S_1(A)$.*

Definition 18 (Ensemble). *By an ensemble, we mean a collection of states $\rho_n \in S(A)$ and coefficients $p_n \geq 0$ such that $\sum_n p_n = 1$. The state associated to the ensemble is $\sum_n p_n \rho_n$.*

Generically we use $\sigma, \rho \in S(V)$ as quantum states, and point out that the unfortunate clash $\sigma(\rho)$ will never be the multiplication of states but rather the spectrum of ρ . We will also abuse notation and refer to unit vectors themselves as pure states, such as by writing $|x\rangle \in S_1(A)$ when we really mean $|x\rangle\langle x| \in S_1(A)$.

A quantum channel $\mathcal{N}_{A \rightarrow B}$ is a map sending states from Alice to Bob (and we will drop the subscript $A \rightarrow B$ if the spaces are clear). It turns out such maps have simple characterizing features.

Definition 19 (Quantum Channel). *A quantum channel $\mathcal{N}: S(A) \rightarrow S(B)$ is a linear operator $\mathcal{N} \in L(A, B)$ which completely-positive, meaning $\mathbb{1}_k \otimes \mathcal{N} \geq 0$ for all $k \geq 0$, and trace preserving, meaning $\text{tr } \mathcal{N}(\rho) = 1$ for all $\rho \in S(A)$.*

We will use \mathcal{N} for a generic channel, with A and B being its respective input and output spaces. In the context of quantum channels we may revisit the diamond norm, defining a distance metric.

Definition 20 (Diamond Metric). *For two quantum channels $\mathcal{N}_{A \rightarrow B}$ and $\mathcal{M}_{A \rightarrow B}$ we define the diamond metric*

$$d_\diamond(\mathcal{N}, \mathcal{M}) = \max_{\rho \in S(A \otimes A)} \|(\mathcal{N} \otimes \mathbb{1}_A)\rho - (\mathcal{M} \otimes \mathbb{1}_A)\rho\|_1.$$

Unlike other metrics, such as those induced by the operator norm, the diamond metric may in fact be efficiently computed in terms using a semidefinite program, a term we will define and a fact we will revisit later. It also has a very physical meaning, being used in the optimal protocol for channel discrimination.

The following theorem provides a useful operational interpretation of a channel, in that the process of sending a state not only entails behaviour between the spaces of Alice and Bob, but also an ambient environment system. It also allows us to derive two related channels given a starting one.

Theorem 6 (Stinespring). *For every $\mathcal{N}_{A \rightarrow B}$ there exists a so-called Stinespring isometry $V: A \rightarrow B \otimes E$, where E is an ancillary finite-dimensional Hilbert space, such that $\mathcal{N}(\rho) = \text{tr}_E(V\rho V^*)$. Moreover, all isometries of this form induce a quantum channel.*

Returning to our generic channel \mathcal{N} , we will let it have Stinespring isometry V with associated environment E .

Note, though, that several choices of V are possible. Not only do any unitarily equivalent isometries give the same channel, but the size of the environment is also unbounded above. Also, the Stinespring isometry is not really an isometry since its image is often too small in dimension, so the term partial isometry is sometimes used.

Definition 21 (Complementary and Conjugate Channel). *Given a channel $\mathcal{N}_{A \rightarrow B}$ we associate a complementary channel $\mathcal{N}_{A \rightarrow E}$, where E is the Stinespring environment from the isometry $V: A \rightarrow B \otimes E$. We also associate the conjugate channel $\tilde{\mathcal{N}}$, whose Stinespring isometry is \tilde{V} .*

The complementary channel is a natural channel to consider in the context of noise, as it allows us to look at the ambient environment and study what information was lost to it. This is important in the study of the additivity of quantum capacity, in particular, and one of the first instances of their thorough examination is [Kin+07]. Note that since the Stinespring representation is not unique, complementary channels are not either. In regards to the conjugate channel, they are important in constructing counterexamples to the additivity of classical capacity, and we will see them again in [SECTION 3.3](#).

To assist in explicit calculations, there are several operators and matrices related to a quantum channel. In particular, we have the Kraus and Choi-Jamiołkowski operators.

Theorem 7 (Kraus). *There exists a finite collection $K_i \in L(A, B)$ such that $\sum_i K_i^* K_i = \mathbb{1}_V$ and*

$$\mathcal{N}(\rho) = \sum_i K_i \rho K_i^*.$$

We call these K_i Kraus operators, and they are unique up to unitaries.

Actually, Kraus' theorem is more general in that it only requires that \mathcal{N} be trace non-increasing. There is a connection between both Kraus' and Stinespring's theorems (sometimes respectively called the operator-sum and system-environment representations of a quantum channel).

Proposition 2. Let $\{K_i\}_{i=1}^k$ form a set of Kraus operators for $\mathcal{N}_{A \rightarrow B}$. Then, for an ancillary space E of dimension k , form orthonormal basis $|i\rangle_E$ and define $V: A \rightarrow B \otimes E$ via

$$V|x\rangle = \sum_{i=1}^k K_i|x\rangle \otimes |i\rangle_E.$$

Then, V is a valid Stinespring isometry. On the other hand, suppose we have some Stinespring isometry $V: A \rightarrow B \otimes E$. Then, for an orthonormal basis $|i\rangle_E$ define

$$K_i = (\mathbb{1}_B \otimes \langle i|_E)V,$$

which forms a valid set of Kraus operators.

Proof. Let us suppose we start with our Kraus operators and define V . Witness that

$$\begin{aligned} \langle x|V^*V|x\rangle &= \left(\sum_{i=1}^k \langle x|K_i^* \otimes \langle i|_E \right) \left(\sum_{i=1}^k K_i|x\rangle \otimes |i\rangle_E \right) \\ &= \sum_{i=1}^k \langle x|K_i^*K_i|x\rangle \otimes \langle i|i\rangle_E \\ &= \langle x|x\rangle, \end{aligned}$$

where in the last two equalities we used the completeness relation $\sum_i K_i^*K_i = \mathbb{1}_A$ and the orthonormality of our basis for E . In all, V is an isometry. To see it behaves as intended in Stinespring form, we write

$$\text{tr}_E(V|x\rangle\langle x|V^*) = \text{tr}_E\left(\sum_{i,j=1}^k K_i|x\rangle\langle x|K_j^* \otimes |i\rangle\langle j|_E\right).$$

To take the partial trace, we examine the summands

$$(\mathbb{1}_B \otimes \langle \ell|_E) \left(\sum_{i,j=1}^k K_i|x\rangle\langle x|K_j^* \otimes |i\rangle\langle j|_E \right) (\mathbb{1}_B \otimes |\ell\rangle_E) = K_\ell|x\rangle\langle x|K_\ell^*.$$

Summing over ℓ , we recover the Kraus operator form of \mathcal{N} as needed.

In fact, were to start with the Stinespring isometry V we see defining K_i as stated would indeed return the action of our channel via their conjugation due to the above equation. It remains only to check the completeness equation:

$$\sum_{\ell=1}^k (\mathbb{1}_B \otimes \langle \ell|_E)(\mathbb{1}_B \otimes |\ell\rangle_E) = \mathbb{1}_B.$$

So, we indeed have produced a set of Kraus operators. □

Theorem 8 (Choi-Jamiołkowski). *The Choi-Jamiołkowski matrix of a channel is*

$$\mathcal{J}_{\mathcal{N}} = \sum_{i,j} |i\rangle \langle j| \otimes \mathcal{N}(|i\rangle \langle j|) \in L(A \otimes B).$$

In particular, $\mathcal{J}_{\mathcal{N}}$ is positive-semidefinite and $\text{tr}_B(\mathcal{J}_{\mathcal{N}}) = \mathbb{1}_A$.

The importance of the Choi-Jamiołkowski matrix is its relation to the transfer matrix below, which behaves nicely under the composition of channels. This is crucial for the later analysis of the degradability of quantum channels in [SECTION 4.2](#).

Definition 22 (Transfer Matrix). *For a quantum channel \mathcal{N} with Kraus operators K_i , we define its transfer matrix*

$$\mathcal{T}_{\mathcal{N}} = \sum_i \bar{K}_i \otimes K_i.$$

Proposition 3. *Transfer matrices are multiplicative under composition. That is, for two channels $\mathcal{N}_{A \rightarrow B}$ and $\mathcal{M}_{B \rightarrow C}$ we have*

$$\mathcal{T}_{\mathcal{M} \circ \mathcal{N}} = \mathcal{T}_{\mathcal{M}} \mathcal{T}_{\mathcal{N}}.$$

Proof. Let K_i and J_i respectively be the Kraus operators of \mathcal{M} and \mathcal{N} . Then, for an input state $\rho \in S(A)$ we see

$$\mathcal{M} \circ \mathcal{N} = \sum_i J_i \left(\sum_j K_j \rho K_j^* \right) J_i^* = \sum_{i,j} J_i K_j \rho (J_i K_j)^*,$$

and so $J_i K_j$ form a set of Kraus operators of $\mathcal{M} \circ \mathcal{N}$. Therefore,

$$\mathcal{T}_{\mathcal{M} \circ \mathcal{N}} = \sum_{i,j} \overline{J_i K_j} \otimes J_i K_j = \sum_{i,j} (\bar{J}_i \otimes J_i)(\bar{K}_j \otimes K_j) = \mathcal{T}_{\mathcal{M}} \mathcal{T}_{\mathcal{N}},$$

as claimed. □

There is also a nice relationship between the transfer matrix and the Choi-Jamiołkowski matrix given by involution.

Proposition 4. *Define the the involution map*

$$\vartheta: L(A \otimes B) \rightarrow L(A \otimes A, B \otimes B) \text{ by } |i\rangle \langle j| \otimes |r\rangle \langle s| \mapsto |s\rangle \langle j| \otimes |r\rangle \langle i|,$$

extended by linearity from an orthonormal basis. Then, $\mathcal{T}_{\mathcal{N}} = \vartheta(\mathcal{J}_{\mathcal{N}})$.

Proof. Let \mathcal{N} have Kraus operators K_a . We know $K_a \in L(A, B)$ and so we may write them as

$$K_a = \sum_{i,r} \kappa_{ajr} |r\rangle \langle j|$$

for orthonormal bases of A and B . We then have

$$\mathcal{T}_{\mathcal{N}} = \sum_a \bar{K}_a \otimes K_a = \sum_{a,i,j,r,s} \kappa_{air} |r\rangle \langle i| \otimes \kappa_{ajs} |s\rangle \langle j|.$$

We also have

$$\mathcal{J}_{\mathcal{N}} = \sum_{i,j} |i\rangle \langle j| \otimes \mathcal{N}(|i\rangle \langle j|) = \sum_{i,j} |i\rangle \langle j| \otimes \left[\sum_{a,k,r} \kappa_{akr} |r\rangle \langle k| \right] |i\rangle \langle j| \left[\sum_{a,\ell,s} \bar{\kappa}_{b\ell s} |\ell\rangle \langle s| \right].$$

Due to orthonormality, we then have

$$\mathcal{J}_{\mathcal{N}} = \sum_{i,j} |i\rangle \langle j| \otimes \left[\sum_{a,r} \kappa_{air} |r\rangle \right] \sum [\bar{\kappa}_{ajs} \langle s|] = \sum_{a,i,j,r,s} \kappa_{air} \bar{\kappa}_{ajs} |i\rangle \langle j| \otimes |r\rangle \langle s|.$$

Applying the involution yields

$$\vartheta(\mathcal{J}_{\mathcal{N}}) = \sum_{a,i,j,r,s} \kappa_{air} \bar{\kappa}_{ajs} |s\rangle \langle j| \otimes |r\rangle \langle i| = \sum_{a,s,j} \bar{\kappa}_{ajs} |s\rangle \langle j| \otimes \sum_{a,r,i} \kappa_{air} |r\rangle \langle i|,$$

which we see is the desired result. □

We will spend time not just analyzing channels, but also their subchannels. This is as the behavior of a channel overall can be greatly informed by how it performs on a restricted part of its domain.

Definition 23 (Subchannel). Consider $\mathcal{N}_{A \rightarrow B}$ and a subspace $\hat{A} \subseteq A$. Consider then some isometry $\hat{U}: \mathbb{C}^{|\hat{A}|} \rightarrow \hat{A}$ and let $|\hat{i}\rangle$ form the standard basis of $\mathbb{C}^{|\hat{A}|}$. Define $\hat{U}|\hat{i}\rangle = |i\rangle$ and extend this set to a basis of A : $\{|i\rangle\}$. Consider the embedding

$$\iota: S(\mathbb{C}^{|\hat{A}|}) \rightarrow S(A) \text{ by } |\hat{i}\rangle\langle\hat{j}| \mapsto U|\hat{i}\rangle\langle\hat{j}|U^* = |i\rangle\langle j|,$$

extended by linearity. We then call $\hat{\mathcal{N}} = \mathcal{N} \circ \iota$ a subchannel of \mathcal{N} restricted to \hat{A} .

For example, suppose $\mathcal{N}_{\mathbb{C}^3 \rightarrow \mathbb{C}^3} = \mathbb{1}_3$ and we want to consider its subchannel which acts on $|0\rangle$ and $|2\rangle$. That is, our subspace of interest is $\hat{A} = \text{span}\{|0\rangle, |2\rangle\} \subseteq A = \mathbb{C}^3$. One possible isometric embedding U is

$$U \begin{pmatrix} 1 \\ 0 \end{pmatrix} = |0\rangle \text{ and } U \begin{pmatrix} 0 \\ 1 \end{pmatrix} = |2\rangle,$$

so that our subchannel $\hat{\mathcal{N}}$ is the qubit channel

$$\hat{\mathcal{N}} \left(\begin{pmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{pmatrix} \right) = \mathcal{N} \left(\begin{pmatrix} \rho_{00} & 0 & \rho_{01} \\ 0 & 0 & 0 \\ \rho_{10} & 0 & \rho_{11} \end{pmatrix} \right) = \begin{pmatrix} \rho_{00} & 0 & \rho_{01} \\ 0 & 0 & 0 \\ \rho_{10} & 0 & \rho_{11} \end{pmatrix}.$$

For our computational purposes, however, we might notice that our description of \hat{A} is very simple so it is easy to see

$$\begin{aligned} S(\hat{A}) &= \{\rho \in S(\mathbb{C}^3) : \rho_{2i} = \rho_{i1}\} \\ &= \left\{ \begin{pmatrix} \rho_{00} & 0 & \rho_{02} \\ 0 & 0 & 0 \\ \rho_{20} & 0 & \rho_{22} \end{pmatrix} : \rho_{02} = \bar{\rho}_{20}, \rho_{00} + \rho_{22} = 1, \rho_{00} \geq 0, \rho_{22} \geq 0 \right\}. \end{aligned}$$

If we would like, we can then understand our subchannel as an ordinary restriction: $\hat{\mathcal{N}} = \mathcal{N}|_{S(\hat{A})}$. This would give us a qutrit channel

$$\hat{\mathcal{N}} \left(\begin{pmatrix} \rho_{00} & 0 & \rho_{02} \\ 0 & 0 & 0 \\ \rho_{20} & 0 & \rho_{22} \end{pmatrix} \right) = \mathcal{N} \left(\begin{pmatrix} \rho_{00} & 0 & \rho_{02} \\ 0 & 0 & 0 \\ \rho_{20} & 0 & \rho_{22} \end{pmatrix} \right) = \begin{pmatrix} \rho_{00} & 0 & \rho_{02} \\ 0 & 0 & 0 \\ \rho_{20} & 0 & \rho_{22} \end{pmatrix}.$$

For very simple subspaces then we may view these as more-or-less equivalent definitions of $\hat{\mathcal{N}}$. In the event that \hat{A} were not so easy to describe, however, understanding the subchannel formally through an embedding provides a definition more rigorous than “just set some matrix entries to zero”.

Proposition 5. Suppose $\hat{T} \in L(A, B)$ satisfies $\mathcal{J}_{\hat{T}} \geq 0$ and that $\text{tr}_B \mathcal{J}_{\hat{T}}$ is a diagonal binary matrix. That is,

$$\text{tr}_B \mathcal{J}_{\hat{T}} = \text{diag}(d_1, \dots, d_{|A|})$$

where $d_i \in \{0, 1\}$. Then, \hat{T} is a subchannel of a quantum channel.

Proof. Without loss of generality suppose

$$\text{tr}_B (\mathcal{J}_{\hat{T}}) = \text{diag}(\underbrace{1, \dots, 1}_{m \text{ times}}, \underbrace{0, \dots, 0}_{n \text{ times}}).$$

Then, define in the standard basis

$$T: S(\mathbb{C}^{m+n}) \rightarrow S(\mathbb{C}^{m+n}) \text{ by } |i\rangle\langle j| \mapsto \begin{cases} \hat{T}(|i\rangle\langle j|) & 1 \leq i, j \leq m \\ |i\rangle\langle j| & \text{otherwise.} \end{cases}$$

Indeed, we see

$$\begin{aligned} \mathcal{J}_T &= \sum_{i,j} |i\rangle\langle j| \otimes T(|i\rangle\langle j|) \\ &= \sum_{1 \leq i, j \leq m} |i\rangle\langle j| \otimes \hat{T}(|i\rangle\langle j|) + \sum_{m < i \text{ or } n < j} |i\rangle\langle j| \otimes |i\rangle\langle j| \\ &= \mathcal{J}_{\hat{T}} + \sum_{m < i \text{ or } m < j} |i\rangle\langle j| \otimes |i\rangle\langle j|. \end{aligned}$$

From this it is easy to see that $\mathcal{J}_T \geq 0$. In terms of its partial trace, we have

$$\text{tr}_B(\mathcal{J}_T) = \text{tr}_B(\mathcal{J}_{\hat{T}}) + \text{tr}_B \left(\sum_{m < i \text{ or } m < j} |i\rangle\langle j| \otimes |i\rangle\langle j| \right) = \text{tr}_B(\mathcal{J}_{\hat{T}}) + \sum_{i > m} |i\rangle\langle i| = \mathbb{1}_{m+n}.$$

Therefore, we see that \hat{T} is a subchannel of T . \square

The recovery of the parent channel above is not unique, of course. We will frequently make use of this result when we are studying the compositions of channels, and in particular when a channel has a domain of smaller dimension than its codomain.

2.4 Classical Information Theory

As mentioned at the very beginning, classical information theory was established by Shannon in [Sha48] with his key definition: entropy. We give a brief overview of his results here, to motivate our definitions in quantum Shannon theory. For this and the following section, we use [Wil17] as our main reference.

Definition 24 (Shannon Entropy). *Let X be a discrete random variable with finite support and probability masses p_x . Define its Shannon entropy as*

$$H(X) = - \sum_x p_x \log p_x.$$

Entropy may be thought of as measuring how uniform such a probability distribution is, in that its entropy attains the maximal value of $\log n$ (where n is the size of its support) when $p_x = 1/n$ for all x .

Suppose now that we receive messages comprising a string whose symbols are drawn according to X .

Definition 25 (Compression Rate). *By a compression rate R we mean a protocol compressing an n -symbol string to an nR -symbol string (which we call a codeword) and a protocol to decompress the codeword to the original string. Such a protocol need not be perfect and is allowed to fail.*

We say R is achievable if there is an asymptotically vanishing error rate, in that the probability of error goes to 0 as $n \rightarrow \infty$.

Theorem 9 (Shannon Source Coding). *All compression rates $R > H(X)$ are achievable.*

This theorem is easiest to understand in the binary case $n = 2$ where $H(X) \leq 1$ always. In this case, we see that as long as X is not uniform then $H(X) < 1$ and so there is an achievable rate $R < 1$. That is, it is possible to compress an n -symbol string to an nR -symbol codeword string, and $nR < n$. Thus, compression is possible.

Definition 26 (Classical Channel). *A classical channel consists of two discrete random variables with finite support X, Y and is characterized by the conditional distribution $Y|X$, codifying Alice sending a string in X to Bob who receives a string in Y .*

Note that the conditional distribution is sufficient to describe the channel since the distribution of X itself is not terribly important, as Alice can “modify” this distribution (indeed, this is what she does by sending Bob a codeword instead of the original string).

Definition 27 (Conditional Entropy and Mutual Information). *The conditional entropy of X given Y and the mutual information between X and Y are respectively defined as*

$$H(X|Y) = H(XY) - H(Y) \text{ and } I(X : Y) = H(X) - H(X|Y).$$

Conditional entropy answers the following question: If I know Y , what is my remaining ignorance about X ? In the context of compression, we see this lets us encode X given this information with an asymptotic rate of $H(X|Y)$.

The inverse is the answered by mutual information: By knowing Y , how much has my compression burden been reduced? It is precisely the difference between $H(X)$ and $H(X|Y)$, since these are the best achievable rates.

Definition 28 (Communication Rate). *By a communication rate R for a channel given by $Y|X$ we mean a protocol encoding an n -symbol output string in Y to an nR -symbol input codeword in X , and a protocol to decompress the codeword to the original string. Such a protocol need not be perfect and is allowed to fail.*

We say R is achievable if there is an asymptotically vanishing error rate, in that the probability of error goes to 0 as $n \rightarrow \infty$.

Theorem 10 (Shannon Noisy Coding). *Define the capacity of a channel given by $Y|X$ to be the supremum over all achievable rates R , where the optimization is done over all codeword distributions X . Then, channel capacity is given by $\sup_X I(X : Y)$.*

The proofs of Shannon’s source and noisy coding theorems depend on looking at “typical” n -symbol strings, whose complement occurs with some small probability.

2.5 From Classical to Quantum Information

It is straightforward to port the definitions from classical information theory over to the quantum setting. We start with the analogue of entropy, where we will see one of the many reasons why quantum states are required to be positive semidefinite and self-adjoint: it allows us to invoke the spectral theorem and write

$$S(A) \ni \rho = \sum_{\lambda \in \sigma(\rho)} \lambda |\lambda\rangle \langle \lambda|.$$

Definition 29 (von Neumann Entropy). We define the von Neumann entropy

$$S(\rho) = - \sum_{\lambda \in \sigma(\rho)} \lambda \log \lambda,$$

with the convention that $0 \log 0 = 0$.

It is unfortunate that S is now overloaded to be both the set of states and also entropy, however the symbol set of Hilbert spaces and states will never clash. Note too that the convention of handling the kernel is justified as $\lim_{x \rightarrow 0^+} x \log x = 0$. Interestingly, although we write as though von Neumann entropy is a “quantized” version of the Shannon entropy, historically Shannon actually “classicalized” the von Neumann entropy, with the former being introduced in 1948 but the latter in 1927 [Pet01]. We may slightly generalize this quantity, in a way that often makes it easier to study as we recover the von Neumann entropy in the limit.

Definition 30 (p -Rényi Entropy). For $p \in (0, \infty) \setminus \{1\}$ we define

$$S_p(\rho) = \frac{1}{1-p} \log \text{tr}(\rho^p) = \frac{1}{1-p} \log \sum_{\lambda \in \sigma(\rho)} \lambda^p,$$

and set $S_1(\rho) = S(\rho)$.

The definition of S_1 is justified as we do indeed recover the von Neumann entropy in the limit $p \rightarrow 1$. We may also take limits to define S_0 and S_∞ , however these quantities will not be particularly useful for us. Of special interest is the case $p > 1$ since here we may relate the p -Rényi entropy to the Schatten p -norm, readily seen as $\|\rho\|_p = (\text{tr}(|\rho|^p))^{1/p}$. Since states are always positive, this means

$$S_p(\rho) = \frac{p}{1-p} \log \|\rho\|_p.$$

There is a nice representation of how the entropy of a channel acting on a pure state is connected to its Stinespring form and Schmidt decompositions.

Proposition 6. Let $\rho = |x\rangle\langle x|$ be pure and define $|y\rangle = V|x\rangle$ as the output of a Stinespring isometry. Then,

$$S_p(\mathcal{N}(\rho)) = \begin{cases} \frac{1}{1-p} \log \sum_i \mu_i^2(y) & p \neq 1 \\ - \sum_i \mu_i^2(y) \log \mu_i^2(y) & p = 1. \end{cases}$$

Proof. We know $|y\rangle$ is pure and so admits some Schmidt decomposition

$$|y\rangle = \sum_i \mu_i(y) |i\rangle_B \otimes |i\rangle_E.$$

Notice that

$$\mathcal{N}(\rho) = \text{tr}_E |y\rangle \langle y| = \sum_i (\mathbb{1}_B \otimes \langle i|_E) \left(\sum_{j,k} \mu_j(y) \mu_k(y) |j\rangle \langle k|_B \otimes |j\rangle \langle k|_E \right) (\mathbb{1}_B \otimes |i\rangle_E).$$

Due to orthonormality all terms in the summands vanish except for when $i = j = k$, so

$$\mathcal{N}(\rho) = \sum_i \mu_i^2(y) |i\rangle \langle i|_B.$$

Of course, this is diagonal, and so referring to the definition of entropies in terms of spectra, we are done. \square

There is a distance measure used between quantum states that helps characterize a key fact about quantum information: processing your data can only lose information.

Definition 31 (Umegaki Relative Entropy). *The Umegaki relative entropy between two states $\rho, \sigma \in S(A)$ is*

$$S(\rho \parallel \sigma) = \text{tr}(\rho(\log \rho - \log \sigma)).$$

Theorem 11 (Data Processing Inequality). *For any states and channels, $S(\mathcal{N}(\rho) \parallel \mathcal{N}(\sigma)) \leq S(\rho \parallel \sigma)$.*

This notion of relative entropy is inspired by the classical notion of Kullback-Leibler divergence. However, the “correct” way to quantize this quantity is not so straightforward, as many plausible definitions with similar properties exist. The above definition, introduced in [Ume54], happens to be the most natural, as shown by [HP91].

When a state is multipartite, say $\rho_{AB} \in S(A \otimes B)$, we often use the notation for the entropy of a subsystem as follows: $S(A) = S(\rho_A)$. We note this does overload the notation $S(A)$ as both state space and entropy of a subsystem, however their uses will always be clear from context. This notation allows us to immediately port over conditional entropy and mutual information by using von Neumann instead of Shannon entropy. For instance,

$$I(A : B) = S(A) + S(B) - S(AB) = S(\rho_A) + S(\rho_B) - S(\rho_{AB}).$$

Proposition 7. For all states $\rho \in S(A \otimes B)$,

$$I(A : B) = S(\rho_{AB} \| \rho_A \otimes \rho_B).$$

Proof. We see that

$$S(\rho_{AB} \| \rho_A \otimes \rho_B) = \text{tr}(\rho_{AB} \log \rho_{AB}) - \text{tr}(\rho_{AB} \log(\rho_A \otimes \rho_B)).$$

Diagonalize our state as $\rho_A = \text{diag}(\lambda_1, \dots, \lambda_n)$. Then,

$$\log(\rho_A \otimes \rho_B) = \log \text{diag}(\lambda_1 \rho_B, \dots, \lambda_n \rho_B),$$

seen via the Kronecker product. From here, we see

$$\begin{aligned} \log \text{diag}(\lambda_1 \rho_B, \dots, \lambda_n \rho_B) &= \text{diag}(n \log \lambda_1 + \log \rho_B, \dots, n \log \lambda_n + \log \rho_B) \\ &= \text{diag}(\log \lambda_1, \dots, \log \lambda_n) \otimes \mathbb{1}_B + \mathbb{1}_A \otimes \log \rho_B \\ &= \log \rho_A \otimes \mathbb{1}_B + \mathbb{1}_A \otimes \log \rho_B. \end{aligned}$$

Thus,

$$\text{tr}(\rho_{AB} \log(\rho_A \otimes \rho_B)) = \text{tr}(\rho_{AB}(\log \rho_A \otimes \mathbb{1}_B)) + \text{tr}(\rho_{AB}(\mathbb{1}_A \otimes \log \rho_B)).$$

Since

$$\text{tr}_B(\rho_{AB}(\log \rho_A \otimes \mathbb{1}_B)) = (\text{tr}_B \rho_{AB}) \log \rho_A = \rho_A \log \rho_A,$$

we see that

$$S(\rho_{AB} \| \rho_A \otimes \rho_B) = -S(AB) + S(A) + S(B),$$

as needed. □

2.6 Majorization and Convexity

Convexity is a concept which arises naturally in quantum information due to its mathematical underpinnings, revolving around concepts like positive matrices and probability vectors. These spaces are often examples of convex sets or cones.

Definition 32 (Majorization). Given some $x \in \mathbb{R}^n$, let x^\downarrow be the vector whose entries are of x , but reordered from largest to smallest. Then, for $x, y \in \mathbb{R}^n$ we say y majorizes x and write $x \leq y$ if

$$\sum_{i=1}^k x_i^\downarrow \leq \sum_{i=1}^k y_i^\downarrow$$

for all $k = 1, \dots, n$, with saturation at $k = n$.

Majorization is essentially a ranking of disorder or entropy between two vectors. For example, if x is a maximally-mixed vector (i.e. $x_i = 1/n$ for all i) then it will be majorized by every other vector. So, majorization is central in characterizing certain types of quantum information processing [Nie99; JKM24].

Definition 33 (Schur-Convex). We say $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is Schur-convex if $f(x) \leq f(y)$ whenever $x \leq y$. If $-f$ is Schur-convex, we say f is Schur-concave.

Lemma 1. The p -Rényi entropy is Schur-concave in its eigenvalues.

Proof. Take two states ρ and σ , letting x and y respectively be the vectors of their eigenvalues, and suppose $x \leq y$. Since these are states we know that $x_i, y_i \in [0, 1]$ for all i , and so for all $p > 1$ we have that

$$\sum_{i=1}^k (x_i^\downarrow)^p \leq \sum_{i=1}^k (y_i^\downarrow)^p$$

for all $1 \leq k \leq n$. Of course, this holds true for $k = n$ in particular, and as the logarithm is monotonic and $1 - p < 0$ we have that

$$S_p(\rho) = \frac{1}{1-p} \log \sum_{i=1}^n (x_i^\downarrow)^p \geq \frac{1}{1-p} \log \sum_{i=1}^n (y_i^\downarrow)^p = S_p(\sigma).$$

For $p < 1$ we note that now that $x_i^p > x_i$, but $1 - p > 0$ and so our inequality still goes as desired. For $p = 1$, it stands only to notice that $x_i \log x_i < x_i$ for all i since the logarithm is negative on our domain, and so the proof follows the same logic as $p > 1$. \square

Lemma 2. Von Neumann entropy is concave in ensembles. That is, given some ensemble $\rho = \sum_x p(x)\rho(x)$ then

$$S(\rho) \geq \sum_x p(x)S(\rho(x)).$$

2.7 Semidefinite Programming

Optimization constraints in quantum information often take the form of matrix inequalities (for instance, being a state imposes constraints on the positivity and trace of a matrix). Therefore, semidefinite programming is abundant, being the area of mathematical optimization purpose-suited for such problems, and conveniently being very computationally tractable. A reference for semidefinite programming with applications specifically for quantum information is [SC23].

Definition 34 (Semidefinite Program). Fix some $c \in \mathbb{R}^n$ and $A_i, B \in L(\mathbb{R}^m)$ be symmetric for $i = 1, \dots, n$. Then, a semidefinite program is an optimization problem of the form

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & c^\top x \\ \text{s.t.} \quad & x_1 A_1 + \dots + x_n A_n - B \geq 0. \end{aligned}$$

We see that this generalizes other optimization problems (for example, we have a linear program if the A_i and B are diagonal). Often we are not concerned with any particular objective function, but merely whether some x exists satisfying the constraints. We call this a feasibility problem and use the more suggestive notation

$$\begin{aligned} \text{find} \quad & x \in \mathbb{R}^n \\ \text{s.t.} \quad & x_1 A_1 + \dots + x_n A_n - B \geq 0, \end{aligned}$$

where we may imagine some hidden constant objective function.

If we have a set of indexed variable with similar constraints, we will sometimes write the constraints as an indexed set too. For example, if above we require that each x_i is positive, we would write

$$\begin{aligned} \text{find} \quad & x \in \mathbb{R}^n \\ \text{s.t.} \quad & x_1 A_1 + \dots + x_n A_n - B \geq 0 \\ & \{x_i \geq 0\}_{i=1}^n. \end{aligned}$$

Several very efficient toolboxes exist for solving semidefinite programs, and we will employ them heavily in our analysis of degradability and coherent information. We use CVX as our solver within MATLAB [GB08; GB14].

2.8 Irreducible Representations of Symmetric Groups

All of the spaces we will look at in our forthcoming discussion of classical capacity are related to representation theory. In particular, we will use representations of symmetric groups S_n , where for reference we cite [Ful97; FH04].

Definition 35 (Representation). Let G be a group. We say that $r: G \rightarrow L(V)$ is a representation of G if $r(g_1 g_2) = r(g_1) r(g_2)$ for all $g_1, g_2 \in G$. Given some $W \subseteq V$ such that $r(g)w \in W$ for all $g \in G$ and $w \in W$, we call $r|_W: r^{-1}(W) \rightarrow L(W)$ a subrepresentation of V . If no non-trivial subrepresentations exist, then we say r is irreducible.

We will work with the tensor product space

$$W_{nk} = \underbrace{\mathbf{C}^k \otimes \cdots \otimes \mathbf{C}^k}_{n \text{ times}}$$

using the representation $r: S_n \rightarrow L(W_{nk})$ defined by

$$r(\sigma): e_{i_1} \otimes \cdots \otimes e_{i_n} \mapsto e_{i_{\sigma(1)}} \otimes \cdots \otimes e_{i_{\sigma(n)}}.$$

It is important to note here that σ acts on the order of the tensors (i_1, \dots, i_n) , and not of the basis $(1, \dots, k)$. For example, if $\sigma = \begin{pmatrix} 1 & 3 & 4 & 2 \end{pmatrix} \in S_4$ and $k = 5$, then

$$r(\sigma)(e_5 \otimes e_1 \otimes e_2 \otimes e_1) = e_2 \otimes e_5 \otimes e_1 \otimes e_1.$$

Definition 36 (Partition). We say $\lambda = (\lambda_1, \dots, \lambda_m)$ is a partition of n , written $\lambda \vdash n$, if $1 \leq \lambda_{i+1} \leq \lambda_i \leq n$ for all $i = 1, \dots, m-1$ and $\sum_i \lambda_i = n$. We say that $|\lambda| = m$.

These partitions are intimately related to S_n and its irreducible representations. We will use these as the starting point for our search for non-additive subspaces.

Theorem 12. Let Λ be the set of partitions $\lambda \vdash n$. Then, there is a bijection between Λ and irreducible representations of S_n .

We will now get into how to explicitly construct the irreducible representations, which we will denote by S_n^λ in accordance their corresponding to partitions.

To start, we introduce the concept of Young diagrams which allow us to count the dimension of these subspaces. These are diagrams of squares ordered side-by-side, which we may view as an alternative representation of a binary matrix with non-increasing rows. For instance,

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \cong \begin{array}{|c|c|c|c|} \hline \square & \square & \square & \square \\ \hline \square & \square & & \\ \hline \square & & & \\ \hline \end{array}.$$

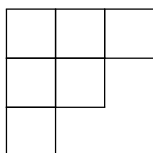
We may index a Young diagram the same way we index a matrix, by its i -th row and j -th column. Since there is an exact correspondence between partitions and diagrams, we will often interchange our use of λ to mean a partition or a diagram.

Definition 37 (Young Diagram). Given a partition λ we define its Young diagram to be the diagram whose i -th row consists of $\lambda_i - \lambda_{i-1}$ boxes.

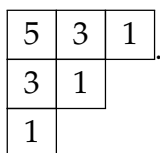
Definition 38 (Hook-Length). *Given a Young diagram, we say the hook-length of its (i, j) -th entry $h(i, j)$ is the number of boxes to its right and also below, plus one.*

We now pause to make a remark on the convention for drawing Young diagrams. We have defined our partitions such that they are non-increasing, and therefore draw our diagrams with the longest rows at the top. We could alternatively define partitions to be non-decreasing, and have the longest rows at the bottom. These are equivalent as diagrams, however we would have to change our definition of hook-length in order to still compute the dimension correctly.

For instance with $n = 6$ and the partition $\lambda = (3, 2, 1)$ the associated Young tableau is



Below we inscribe the diagram's hook-lengths:



Note that Young diagrams with entries inscribed are also referred to as Young tableaux, which we will introduce in a moment. The importance of hook-lengths is established in the following result.

Theorem 13. *The dimension of the irreducible representation S_n^λ is given by*

$$H(\lambda) = \frac{n!}{\prod_{(i,j)} h(i, j)}.$$

Definition 39 (Young Tableaux). *A Young tableau is a Young diagram $\lambda \vdash n$ where each box holds a single number between 1 and n with no repeats. We call a tableau standard if it is strictly increasing in both the rows and columns.*

Permutations naturally act on tableaux. For instance with the permutation $\sigma = (2\ 5\ 1)(4\ 6) \in S_6$, we may act on the following non-standard tableau:

$$\sigma \left(\begin{array}{|c|c|c|} \hline 1 & 5 & 2 \\ \hline 4 & 6 & \\ \hline 3 & & \\ \hline \end{array} \right) = \begin{array}{|c|c|c|} \hline 2 & 1 & 5 \\ \hline 6 & 4 & \\ \hline 3 & & \\ \hline \end{array}.$$

Definition 40 (Row and Column Group). Let $\sigma \in S_n$. We say that σ is row-preserving for λ if each row of $\sigma(\lambda)$ contains exactly the same entries as λ , up to permutation. We write $R(\lambda)$ for the set of all row-preserving permutations. Likewise, we define column-preserving and $C(\lambda)$.

In our prior example our permutation is row-preserving but not column-preserving.

Definition 41 (Young Symmetrizer). Let us define the row symmetrizer and column symmetrizer of λ by

$$r_\lambda = \sum_{\sigma \in R(\lambda)} \sigma \text{ and } c_\lambda = \sum_{\sigma \in C(\lambda)} \text{sgn}(\sigma)\sigma,$$

which we understand as acting on W_{nk} . We then define the Young symmetrizer of λ as the product

$$y_\lambda = r_\lambda c_\lambda = \sum_{\substack{\sigma \in R(\lambda) \\ \tau \in C(\lambda)}} \text{sgn}(\tau)\sigma\tau.$$

Let us return to our example. We see that

$$R(\lambda) = \{ \mathbb{1}, (1\ 2\ 5), (1\ 5\ 2), (4\ 6), (1\ 2\ 5)(4\ 6), (1\ 5\ 2)(4\ 6) \},$$

where $\mathbb{1}$ is the empty permutation. We also have

$$C(\lambda) = \{ \mathbb{1}, (1\ 3\ 4), (1\ 4\ 3), (5\ 6), (1\ 3\ 4)(5\ 6), (1\ 4\ 3)(5\ 6) \}.$$

From here, we simply take products of permutations to get the Young symmetrizer. It is important to note that the convention for the composition of permutations is to go left-to-right (unlike traditional function composition which is right-to-left). That is, if $\sigma = (1\ 3)$ and $\tau = (1\ 2)$, then $\sigma\tau = (1\ 2\ 3)$.

The last point to clear up is that many standard tableaux may exist for a single diagram. For instance, the following two tableaux are both standard:

1	2	5
3	6	
4		

1	2	4
3	5	
6		

Let us now recall the tensor product structure of W_{nk} . What the row and column symmetrizers aim to capture is a subspace which enforces a particular set of symmetries on the tensor indices (and this is why we require $k = |\lambda|$, so the symmetries are fully enforced with no “extraneous” tensors). For this reason, in the following theorem, we may take any standard tableau, as the symmetries enforced by the various other tableaux will be the same, just with a reshuffling of our tensor powers. For a canonical standard tableau we may simply fill the entries from 1 to n left-to-right and top-to-bottom.

Theorem 14. *Let $k = |\lambda|$. Then, the vector space $\text{span}(y_\lambda(W_{nk})) = W_{nk}^\lambda$ is isomorphic to an irreducible representation of S_n with dimension $H(\lambda)$.*

To illustrate two simple examples we will consider the partitions $\lambda = (n)$ and $\lambda^\top = (1, 1, \dots, 1)$. The Young diagram corresponding to the first is a single row of n boxes, while to the latter is a single column of n boxes. Inscripting the entries however we wish, it is clear that $R(\lambda) = S_n$, since every permutation will preserve its singular row. On the other hand, only the identity permutation e will preserve all of the columns simultaneously, so $C(\lambda) = \{e\}$. For identical reasons, $R(\lambda^\top) = \{e\}$ and $C(\lambda^\top) = S_n$.

For our first partition, we therefore see that

$$y_\lambda = r_\lambda c_\lambda = \sum_{\sigma \in R(\lambda)} \sigma,$$

since $\text{sgn}(e) = 1$. Notice that y_λ is equivalent symmetrizing a tensor, as $\sigma(y_\lambda w) = y_\lambda w$ for any $\sigma \in S_n$ and $w \in W_{nk}$. For example, if $n = 2$ then

$$y_\lambda(e_1 \otimes e_2) = e_1 \otimes e_2 + e_2 \otimes e_1.$$

Recalling also that we must have $k = |\lambda| = 1$, we see

$$W_{nk}^\lambda = \text{span}(y_\lambda(W_{nk})) = \text{span}(y_\lambda\{e_1 \otimes \dots \otimes e_1\}) = \{e_1 \otimes \dots \otimes e_1\}.$$

This trivial subspace corresponds to the fact that S_n acts trivially on symmetric tensors.

On the other hand, for λ^\top we have

$$y_{\lambda^\top} = \sum_{\sigma \in S_n} \text{sgn}(\sigma) \sigma,$$

Back to our example with $n = 2$, we have

$$y_{\lambda^\top}(e_1 \otimes e_2) = e_1 \otimes e_2 - e_2 \otimes e_1.$$

This antisymmetrizes a tensor, in the sense that $\sigma(y_{\lambda^\top} w) = \text{sgn}(\sigma) y_{\lambda^\top} w$ for all $\sigma \in S_n$ and $w \in W_{nk}$. We have

$$W_{nk}^\lambda = \text{span} \left\{ \sum_{\sigma \in S_n} \text{sgn}(\sigma) e_{i_{\sigma(1)}} \otimes \cdots \otimes e_{i_{\sigma(n)}} : i_j \neq i_k \text{ for } j \neq k \right\}.$$

We show that matching indices are forbidden in [PROPOSITION 12](#) in [SECTION 3.4](#), where we discuss antisymmetric spaces in more detail. From this fact, it is also clear that we do indeed require $k = n$, for otherwise we would be pigeonholed into repeated indices.

3 Classical Capacity

Classical capacity is concerned with using a quantum channel to communicate classical bits. On his end, Bob may perform measurements and ascribe classical information to their outcomes. Alice, on the other hand, has control of the ensembles she passes to Bob through the channel. The optimal protocol then corresponds to dually optimizing Bob to create the best set of measurements for a given distribution, and Alice to find the best ensembles to send for Bob to optimize. There is no sole reference for this section, however we note that [AS17] provides an excellent overview of much of the material.

3.1 Holevo Capacity and Minimum Output Entropy

For classical capacity, the quantity which is to be regularized is known as the Holevo chi.

Definition 42 (Holevo Chi). *Let $\rho = \sum_x p(x)\rho(x)$ be an ensemble representation. Then, we define the Holevo chi as*

$$\chi(\rho) = S(\rho) - \sum_x p(x)S(\rho(x)).$$

We define the Holevo capacity of a channel by

$$\chi(\mathcal{N}) = \sup_{\rho \in \mathcal{S}(A)} \chi(\mathcal{N}(\rho)).$$

The Holevo chi was originally introduced in [Hol73]. Later work by [SW97] culminated in the following theorem, establishing the Holevo chi as the regularizing quantity for classical capacity.

Theorem 15 (Holevo-Schumacher-Westmoreland). *The classical capacity of a quantum channel \mathcal{N} is given by regularizing the Holevo capacity:*

$$C(\mathcal{N}) = \lim_{n \rightarrow \infty} \frac{1}{n} \chi(\mathcal{N}^{\otimes n}).$$

Holevo capacity is indirectly known to non-additive. Specifically, [Sho04] established that its additivity is equivalent to the additivity of another quantity, the minimum output von Neumann entropy, which is the quantity whose additivity is directly studied.

Definition 43 (Minimum Output p -Entropy). *The minimum output p -Rényi entropy of a channel is*

$$S_p^{\min}(\mathcal{N}) = \inf_{\rho \in S(A)} S_p(\mathcal{N}(\rho)).$$

By additivity, we mean the equality

$$S_p^{\min}(\mathcal{N} \otimes \mathcal{N}) \stackrel{?}{=} 2S_p^{\min}(\mathcal{N}).$$

Recall that von Neumann entropy is the same as 1-Rényi entropy. This conjecture was slowly wittled away by first showing non-additivity of p -entropy for $p > 1$ [HW08], and later the $p = 1$ case [Has09]. Unfortunately, all proofs for the von Neumann case are non-constructive probabilistic arguments, and therefore no actual examples of a channel with non-additive Holevo capacity are presently known.

The probabilistic nature of these arguments was later formalized into statements about concentration of measure [BH10]. Later still, this was rephrased into statements about random convex bodies, connecting quantum information theory with asymptotic geometric analysis [ASW10]. This connection arises from Stinespring dilation, as presented here.

Proposition 8. *Consider the image of the Stinespring isometry $V(A) = W \subseteq B \otimes E$ associated to \mathcal{N} . Then, for all $p > 1$,*

$$S_p^{\min}(\mathcal{N}) = \min_{\substack{y \in W \\ \|y\|=1}} S_p(\text{tr}_E |y\rangle \langle y|).$$

Proof. Diagonalizing our input states, so that $S(A) \ni \rho = \sum_i p_i |i\rangle \langle i|$, we see

$$\begin{aligned} S_p(\mathcal{N}(\rho)) &= S_p\left(\mathcal{N}\left(\sum_i p_i |i\rangle \langle i|\right)\right) \\ &= \frac{p}{1-p} \log \left\| \mathcal{N}\left(\sum_i p_i |i\rangle \langle i|\right) \right\|_p \\ &\geq \frac{p}{1-p} \log \left(\sum_i p_i \|\mathcal{N}(|i\rangle \langle i|)\|_p \right) \end{aligned}$$

due to convexity of the ℓ^p -norm and $p > 1$. From here, concavity of the logarithm means

$$\inf_{\rho \in S(A)} S_p(\mathcal{N}(\rho)) \geq \inf_{\substack{i \in A \\ \langle i|i \rangle = 1}} \frac{p}{1-p} \log \|\mathcal{N}(|i\rangle \langle i|)\|_p = \min_{\substack{i \in A \\ \langle i|i \rangle = 1}} S_p(\mathcal{N}(|i\rangle \langle i|)).$$

We know a minimum will exist as we are now optimizing over a compact set. From here, we rewrite our channel in its Stinespring form so

$$S_p(\mathcal{N}(|i\rangle\langle i|)) = S_p(\text{tr}_E V |i\rangle\langle i| V^*).$$

However, V is an isometry, so we know $V|i\rangle$ is itself a unit vector again and is bijective onto its image. So, we may switch our optimization set accordingly to

$$\min_{\substack{i \in A \\ \langle i|i \rangle = 1}} S_p(\mathcal{N}(|i\rangle\langle i|)) = \min_{\substack{y \in W \\ \langle y|y \rangle = 1}} S_p(\text{tr}_E |y\rangle\langle y|),$$

as needed. □

In this way, studying the minimum output entropy of \mathcal{N} is equivalent to studying the Schmidt coefficients of unit vectors in $V(A)$, which is simply some $|A|$ -dimensional subspace of $B \otimes E$. By studying random subspaces of this form, tools from asymptotic geometric analysis may be borrowed. For similar reasoning, random matrix theory and free probability theory have also recently been employed [BCN16].

Unfortunately, all such approaches only provide enough machinery to prove a very particular statement involving a channel and its conjugate:

$$S_p^{\min}(\mathcal{N} \otimes \bar{\mathcal{N}}) \leq S_p((\mathcal{N} \otimes \bar{\mathcal{N}})|\Phi\rangle\langle\Phi|) < 2S_p^{\min}(\mathcal{N}),$$

where

$$|\Phi\rangle = \frac{1}{\sqrt{|A|}} \sum_{i=1}^{|A|} |i\rangle \otimes |i\rangle$$

is the maximally-entangled state. The reason the conjugate is used is because it does not change the output entropy.

Lemma 3. *A channel and its conjugate have the same minimum output p -entropy:*

$$S_p^{\min}(\mathcal{N}) = S_p^{\min}(\bar{\mathcal{N}}).$$

Proof. Due to [PROPOSITION 2](#), we know that given V the operators

$$K_i = (1_B \otimes \langle i|_E)V$$

form a set of Kraus operators for \mathcal{N} , where we take $|i\rangle_E$ to be the computational basis in which we are taking the complex conjugation of V . Since $\overline{|i\rangle_E} = |i\rangle_E$, it is clear that \bar{K}_i are therefore a set of Kraus operators for $\bar{\mathcal{N}}$. Now, given an arbitrary state $\rho \in S(A)$ we see

$$\mathcal{N}(\rho) = \sum_i K_i \rho K_i^* = \overline{\sum_i \bar{K}_i \bar{\rho} \bar{K}_i^*} = \overline{\bar{\mathcal{N}}(\bar{\rho})}.$$

In particular, $S_p(\mathcal{N}(\rho)) = S_p(\overline{\bar{\mathcal{N}}(\bar{\rho})})$. Since the complex conjugate of a state is another state with the same eigenvalues, we further know that $S_p(\mathcal{N}(\rho)) = S_p(\bar{\mathcal{N}}(\bar{\rho}))$. Since $\bar{\rho}$ is again just some state, when optimizing over all states we then see

$$\inf_{\rho \in S(A)} S_p(\bar{\mathcal{N}}(\bar{\rho})) = \inf_{\rho \in S(A)} S_p(\bar{\mathcal{N}}(\rho)) = S_p^{\min}(\bar{\mathcal{N}}),$$

however we know too the leftmost term is equal to $S_p^{\min}(\mathcal{N})$, precisely as needed. \square

Within this framework the lowest known dimension where non-additivity occurs is when $|B| = 183$, and such a violation almost surely does not occur when $|B| \leq 182$ [BCN16]. Therefore, such arguments are condemned to stay probabilistic since it is impossible to perform exhaustive search over such a large dimension. This intractability is exacerbated by the fact that although explicit constraints on the environment dimension do not currently exist, it is generally believed $|E| \geq 2|B|$ is necessary.

Although probabilistic arguments may demonstrate harsh lower bounds for the channel-conjugate pair, they do not disprove the existence of non-additive channels in low dimensions. Research is active in search of such channels by considering clever constructions of subspaces $W \subseteq B \otimes E$, drawing inspiration from quantum group theory [BC18] and representation theory [MSH18]. Recently, explicit low-dimensional counterexamples for $p > 2$ have been found [GHP10; SS24], and we continue this line of work. Examples near $p \approx 0$ also exist [Cub+08].

In SECTION 5.2 we explore several subspace constructions, some of which are novel and others which generalize prior work. One approach is to build W within a larger ambient space consisting of more than two tensor products, which could allow for more intricate subspace construction. Recall our multipartite tensor product

$$W_{nk} = \underbrace{\mathbb{C}^k \otimes \dots \otimes \mathbb{C}^k}_{n \text{ times}}.$$

In order to then extract B and E from here, we require a bipartition.

Definition 44 (Bipartition). Consider some $W \subseteq W_{nk}$ for $n \geq 3$, and define index sets $P, Q \subseteq \{1, \dots, n\}$ such that $P \cup Q = \{1, \dots, n\}$. With $m = |P|$, by a bipartition of Bob into m components we mean the identification

$$W \ni w_1 \otimes \dots \otimes w_n \cong (w_{i_1} \otimes \dots \otimes w_{i_m}) \otimes (w_{i_{m+1}} \otimes \dots \otimes w_{i_n}) \in (\mathbb{C}^k)^{\otimes m} \otimes (\mathbb{C}^k)^{\otimes (n-m)},$$

and we correspondingly write $W \subseteq B \otimes E$. By $W_{nk}^{(m)}$ we mean a bipartition $W_{nk} = B \otimes E$ where Bob holds m components.

We will see that bipartitions are critical in describing the behavior of W as channel. Genuinely different non-additive behaviors can arise depending on the choice of bipartition, visible in the content of our later [THEOREM 18](#).

The necessity of bipartitioning stems from the fact that output entropy reduces, as we have seen, to the evaluation of Schmidt coefficients. No canonical generalization of the Schmidt decomposition exists to more than two tensor powers. Ideas to generalize it have been approached from a quantum perspective, to find decompositions with nice entropic behavior, of which a few examples are [\[CHS00; GF15\]](#). Multilinear algebra perspectives have been taken as well, such as using tensor decomposibility [\[Rob16\]](#) or higher-order singular value decompositions [\[Ana+14\]](#). We make a brief mention of this as bypassing bipartitions and directly using these larger decompositions could allow for a study of other types of output entropies or related quantities, for future investigation.

3.2 Numerical Evaluation and Bounds

Numerical searches must be performed on our subspaces $W \subseteq B \otimes E$ to evaluate output entropy. The first method is very straightforward: directly compute it.

Proposition 9. Let $B = \mathbb{C}^k$ and $E = \mathbb{C}^d$. For all $p > 1$ and a channel \mathcal{N} associated with the subspace $W \subseteq B \otimes E$ we have

$$S_p^{\min}(\mathcal{N}) = \frac{2p}{1-p} \log \sup_{\substack{X \in M_{kd}(\mathbb{C}) \\ \|X\|_2=1 \\ X \in W}} \|X\|_{2p}.$$

Proof. Take some unit $w \in W$ and let $|i\rangle_B$ and $|i\rangle_E$ respectively form orthonormal bases of \mathbb{C}^k and \mathbb{C}^d . Let us write $w = \sum_{i,j} \lambda_{ij} |ij\rangle_{BE}$, which we may represent lexicographically

via the matrix $X_w = (\lambda_{ij})$. Pushing around Dirac notation shows

$$\begin{aligned}
\mathrm{tr}_E |w\rangle\langle w| &= \mathrm{tr}_E \left[\left(\sum_{i,j} \lambda_{ij} |i\rangle_B \otimes |j\rangle_E \right) \left(\sum_{i,j} \bar{\lambda}_{ij} \langle i|_B \otimes \langle j|_E \right) \right] \\
&= \mathrm{tr}_E \left[\sum_{i,j,k,\ell} \lambda_{ij} \bar{\lambda}_{k\ell} |i\rangle \langle k|_B \otimes |j\rangle \langle \ell|_E \right] \\
&= \sum_{i,j,k,\ell} \lambda_{ij} \bar{\lambda}_{k\ell} \mathrm{tr} (|j\rangle \langle \ell|_E) |i\rangle \langle k|_B \\
&= \sum_{i,j,k} \lambda_{ij} \bar{\lambda}_{kj} |i\rangle \langle k|.
\end{aligned}$$

From here it is easy to see that $\mathrm{tr}_E |w\rangle\langle w| = X_w X_w^*$ as a matrix. Therefore,

$$S_p(\mathrm{tr}_E |w\rangle\langle w|) = \frac{1}{1-p} \log \mathrm{tr} ((X_w X_w^*)^p) = \frac{1}{1-p} \log \mathrm{tr} (|X_w|^{2p}),$$

as trace is cyclic. Invoking the definition of the Schatten norm and noting that for $p > 1$ entropy decreases in the logarithm, we get

$$S_p^{\min}(\mathcal{N}) = \frac{2p}{1-p} \log \sup_{\|w\|=1} \|X_w\|_{2p}.$$

To close, we note that for all w with $\|w\| = \|w\|_2 = 1$ we necessarily have $\|X_w\|_2 = 1$, and that it is sufficient to only look at unit vectors w due to [PROPOSITION 8](#). \square

Of course, this method is only straightforward mathematically. In reality, we face several hiccoughs. First, our constraint $X \in W$ which imposes the structure of our tensor product has no reason to be generically well-behaved (such as given by matrix inequalities). Due to this we have a few choices: we can either use a global optimizer, which is computationally intractable, or a stochastic optimizer which is not guaranteed to find the global maximum required to actually realize our output entropy.

In terms of directly handling the capacity itself, we can therefore only hope for bounds. A lower bound is trivial to attain: the Holevo capacity on a single symbol. This is straightforward to implement as channels themselves are simply linear maps, and entropies are easy to compute. However, it is important to note that in contrast to

minimum output entropy, it is insufficient to simply look at pure states as their Holevo capacity always vanishes. However, we may look over ensembles of pure states, with a proof we reproduce from [SW97].

Proposition 10. *Let $\rho = \sum_x p(x)\rho(x)$ be an ensemble representation. Then, for any \mathcal{N} there is another ensemble $\sigma = \sum_y q(y)\sigma(y)$ such that each $\sigma(y)$ is pure and $\chi(\mathcal{N}(\rho)) \leq \chi(\mathcal{N}(\sigma))$.*

Proof. Our first step is to write the spectral decompositions $\rho(x) = \sum_i \lambda_i(x) |\lambda_i(x)\rangle \langle \lambda_i(x)|$. Denote the new ensemble $\sigma = \sum_{i,x} q(x,i)\sigma(x,i)$ with $q(x,i) = p(x)\lambda_i(x)$ and $\sigma(x,i) = |\lambda_i(x)\rangle \langle \lambda_i(x)|$. Note that $\sigma = \rho$ and therefore $S(\mathcal{N}(\rho)) = S(\mathcal{N}(\sigma))$. Now, notice that

$$S(\mathcal{N}(\rho(x))) = S\left(\sum_i \mathcal{N}(\lambda_i(x)\sigma(x,i))\right) \geq \sum_i \lambda_i(x) S(\mathcal{N}(\sigma(x,i)))$$

due to the linearity of \mathcal{N} and the concavity of S in ensembles from LEMMA 2. In all,

$$\begin{aligned} \chi(\mathcal{N}(\sigma)) &= S(\mathcal{N}(\sigma)) - \sum_{x,i} q(x,i) S(\mathcal{N}(\sigma(x,i))) \\ &= S(\mathcal{N}(\rho)) - \sum_{x,i} p(x)\lambda_i(x) S(\mathcal{N}(\sigma(x,i))) \\ &\geq S(\mathcal{N}(\rho)) - \sum_{x,i} p(x) S(\mathcal{N}(\rho(x))) \\ &= \chi(\mathcal{N}(\rho)), \end{aligned}$$

as claimed. □

So, mixed ensembles can be further improved by pure ensembles. There is also a separate result from [Dav78] concerning the maximum number of codewords necessary to optimize information-theoretic quantities. In all, our optimization can be reduced to ensembles of $|A|^2$ -many pure states.

With a lower bound addressed, we close by presenting one of the two upper-bounds given in [WXD18].

Theorem 16. *Let $\mathcal{N}_{A \rightarrow B}$ have $|A| = a$ and $|B| = b$. Define $\beta(\mathcal{N})$ to be the solution to the SDP*

$$\begin{aligned} \min_{\substack{S_B \in M_b^*(\mathbb{C}) \\ R_{AB} \in M_{ab}^*(\mathbb{C})}} \quad & \text{tr } S_B \\ \text{s.t.} \quad & -R_{AB} \leq \mathcal{J}_{\mathcal{N}}^{\top B} \leq R_{AB} \\ & -\mathbb{1}_A \otimes S_B \leq R_{AB}^{\top B} \leq \mathbb{1}_A \otimes S_B. \end{aligned}$$

Then, $C(\mathcal{N}) \leq \log \beta(\mathcal{N})$.

This is actually an upper-bound on another type of capacity as well, one related to assisting classical communication via some additional codes on top of the channel. The quantity β is additive as well, meaning this SDP provides a multiplicative upper bound on $C(\mathcal{N})$.

3.3 Non-Additivity via Dimensional Bounds on Schmidt Coefficients

Fortunately, we do not need to necessarily evaluate classical capacity in order to find non-additive behavior, as we can chain together some inequalities which yield comparably easier optimization problems. We adhere to [SS24] for our general presentation.

First, we recall our channel-conjugate pair, and have the following general aim: find an upper bound $c \geq S_p^{\min}(\mathcal{N} \otimes \bar{\mathcal{N}})$ and lower bounds $C \leq S_p^{\min}(\mathcal{N})$ and $C' \leq S_p^{\min}(\bar{\mathcal{N}})$. If $c < C + C'$, then we have

$$S_p^{\min}(\mathcal{N} \otimes \bar{\mathcal{N}}) \leq c < C + C' \leq S_p^{\min}(\mathcal{N}) + S_p^{\min}(\bar{\mathcal{N}}).$$

Ideally, all of c, C, C' are functions of p so that we may take $p \rightarrow 1$ and approach the von Neumann entropy case. Moreover, we would like them to depend on the dimension of \mathcal{N} (or rather, of Alice's input space A) so that we may find low-dimensional examples, or at the very least derive some sort of dimensional bounds. Since we already know conjugating a channel does not change its capacity, we may take $C = C'$.

We will now move on to get our upper bound c on the product channel, originally due to [HW08]. As working with the channel itself directly is tricky, we instead derive our upper bound in terms of Schmidt coefficients of the channel output, as hinted.

Lemma 4. Denote by $|\Phi\rangle_W \in W \otimes W$ the maximally-entangled state. Then, we have the following bound on the maximal Schmidt coefficient:

$$\mu_1^2((\mathcal{N} \otimes \bar{\mathcal{N}})(|\Phi\rangle\langle\Phi|_A)) \geq \frac{\dim A}{\dim B \dim E}.$$

Proof. As shorthand, let us write $(\mathcal{N} \otimes \bar{\mathcal{N}})(|\Phi\rangle\langle\Phi|_A) = \mathcal{N}(\Phi_A)$. Now, as $|\Phi\rangle_A$ is pure we use the techniques of PROPOSITION 6 to see

$$\sigma(\mathcal{N}(\Phi_A)) = \{\mu_i^2(\mathcal{N}(\Phi_A))\}_i.$$

Due to [PROPOSITION 1](#) we know that

$$\mu_1^2(\mathcal{N}(\Phi_A)) = \|\mathcal{N}(\Phi_A)\|_\infty \geq {}_B \langle \Phi | (\mathcal{N} \otimes \bar{\mathcal{N}})(|\Phi\rangle \langle \Phi|_A) |\Phi\rangle_B$$

We make use of the states $|\Phi\rangle_B$ and $|\Phi\rangle_E$ as well. We begin with a bound on the ∞ -norm, where by definition we have

$$\begin{aligned} & \|(\mathcal{N} \otimes \bar{\mathcal{N}})(|\Phi\rangle \langle \Phi|_A)\|_\infty \\ & \geq {}_B \langle \Phi | (\mathcal{N} \otimes \bar{\mathcal{N}})(|\Phi\rangle \langle \Phi|_A) |\Phi\rangle_B \\ & = \text{tr}((\mathcal{N} \otimes \bar{\mathcal{N}})(|\Phi\rangle \langle \Phi|_A) |\Phi\rangle \langle \Phi|_B) \end{aligned}$$

using the cyclicity of trace. Rewriting the channel in Stinespring form and cycling the partial trace, we see

$$\begin{aligned} & \text{tr}((\mathcal{N} \otimes \bar{\mathcal{N}})(|\Phi\rangle \langle \Phi|_A) |\Phi\rangle \langle \Phi|_B) \\ & = \sum_i \text{tr}((V \otimes \bar{V}) |\Phi\rangle \langle \Phi|_A (V \otimes \bar{V})^* (\mathbb{1}_{BB} \otimes |ii\rangle_{EE}) |\Phi\rangle \langle \Phi|_B (\mathbb{1}_{BB} \otimes \langle ii|_{EE})). \end{aligned}$$

Note that we are implicitly switching the ordering of some tensor products for the multiplication to carry through. Dividing through by the normalizing factor, we find that

$$\begin{aligned} & \text{tr}((\mathcal{N} \otimes \bar{\mathcal{N}})(|\Phi\rangle \langle \Phi|_A) |\Phi\rangle \langle \Phi|_B) \\ & \geq \text{tr}((V \otimes \bar{V}) |\Phi\rangle \langle \Phi|_A (V \otimes \bar{V})^* (|\Phi\rangle \langle \Phi|_B \otimes |\Phi\rangle \langle \Phi|_E)) \\ & = |\langle \langle \Phi|_B \otimes \langle \Phi|_E \rangle (V \otimes \bar{V}) |\Phi\rangle_A|^2. \end{aligned}$$

Now, up to conjugation by a unitary we may suppose that V (and therefore \bar{V}) is an orthogonal projection with respect to our fixed basis. From there, we pull out the normalizing factors of our maximally-mixed states to get

$$|\langle \langle \Phi|_B \otimes \langle \Phi|_E \rangle (V \otimes \bar{V}) |\Phi\rangle_A|^2 = \frac{(\dim A)^2}{\dim A \dim B \dim E},$$

which is precisely our desired lower bound. \square

Lastly, we need a lower bound for the output entropy of a single copy of the channel.

Lemma 5. *Suppose $W = V(A) \subseteq B \otimes E$ is such that all unit vectors $y \in W$ are such that $\mu_1^2(y) \leq \mu$ for some $\mu \in (0, 1)$. Then,*

$$\frac{1}{1-p} \log((1-\mu)^p + \mu^p) \leq S_p^{\min}(\mathcal{N}).$$

Proof. We know from Stinespring dilation that

$$S_p^{\min}(\mathcal{N}) = \min_{\text{unit } y \in W} S_p(\text{tr}_E |y\rangle\langle y|).$$

Take any such unit vector y , writing its Schmidt decomposition

$$|y\rangle = \sum_i \mu_i(u) |i\rangle_B \otimes |i\rangle_E,$$

whence it is straightforward to compute due to orthonormality that

$$\text{tr}_E |y\rangle\langle y| = \sum_i \mu_i(y)^2 |i\rangle\langle i|_B.$$

Now, define the equidimensional vectors

$$\mu^2(y) = (\mu_1^2(y) \ \mu_2^2(y) \ \cdots)^\top \text{ and } m = (\max\{A, 1 - A\} \ \min\{\mu, 1 - \mu\} \ 0 \ \cdots)^\top.$$

Due to the assumption that $\mu_1^2(u) \leq \mu$, we know $\mu^2(y) \leq m$, and due to Schur-concavity courtesy of [LEMMA 1](#) we get

$$\frac{1}{1-p} \log((1-\mu)^p + \mu^p) = S_p(m) \leq S_p(\text{tr}_E |y\rangle\langle y|).$$

We did this for a particular choice of y , and upon minimizing we see indeed that we indeed bound $S_p^{\min}(\mathcal{N})$ from below. \square

We now have all the results necessary, which we bundle up in the following theorem.

Theorem 17. *Let $W \subseteq B \otimes E$ be such that some $\mu \in (0, 1)$ exists to bound the Schmidt coefficients for all unit vectors $y \in W$ by $\mu_1^2(y) \leq \mu$. If*

$$\frac{\dim W}{\dim B \dim E} > ((1-\mu)^p + \mu^p)^{\frac{2}{p}},$$

then associated with W is a channel $\mathcal{N}_{A \rightarrow B}$ with non-additive p -entropy, where $A \cong \mathbb{C}^{\dim W}$.

Proof. From the techniques of [PROPOSITION 6](#) we see

$$S_p^{\min}(\mathcal{N} \otimes \bar{\mathcal{N}}) \leq S_p((\mathcal{N} \otimes \bar{\mathcal{N}})(|\Phi\rangle\langle\Phi|_A)) = \frac{p}{1-p} \log \sum_i \mu_i^2((\mathcal{N} \otimes \bar{\mathcal{N}})(|\Phi\rangle\langle\Phi|_A)).$$

Since $p > 1$, and so $1 - p < 0$, we use our lower bound from [LEMMA 4](#) to see

$$S_p^{\min}(\mathcal{N} \otimes \bar{\mathcal{N}}) \leq \frac{p}{1-p} \log \mu_1^2((\mathcal{N} \otimes \bar{\mathcal{N}})(|\Phi\rangle\langle\Phi|_A)) \leq \frac{p}{1-p} \log \frac{\dim W}{\dim B \dim E}.$$

Now, due to our hypothesis and [LEMMA 5](#) we have

$$\frac{1}{1-p} \log((1-\mu)^p + \mu^p) = S_p^{\min}(\mathcal{N}) = S_p^{\min}(\bar{\mathcal{N}}),$$

where the last equality is proven in [LEMMA 3](#). Next, the logarithm is monotonic and so

$$\log \frac{\dim W}{\dim B \dim E} > \log\left(\left((1-\mu)^p + \mu^p\right)^{\frac{2}{p}}\right) = \frac{2}{p} \log((1-\mu)^p + \mu^p).$$

As $1 - p < 0$ we have

$$\frac{p}{1-p} \log \frac{\dim W}{\dim B \dim E} < 2 \left(\frac{p}{(1-p)p} \log((1-\mu)^p + \mu^p) \right),$$

and from here it is clear that we have non-additivity. □

The last step is to provide a general technique for finding such an upper bound μ .

Proposition 11. *Let P_W be the projection onto $W \subseteq B \otimes E$. Then,*

$$\mu = \sup_{\|x \otimes y\|=1} \|P_W(x \otimes y)\|^2$$

satisfies $\mu \geq \mu_1^2(w)$ for all unit $w \in W$.

Proof. Take some such w with Schmidt decomposition

$$w = \sum_i \mu_i e_i \otimes f_i$$

and take an arbitrary unit vector $x \otimes y \in B \otimes E$. Witness that

$$|\langle w, x \otimes y \rangle|^2 = \left| \sum_i \mu_i \langle e_i, x \rangle_B \langle f_i, y \rangle_E \right|^2 \leq \mu_1^2 \sum_i |\langle e_i, x \rangle_B| |\langle f_i, y \rangle_E|.$$

Through Hölder's inequality followed by Bessel's inequality, we have

$$\mu_1^2 \sum_i |\langle e_i, x \rangle_B| |\langle f_i, y \rangle_E| \leq \mu_1^2 \sum_i |\langle e_i, x \rangle_B|^2 \sum_i |\langle f_i, y \rangle_E|^2 \leq \mu_1^2 \|x\|_B \|y\|_B,$$

where using the fact we have a unit vector, we see $|\langle w, x \otimes y \rangle| \leq \mu_1$. Note that by taking $x = e_1$ and $y = f_1$, we actually get equality.

Now, let w_i form an orthonormal basis of W so that $P_W = \sum_i |w_i\rangle\langle w_i|$. Write $w = \sum_i c_i w_i$, and witness that

$$\mu_1^2 = \sup_{\|x \otimes y\|=1} |\langle w, x \otimes y \rangle|^2 = \sup_{\|x \otimes y\|=1} \left| \sum_i c_i \langle w_i, x \otimes y \rangle \right|^2 \leq \sup_{\|x \otimes y\|=1} \left(\sum_i |c_i| |\langle w_i, x \otimes y \rangle| \right)^2.$$

Applying Hölder's once more, we have

$$\left(\sum_i |c_i| |\langle w_i, x \otimes y \rangle| \right)^2 \leq \sum_i |c_i|^2 \sum_i |\langle w_i, x \otimes y \rangle|^2 = \left\| \sum_i \langle w_i, x \otimes y \rangle w_i \right\|^2,$$

where at the end we recall $\|w\| = 1$ and also apply the Pythagorean theorem. Of course, $P_W = \sum_i |w_i\rangle\langle w_i|$, so we have

$$\mu_1^2 \leq \sup_{\|x \otimes y\|=1} \|P_W(x \otimes y)\|^2$$

as desired. □

We will frequently use the phrase that some construction cannot yield non-additivity in the context of [THEOREM 17](#). By this, we mean that the subspace W in question cannot be used to construct a channel-conjugate pair which violates the inequality in the theorem.

3.4 Antisymmetric Spaces

We have already seen a simple example of an antisymmetric space in [SECTION 2.8](#), the antisymmetric representation of a symmetric group. However, we may generalize this by simply antisymmetrizing more tensors. These spaces will form the foundation for several of the subspaces W that others have and we will study to look for non-additivity.

Definition 45 (Antisymmetric Space). Let y_\wedge be the Young symmetrizer corresponding to the partition $\lambda = (1, \dots, 1)$ with $|\lambda| = n$. Then, we define the antisymmetric space $W_{nk}^\wedge = y_\wedge(W_{nk})$.

Proposition 12. Let $w = e_{i_1} \otimes \dots \otimes e_{i_n} \in W_{nk}$ be a basis tensor. If $i_a = i_b$ for any $a \neq b$, then $y_\wedge w = 0$.

Proof. Define the permutation $\sigma = (a \ b)$, and notice that $\sigma(w) = w$. Within the Young symmetrizer we then have

$$y_\wedge w = \sum_{\tau \in S_n} \text{sgn}(\tau) \tau(w) = \frac{1}{2} \left(\sum_{\tau \in S_n} \text{sgn}(\tau\sigma) (\tau\sigma)(w) + \sum_{\tau \in S_n} \text{sgn}(\tau) \tau(w) \right).$$

For any $\tau \in S_n$ we have

$$\text{sgn}(\tau\sigma) (\tau\sigma)(w) + \text{sgn}(\tau) \tau(w) = -\text{sgn}(\tau) \tau(w) + \text{sgn}(\tau) \tau(w) = 0$$

since the sign of a permutation is multiplicative. □

An immediate consequence of this proposition is that $W_{nk}^\wedge = \emptyset$ when $k < n$, and that $\dim W_{nk}^\wedge = \binom{k}{n}$ for $k \geq n$.

The following result from [GOY25] is standard, but will be very helpful when working with antisymmetric spaces numerically later.

Proposition 13. The set

$$\left\{ \frac{1}{n!} y_\wedge (e_{i_1} \otimes \dots \otimes e_{i_n}) : 1 \leq i_j < i_{j+1} \leq k \right\}$$

forms an orthonormal basis of W_{nk}^\wedge .

As shown in [SS24], the corresponding upper bound of PROPOSITION 11 for W_{2k}^\wedge is $1/2$. Thus, the non-additivity condition in the language of THEOREM 17 is

$$\frac{k(k-1)}{2} = \frac{\binom{k}{2}}{k^2} > \left(\left(1 - \frac{1}{2}\right)^p + \left(\frac{1}{2}\right)^p \right)^{\frac{2}{p}},$$

which for any fixed p may be satisfied for some sufficiently large k . Unfortunately, we will show that this does not extend to $n \geq 3$.

Definition 46 (Shuffle). Fix some integers $2 \leq m < n$. A shuffle is a permutation $\sigma \in S_n$ where

$$\sigma(1) < \cdots < \sigma(m) \text{ and } \sigma(m+1) < \cdots < \sigma(n).$$

We denote the set of such shuffles by $S_n^{(m)}$.

Shuffles are essentially mixings between bipartitions of n that preserve their internal order. As such,

$$|S_n^{(m)}| = \binom{n}{m}.$$

Moreover, every permutation in S_n may be written as the product of three distinct permutations given some bipartition: a shuffle and two permutations. The shuffle is required to send the integers to their respective partitions, and then the two permutations are needed to internally permute them. We now state our first original result.

Theorem 18. Let $W = W_{nk}^\wedge$ be bipartitioned into $W = B \otimes E$ where B holds m components. Then,

$$\sup_{\|x \otimes y\|=1} \|P_W(x \otimes y)\|^2 = \frac{1}{\binom{n}{m}}.$$

Proof. For any $x \otimes y \in W$, we know that we may embed it back into W_{nk} so that

$$P_W(x \otimes y) = \frac{1}{n!} y_\wedge(x \otimes y).$$

Now, suppose that $x \in W_{mk}^\wedge$ and $y \in W_{(n-m)k}^\wedge$. Take an arbitrary permutation $\sigma \in S_n$, and decompose into $\sigma = \tau \sigma_B \sigma_E$, where $\sigma_B \in S_m$, $\sigma_E \in S_{n-m}$, and $\tau \in S_n^{(m)}$ (technically, every index σ_E must have m added to it). Due to our antisymmetry assumptions, $\sigma_B(x \otimes y) = x \otimes y$, and likewise for σ_E . Therefore,

$$\frac{1}{n!} y_\wedge(x \otimes y) = \frac{m!(n-m)!}{n!} \sum_{\tau \in S_n^{(m)}} \text{sgn}(\tau) \tau(x \otimes y).$$

where we also use the fact that $|S_m| = m!$. Now, notice that the summands corresponding to distinct shuffles are orthogonal, and that as P_W is orthogonal each summand has unit norm. Thus,

$$\|P_W(x \otimes y)\|^2 = \langle P_W(x \otimes y), P_W(x \otimes y) \rangle_{BE} = \frac{1}{\binom{n}{m}^2} \binom{n}{m} = \frac{1}{\binom{n}{m}}.$$

Now that such a norm is achievable, we will show it is indeed an upper bound when $x \otimes y$ does not comprise individually antisymmetrized components. To start, we see

$$\begin{aligned} y_{\wedge}(x \otimes y) &= \sum_{\substack{\tau \in S_n^{(m)} \\ \sigma_B \in S_m \\ \sigma_E \in S_{n-m}}} \text{sgn}(\tau \sigma_B \sigma_E) \tau \sigma_B \sigma_E (x \otimes y) \\ &= \sum_{\substack{\tau \in S_n^{(m)} \\ \sigma_B \in S_m \\ \sigma_E \in S_{n-m}}} \text{sgn}(\tau) [\text{sgn}(\sigma_B)(\sigma_B)(x) \otimes (\text{sgn}(\sigma_E)\sigma_E)(y)], \end{aligned}$$

using multiplicativity of the sign and tensor multilinearity. That is,

$$P_W(x \otimes y) = \frac{m!(n-m)!}{n!} (P_B x \otimes P_E y),$$

where P_B and P_E are the respective projections onto W_{mk}^{\wedge} and $W_{(n-m)k}^{\wedge}$. Therefore,

$$\|P_W(x \otimes y)\| \leq \frac{m!(n-m)!}{n!} \|P_B x\| \|P_E y\| \leq \frac{1}{\binom{n}{m}},$$

as established. □

Note that applying this theorem with $n = 2$ gives as a special case the aforementioned bound. The relevant inequality to consider for non-additivity in this more general case is therefore

$$\frac{\binom{k}{n}}{k^n} > \left(\left(1 - \frac{1}{\binom{n}{m}} \right)^p + \frac{1}{\binom{n}{m}^p} \right)^{\frac{2}{p}}.$$

We would like to take a brief digression to discuss the physical interpretation of this bound. Shuffles and antisymmetrization of individual bipartitions naturally arise in the study of intermolecular interactions, see for example [SC78; LS97; PKJ01]. This is as antisymmetries must be enforced to correctly model systems of fermionic particles. When two such systems interact, the total system must then enforce the antisymmetries in aggregate. However, this can be broken down into antisymmetrizing the individual systems and then handling their interactions via shuffles, which is precisely the decomposition we employed in the above theorem.

Theorem 19. *Bipartitions of antisymmetric spaces W_{nk}^\wedge such that Bob holds m components only have non-additive p -Rényi entropy in the context of [THEOREM 17](#) if $n = 2$ and $k \geq 3$.*

Proof. We will take our inequality in [THEOREM 17](#), and through a series of steps which drive the left-hand side higher and the right-hand side lower, end up with an inequality which can still never be satisfied if $n \geq 3$.

To start, define the function

$$f(\mu, p) = ((1 - \mu)^p + \mu^p)^{2/p}$$

for $\mu \in (0, 1/2)$. Note that we may consider $\mu \in (0, 1)$, however f is symmetric in μ about $1/2$. It is a straightforward computation that

$$\begin{aligned} & \frac{\partial f}{\partial p} \\ &= \frac{2((1 - \mu)^p + \mu^p)^{2/p-1} (p(1 - \mu)^p \log(1 - \mu) + p\mu^p \log(\mu) - ((1 - \mu)^p + \mu^p) \log((1 - \mu)^p + \mu^p))}{p^2}, \end{aligned}$$

which we may verify numerically satisfies $\partial f / \partial p \leq 0$ for all μ .

Next, we factor f as

$$f(x, p) = (1 - \mu)^2 \left(1 + \left(\frac{\mu}{1 - \mu} \right)^p \right)^{\frac{2}{p}}$$

so that

$$\lim_{p \rightarrow \infty} f(\mu, p) = (1 - \mu)^2 \lim_{p \rightarrow \infty} \left(1 + \left(\frac{\mu}{1 - \mu} \right)^p \right)^{\frac{2}{p}} = (1 - \mu)^2 \cdot 1.$$

Returning to our additivity inequality, with $\mu = 1/\binom{n}{m}$ we are thus left to examine

$$\frac{\binom{k}{n}}{k^n} > f(\mu, p) \geq \left(1 - \frac{1}{\binom{n}{m}} \right)^2.$$

Our goal now is to maximize the right-hand side while minimizing the left-hand side. If the inequality ever holds, then there must have been some p as $p \rightarrow \infty$ such that non-additivity began to occur.

Witness that

$$\binom{k}{n} = \frac{k^n}{n!} \prod_{i=0}^{n-1} \left(1 - \frac{i}{k}\right)$$

and so

$$\frac{\binom{k}{n}}{k^n} = \frac{1}{n!} \prod_{i=0}^{n-1} \left(1 - \frac{i}{k}\right) < \frac{1}{n!},$$

with us approaching this bound as $k \rightarrow \infty$. On the other hand,

$$\left(1 - \frac{1}{\binom{n}{m}}\right)^2 \geq \left(1 - \frac{1}{\binom{n}{1}}\right) = \left(1 - \frac{1}{n}\right)^2.$$

By symmetry, we could also pick $m = n - 1$.

In all, the widest inequality we could have is

$$\frac{1}{n!} > \left(\frac{n-1}{n}\right)^2.$$

We will now prove that

$$n^2 \leq n!(n-1)^2,$$

for all $n \geq 3$, so we cannot have any additivity even while both $k \rightarrow \infty$ and $p \rightarrow \infty$. When $n = 3$, this becomes $9 \leq 24$. Supposing now this is true for some $n \geq 3$, we see

$$\begin{aligned} (n+1)^2 &= n^2 + 2n + 1 \\ &\leq n!(n-1)^2 + 2n + 1 \\ &< n!(n-1)^2 + n^2 \\ &\leq n!(n-1)^2 + n!(n-1)^2 \\ &< (n+1)!n^2. \end{aligned}$$

So, our claim holds.

We are left to examine the $n = 2$ case. Here, we consider the less wide inequality

$$\frac{\binom{k}{2}}{k^2} = \frac{k-1}{2k} > \left(1 - \frac{1}{2}\right)^2 = \frac{1}{4},$$

since only $m = 1$ is possible. Equivalently, we need $2k - 4 > 0$, which we see is only true for $k > 2$. □

Note that as a consequence of the above theorem, we have the following telling us that subspaces which are too small cannot hope to yield non-additivity within the current mathematical frameworks.

Corollary 1. *Suppose W is such that*

$$\frac{\dim W}{\dim B \dim E} \leq \frac{1}{4}.$$

Then, W cannot yield non-additivity in the context of [THEOREM 17](#).

Proof. In the language of the previous theorem, non-additivity is equivalent to finding μ and p so that

$$\frac{\dim W}{\dim B \dim E} > f(\mu, p) \geq \begin{cases} \mu^2 & \mu \geq 1/2 \\ (1 - \mu)^2 & \mu \leq 1/2. \end{cases}$$

However, we see that $f(\mu, p) \geq 1/4$, with the inequality saturated when $\mu = 1/2$. □

4 Quantum Capacity

Quantum capacity may be thought of as Alice trying to transmit a quantum state to Bob in a way that preserves some shared entanglement between them. In particular, Alice ought not be able to touch this shared resource. This is a sensible notion as it means that Alice is preserving quantum information – their entanglement – and efficiently communicating quantum information can be understood as efficiently preserving quantum information in the presence of noise.

4.1 Coherent Information

Quantum capacity has a regularizing formula given by a quantity known as the coherent information, achieved through the cumulative work of [Llo97; Sho02; Dev05].

Definition 47 (Coherent Information). *For a channel $\mathcal{N}_{A \rightarrow B}$ we say the coherent information of a state $\rho \in S(A)$ with respect to \mathcal{N} is $I_c(\mathcal{N}, \rho) = S(\mathcal{N}(\rho)) - S(\mathcal{N}^c(\rho))$. The coherent information of \mathcal{N} is given by the optimization*

$$Q^{(1)}(\mathcal{N}) = \max_{\rho \in S(A)} I_c(\mathcal{N}, \rho).$$

Theorem 20 (Lloyd-Shor-Devetak). *The quantum capacity of a channel is given by regularizing the coherent information:*

$$Q(\mathcal{N}) = \lim_{n \rightarrow \infty} \frac{1}{n} Q^{(n)}(\mathcal{N}).$$

Coherent information is also non-additive in a very strong sense. For every $n \geq 1$ there exists some channel \mathcal{N} so $Q^{(n)}(\mathcal{N}) = 0$ but $Q(\mathcal{N}) > 0$ [Cub+15]. Moreover, the gap $Q(\mathcal{N}) - Q^{(1)}(\mathcal{N})$ may be arbitrarily large [SY08]. Due to this severe non-additivity, not only are we generically unable to explicitly evaluate the quantum capacity, but we cannot even hope to estimate it from small symbols. It is difficult to even tell whether the capacity is zero, outside of a few special cases we will explore next section.

Fortunately, coherent information for one or two symbols is at least fairly tractable computationally for small qudit channels (up to roughly $d = 4, 5$ in our experience).

Also demonstrated in [SY08] is the phenomenon of superactivation, where two channels \mathcal{N} and \mathcal{M} may each have zero quantum capacity, but nevertheless $Q(\mathcal{N} \otimes \mathcal{M}) > 0$. The relevant channels are not necessarily abstract constructions either, but can be physically relevant channels from quantum optics [SSY11]. Instances of activation, where only one of these channels has zero quantum capacity, have also been studied [Lim+19].

In SECTION 6.3 we find a new example of the activation of coherent information. Our choice of zero capacity activator is the following, originally introduced for studying error correcting codes [GBP97]. The proposition is due to [BDS97], so in particular we know erasure channels have zero quantum capacity for $p \geq 1/2$.

Definition 48 (Erasure Channel). *The erasure channel $\mathcal{E}_{A \rightarrow B}^p$ with erasure probability p is*

$$\mathcal{E}_p(\rho) = (1 - p)\rho + p |e\rangle \langle e|,$$

where $B = A \oplus \text{span}(|e\rangle)$.

Proposition 14. *The quantum capacity of an erasure channel \mathcal{E}^p is $Q(\mathcal{E}^p) = \max\{0, 1 - 2p\}$.*

In light of activation being achieved by a combination with another channel, we actually define two types of additivity.

Definition 49 (Weak and Strong Additivity). *We say \mathcal{N} has weakly additive quantum capacity if $Q^{(n)}(\mathcal{N}) = nQ^{(1)}(\mathcal{N})$. We say it is strongly additive if for all quantum channels \mathcal{M} we have*

$$Q^{(1)}(\mathcal{N} \otimes \mathcal{M}) = Q^{(1)}(\mathcal{N}) + Q^{(1)}(\mathcal{M}).$$

Often, we might consider strongly additive with respect to some particular class of channels as well. For example, any degradable channel, defined in the next section, is strongly additive with respect to all other degradable channels.

4.2 Degradability

Additivity of quantum capacity seems to be very strongly related to the notion of degradability, introduced by [DS05], which broadly aims to describe how strongly correlated the channel output is with its environment noise.

Definition 50 (Degradable). *We say \mathcal{N} is degradable if there exists some other channel \mathcal{D} , a degrading map, such that $\mathcal{N}^c = \mathcal{D} \circ \mathcal{N}$. If the complementary channel is degradable, then we say \mathcal{N} is antidegradable.*

Degradability can be thought of as being the closest thing to preserving information through transmission, as Bob may construct a new channel which processes his output to understand exactly what happened in the environment. The only information truly lost is the entanglement between Bob and his environment, though we should always expect this to be irrecoverable due to the non-uniqueness of the Stinespring dilation. Degradability also implies coherent information plays nicely with channel inputs, thanks to [YHD08].

Theorem 21. *If \mathcal{N} is degradable, then $I_c(\mathcal{N}, \rho)$ is concave in ρ .*

This concept may be generalized to provide a more information-theoretic perspective, introduced in [CLS17].

Definition 51 (Informationally Degradable). *We call a channel $\mathcal{N}_{A \rightarrow B}$ informationally degradable if $I(W : B) \geq I(W : E)$ for all states with an arbitrary finite ancillary system $\rho_{WA} \in S(W \otimes A)$.*

Proposition 15. *Degradability implies informational degradability.*

Proof. Suppose \mathcal{N} is degradable with degrading map \mathcal{D} . Let W be an arbitrary finite-dimensional ancilla and take some starting state $\rho \in S(W \otimes A)$. Due to PROPOSITION 7 we know that

$$I(W : B) = S(\rho_{WB} \| \rho_W \otimes \rho_B) = S((\mathbb{1}_B \otimes \mathcal{N})\rho_{WA} \| \rho_W \otimes \mathcal{N}(\rho_A)).$$

We also have

$$\begin{aligned} I(W : E) &= S((\mathbb{1}_B \otimes \mathcal{N}^c)\rho_{WA} \| \rho_W \otimes \mathcal{N}^c(\rho_A)) \\ &= S((\mathbb{1}_B \otimes (\mathcal{D} \circ \mathcal{N}))\rho_{WA} \| \rho_W \otimes (\mathcal{D} \circ \mathcal{N})(\rho_A)). \end{aligned}$$

Due to information processing, we have $I(W : E) \leq I(W : B)$. Since W was arbitrary, we are done. \square

Therefore, if a channel is informationally degradable but not degradable, Bob may not be able to exactly process his state to find the complementary output. However, he still holds more information than the environment in the sense of having more mutual information than the environment with arbitrary ancillas.

Degradability and informational degradability are relevant in the context of quantum capacity as channels with these properties are known to have additive coherent information, respectively proven in [DS05; CLS17]. Several other channels with degradable-like

structures have been studied with similar additivity results derived, such as in [Brá+10; Wat12; GJL18]. However, these lack any unifying framework.

Antidegradable channels are also one of only two classes where the quantum capacity is known to be zero. A proof of this is contained in [SS12], which also proves that PPT (positive partial transpose) channels have no capacity, the second such class.

Degradability is not a mere abstract curiosity. It is very easy to explicitly construct degradable channels in low dimensions. Notice that if a degrading map \mathcal{D} exists, then

$$\mathcal{N}^c = \mathcal{D} \circ \mathcal{N} \iff \mathcal{T}_{\mathcal{N}^c} = \mathcal{T}_{\mathcal{D} \circ \mathcal{N}} = \mathcal{T}_{\mathcal{D}} \mathcal{T}_{\mathcal{N}}.$$

Given a channel then, we may compute the transfer matrix of \mathcal{N} and \mathcal{N}^c , and then examine the matrix equation $\mathcal{T}_{\mathcal{N}^c} = D \mathcal{T}_{\mathcal{N}}$, where D is the unknown matrix. The existence of such a D is an elementary linear algebra problem. If found, it remains only to check that D is a valid transfer matrix, in the sense that it corresponds to some channel $\mathcal{D}_{B \rightarrow E}$. Now, recall that

$$D \text{ is a transfer matrix} \iff \mathfrak{D}(D) \text{ is a Choi-Jamiołkowski matrix.}$$

Checking if $\mathfrak{D}(D)$ is a Choi-Jamiołkowski matrix amounts to verifying if it positive-semidefinite and $\text{tr}_B(D) = \mathbb{1}_A$, which again is a mere linear algebra problem. All of these checks are also easily implementable numerically.

An important note is that multiple possible solutions D may exist. This problem has been noted before, such as in [CRS08; Brá15], where they derived some conditions for when degrading maps are or are not unique. To see an example relevant to us, let $0 \leq s, t \leq 1$ and $1 - s - t \geq 0$, and consider the qubit channel

$$\mathcal{N}(\rho) = \begin{pmatrix} \rho_{00} + s\rho_{11} & 0 & \sqrt{1-s-t}\rho_{01} \\ 0 & t\rho_{11} & 0 \\ \sqrt{1-s-t}\rho_{10} & 0 & (1-s-t)\rho_{11} \end{pmatrix}$$

with complement

$$\mathcal{N}^c(\rho) = \begin{pmatrix} \rho_{00} + (1-s-t)\rho_{11} & 0 & \sqrt{s}\rho_{01} \\ 0 & t\rho_{11} & 0 \\ \sqrt{s}\rho_{10} & 0 & s\rho_{11} \end{pmatrix}.$$

We will see later in SECTION 6.2 that this is \mathcal{N}_{02} , a subchannel of the ReMAD channel $\mathcal{N}_{r,s,t}$

restricted to kets $|0\rangle$ and $|2\rangle$. One possible choice of transfer matrix is

$$D_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & \frac{1-2s-t}{t} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sqrt{\frac{s}{1-s-t}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sqrt{\frac{s}{1-s-t}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{s}{t} & 0 & 0 & 0 & 0 \end{pmatrix}$$

which induces the qutrit degrading channel

$$\mathcal{D}_1(\rho) = \begin{pmatrix} \rho_{00} + \frac{1-2s-t}{t}\rho_{11} & 0 & \sqrt{\frac{s}{1-s-t}}\rho_{02} \\ 0 & \rho_{11} & 0 \\ \sqrt{\frac{s}{1-s-t}}\rho_{20} & 0 & \frac{s}{t}\rho_{11} \end{pmatrix}.$$

However, another possible transfer matrix is

$$D_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 - \frac{s}{1-s-t} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sqrt{\frac{s}{1-s-t}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sqrt{\frac{s}{1-s-t}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{s}{1-s-t} \end{pmatrix}$$

which induces another qutrit degrading channel

$$\mathcal{D}_2(\rho) = \begin{pmatrix} \rho_{00} + \left(1 - \frac{s}{1-s-t}\right)\rho_{22} & 0 & \sqrt{\frac{s}{1-s-t}}\rho_{02} \\ 0 & \rho_{11} & 0 \\ \sqrt{\frac{s}{1-s-t}}\rho_{20} & 0 & \frac{s}{1-s-t}\rho_{22} \end{pmatrix}.$$

For \mathcal{D}_1 to be a valid quantum channel, the necessary condition turns out to be $s = 1 - t$. However, we also see that we require $\rho_{00} + \rho_{11} = 1$ for its output to be a state. This is

because \mathcal{D}_1 is best understood as a subchannel which is restricted to the same domain as \mathcal{N} . Indeed, we may compute that

$$\text{tr}_B(\mathcal{J}_{\mathcal{D}_1}) = \text{diag}(1, 1, 0),$$

and from [PROPOSITION 5](#) we know we can therefore extend \mathcal{D}_1 to a full channel, with one possible extension being

$$\mathcal{D}_1(\rho) = \begin{pmatrix} \rho_{00} + \frac{1-2s-t}{t}\rho_{11} & 0 & 2\sqrt{\frac{s}{1-s-t}}\rho_{02} \\ 0 & \rho_{11} & \rho_{12} \\ 2\sqrt{\frac{s}{1-s-t}}\rho_{20} & \rho_{21} & \frac{s}{t}\rho_{11} + \rho_{22} \end{pmatrix}.$$

For \mathcal{D}_2 , the necessary condition is simply that $s \leq 1 - s - t$. In all, we see these two degrading channels thus cover degradability in different parametre spaces since $(1-t) \leq 1 - (1-t) - t$ is only true for $t = 1$.

4.3 Numerical Evaluation and Bounds

Just as for classical capacity, the computational intractability of coherent information for many symbols means all we may hope for are bounds. Our first is an SDP from [\[FF21\]](#).

Theorem 22. *For a channel $\mathcal{N}_{A \rightarrow B}$ let $a = |A|$ and $b = |B|$. Fix some integer $\ell \geq 1$ and define $S_\ell(\mathcal{N})$ as the solution to the SDP*

$$\begin{aligned} & \max_{\substack{\rho \in \mathcal{S}(A) \\ K \in M_{ab}(\mathbb{C}) \\ \{W_i \in M_{ab}^*(\mathbb{C})\}_{i=1}^\ell \\ \{Z_i \in M_{ab}(\mathbb{C})\}_{i=0}^\ell}} \text{tr} \left(\left(K^H - \sum_{i=1}^\ell W_i \right) \odot \mathcal{J}_{\mathcal{N}} \right) \\ & \text{s.t.} \quad \begin{pmatrix} \rho \otimes \mathbb{1}_B & K \\ K^* & Z_\ell^H \end{pmatrix} \geq 0 \\ & \quad (\rho \otimes \mathbb{1}_B) \pm (Z_0^H)^{\top_B} \geq 0 \\ & \quad \left\{ \begin{pmatrix} W_i & Z_i \\ Z_i^* & Z_{i-1}^H \end{pmatrix} \geq 0 \right\}_{i=1}^\ell. \end{aligned}$$

Here, $X^H = X + X^*$ is the Hermitian part of a matrix X . Defining

$$R_\ell(\mathcal{N}) = 2^\ell \ell - (2^\ell + 1) \log(2^\ell + 1) + (2^\ell + 1) \log S_\ell(\mathcal{N}),$$

we have $R_\ell(\mathcal{N}) \geq Q(\mathcal{N})$.

The letter R above references the Rains bound, originally presented in the context of an SDP bound for a problem known as entanglement distillation [Rai02]. A related quantity, the Rains information, establishes a connection to quantum capacity [TWW07], with R_ℓ being a bound approaching the Rains information from above as $\ell \rightarrow \infty$.

Studying the degradability of channels is also an area in which numerical techniques may be used. When solving for possible transfer matrices, we have just seen that multiple possible solutions may exist. Most techniques for solving matrix equations with a numerical or symbolic linear algebra system are constrained to finding only one possible solution due to the algorithmic processes employed under-the-hood.

A channel \mathcal{N} often depends on some parameter set $c = (c_1, \dots, c_m)$, and so we may wish to work specifically with a symbolic computer algebra system to find the general analytic form of a degrading map. In this case, our symbolic output will give one family of potential transfer matrices $D = D(c)$. However, this family may not be defined for all choices of c . Despite this, some other family might exist which would be defined for these singular choices of c . Moreover, even at points where $D(c)$ is well-defined, these transfer matrices might not yield valid quantum channels. Yet, another family could.

The former occurred in previous example, where $c = (s, t)$. Our first solution $D_1 = D_1(s, t)$ is not defined when $t = 0$ as it contains the entry s/t . However, $D_2(s, t)$ is defined for $t = 0$, and in fact yields a degrading map so long as $s \leq 1/2$. The reverse happens when $1 - s - t = 0$, where $D_2(s, t)$ is always undefined while $D_1(s, t)$ is not.

A partial remedy to this problem exists, in that if we have one family of transfer matrices $D(\vartheta)$, it is possible to bootstrap several other families.

Proposition 16. Consider the matrix equation $XA = B$, where A and B are known, and suppose we have a known solution $X = D$. Let $k_i \in \ker A^\top$, and define

$$K = \begin{pmatrix} k_1 \\ \dots \\ k_m \end{pmatrix},$$

where m is the number of rows of X . Then, $D + K$ is another solution to the equation. Moreover, every solution X may be written as $D + K$ for some such K .

Proof. It is straightforward to see that $KA = 0$ due to the constructions of rows of K belonging to the kernel, and so $(D + K)A = B$. Moreover, suppose we have some solution X . Then, $DA = B = XA$, and so $(X - D)A = 0$. In particular, this means each row of $D - X$ belongs to $\ker A^\top$. So, we may take $K = X - D$ and $D + K = X$. \square

Theorem 23. For a channel $\mathcal{N}_{A \rightarrow B}$ let $a = |A|$ and $b = |B|$. Let D be such that $D\mathcal{T}_{\mathcal{N}} = \mathcal{T}_{\mathcal{N}^c}$, and let $\{e_i\}_{i=1}^n$ form a basis for $\ker \mathcal{T}_{\mathcal{N}}^\top$. Define the SDP feasibility problem

$$\begin{aligned} \text{find} \quad & \{c_{ij} \in \mathbb{C}\}_{i=1, j=1}^{n, ab} \\ \text{s.t.} \quad & K = \begin{pmatrix} c_{11}e_1 + \cdots + c_{1n}e_n \\ \cdots \\ c_{ab,1}e_1 + \cdots + c_{ab,n}e_n \end{pmatrix}, \\ & \text{tr}_B(\vartheta(D + K)) = \mathbb{1}_a \\ & \vartheta(D + K) \geq 0. \end{aligned}$$

Then, the feasibility of this SDP implies the degradability of \mathcal{N} .

This is an imperfect solution, however. Continuing with the same example,

$$\ker \mathcal{T}_{\mathcal{N}}^\top = \left\{ \begin{pmatrix} (0 \ 0 \ 0 \ 0 \ -\frac{1-s-t}{t} \ 0 \ 0 \ 0 \ 1)^\top \\ (0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0)^\top \\ (0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0)^\top \\ (0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0)^\top \\ (0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0)^\top \end{pmatrix} \right\}.$$

Looking at the last row of both D_1 and D_2 , it is clear that no such K exists so that $D_1 = D_2 + K$. Therefore, there is no way to bootstrap our way to the degradability region of D_1 if we only have access to D_2 .

Moving away from symbolic computation, we can reformulate the above result into a helpful corollary when we are working numerically.

Corollary 2. For a channel $\mathcal{N}_{A \rightarrow B}$ let $a = |A|$ and $b = |B|$. Then, the feasibility of the following SDP is equivalent to the degradability of \mathcal{N} :

$$\begin{aligned} \text{find} \quad & K, D \in M_{ab}(\mathbb{C}) \\ \text{s.t.} \quad & D\mathcal{T}_{\mathcal{N}} = \mathcal{T}_{\mathcal{N}^c} \\ & K\mathcal{T}_{\mathcal{N}} = 0 \\ & \text{tr}_B(\vartheta(D + K)) = \mathbb{1}_a \\ & \vartheta(D + K) \geq 0. \end{aligned}$$

The key difference between [THEOREM 23](#) and [COROLLARY 2](#) is in philosophy. In the former case, we are working symbolically with some fixed family of potential transfer matrices $D(c)$. We use this family as an ansatz to generate potential degrading maps. This is helpful in the event that $D(c)$ merely satisfies the transfer matrix equation but does not correspond to a valid degrading map. For example, it might be that $\vartheta(D(c)) < 0$, in which case incorporating elements from its kernel will ensure the equation remains satisfied, but might result in a different Choi-Jamiołkowski matrix which is positive. Since our null space is explicitly described, this new solution is analytic as well.

In the latter case, we are working numerically. At each point of our parameter space we need only find some D and K which construct a valid degrading map. These need not belong to any particular or consistent family, and we are not concerned with deriving an analytical expression for them.

The SDP in [COROLLARY 2](#) is also able to study the degradability of subchannels with slight modification. As stated, we could not find the solution \mathcal{D}_1 since $\text{tr}_B \vartheta(D_1) \neq \mathbb{1}_3$ due to its subchannel structure. However, this is easily fixed with slack variables. Since we are studying a qubit subchannel of a qutrit channel, our modified SDP would be

$$\begin{aligned}
&\text{find} && K, D \in M_{9,4}(\mathbb{C}), \{c_i\}_{i=0}^3 \subseteq \mathbb{R} \\
&\text{s.t.} && D\mathcal{T}_N = \mathcal{T}_N^c \\
&&& K\mathcal{T}_N = 0 \\
&&& \text{tr}_B(\vartheta(D + K)) = c_0 \mathbb{1}_3 \\
&&& \text{tr}_B(\vartheta(D + K)) = c_1 \text{diag}(0, 1, 1) \\
&&& \text{tr}_B(\vartheta(D + K)) = c_2 \text{diag}(1, 0, 1) \\
&&& \text{tr}_B(\vartheta(D + K)) = c_3 \text{diag}(1, 1, 0) \\
&&& \sum_{i=0}^3 c_i = 1 \\
&&& \{c_i \geq 0\}_{i=0}^3 \\
&&& \vartheta(D + K) \geq 0.
\end{aligned}$$

Alternatively, let $D_n(\mathbb{C}) \subseteq M_n(\mathbb{C})$ be the set of diagonal matrices with complex entries,

which we know is convex. We see the above is equivalent to the optimization problem

$$\begin{aligned}
&\text{find} && K, D \in M_{9,4}(\mathbb{C}), P \in D_3(\mathbb{C}) \\
&\text{s.t.} && D\mathcal{T}_N = \mathcal{T}_{N^c} \\
&&& K\mathcal{T}_N = 0 \\
&&& P = \text{tr}_B(\mathfrak{S}(D + K)) \\
&&& \text{tr}(P) = 2 \\
&&& \{P_{ii} \in \{0, 1\}\}_{i=1}^3 \\
&&& \mathfrak{S}(D + K) \geq 0.
\end{aligned}$$

This is no longer an SDP and instead requires a mixed-integer solver. As at the time of writing we did not have a license for any such solver, like Gurobi or MOSEK. Thus, we opted for the slack variable approach. This is likely the most desirable solution regardless, as mixed-integer programming loses several of the computational advantages that semidefinite programs have.

An SDP also exists to show if a channel is non-degradable due to [Wan21]. It stems from the following weaker notion of degradability introduced by [Sut+17].

Definition 52 (ε -Degradable). *We call \mathcal{N} ε -degradable for $\varepsilon > 0$ if there is some other channel $\tilde{\mathcal{D}}$ such that $\|\mathcal{N}^c - \tilde{\mathcal{D}} \circ \mathcal{N}\|_{\diamond} \leq \varepsilon$.*

Theorem 24. *The minimal ε such that \mathcal{N} is ε -degradable is given by the SDP*

$$\begin{aligned}
&\min_{\tilde{\mathcal{D}}_{B \rightarrow E}, Z \in S^+(A \otimes E)} && \varepsilon \\
&\text{s.t.} && \text{tr}_E Z \leq \varepsilon \mathbb{1}_A \\
&&& Z \geq \mathcal{J}_{N^c} - \mathcal{J}_{\tilde{\mathcal{D}} \circ \mathcal{N}} \\
&&& \text{tr}_E(\mathcal{J}_{\tilde{\mathcal{D}}}) = \mathbb{1}_B \\
&&& \mathcal{J}_{\tilde{\mathcal{D}}} \geq 0.
\end{aligned}$$

Note that degradability is equivalent to 0-degradability, and thus if some $\varepsilon > 0$ is the optimal solution, we can be certain that \mathcal{N} is not degradable.

4.4 The Platypus Family

One particular channel has spawned a great deal of research due to its unique properties.

Definition 53 (Platypus Channel). *Let $s \in [0, 1]$. Then, the Platypus channel \mathcal{N}_s is a qutrit-to-qutrit channel given by the Stinespring isometry*

$$V_s |i\rangle = \begin{cases} \sqrt{s} |00\rangle + \sqrt{1-s} |11\rangle & i = 0 \\ |20\rangle & i = 1 \\ |21\rangle & i = 2 \end{cases}.$$

First presented in this form in [Sid21], and later thoroughly studied in [Led+23], the Platypus channel earns its namesake from its peculiar properties. Of importance to us, it is neither degradable nor antidegradable. However, it is conjectured to be weakly additive, which we will discuss more in SECTION 4.5.1.

Several generalizations of the channel exist. We will be looking at two.

Definition 54 (Generalized Platypus Channel). *Let $s, t \in [0, 1]$ with $s + t < 1$. Then, the generalized Platypus channel $\mathcal{N}_{s,t}$ is a qutrit-to-qutrit channel with Stinespring isometry*

$$V_{s,t} |i\rangle = \begin{cases} \sqrt{s} |00\rangle + \sqrt{1-s-t} |11\rangle + \sqrt{t} |22\rangle & i = 0 \\ |20\rangle & i = 1 \\ |21\rangle & i = 2. \end{cases}$$

For $t = 0$ we recover the original Platypus channel. This extra parameter, first studied in [SW24], allows for degradability to appear. Specifically,

- For $t \geq 1/2$ the channel is always antidegradable, and otherwise neither degradable nor antidegradable;
- For $t < 1/2$, with either $s + t = 1$ or $s = 0$, coherent information is strongly additive with degradable channels;
- For $t < 1/2$, but $s + t < 1$ and $s = 0$, coherent information is not strongly additive with degradable channels, but is conjectured to be weakly additive.

Definition 55 (ReMAD Channel). *Define the matrix*

$$\Gamma = \begin{pmatrix} \gamma_{0,0} & 0 & 0 & \cdots & 0 \\ \gamma_{1,0} & \gamma_{1,1} & 0 & \cdots & 0 \\ \gamma_{1,0} & \gamma_{1,1} & \gamma_{1,2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma_{d-1,0} & \gamma_{d-1,1} & \gamma_{d-1,2} & \cdots & \gamma_{d-1,d-1} \end{pmatrix}$$

of transition probabilities (that is, $\gamma_{j,k} \geq 0$ and $\sum_{k=0}^j \gamma_{j,k} = 1$). Then, the ReMAD (resonant multilevel amplitude damping) channel \mathcal{N}_Γ is a qudit to qudit channel with Stinespring isometry

$$U |j\rangle = \sum_{k=0}^j \sqrt{\gamma_{j,j-k}} |j-k\rangle |k\rangle.$$

This channel was introduced in [CG23], where the (anti-)degradability of the qutrit case (that is, $d = 3$) was characterized by assuming the (anti-)degrading maps are also ReMAD channels. This assumption was supported by numerical evidence, and we will validate this assumption in SECTION 6.1 through an analytical derivation of the same (anti-)degrading maps.

It will be helpful to write Γ as

$$\Gamma = \begin{pmatrix} 1 & 0 & 0 \\ r & 1-r & 0 \\ s & t & 1-s-t \end{pmatrix}$$

so that the isometry is

$$V_{r,s,t} |i\rangle = \begin{cases} |00\rangle & i = 0 \\ \sqrt{r} |01\rangle + \sqrt{1-r} |10\rangle & i = 1 \\ \sqrt{s} |02\rangle + \sqrt{t} |11\rangle + \sqrt{1-s-t} |20\rangle & i = 2. \end{cases}$$

Correspondingly, we write $\mathcal{N}_{r,s,t}$ for the channel. With the unitaries

$$U_B |i\rangle = \begin{cases} |2\rangle & i = 0 \\ |0\rangle & i = 1 \\ |1\rangle & i = 2 \end{cases} \quad U_E |i\rangle = \begin{cases} |1\rangle & i = 0 \\ |0\rangle & i = 1 \\ |2\rangle & i = 2, \end{cases}$$

we may compute that

$$(U_B \otimes U_E) V_{r,s,t} |i\rangle = \begin{cases} |21\rangle & i = 0 \\ \sqrt{r} |20\rangle + \sqrt{1-r} |01\rangle & i = 1 \\ \sqrt{t} |00\rangle + \sqrt{1-s-t} |11\rangle + \sqrt{s} |22\rangle & i = 2. \end{cases}$$

Therefore, it is clear that, up to relabeling the qutrit that Alice holds, $\mathcal{N}_{1,s,t} = \mathcal{N}_{t,s}$. That is, the ReMAD channel is a generalization of the generalized Platypus channel.

4.5 Potential Theoretical Frameworks for Non-Additivity

We have identified four main theoretical approaches currently employed to tackle the non-additivity of quantum capacity, and we briefly discuss them here.

4.5.1 Spin Alignment Conjecture

The spin alignment conjecture is concerned with minimizing the entropy of certain bipartitions that may be formed given a starting state. We start by introducing the notation $I_n = \{1, \dots, n\}$, and for any $I \subseteq I_n$ we define its complement $I^c = I_n \setminus I$. Then, consider two states $\rho, \sigma \in S(A)$. We understand the product $\sigma^{\otimes I} \otimes \rho^{\otimes I^c}$ as implicitly involving reordered tensors powers. For example, if $n = 3$ and $I = \{1, 3\}$,

$$\sigma^{\otimes I} \otimes \rho^{\otimes I^c} = \sigma \otimes \rho \otimes \sigma.$$

We now state the spin-alignment conjecture.

Conjecture 1. Fix $n \geq 2$ and a state $\rho \in S(A)$. Denote by $|\lambda\rangle$ the eigenvector associated with the maximal eigenvalue of ρ . Then, for any probability distribution $(p_I)_{I \subseteq I_n}$ and an indexed set of pure states $|\psi_I\rangle \in S(A)$,

$$S\left(\sum_{I \subseteq I_n} p_I |\psi_I\rangle \langle \psi_I|^{\otimes I} \otimes \rho^{\otimes I^c}\right) \geq S\left(\sum_{I \subseteq I_n} p_I |\lambda\rangle \langle \lambda|^{\otimes I} \otimes \rho^{\otimes I^c}\right).$$

Specifically, this is a family of conjectures, those of the spin alignment conjectures at n levels. The origin of the conjecture is in [Led+23], where the conjecture at n levels implies n -symbol additivity of the coherent information of the Platypus channel, among other related results. Technically, it is more accurate therefore to state that the spin-alignment conjecture is a framework for additivity, not non-additivity.

This conjecture has been generalized and connected to the study of majorization and various matrix norms [AK24]. Recently, it has been positively resolved to two levels, implying $Q^{(2)}(\mathcal{N}_s) = 2Q^{(1)}(\mathcal{N}_s)$ [Alh25].

4.5.2 ε -Log Singularities

Since additivity hinges on entropy quantities, which are related to the spectra of the states in question, small perturbations of the spectra in question can be used to isolate non-additive effects.

Definition 56 (Perturbation). Fix some state ρ . Then, we say a family of states $\rho(\varepsilon)$ for $\varepsilon \geq 0$ is a perturbation of ρ if there exists some $c > 0$ so $\|\rho(\varepsilon) - \rho\|_1 \leq \varepsilon c$.

Definition 57 (ε -Log Singularity Rate). Let $\rho(\varepsilon)$ be a perturbation of some state $\rho(0)$. Then,

$$\lim_{\varepsilon \rightarrow 0^+} \frac{S(\rho(\varepsilon)) - S(\rho(0))}{\varepsilon |\log \varepsilon|}$$

is the ε -log singularity rate of $\rho(\varepsilon)$.

Since coherent information is given as the difference of two entropies, the output of the channel versus its complement, quantum capacity of a channel \mathcal{N} can be studied by looking at the ε -log singularity rates of $\mathcal{N}(\rho(\varepsilon))$ versus $\mathcal{N}^c(\rho(\varepsilon))$. This argument was first used by [Sid21] to identify positive quantum capacity.

Proposition 17. Fix some channel \mathcal{N} . Let $\rho(\varepsilon)$ be a perturbation of a pure state ρ . Let r_B and r_E be the respective ε -log singularity rates of $\mathcal{N}(\rho(\varepsilon)) = \rho_B(\varepsilon)$ and $\mathcal{N}^c(\rho(\varepsilon)) = \rho_E(\varepsilon)$. If $r_B - r_E > 0$, then $Q(\mathcal{N}) > 0$.

Proof. First, note that looking at the Schmidt decomposition of ρ and the Stinespring form \mathcal{N} immediately implies $S(\mathcal{N}(\rho)) = S(\mathcal{N}^c(\rho))$. Then, notice that

$$\begin{aligned} Q^{(1)}(\mathcal{N}) &\geq S(\mathcal{N}(\rho(\varepsilon))) - S(\mathcal{N}^c(\rho(\varepsilon))) \\ &= [S(\rho_B(\varepsilon)) - S(\rho_B(0))] - [S(\rho_E(\varepsilon)) - S(\rho_E(0))]. \end{aligned}$$

From here, the positivity of the difference in rates immediately implies $0 < Q^{(1)}(\mathcal{N}) \leq Q(\mathcal{N})$. \square

The same work also examined the non-additivity of coherent information between two channels \mathcal{N} and \mathcal{M} . This involves an identical argument to the above, but using $\rho(0) = \rho_{\mathcal{N}} \otimes \rho_{\mathcal{M}}$, where

$$\rho_{\mathcal{N}} = \arg \max Q^{(1)}(\mathcal{N}) \text{ and } \rho_{\mathcal{M}} = \arg \max Q^{(1)}(\mathcal{M}).$$

Although it may initially seem that using the ε -log singularity rate is no more straightforward than computing the coherent information directly, that is not the case. Structural theorems may be proved pertaining to the ranks of certain inputs and outputs by using the singularity rate as a starting point. This allows for an understanding of coherent information from an algebraic perspective, and has been used to show the positivity of capacity for several important classes of channels [SD22]. Moreover, the rates for generic perturbations may be computed analytically, allowing for non-additivity to be demonstrated when it is numerically impossible due to the required perturbations being excessively small in practice, such as in [SW24].

4.5.3 Information Leakage

Another possible approach stems from [Sid25]. Here, we begin with a simple channel with isometry

$$V_p |i\rangle = \begin{cases} |00\rangle + \sqrt{\frac{1-p}{2}} (|11\rangle + |22\rangle) & i = 0 \\ |01\rangle & i = 1, \end{cases}$$

for $p \in [0, 1]$. The corresponding channel \mathcal{N} turns out to be degradable for $p \leq 1/3$ and antidegradable for $p \geq 1/3$, and therefore it is always additive.

We may then extend this to a larger channel $\tilde{\mathcal{N}}$ with the new isometry

$$\tilde{V}_p |i\rangle = \begin{cases} V_p |i\rangle & i = 0 \\ V_p |i\rangle & i = 1 \\ |0i\rangle & i \geq 2. \end{cases}$$

It turns out that $\tilde{\mathcal{N}}$ is antidegradable for $p \geq 1/2$, and is never degradable. However, $Q^{(2)}(\tilde{\mathcal{N}}) > 2Q^{(1)}(\tilde{\mathcal{N}})$ for $p_0 \leq p \leq 1/2$, where $p_0 \approx 0.23$.

This indicates that channels may exhibit non-additivity near degradable subchannels, even when the full channel only seems to perform a “useless” action (in this case, sending the extra ket to the environment, which will be immediately traced out and lost). This is the motivation for the specific parametre space over which we search for non-additivity in ReMAD channels in SECTION 6.3.

4.5.4 Private Channels

Finally, we briefly discuss an approach to get quantum capacity amplification via private channels, which also draws upon several of the previously covered frameworks. Such channels arise by first considering multipartite states designed to securely communicate keys in the presence of eavesdroppers, first discussed by [Hor+05]. Then, [WW25] shows that by viewing these states as Choi-Jamiołkowski matrices, one can recover quantum channels with interesting information-theoretic properties, ostensibly induced by the privacy concerns.

To illustrate, consider the channel

$$\mathcal{N}_{A'A \rightarrow B'B}(\rho) = \sum_{i,j=0}^{d-1} K_{ij} \rho K_{ij}^*$$

where $A' \cong B' \cong \mathbb{C}^2$ and $A \cong B \cong \mathbb{C}^d$ and the Kraus operators are

$$K_{ij} = \frac{1}{\sqrt{d}} \left(|0j\rangle \langle 0i|_{B'BA'A} + |1i\rangle \langle 1j|_{B'BA'A} \right).$$

It turns out the corresponding Choi-Jamiołkowski matrix is the so-called Werner state first introduced in [Wer89]. Due to [WW25], this channel has non-additive coherent information when combined with erasure channels, however weak additivity holds if the spin alignment conjecture is true.

A heuristic understanding of this behavior may be as follows. The privacy structure ensures that \mathcal{N} leaks the majority of its quantum information to the environment to ensure that the key can be communicated securely even in the presence of adversaries. However, when combined with a simple auxiliary channel, such as an erasure channel, some of this “lost” quantum information then becomes recoverable with positive probability. This recovery mechanism allows for a net increase in the overall capacity, resulting in quantum capacity amplification.

5 Numerical Searches for Non-Additive Classical Capacity

We perform numerical searches channel with non-additive classical capacity, using the fundamental inequality [THEOREM 17](#).

5.1 Numerical Framework and Challenges

5.1.1 Optimization of Bounds on Schmidt Coefficients

There are several challenges associated with the optimization problem presented in [PROPOSITION 11](#). In general, no analytical solutions exist and no easily exploitable structure, such as convexity, is present. Therefore, either a global optimizer must be employed, or a large amount of local optimizations must be performed. The only two pleasant characteristics are a constraint to a unit sphere and differentiability.

Empirically, the loss landscape is very ill-conditioned with several local minima. On top of simply being very computationally expensive, we find that global optimizers – of which we tested differential evolution, basinhopping, and simplicial homology global optimization – are no less prone to getting caught in these local minima than any local optimizer. Moreover, although this latter-most optimizer is capable of taking in the aforementioned constraints, this appears to only slow down the optimization without yielding more consistent results. For these reasons we instead perform a large amount of local optimizations, taking advantage of differentiability via PyTorch [[Pas+19](#)].

A technique we employ to save time is to initially use a small number of optimizations with relatively few random starts. Given the dimension of our subspace we can determine beforehand a range into which our optimal μ must fall for non-additivity to be possible. So, we can quickly discard a subspace if it falls out of this range. After performing these haphazard optimizations on a wide range of subspaces, we then re-examine any non-discarded ones with more computational resources.

Consider for example the later [TABLE 5.3](#), where for the subspace W in the first row we compute a value of $\mu \approx 0.9$. We know that $\mu \in [1/3, 2/3]$ is required for non-additivity.

So, it does not matter if the true optimal value is $\mu = 0.91$ or $\mu = 0.99$ – knowing $\mu > 2/3$ is sufficient to eliminate W .

Our code, which consumes a projection onto an arbitrary subspace W , and searches over its bipartitions and even smaller subspaces, is located in [LISTING C.1](#).

5.1.2 Numerically Constructing Orthonormal Bases

In order to construct our projection P_W we require a basis set. We have nice bases for the antisymmetric spaces W_{nk}^\wedge thanks to [PROPOSITION 13](#). However, we lack a general such expression for irreducible representations W_{nk}^λ and the to-be-introduced maximal completely entangled spaces W_{nk}^\otimes . We are thus forced to numerically compute a basis by orthonormalizing their spanning sets with the help of SciPy [[Vir+20](#)].

Unfortunately, such a task becomes numerically unstable even for relatively small tensor powers and dimensions. For instance, consider $n = 5$ and a partition λ with $|\lambda| = k = 3$. Any orthonormal basis of $W_{5,3}^\lambda$ will be given in terms of the $5^3 = 125$ basis tensors of $W_{5,3}$. From the built-in `orth`, which is already very stable as it makes use of the singular value decomposition, several coefficients of our basis elements are on the order of 10^{-17} . Once we reach $n = 6$, the case with $k = 3$ begins to fail, returning orthonormal bases of incorrect size.

One remedy we employ is by changing the underlying singular value decomposition technique employed by `orth`, as we find that some of these engines perform better than others. For instance, we find `gesdd` begins to fail for W_{nk}^\otimes when $n = 3$ and $k \geq 6$, and so we switch to `gesvd`. Another possible remedy could be to use sparse matrix versions of singular value decompositions. For example, we find that only 1 – 2% of the symmetrized vector coefficients for our spanning sets of W_{nk}^λ are non-zero. However, we did not investigate this in detail.

5.1.3 Solving for Minimal Non-Additive p

Numerical precision can become an issue for determining when exactly non-additivity occurs. In the proof of [THEOREM 19](#), we show the function

$$f(\mu, p) = ((1 - x)^p + \mu^p)^{\frac{2}{p}}$$

is decreasing in p for a fixed μ . In the event that we compute a Schmidt bound which is very close to the relative dimension of W in $B \otimes E$, we may need a very large choice of p

for non-additivity to occur. For example, suppose we have

$$c = \frac{\dim W}{\dim B \dim E} = 0.5625000000000001$$

and we compute a Schmidt bound of $\mu = 0.25$. Since

$$0.5625000000000001 = \frac{\dim W}{\dim B \dim E} > (1 - \mu)^2 = 0.5625,$$

we know that non-additivity must occur somewhere. A naïve computation will likely show that non-additivity is achieved for $p = 32$, because $f(\mu, 32) = 0.5625$. However, the reality here is that we are running into issues with machine precision, readily seen as $f(\mu, 31) = c$ exactly. So, it is impossible to tell where exactly this non-additivity occurs, even using 64-bit precision.

This example may seem contrived due to the specific choice of the relative dimension of W , and to a certain extent that is a valid remark. However, we genuinely do face this numerical issue, but rather due to our numerical and inexact evaluation of μ . It is easy to find false candidates for non-additivity if $(1 - \mu)^2$ and c are very close due to fluctuations in the the rightmost decimal places of μ from our stochastic optimizer, specifically after the 4th or 5th position.

To further complicate matters, $f(\mu, p)$ may also suddenly vanish. For example, we find that $f(\mu, p) = 0$ numerically for $p \geq 2590$, since $(1 - \mu)^{2590} = 0$ due to the limits of machine precision (and $\mu^p = 0$ for $p \ll 2590$ since $\mu \leq 1 - \mu$). Without being mindful of this, a naïve numerical evaluation of f to find the first p such that the inequality holds will lead to erroneous reporting.

We characterize the limits of numerical analysis in [TABLE 5.1](#), where we in fact consider the equivalent inequality

$$\frac{p}{2} \log c > \log ((1 - \mu)^p + \mu^p)$$

as it is marginally more stable due to avoiding one exponentiation. We do not report any non-additivity that we believe is sufficiently tenuous to be affected by numerical precision or stochasticity of our optimizer.

Table 5.1: Minimal Schmidt bound μ for which non-additive p -Rényi entropy may still be numerically computed, grouped in terms of relative size c of subspace W . Up to 64-bit precision, μ_{\min} is the smallest value of μ so $\mu^p > 0$ and non-additivity occurs for $p \in [1, 100]$. We also give $\mu_{\max} = 1 - \mu_{\min}$ to account for symmetry. Values of p which appear to violate monotonicity are in bold, indicating the challenges imposed by numerical precision in these edge cases.

c	μ_{\min}	μ_{\max}	p
0.30	0.4555	0.5445	13.87322
0.35	0.4109	0.5891	9.02101
0.40	0.3713	0.6287	6.22356
0.45	0.3317	0.6683	5.53038
0.50	0.2971	0.7029	4.26782
0.55	0.2624	0.7376	3.77269
0.60	0.2278	0.7722	3.67367
0.65	0.1981	0.8019	2.98050
0.70	0.1634	0.8366	4.63916
0.75	0.1387	0.8613	2.38645
0.80	0.1090	0.8910	2.26257
0.85	0.0793	0.9207	2.36159
0.90	0.0545	0.9455	1.79220
0.95	0.0298	0.9702	1.44561

5.2 Numerical Experiments

Due to [SS24], we know antisymmetric subspaces comprise explicit examples of non-additive subspaces. So, we continue this work and numerically test several other subspaces connected in some way to antisymmetry.

Overall, we hope to find a subspace with a large relative dimension c which also has a very large Schmidt bound μ for $\mu \in [0, 1/2]$ (if $\mu \in [1/2, 1]$, then by symmetry we hope that μ is very small).

5.2.1 Irreducible Representations of Symmetric Groups

We initially explored subspaces W_{nk}^λ associated to irreducible representations λ of S_n due to their natural connection with antisymmetric spaces. Indeed, antisymmetric spaces are one particular representation. Unfortunately, all other representations prove to be a poor choice due to their extraordinary small relative dimensions, in the following original result.

Proposition 18. *Suppose $\lambda \vdash n$ is a partition such that $|\lambda| = 2$. Then, asymptotically in n , W_{nk}^λ has its dimension maximized when $\lambda = (n/2, n/2)$.*

Proof. Let us consider a generic two-row partition $\lambda = (n - m, m)$. The hook-lengths of the individual boxes are

$$\lambda = \begin{array}{|c|c|} \hline d+1 & d \\ \hline m & m-1 \\ \hline \end{array} \cdots \begin{array}{|c|c|c|} \hline d+1-(m-1) & d-m & d-(m+1) \\ \hline 1 & & \\ \hline \end{array} \cdots \begin{array}{|c|} \hline 1 \\ \hline \end{array},$$

where $d = n - m$ is the length of the first row. So,

$$\begin{aligned} \prod_{(i,j) \in \lambda} h(i,j) &= \prod_{i=0}^{m-1} (d+1-i) \prod_{i=m}^{d-1} (d-i) \prod_{i=0}^{m-1} (m-i) \\ &= \prod_{i=1}^m (n-m+2-i) \prod_{i=m+1}^{n-m} (n-m+1-i) \prod_{i=1}^m (m+1-i). \end{aligned}$$

By examining the first and last indices, we see

$$\prod_{i=1}^m (n-m+2-i) = \frac{(n-m+1)!}{(n-2m+1)!}.$$

Moreover,

$$\prod_{i=m+1}^{n-m} (n-m+1-i) = (n-2m)(n-2m-1) \cdots 1 = (n-2m)!$$

and

$$\prod_{i=1}^m (m+1-i) = m(m-1)\cdots 1 = m!.$$

Therefore,

$$\prod_{(i,j)\in\lambda} h(i,j) = \frac{(n-m+1)!(n-2m)!m!}{(n-2m+1)!} = \frac{(n-m+1)!m!}{n-2m+1},$$

and so

$$\dim W_{nk}^\lambda = \frac{n!(n-2m+1)}{(n-m+1)!m!}.$$

Now, we take logarithms and write

$$\log \dim W_{nk}^\lambda = \log(n!) + \log(n-2m+1) - \log((n-m+1)!) - \log(m!).$$

Let us now write $m = xn$ for $x \in (0, 1)$, taking a continuous approximation as we are considering $n \rightarrow \infty$. This allows us to differentiate with respect to m , yielding

$$\frac{d}{dm} \log \dim W_{nk}^\lambda = -\frac{2}{n-2m+1} + \psi(n-m+2) - \psi(m+1),$$

where we use the digamma function

$$\psi(z) = \frac{d}{dz} \log \Gamma(z) \approx \log z - \frac{1}{2z} - \frac{1}{12z^2}.$$

Searching for its critical points, we are left to examine the equation

$$\psi(n-m+2) - \psi(m+1) = \frac{2}{n-2m+1},$$

which becomes

$$\log\left(\frac{n-m+2}{m+1}\right) - \frac{1}{2(n-m+2)} + \frac{1}{2(m+1)} = \frac{2}{n-2m+1}$$

after our asymptotic expansion up to the second term.

In terms of x , this becomes

$$\log\left(\frac{1-x+\frac{2}{n}}{x+\frac{1}{n}}\right) - \frac{1}{2n(1-x)+4} + \frac{1}{2(nx+1)} = \frac{2}{n(1-2x)+1}.$$

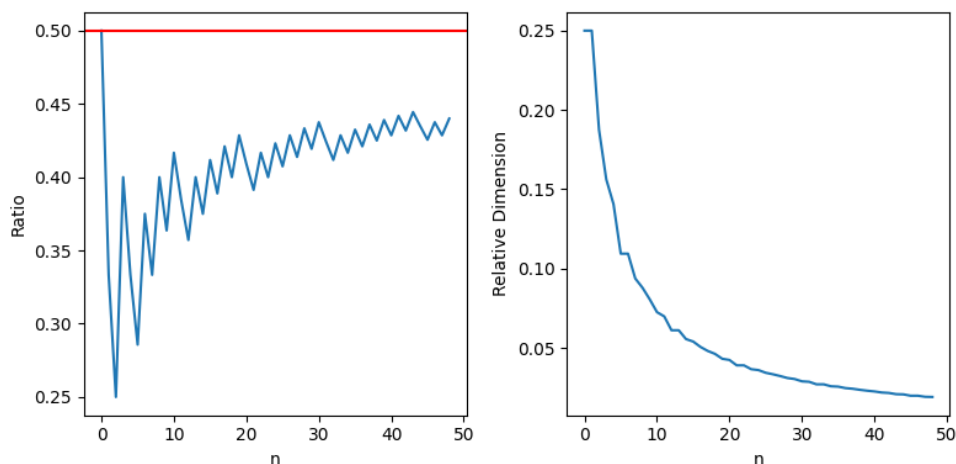


Figure 5.1: Descriptions of the irreducible representation W_{nk}^λ of maximal dimension relative to W_{nk} for $n = 2, \dots, 50$. As λ always satisfies $|\lambda| = 2$, the left plot shows the ratio λ_2/λ_1 of its two rows. The right plot is the relative dimension within the corresponding ambient space W_{n2} .

As $n \rightarrow \infty$, we are left with

$$\log\left(\frac{1-x}{x}\right) = 0,$$

which is attained when $x = 1/2$, as claimed. \square

Note that in the statement of the theorem, we are assuming n is always even. In reality, our result shows asymptotic maximization on $\lambda = (n - m, m)$, where m is the closest integer to $n/2$. This forms the basis for our conjecture.

Conjecture 2. For $n \geq 2$, there are no irreducible representations W_{nk}^λ with relative dimension exceeding $1/4$ except for the trivial representation $\lambda = (n)$.

Our conjecture is true if the partition λ in [PROPOSITION 18](#) is also the partition which maximizes $\dim W_{nk}^\lambda / |\lambda|^n$ among all partitions. The proof would simply be to witness that $\dim W_{nk}^\lambda / 2^n \rightarrow 0$ as $n \rightarrow \infty$, and then check smaller choices of n by hand. We illustrate our evidence for this conjecture in [FIGURE 5.1](#), where we check several dimensions to verify this is indeed the maximizing partition, and also that our asymptotics hold.

Few tools seem to exist which help to characterize the asymptotic behavior of irreducible representations of symmetric groups. The most developed theory is related

to the Plancherel measure

$$\mu(\lambda) = \frac{\left(\dim W_{nk}^\lambda\right)^2}{n!},$$

a measure which naturally defines a probability distribution on the irreducible representations (and may be defined for many groups, not just the symmetric group). Several statistics are known for this distribution, such as the typical shape of asymptotically large sampled partitions [VK76; LS77] as well as asymptotic bounds on their dimensions [VK85; Buf12]. Outside of the Plancherel measure, there are also results from character and representation theory which give bounds on the dimensions of irreducible representations of generic groups, although we are only aware of the existence of such results and do not know any particular details.

Unfortunately, we could not find a way to directly apply any of the above work to our conjecture. As we are concerned with the relative dimension, it is crucial to preserve information about $|\lambda| = k$, which the asymptotic dimensional bounds do not consider. Indeed, the most relevant bound is

$$\log \dim W_{nk}^\lambda \lesssim -\frac{c}{2}n + \frac{1}{2} \log(n!)$$

for some constant c , however this is not sharp enough in light of the lack of k . Moreover, other statistics, such as expected dimensions or variance, are not applicable as in our context we have uniformly- rather than Plancherel-distributed partitions. A possible approach could rely on statistics recently developed by [GOT24] on the individual hook lengths $h(i, j)$, rather than the total product.

In spite of this, we provide in [LISTING C.2](#) a framework for working with partitions and the associated subspaces W_{nk}^λ , as well as some bounds in [TABLE 5.2](#). We are also careful to remark that this conjecture does not imply W_{nk}^λ and its subspaces are necessarily additive. Their small dimension merely implies they cannot be additive within the framework established by [THEOREM 17](#). Indeed, it could be possible that these subspaces yield non-additivity through some other construction (which, at present, is unknown to the literature).

Table 5.2: Bounds of the Schmidt coefficients for subspaces and bipartitions of $W_{5,3}^\lambda$ with $\lambda = (1, 2, 5)$. For each dimension $3 \leq w \leq 6 = \dim W_{5,3}^\lambda$, we search across all subspaces $W \subseteq W_{5,3}^\lambda$ of dimension w , and across all bipartitions where Bob holds m components. The largest Schmidt coefficient found for each w and m pair is reported as μ_{\max} . We also include the relative dimension c and $N = \binom{6}{w}$, the number of subspaces W . This optimization took approximately 14 hours, with 25 trials of 1000 iterations each for every subspace W , and a learning rate of 10^{-4} .

w	m	μ_{\max}	c	N
3	1	0.42362076	0.0123	20
3	2	0.54235047	0.0123	20
4	1	0.49491119	0.0165	15
4	2	0.56314707	0.0165	15
5	1	0.47924414	0.0206	6
5	2	0.53285229	0.0206	6
6	1	0.4412941	0.0247	1
6	2	0.42078349	0.0247	1

5.2.2 Maximal Completely Entangled Spaces

A class of spaces with bonafide quantum structures are completely entangled spaces.

Definition 58 (Completely Entangled Space). *We call $W \subseteq W_{nk}$ completely entangled if it contains no decomposable tensors. That is, if there is no $w \in W$ such that $w = a_1 \otimes \cdots \otimes a_n$ for any $a_i \in \mathbb{C}^k$.*

Theorem 25. *Let W be completely entangled. Then,*

$$\dim W \leq k^n - nk + n - 1.$$

Moreover, there is some W which saturates the inequality.

Spaces which saturate the above inequality, so-called maximal completely entangled spaces, were originally explored in detail by [Par04]. In fact, the dimension bound was derived for a more general setting, that of completely entangled spaces within $\mathbb{C}^{k_1} \otimes \cdots \otimes \mathbb{C}^{k_n}$ where the k_i need not be equal. This work was continued by [Bha06], which provides us the explicit construction below necessary for numerical experiments.

Theorem 26. Let $\{e_i\}_{i=1}^k$ form an orthonormal basis of \mathbb{C}^k . Then, the completely entangled space of maximal dimension in W_{nk} , denoted by W_{nk}^\otimes , is spanned by

$$\left\{ e_{i_1} \otimes \cdots \otimes e_{i_n} - e_{j_1} \otimes \cdots \otimes e_{j_n} : \sum_{\ell=1}^n i_\ell = \sum_{\ell=1}^n j_\ell \right\}.$$

From here, we see the precedent for exploring W_{nk}^\otimes – it contains W_{nk}^\wedge as a subspace. However, in general W_{nk}^\otimes is significantly larger for $k > 2$.

A distinct initial advantage of using W_{nk}^\otimes seems to be that our search space would be relatively high-dimensional compared to the ambient space. In fact,

$$\frac{\dim W_{nk}^\otimes}{\dim B \dim E} = \frac{k^n - nk + n - 1}{k^n} \rightarrow 1$$

as $k \rightarrow \infty$ for any fixed $n > 2$. However, in this sense $W_{nk}^\otimes \rightarrow W_{nk}$ at the same time, and so we expect the Schmidt bound to tend to one simultaneously: $\mu \rightarrow 1$. These behaviors counteract each other when it comes to enforcing the non-additivity inequality.

For this reason, W_{nk}^\otimes and its subspaces are not very promising. Indeed, our numerical results in [TABLE 5.3](#) show the futility of examining the full space. The reliability of the numerical optimization also rapidly becomes of concern, as the largest spaces in our table have $\dim W_{nk}^\otimes \approx 500$. Factoring in complex coefficients, this would become a 1000-dimensional optimization problem, which as we have discussed before is already very poorly behaved.

Table 5.3: Schmidt bound μ computed for W_{nk}^{\otimes} . Performed 50 local optimizations running for 5000 maximum iterations for each with a learning rate of 10^{-4} , taking about 7 hours. Also given is the relative dimension $c = \dim W_{nk}^{\otimes} / k^n$ as well as the maximum Schmidt coefficient $\mu \in [1/2, 1]$ which could result in non-additivity $\mu_{\max} = \sqrt{c}$ for the particular relative dimension. In all cases, $\mu > \mu_{\max}$, meaning no non-additivity occurs.

n	k	c	μ	μ_{\max}
2	3	0.4444	0.90450752	0.6666
2	4	0.5625	0.97940618	0.7500
2	5	0.6400	0.98545271	0.8000
2	6	0.6944	0.98436028	0.8333
2	7	0.7347	0.98052782	0.8571
2	8	0.7656	0.98381998	0.8750
2	9	0.7901	0.98064214	0.8888
2	10	0.8100	0.98337519	0.9000
2	11	0.8264	0.98203647	0.9091
2	12	0.8403	0.97941673	0.9167
2	13	0.8521	0.98221046	0.9231
2	14	0.8622	0.98110616	0.9286
2	15	0.8711	0.98056281	0.9333
2	16	0.8789	0.98067427	0.9375
2	17	0.8858	0.97737449	0.9412
2	18	0.8920	0.97745657	0.9444
2	19	0.8975	0.97781932	0.9474
2	20	0.9025	0.97887099	0.9500
2	21	0.9070	0.97854578	0.9524
2	22	0.9111	0.97261387	0.9545
2	23	0.9149	0.97820729	0.9565
2	24	0.9184	0.97448325	0.9583
3	3	0.7407	0.98509163	0.8607
3	4	0.8438	0.98436832	0.9186
3	5	0.8960	0.98304224	0.9466
3	6	0.9259	0.98559076	0.9623
3	7	0.9446	0.98549736	0.9719
3	8	0.9570	0.98519462	0.9783
4	4	0.9492	0.98593330	0.9743
4	5	0.9728	0.99175471	0.9863

5.2.3 Extensions of Antisymmetric Spaces

Antisymmetric spaces W_{2k}^\wedge have already been studied in great detail by [SS24]. In particular, they explored the subspaces $W \subseteq W_{2k}^\wedge$ and found that they consistently yield non-additivity. Additionally, they explored small extensions of W_{2k}^\wedge , with the goal being to add a small amount vectors such that the Schmidt bound μ might not be perturbed too much, yet the space could become larger and yield non-additivity faster. We summarize their core results here.

Theorem 27. *For all $k \geq 7$ there exists some $p_0(k) \in (2, \infty)$ such that for each $p \in [p_0(k), \infty)$ there exists a non-trivial sequence of subspaces $(X_m)_{m=1}^{n_0}$ with $\dim X_m = m$ and $n_0 < \lceil k/2 \rceil$ such that $W_{2k}^\wedge \oplus X_m$ have non-additive p -Rényi entropy. In particular,*

$$p_0(k) = \frac{2 \log 2}{\log 2 + \log(k^2 - k + 2) - 2 \log(k + 2)}.$$

The above spaces X_n are constructed using a generalization of Bell states introduced in [Ben+93], which have easy-to-compute Schmidt bounds and are sufficiently numerous to find several which are orthogonal to W_{2k}^\wedge .

This construction is favourable in that it is explicit and has a well-understood structure. However, we find that it is not optimal, in the sense that non-additivity can be achieved for $p < p_0(k)$ while using extensions of the same size as X_m . We do this through a slightly less pleasant construction, however: a subspace we denote by W_{nk}^\wedge .

Let $d = \dim W_{nk}^\wedge$ with $\{v_i\}_{i=1}^d$ being an orthonormal basis of W_{nk}^\wedge . Next, let w_i form an orthonormal basis for $W_{nk}^\otimes = W_0$. Now, check if $v_1 \in W_0$, in which case we may uniquely write $v_1 = \sum_i c_i w_i$ where $c_i = \langle v_1, w_i \rangle$. Let j be some index such that $c_j \neq 0$, and define $W_1 = \text{span}\{w_i : i \neq j\}$. If $v_1 \notin W_0$, then define $W_1 = W_0$. Next, check if $v_2 \in W_1$ and similarly define W_2 . Repeat this until we arrive $W_d = W_{nk}^\wedge$. By construction, $W_{nk}^\wedge \cap W_{nk}^\wedge = \{0\}$ and W_{nk}^\wedge is completely entangled. Moreover, we have an orthonormal basis $\{\tilde{w}_i\}$ of W_{nk}^\wedge with we can work.

It is important to note that W_{nk}^\wedge is ill-defined, even in dimension. It depends not just on the choice of basis of used, but also on the indices removed. To best illustrate this we consider a toy example and let $\{e_1, e_2, e_3\}$ be an orthonormal basis for \mathbb{C}^3 and set $W_0 = \mathbb{C}^3$. Consider the two orthonormal bases

$$W_0 = \text{span}\{e_1, e_2, e_3\} = \text{span}\{f_+, f_-, e_3\},$$

where $f_{\pm} = (e_1 \pm e_2)/\sqrt{2}$. Suppose we wish to perform the above process on the space $\text{span}\{v_1, v_2\}$ where $v_1 = f_+$ and $v_2 = f_-$. With the first basis, we clearly see

$$v_1 = \frac{1}{\sqrt{2}}e_1 + \frac{1}{\sqrt{2}}e_2$$

and so we may choose to remove either e_1 or e_2 . Since $v_2 \notin \text{span}\{e_1, e_3\}$ and $v_2 \notin \text{span}\{e_2, e_3\}$, our process halts after either removal. On the other hand, with our second orthonormal basis we would have to remove both f_+ and f_- before stopping. So, we have three possible outcomes:

$$W_2 = \begin{cases} \text{span}\{e_1, e_3\} \\ \text{span}\{e_2, e_3\} \\ \text{span}\{e_3\}. \end{cases}$$

In light of this, we clarify that our calculations are always produced using the same starting spanning set from [THEOREM 26](#), we always use the same orthonormalization algorithm, and we always pick the minimal non-zero index c_j . This ensures our results are consistent and reproducible, with the procedure given in [LISTING C.3](#).

Our non-additive subspaces are of the form $W_{2k}^{\wedge} \oplus \{e\}$, where $e \in W_{2k}^{\bar{\wedge}}$. The relevant comparison to [THEOREM 27](#) is $W_{2k}^{\wedge} \oplus X_1$. We report our results in [TABLE 5.4](#), which indicates our novel construction is able to achieve non-additive p -entropy for $p < p_0(k)$.

Table 5.4: Non-additive extensions of W_{2k}^\wedge via the inclusion of a single vector from W_{2k}^\wedge . “Best Index” refers to which vector yielded the corresponding smallest values p_{\min} and μ_{\min} . The total number of non-additive examples found for each choice of k is reported in the last column, and p_{\max} is the largest value of non-additive p among all these examples. For comparison, $p_0(k)$ is given as the previously best known value of p_{\min} . Performed for 10000 iterations and 240 initial starts with a learning rate of 10^{-4} . Searching over a single index took approximately 15 minutes.

k	$p_0(k)$	p_{\min}	p_{\max}	μ_{\min}	Best Index	Number of Examples
7	16.725	-	-	-	-	0
8	9.340	-	-	-	-	0
9	6.882	3.908	3.908	0.6643585	8	1
10	5.656	3.014	3.193	0.64445359	4	4
11	4.920	2.444	6.038	0.59323102	8	8
12	4.431	2.709	5.310	0.63146943	54	7
13	4.082	2.553	4.448	0.61861765	0	17

6 ReMAD Channel Additivity and Degradability

In order to further understand the structure embedded in the Platypus family, we study the degradability and additivity of ReMAD channels and their subchannels. To recall, for $r, s, t \in [0, 1]$ such that $1 - s - t \geq 0$ we define the qutrit ReMAD channel

$$\mathcal{N}_{r,s,t}(\rho) = \begin{pmatrix} \rho_{00} + r\rho_{11} + s\rho_{22} & \sqrt{1-r}\rho_{01} + \sqrt{rt}\rho_{12} & \sqrt{1-s-t}\rho_{02} \\ \sqrt{1-r}\rho_{10} + \sqrt{rt}\rho_{21} & t\rho_{22} + (1-r)\rho_{11} & \sqrt{(1-r)(1-s-t)}\rho_{12} \\ \sqrt{1-s-t}\rho_{20} & \sqrt{(1-r)(1-s-t)}\rho_{21} & (1-s-t)\rho_{22} \end{pmatrix}$$

and its complement

$$\mathcal{N}_{r,s,t}^c(\rho) = \begin{pmatrix} \rho_{00} + (1-r)\rho_{11} + (1-s-t)\rho_{22} & \sqrt{r}\rho_{01} + \sqrt{(1-r)t}\rho_{12} & \sqrt{s}\rho_{02} \\ \sqrt{r}\rho_{10} + \sqrt{(1-r)t}\rho_{21} & r\rho_{11} + t\rho_{22} & \sqrt{rs}\rho_{12} \\ \sqrt{s}\rho_{20} & \sqrt{rs}\rho_{21} & s\rho_{22} \end{pmatrix}$$

through their Stinespring isometry

$$V_{r,s,t} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \sqrt{r} & 0 \\ 0 & 0 & \sqrt{s} \\ 0 & \sqrt{1-r} & 0 \\ 0 & 0 & \sqrt{t} \\ 0 & 0 & 0 \\ 0 & 0 & \sqrt{1-s-t} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

We develop a framework in `Mathematica` to aid in our symbolic studies of the degradability of both the ReMAD channel and its subchannels, available in [LISTING A.1](#). This framework implements, among other things, the symbolic computation of Kraus operators, transfer matrices, and Choi-Jamiołkowski matrices. Since symbolic analysis alone is prone to missing degradability regions, as discussed in [SECTION 4.3](#), we verify all our regions via the SDP in [COROLLARY 2](#) with an example implementation in [LISTING A.2](#).

6.1 Full Channel Analysis

6.1.1 Degradability

We begin with the study of degradability. As mentioned, [CG23] already derived the degradability regions for the qutrit ReMAD channel. However, they did so by assuming that any possible degrading map \mathcal{D} is itself another ReMAD channel, writing $\mathcal{D} = \mathcal{N}_{\tilde{r}, \tilde{s}, \tilde{t}}$. They cite the intractability of an exact analytical approach as the reason for this assumption, but provided numerical evidence supporting its validity. By applying a composition rule for $\mathcal{D} \circ \mathcal{N}_{r,s,t}$, they then derive the necessary constraints on $\tilde{r}, \tilde{s}, \tilde{t}$. We verify the validity of their assumption and results by rederiving the same degradability region, but analytically for the first time.

Before we begin, we state that

$$\mathcal{T}_{\mathcal{N}_{r,s,t}^c} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1-r & 0 & 0 & 0 & 1-s-t \\ 0 & \sqrt{r} & 0 & 0 & 0 & \sqrt{t(1-r)} & 0 & 0 & 0 \\ 0 & 0 & \sqrt{s} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sqrt{r} & 0 & 0 & 0 & \sqrt{t(1-r)} & 0 \\ 0 & 0 & 0 & 0 & r & 0 & 0 & 0 & t \\ 0 & 0 & 0 & 0 & 0 & \sqrt{rs} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sqrt{s} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sqrt{rs} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & s \end{pmatrix}.$$

We omit writing $\mathcal{T}_{\mathcal{N}_{r,s,t}}$ for space, but note it is identical to the above except for the substitutions $r \leftrightarrow 1-r$ and $s \leftrightarrow 1-s-t$.

Theorem 28. *The ReMAD channel $\mathcal{N}_{r,s,t}$ is degradable and antidegradable under the respective disjoint constraints*

$$\begin{cases} 1-s-t \neq 0 \\ r \leq \frac{1}{2} \\ (1-r)(1-2s) + (3r-2)t > 0 \end{cases} \quad \begin{cases} s \neq 0 \\ r \geq \frac{1}{2} \\ r(2s-1) + (1-r)t > 0. \end{cases}$$

Moreover, the degrading or antidegrading map is always another ReMAD channel.

Proof. For some generic parametres k_1, k_2, ℓ_1, ℓ_2 , and α , define the matrix

$$D = \begin{pmatrix} 1 & 0 & 0 & 0 & \ell_2 & 0 & 0 & 0 & \alpha \\ 0 & \sqrt{\ell_1} & 0 & 0 & 0 & \ell_2 \sqrt{k_2} & 0 & 0 & 0 \\ 0 & 0 & \sqrt{k_1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sqrt{\ell_1} & 0 & 0 & 0 & \ell_2 \sqrt{k_2} & 0 \\ 0 & 0 & 0 & 0 & \ell_1 & 0 & 0 & 0 & \ell_2 k_2 \\ 0 & 0 & 0 & 0 & 0 & \sqrt{\ell_1 k_1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sqrt{k_1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sqrt{\ell_1 k_1} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & k_1 \end{pmatrix}$$

whose involution assuming a $\mathbb{C}^3 \otimes \mathbb{C}^3$ structure is

$$\vartheta(D) = \begin{pmatrix} 1 & 0 & 0 & 0 & \sqrt{\ell_1} & 0 & 0 & 0 & \sqrt{k_1} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \ell_2 & 0 & 0 & 0 & \ell_2 \sqrt{k_2} & 0 \\ \sqrt{\ell_1} & 0 & 0 & 0 & \ell_1 & 0 & 0 & 0 & \sqrt{\ell_1 k_1} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \alpha & 0 & 0 \\ 0 & 0 & 0 & \ell_2 \sqrt{k_2} & 0 & 0 & 0 & \ell_2 k_2 & 0 \\ \sqrt{k_1} & 0 & 0 & 0 & \sqrt{\ell_1 k_1} & 0 & 0 & 0 & k_1 \end{pmatrix}.$$

Within this proof, note that we will now write $\mathcal{N} = \mathcal{N}_{r,s,t}$ for notational simplicity.

Now, set

$$k_1 = \frac{s}{1-s-t} \quad k_2 = \frac{t}{1-s-t} \quad \ell_1 = \frac{r}{1-r} \quad \ell_2 = \frac{1-2r}{1-r}$$

and

$$\alpha = \frac{(1-r)(1-2s) + (3r-2)t}{(1-r)(1-s-t)}.$$

Of course, we must assume here that $1-r \neq 0$ and $1-s-t \neq 0$. A straightforward calculation then shows that $D\mathcal{T}_{\mathcal{N}} = \mathcal{T}_{\mathcal{N}^c}$ and $\text{tr}_B(\vartheta(D)) = \mathbb{1}_3$. Moving on to positivity, we find that

$$\sigma(\vartheta(D)) = \left\{ 0, \frac{(1-s)(1-2r)}{(1-s-t)(1-r)}, \frac{1-r-2s+2rs-2t+3rt}{(1-r)(1-s-t)}, \frac{1}{1-r} + \frac{s}{1-s-t} \right\}.$$

The last term is always positive, and the second term is non-negative precisely when $r \leq 1/2$. The remaining non-zero term reduces to the requirement that

$$(1-r)(1-2s) + (3r-2)t > 0.$$

Under these conditions, D thus corresponds to the transfer matrix of a quantum channel, which we may explicitly write as

$$\mathcal{D}(\rho) = \begin{pmatrix} \rho_{00} + \frac{(2r-1)(1-s-t)\rho_{11} + ((1-r)(1-2s) + (3r-2)t)\rho_{22}}{(1-r)(1-s-t)} & \sqrt{\frac{r}{1-r}}\rho_{01} + \frac{2r-1}{1-r}\sqrt{\frac{t}{1-s-t}}\rho_{12} & \sqrt{\frac{s}{1-s-t}}\rho_{02} \\ \sqrt{\frac{r}{1-r}}\rho_{10} + \frac{2r-1}{1-r}\sqrt{\frac{t}{1-s-t}}\rho_{21} & \frac{r(1-s-t)\rho_{11} + (1-2r)t\rho_{22}}{(1-r)(1-s-t)} & \sqrt{\frac{rs}{(1-r)(1-s-t)}}\rho_{12} \\ \sqrt{\frac{s}{1-s-t}}\rho_{20} & \sqrt{\frac{rs}{(1-r)(1-s-t)}}\rho_{21} & \frac{s}{1-s-t}\rho_{22} \end{pmatrix}.$$

It is not terribly hard to see that this is indeed another ReMAD channel, especially from the form of D and the known form of \mathcal{T}_N . Explicitly, we have $\mathcal{D} = \mathcal{N}_{r',s',t'}$ with

$$\begin{aligned} r' &= \frac{2r-1}{1-r} \\ s' &= \frac{(1-2s)(1-r) + (3r-2)t}{(1-r)(1-s-t)} \\ t' &= \frac{(1-2r)t}{(1-r)(1-s-t)}. \end{aligned}$$

Alternatively, we may set

$$k_1 = \frac{1-s-t}{s} \quad k_2 = \frac{t}{s} \quad \ell_1 = \frac{1-r}{r} \quad \ell_2 = \frac{2r-1}{r}.$$

and

$$\alpha = \frac{r(2s-1) + (1-r)t}{rs}.$$

Here, we have the assumptions $r \neq 0$ and $s \neq 0$. In this case we have $D\mathcal{T}_{N^c} = \mathcal{T}_N$ partial trace condition by more straightforward computations. We then derive symbolically again the spectrum

$$\sigma(\vartheta(D)) = \left\{ 0, \frac{(2r-1)(s+t)}{rs}, \frac{r(2s-1) + (1-r)t}{rs}, \frac{s+r(1-s-t)}{rs} \right\}.$$

The last term is always positive, and the second term is positive exactly when $r \geq 1/2$. This leaves us with the third eigenvalue which requires

$$r(2s - 1) + (1 - r)t > 0$$

for positivity.

When we satisfy these conditions, we find the antidegrading map

$$\mathcal{D}(\rho) = \begin{pmatrix} \rho_{00} + \frac{(2r-1)s\rho_{11} + (t-r(1-2s+t))\rho_{22}}{rs} & \sqrt{\frac{1-r}{r}}\rho_{01} + \frac{2r-1}{r}\sqrt{\frac{t}{s}}\rho_{12} & \sqrt{\frac{1-s-t}{s}}\rho_{02} \\ \sqrt{\frac{1-r}{r}}\rho_{10} + \frac{2r-1}{r}\sqrt{\frac{t}{s}}\rho_{21} & \frac{s(1-r)\rho_{11} + (2r-1)t\rho_{22}}{rs} & \sqrt{\frac{(1-r)(1-s-t)}{rs}}\rho_{12} \\ \sqrt{\frac{1-s-t}{s}}\rho_{21} & \sqrt{\frac{(1-r)(1-s-t)}{rs}}\rho_{21} & \frac{1-s-t}{s}\rho_{22} \end{pmatrix}.$$

Once more, this is a ReMAD channel $\mathcal{D} = \mathcal{N}_{r',s',t'}$ with

$$\begin{aligned} r' &= \frac{2r-1}{r} \\ s' &= \frac{t-r(1-2s+t)}{rs} \\ t' &= \frac{(2r-1)t}{rs}. \end{aligned}$$

We are now left with the edge cases $1 - s - t = 0$ or $r = 1$, and $s = 0$ or $r = 0$, respectively from our degradability and antidegradability analysis.

Suppose that $1 - s - t = 0$ but that $s \neq 0$ and $r \neq 0$. Then, the equation $X\mathcal{T}_N = \mathcal{T}_N^c$ has no solution in X , meaning degradability is impossible. Solving for a potential antidegrading transfer matrix, we find that it retains the form of D with a valid partial trace and associated spectrum

$$\sigma(\vartheta(D)) = \left\{ 0, \frac{1}{r'}, \frac{r+t-3rt}{rs}, \frac{(2r-1)(1-t)^2}{rs^3} \right\}.$$

As before, the last term imposes that $r \geq 1/2$. Then, because $t = 1 - s$, we see

$$r + t - 3rt = r - 2rt + t - rt = r(2 - 2t - 1) + (1 - r)t = r(2s - 1) + (1 - r)t.$$

Thus, our region stays the same when we impose positivity.

Say now that $s = 0$ but $1 - s - t \neq 0$ and $r \neq 1$. We find that $X\mathcal{T}_{\mathcal{N}^c} = \mathcal{T}_{\mathcal{N}}$ has no solution, so antidegradability is impossible. Moving to degradability, we do find a possible transfer matrix with the same form as D , with a valid partial trace and associated spectrum

$$\sigma(\vartheta(D)) = \left\{ 0, \frac{1}{1-r}, \frac{1-2t+r(3t-1)}{(1-r)(1-t)}, \frac{1-2r}{(1-r)(1-t)} \right\}.$$

First, notice that $1 - t \neq 0$ since $s = 0$. Moving on, our last eigenvalue still necessitates $r \leq 1/2$. Then, witness that

$$1 - 2t + r(3t - 1) = 1 - r + 3tr - 2t = (1 - r)(1 - 2s) + t(3r - 2),$$

and thus we also recover our previous region when needing positivity.

We now wrap up the last edge cases. If both $s = 0$ and $1 = t$, then both $X\mathcal{T}_{\mathcal{N}} = \mathcal{T}_{\mathcal{N}^c}$ and $X\mathcal{T}_{\mathcal{N}^c} = \mathcal{T}_{\mathcal{N}}$ lack solutions, so neither degradability nor antidegradability can occur.

If $r = 0$, then $X\mathcal{T}_{\mathcal{N}^c} = \mathcal{T}_{\mathcal{N}}$ has no solution, so no antidegradability is possible. However, a possible transfer matrix in the form of D does exist, with a valid partial trace and associated spectrum

$$\sigma(\vartheta(D)) = \left\{ 0, \frac{1-s}{1-s-t}, \frac{1-t}{1-s-t}, 2 - \frac{1}{1-s-t} \right\}.$$

The constraint here is $1 - s - t \geq 1/2$ due to the last eigenvalue, and we see

$$2 - 2s - 2t - 1 = (1 - r)(1 - 2s) + (3r - 2)t,$$

the same as our previously derived degrading region.

From [SECTION 4.4](#), we know \mathcal{N} is equivalent to the generalized Platypus channel when $r = 1$, whose degradability conditions are already known. So, through reparametrization, we see that $s \geq 1/2$ implies antidegradability, and that degradability is impossible. Note too that this aligns with our previous antidegradability condition since

$$2s - 1 = r(2s - 1) + (1 - r)t.$$

With all of these edge cases covered, we have explored the full parameter space. To see that the derived regions are disjoint, note that they could only possibly overlap when $r = 1/2$. However, our degradability constraint becomes $1 - 2s - t > 0$, while the antidegradability constraint is $1 - 2s - t < 0$, which of course cannot both be true. \square

We present a visualization of these regions in [FIGURE 6.1](#).

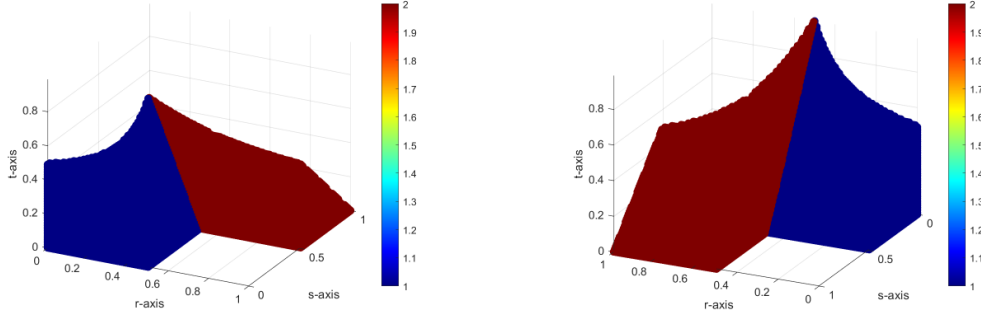


Figure 6.1: Degradable and antidegradable regions of the ReMAD channel $\mathcal{N}_{r,s,t}$, respectively in blue and red.

6.1.2 Coherent Information

We numerically evaluated $Q^{(1)}(\mathcal{N}_{r,s,t})$ across the entire parameter space, visualized in [FIGURE 6.2a](#). We did this on top of the existing MATLAB library developed by [\[Cub13\]](#), with our framework for computing capacities in [LISTING B.1](#). Upon examining our numerical data, we pose the following conjecture.

Conjecture 3. *Single-symbol coherent information for ReMAD channels is always optimized on a diagonal state. That is,*

$$\arg \max Q^{(1)}(\mathcal{N}_{r,s,t}) \ni \text{diag}(\rho_{00}, \rho_{11}, \rho_{22}).$$

This conjecture is known to be true over the regions where the ReMAD channel is degradable, as demonstrated in [\[CG23\]](#), which we briefly reprove here.

Proposition 19. *Suppose r, s, t are such that $\mathcal{N}_{r,s,t}$ is degradable. Then, its coherent information is given by a diagonal state.*

Proof. Define the unitaries

$$U(\vartheta) = \text{diag}(1, e^{-i\vartheta}, e^{-i2\vartheta}).$$

A straightforward calculation will show $U(\vartheta)\mathcal{N}_{r,s,t}(\rho)U^*(\vartheta) = \mathcal{N}_{r,s,t}(U(\vartheta)\rho U^*(\vartheta))$. Then, define the twirling operator

$$T(\rho) = \int_0^{2\pi} U(\vartheta)\rho U^*(\vartheta) d\vartheta = \text{diag}(\rho_{00}, \rho_{11}, \rho_{22}).$$

Since

$$U(\vartheta)\rho U^*(\vartheta) = \begin{pmatrix} \rho_{00} & e^{i\vartheta}\rho_{01} & e^{2i\vartheta}\rho_{02} \\ e^{-i\vartheta}\rho_{10} & \rho_{11} & e^{-i\vartheta}\rho_{12} \\ e^{-2i\vartheta}\rho_{21} & e^{-i\vartheta}\rho_{21} & \rho_{22} \end{pmatrix},$$

we see that $T(\rho) = \text{diag}(\rho_{00}, \rho_{11}, \rho_{22})$. Due to [THEOREM 21](#) we see

$$I_c(\mathcal{N}_{r,s,t}, T(\rho)) \geq \int_0^{2\pi} I_c(\mathcal{N}_{r,s,t}, U(\vartheta)\rho U^*(\vartheta)) d\vartheta = I_c(\mathcal{N}, \rho)$$

due to the invariance of entropy under conjugation by unitaries. Since $T(\rho)$ is diagonal, we see it is sufficient to optimize I_c over such states. \square

This result uses standard techniques which apply immediately to other channels invariant under similar unitary actions, such as in [\[DBF07; Ouy14\]](#). However, degradability is a necessary assumption, as the argument requires concavity. It may be possible that concavity exists as well under some conditions other than degradability. Regardless, it is interesting to observe that off-diagonal terms of ReMAD outputs rely solely on off-diagonal channel inputs.

The basis for our conjecture is that the maximal arguments for coherent information always have vanishingly small off-diagonal entries, by which we mean $|\rho_{ij}| < 10^{-3}$ whenever $i \neq j$ (with most being on the order of 10^{-8}).

Note that, due to the local optimization procedure employed, some optimal states upon a first glance may appear to have non-zero off-diagonal terms. In fact, a naïve pass will show this is the case $\sim 20\%$ of the time! However, these cases are all easily identifiable as spurious or misleading.

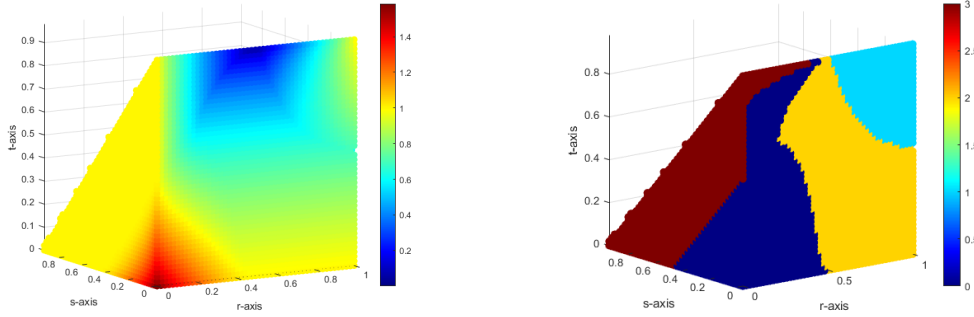
We observe that nearly all of these instances are in the antidegradable regions found in [THEOREM 28](#), where we know $Q^{(1)}(\mathcal{N}_{r,s,t}) = 0$. Then, since it is immediate $\mathcal{N}_{r,s,t}(|0\rangle\langle 0|) = |0\rangle\langle 0| = \mathcal{N}_{r,s,t}^c(|0\rangle\langle 0|)$, we may indeed find a diagonal optimizer here.

The remaining instances outside of our antidegradable region, which account for $\sim 0.1\%$ of our outputs, are simply optimization errors. They are readily identified because they appear as discontinuities in the coherent information. By iterating once more over these points, we find that they do indeed have diagonal optimizers.

In any case, we naturally identify four subspaces, the respective diagonal supports

$$W_{ij} = \text{span}\{|i\rangle\langle i|, |j\rangle\langle j|\} \text{ and } W = \text{span}\{|0\rangle\langle 0|, |1\rangle\langle 1|, |2\rangle\langle 2|\}.$$

These regions are visible in [FIGURE 6.2b](#).



(a) $Q^{(1)}(\mathcal{N}_{r,s,t})$ computed using the natural log, maximized at $\log 2 \approx 1.41$ (b) $\arg \max Q^{(1)}(\mathcal{N}_{r,s,t})$, with associated subspaces $0 = W, 1 = W_{12}, 2 = W_{02}, 3 = W_{01}$

Figure 6.2: Analysis of the coherent information of the ReMAD channel $\mathcal{N}_{r,s,t}$.

6.2 Subchannel Degradability Analysis

As the optimizers of coherent information for the full channel appear to always be diagonal, we have three natural subchannels restrictions to consider: to W_{01} , W_{02} , or W_{12} . Appropriately, we use the notation of \mathcal{N}_{ij} being the restriction of $\mathcal{N}_{r,s,t}$ to W_{ij} . Note that we understand all of the subchannels through a precomposition with an inclusion, and so they are all qubit channels. Specifically, the channels are

$$\mathcal{N}_{01}(\rho) = \begin{pmatrix} \rho_{00} + r\rho_{11} & \sqrt{1-r}\rho_{01} & 0 \\ \sqrt{1-r}\rho_{10} & (1-r)\rho_{11} & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\mathcal{N}_{02}(\rho) = \begin{pmatrix} \rho_{00} + s\rho_{11} & 0 & \sqrt{1-s-t}\rho_{01} \\ 0 & t\rho_{11} & 0 \\ \sqrt{1-s-t}\rho_{10} & 0 & (1-s-t)\rho_{11} \end{pmatrix}$$

$$\mathcal{N}_{12}(\rho) = \begin{pmatrix} r\rho_{00} + s\rho_{11} & \sqrt{rt}\rho_{01} & 0 \\ \sqrt{rt}\rho_{10} & (1-r)\rho_{00} + t\rho_{11} & \sqrt{(1-r)(1-s-t)}\rho_{01} \\ 0 & \sqrt{(1-r)(1-s-t)}\rho_{10} & (1-s-t)\rho_{11} \end{pmatrix}$$

with complements

$$\begin{aligned}\mathcal{N}_{01}^c(\rho) &= \begin{pmatrix} \rho_{00} + (1-r)\rho_{11} & \sqrt{r}\rho_{01} & 0 \\ \sqrt{r}\rho_{10} & r\rho_{11} & 0 \\ 0 & 0 & 0 \end{pmatrix} \\ \mathcal{N}_{02}^c(\rho) &= \begin{pmatrix} \rho_{00} + (1-s-t)\rho_{11} & 0 & \sqrt{s}\rho_{01} \\ 0 & t\rho_{11} & 0 \\ \sqrt{s}\rho_{10} & 0 & s\rho_{11} \end{pmatrix} \\ \mathcal{N}_{12}^c(\rho) &= \begin{pmatrix} (1-r)\rho_{00} + (1-s-t)\rho_{11} & \sqrt{(1-r)t}\rho_{01} & 0 \\ \sqrt{(1-r)t}\rho_{10} & r\rho_{00} + t\rho_{11} & \sqrt{rs}\rho_{01} \\ 0 & \sqrt{rs}\rho_{10} & s\rho_{11} \end{pmatrix}.\end{aligned}$$

We will focus on \mathcal{N}_{12} as it has the most complicated structure, being the only one which includes all three parameters r, s, t . In the rest of this section, we will derive its degradability and antidegradability regions, summarizing our results in [TABLE 6.1](#).

We begin with degradability. Using the same tools from our full channel analysis, we may generate a Choi-Jamiołkowski matrix for our (potential) degrading map \mathcal{D} :

$$\mathcal{J}_{\mathcal{D}} = \begin{pmatrix} \frac{(1-r)(1-s-2t)}{(1-r)s-rt} & 0 & 0 & 0 & \sqrt{\frac{(1-r)}{r}} & 0 & 0 & 0 & 0 \\ 0 & \frac{(1-2r)t}{(1-r)s-rt} & 0 & 0 & 0 & \sqrt{\frac{s}{t}} & 0 & 0 & 0 \\ 0 & 0 & \frac{(1-r)s}{(1-r)s-rt} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{s-r(1-t)}{(1-r)s-rt} & 0 & 0 & 0 & 0 & 0 \\ \sqrt{\frac{(1-r)}{r}} & 0 & 0 & 0 & \frac{r(t-s)}{(1-r)s-rt} & 0 & 0 & 0 & 0 \\ 0 & \sqrt{\frac{s}{t}} & 0 & 0 & 0 & 0 - \frac{rs}{(1-r)s-rt} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

It has spectrum

$$\sigma(\mathcal{J}_{\mathcal{D}}) = \left\{ 0, \frac{(1-r)s}{(1-r)s-rt}, \frac{s-r(1-t)}{(1-r)s-rt}, \right. \\ \left. \frac{(r(s+2t)-t)\sqrt{t} \pm \sqrt{4(1-r)^2s^3 + r(9r-8)ts^2 + 2rst^2 + (1-2r)^2t^3}}{2(rt - (1-r)s)\sqrt{t}}, \right. \\ \left. \left(\pm \sqrt{r(4s^2 + r^2(1+2s-t)(2(3s+4t-1) - r(2s+5t-1)) - r(11s^2 + 2s - 1 + 4(1+s) - 4t^2))} \right. \right. \\ \left. \left. + r(r(1-2s-t) + s + 2t - 1) \right) / (2r((r-1)s + rt)) \right\}$$

and partial trace

$$\text{tr}_B(\mathcal{J}_{\mathcal{D}}) = \begin{pmatrix} \frac{1-r-t}{(1-r)s-rt} & 0 & 0 \\ 0 & \frac{s-r}{(1-r)s-rt} & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The first constraint we identify is on the partial trace, where we need

$$\frac{1-r-t}{(1-r)s-rt} = 1 = \frac{s-r}{(1-r)s-rt}.$$

The first equality implies $(1-r)(1-s-t) = 0$, and therefore $r = 1$ or $1-s = t$. The second implies $r(1-s-t) = 0$. However, note that $r = 0$ is presently inadmissible due to terms proportional to $1/r$ in $\mathcal{J}_{\mathcal{D}}$.

Let us begin with $r = 1$, where we have

$$\sigma(\mathcal{J}_{\mathcal{D}}) = \left\{ 0, \frac{1-s-t}{t}, 1 \pm \frac{s}{t} \right\}.$$

We see this is a non-negative spectrum when $t \geq s$. Our partial trace is

$$\text{tr}_E(\mathcal{J}_{\mathcal{D}}) = \text{diag}\left(1, \frac{1-s}{t}, 0\right).$$

So, whenever $1-s = t$ we therefore have the partial trace condition met, and thus degradability.

Moving on, merely assuming $1-s-t = 0$ is unfortunately insufficient as the spectrum is still far too complicated to handle analytically, despite being in terms of only r and s .

However, let us examine the edge case $s = 0$, which further implies that $t = 1$. We have the great simplification

$$\sigma(\mathcal{J}_{\mathcal{D}}) = \left\{ 0, \frac{1}{r}, 2 - \frac{1}{r} \right\}.$$

We thus see any choice $r \geq 1/2$ works here.

Antidegradability is next. Our new Choi-Jamiołkowski matrix associated to the possible antidegrading map \mathcal{D} is

$$\mathcal{J}_{\mathcal{D}} = \begin{pmatrix} \frac{r(s-t)}{r(1-s)-t} & 0 & 0 & 0 & \sqrt{\frac{r}{(1-r)}} & 0 & 0 & 0 & 0 \\ 0 & \frac{(2r-1)t}{r(1-s)-t} & 0 & 0 & 0 & \sqrt{\frac{1-s-t}{t}} & 0 & 0 & 0 \\ 0 & 0 & \frac{r(1-s-t)}{r(1-s)-t} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{r(1-t)-s}{r(1-s)-t} & 0 & 0 & 0 & 0 & 0 \\ \sqrt{\frac{r}{(1-r)}} & 0 & 0 & 0 & \frac{(1-r)(1-s-2t)}{r(1-t)-t} & 0 & 0 & 0 & 0 \\ 0 & \sqrt{\frac{1-s-t}{t}} & 0 & 0 & 0 & \frac{(1-r)(1-s-t)}{r(s-1)+t} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

For brevity, we do not include the full spectrum in generality here, but remark that it is just as unpleasant as the previously derived one. Now, taking its partial trace yields

$$\text{tr}_E(\mathcal{J}_{\mathcal{D}}) = \text{diag} \left(\frac{r-t}{r(1-s)-t}, \frac{r-s-t}{r(1-s)-t}, 0 \right).$$

We are constrained to have

$$\frac{r-t}{r(1-s)-t} = 1 = \frac{r-s-t}{r(1-s)-t}.$$

Now, the first condition implies $rs = 0$ while the second implies $s(1-r) = 0$. We note however that $r = 1$ is presently inadmissible as it makes the above Choi-Jamiołkowski matrix ill-defined.

We start with $r = 0 = s$, where we find

$$\sigma(\mathcal{J}_{\mathcal{D}}) = \left\{ 0, \frac{1}{t}, \frac{(2t-1)\sqrt{t} + \pm\sqrt{t(2t-1)^2}}{2t^{3/2}} \right\}.$$

Note that the positive version of the radical is always non-negative. We are instead left to consider the negative version, which leads us to the inequality

$$(2t - 1) \geq \sqrt{(2t - 1)^2} = |2t - 1|.$$

We see this is satisfied for $t \geq 1/2$.

Next, we consider $s = 0$, and the special case $t = 1$. Here, we have

$$\sigma(\mathcal{J}_{\mathcal{D}}) = \left\{ 0, \frac{1}{1-r}, \frac{(1-2r)\sqrt{(1-r)} \pm \sqrt{(1-r)(1-2r)^2}}{2(1-r)^{3/2}} \right\}.$$

The analysis here is similar to before. We are left to consider

$$(1 - 2r) \geq \sqrt{(1 - 2r)^2} = |1 - 2r|$$

which entails $r \leq 1/2$.

Table 6.1: Degradable and antidegradable regions of the subchannel \mathcal{N}_{12} . The regions are verified using [COROLLARY 2](#), in particular to ensure the specific edge cases examined in order to have tractable spectra are indeed the only relevant regions.

Type	r	s	t
Degradable	$r = 1$	$s \leq 1/2$	$t = 1 - s$
Degradable	$r \geq 1/2$	$s = 0$	$t = 1$
Antidegradable	$r = 0$	$s = 0$	$t \geq 1/2$
Antidegradable	$r \leq 1/2$	$s = 0$	$t = 1$

6.3 Additivity of Coherent Information

Our reason for characterizing the degradable and antidegradable regions of \mathcal{N}_{12} is to examine the coherent information of the full channel in and around the region where this subchannel is degradable. We recall from [SECTION 4.5.3](#) that simple non-degradable extensions of degradable channels can achieve non-additivity. Moreover, the notion of domination discussed in [\[SW24\]](#) shows that it is possible for subchannels to encapsulate the behavior of the full channel, in regards to the properties of their coherent information.

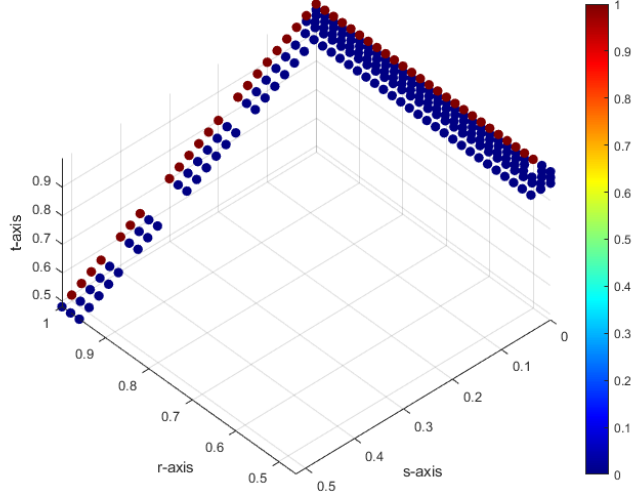


Figure 6.3: Search space for non-additivity of coherent information for $\mathcal{N}_{r,s,t}$. Red indicates points where \mathcal{N}_{12} is degradable. Note that gaps of points are areas missing due to the constraint $1 - s - t \geq 0$. We use a grid resolution of 0.02.

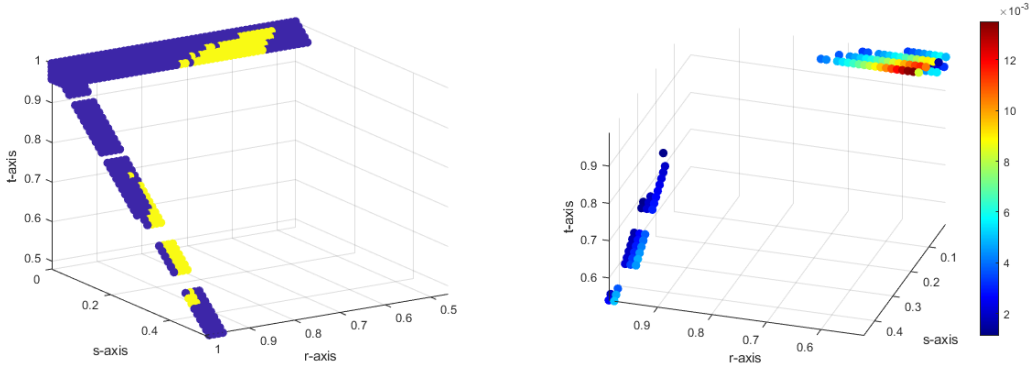
First, we note that there is no overlap between the degradability of \mathcal{N}_{12} and $\mathcal{N}_{r,s,t}$, easily seen by recalling [THEOREM 28](#) and [TABLE 6.1](#). The latter at the very least requires $r \leq 1/2$ while the former requires $r \geq 1/2$. The only possible time both of these could be true is when $r = 1/2$. For \mathcal{N}_{12} to be degradable we need $s = 0 = t - 1$, but $\mathcal{N}_{r,s,t}$ is not degradable here. Therefore, over the region where \mathcal{N}_{12} is degradable, it is non-trivial to ask whether $\mathcal{N}_{r,s,t}$ itself has weakly additive coherent information.

We now slightly broaden our search horizon to include some points close to where \mathcal{N}_{12} is degradable. Specifically, we explore the subset of parameters

$$\begin{cases} r \geq 0.96 \\ s \leq 0.52 \\ t = 1 - s \end{cases} \quad \text{or} \quad \begin{cases} r \geq 0.46 \\ s \leq 0.04 \\ t \geq 0.96. \end{cases}$$

The full channel $\mathcal{N}_{r,s,t}$ is degradable for a small subset of this slightly larger space (namely when $r = 1/2$, $s = 0$, and $t \geq 0.96$), but we omit these points from our investigation. The final space is visualized in [FIGURE 6.3](#).

In this region we look for both weak and strong non-additivity to two levels. Weak non-additivity is of course with two copies of $\mathcal{N}_{r,s,t}$, while strong additivity is with the



(a) Non-additive versus additive regions, (b) $Q^{(1)}(\mathcal{N}_{r,s,t} \otimes \mathcal{E}^{1/2}) - Q^{(1)}(\mathcal{N}_{r,s,t})$ in non-additive regions.

Figure 6.4: Regions of the ReMAD channel $\mathcal{N}_{r,s,t}$ which exhibit strong non-additivity with the erasure channel $\mathcal{E}^{1/2}$. We define non-additivity as an activation greater than 10^{-3} . Optimization took approximately 3 weeks to perform.

equiprobable erasure channel $\mathcal{E}^{1/2}$. Our motivation for this particular erasure channel is that it has already been used to demonstrate strong non-additivity for the generalized Platypus channel in [SW24]. We reparametrize this result into one for ReMAD channels.

Proposition 20. *If $s \geq 1/2$, $t > 0$, and $s + t < 1$, then*

$$Q^{(1)}\left(\mathcal{N}_{1,s,t} \otimes \mathcal{E}^{\frac{1}{2}}\right) > Q^{(1)}(\mathcal{N}_{1,s,t}) = 0.$$

The reason the single-symbol capacity is zero above is because $\mathcal{N}_{1,s,t}$ is antidegradable at those points. Now, note that in our search space whenever $r = 1$ we also have $s + t = 1$. Thus, we are not reexamining any points where strong non-additivity is already known to hold.

Across our search space we do not find any weak non-additivity. That is, the coherent information to two-symbols is additive. However, we do find strong non-additivity with the erasure channel, which we illustrate in [FIGURE 6.4](#), the first time such behavior has been found for ReMAD channels. The largest degree of activation occurs at $(r, s, t) = (0.58, 0.03, 0.96)$, where $Q^{(1)}(\mathcal{N}_{r,s,t} \otimes \mathcal{E}^{1/2}) - Q^{(1)}(\mathcal{N}_{r,s,t}) \approx 0.0022$. Interestingly, in the region where the subchannel \mathcal{N}_{12} is degradable we only ever find additivity. In fact, it even appears that strong non-additivity vanishes as we approach this region.

Conjecture 4. *ReMAD channels are weakly additive to two levels everywhere.*

The basis for this new conjecture is multifaceted. Although our search space is limited, it covers the area where we would most expect weak non-additivity to occur. Moreover, weak additivity of the generalized Platypus channel $\mathcal{N}_{s,t}$, a special case of $\mathcal{N}_{r,s,t}$, has been conjectured by [SW24]. This is on top of the earlier conjecture of weak additivity for the Platypus channel by [Led+23], which due to [Alh25] we recall is true for two levels.

References

- [Sha48] C. Shannon. "A Mathematical Theory of Communication". In: *Bell System Technical Journal* 27 (3 1948), pp. 379–423.
- [Ume54] H. Umegaki. "Conditional Expectation in an Operator Algebra, I". In: *Tohoku Mathematical Journal* 6 (2 1954), pp. 177–181.
- [Hol73] A. S. Holevo. "Bounds for the Quantity of Information Transmitted by a Quantum Communication Channel". In: *Problems of Information Transmission* 9 (1973), pp. 177–183.
- [VK76] A. M. Vershik and S. V. Kerov. "Asymptotics of the Plancherel Measure of the Symmetric Group and the Limiting Form of Young Tableaux". In: *Proceedings of the Russian Academy of Sciences* 233 (6 1976), pp. 1024–1027.
- [LS77] B. F. Logan and L. A. Shepp. "A Variational Problem for Random Young Tableaux". In: *Advances in Mathematics* 26 (2 1977), pp. 206–222.
- [Dav78] E. B. Davies. "Information and Quantum Measurement". In: *IEEE Transactions on Information Theory* 24 (5 1978), pp. 596–599.
- [SC78] W. A. Sokalski and H. Chojnacki. "Approximate Exchange Perturbation Study of Intermolecular Interactions in Molecular Complexes". In: *International Journal of Quantum Chemistry* 13 (5 1978), pp. 679–692.
- [RS81] M. Reed and B. Simon. *Methods of Modern Mathematical Physics. Functional Analysis*. Academic Press, 1981.
- [VK85] A. M. Vershik and S. V. Kerov. "Asymptotic of the Largest and the Typical Dimensions of Irreducible Representations of a Symmetric Group". In: *Functional Analysis and Its Applications* 19 (1985), pp. 21–31.
- [Wer89] R. F. Werner. "Quantum States with Einstein-Podolsky-Rosen Correlations Admitting a Hidden-Variable Model". In: *Physical Review A* 40 (8 1989), pp. 4277–4281.
- [HP91] F. Hiai and D. Petz. "The Proper Formula for Relative Entropy and its Asymptotics in Quantum Probability". In: *Communications in Mathematical Physics* 143 (1 1991), pp. 99–114.

- [Ben+93] C. H. Bennett et al. "Teleporting an Unknown Quantum State via Dual Classical and Einstein-Podolsky-Rosen Channels". In: *Physical Review Letters* 70 (13 1993), pp. 1895–1899.
- [HHH96] M. Horodecki, P. Horodecki, and R. Horodecki. "Separability of Mixed States: Necessary and Sufficient Conditions". In: *Physics Letters A* 223 (1 1996), pp. 1–8.
- [Per96] A. Peres. "Separability Criterion for Density Matrices". In: *Physical Review Letters* 77 (8 1996), pp. 1413–1415.
- [BDS97] C. H. Bennett, D. P. DiVincenzo, and J. A. Smolin. "Capacities of Quantum Erasure Channels". In: *Physical Review Letters* 78 (16 1997), pp. 3217–3220.
- [Ful97] W. Fulton. *Young Tableaux. With Applications to Representation Theory and Geometry*. Cambridge University Press, 1997.
- [GBP97] M. Grassl, T. Beth, and T. Pellizzari. "Codes for the Quantum Erasure Channel". In: *Physical Review A* 56 (1 1997), pp. 33–38.
- [Llo97] S. Lloyd. "Capacity of the Noisy Quantum Channel". In: *Physical Review A* 55 (3 1997), pp. 1613–1622.
- [LS97] V. F. Lotrich and K. Szalewicz. "Symmetry-Adapted Perturbation Theory of Three-Body Nonadditivity of Intermolecular Interaction Energy". In: *Journal of Chemical Physics* 106 (23 1997), pp. 9668–9687.
- [SW97] B. Schumacher and M. D. Westmoreland. "Sending Classical Information via Noisy Quantum Channels". In: *Physical Review A* 56 (1997), pp. 131–138.
- [Nie99] M. A. Nielsen. "Conditions for a Class of Entanglement Transformations". In: *Physical Review Letters* 83 (2 1999), pp. 436–439.
- [CHS00] H. A. Carteret, A. Higuchi, and A. Sudbery. "Multipartite Generalization of the Schmidt Decomposition". In: *Journal of Mathematical Physics* 41 (12 2000), pp. 7932–7939.
- [PKJ01] K. Patkowski, T. Korona, and B. Jeziorski. "Convergence Behavior of the Symmetry-Adapted Perturbation Theory for States Submerged in Pauli Forbidden Continuum". In: *Journal of Chemical Physics* 115 (3 2001), pp. 1137–1152.
- [Pet01] D. Petz. In: *Entropy, von Neumann and the von Neumann Entropy. John von Neumann and the Foundations of Quantum Physics*. Ed. by M. Rédei and M. Stöltzner. Chadwyck-Healey, 2001.

- [Rai02] E. M. Rains. “A Semidefinite Program for Distillable Entanglement”. In: *IEEE Transactions on Information Theory* 47 (7 2002), pp. 2921–2933.
- [Sho02] P. Shor. “The Quantum Channel Capacity and Coherent Information”. In: *MSRI Workshop on Quantum Computation* (2002).
- [FH04] W. Fulton and J. Harris. *Representation Theory. A First Course*. Springer, 2004.
- [Par04] K. R. Parthasarathy. “On the Maximal Dimension of a Completely Entangled Subspace for Finite Level Quantum Systems”. In: *Proceedings Mathematical Sciences* 114 (2004), pp. 365–374.
- [Sho04] P. Shor. “Equivalence of Additivity Questions in Quantum Information Theory”. In: *Communications in Mathematical Physics* 246 (2004), pp. 453–472.
- [Dev05] I. Devetak. “The Private Classical Capacity and Quantum Capacity of a Quantum Channel”. In: *IEEE Transactions on Information Theory* 51 (2005), pp. 44–55.
- [DS05] I. Devetak and P. Shor. “The Capacity of a Quantum Channel for Simultaneous Transmission of Classical and Quantum Information”. In: *Communications in Mathematical Physics* 2 (2005), pp. 287–303.
- [Hor+05] K. Horodecki et al. “Secure Key from Bound Entanglement”. In: *Physical Review Letters* 94 (16 2005), p. 160502.
- [Rom05] S. Roman. *Advanced Linear Algebra*. Springer, 2005.
- [Bha06] B. V. R. Bhat. “A Completely Entangled Subspace of Maximal Dimension”. In: *International Journal of Quantum Information* 4 (2 2006), pp. 325–330.
- [DBF07] A. D’Arrigo, G. Benenti, and G. Falci. “Quantum Capacity of Dephasing Channels with Memory”. In: *New Journal of Physics* 9 (2007), p. 310.
- [Kin+07] C. King et al. “Properties of Conjugate Channels with Applications to Additivity and Multiplicativity”. In: *Markov Processes and Related Fields* 13 (2 2007), pp. 391–423.
- [TWW07] M. Tomamichel, M. M. Wilde, and A. Winter. “Strong Converse Rates for Quantum Communication”. In: *IEEE Transactions on Information Theory* 63 (1 2007), pp. 715–727.
- [Cub+08] T. Cubitt et al. “Counterexamples to Additivity of Minimum Output p -Rényi Entropy for p Close to 0”. In: *Communications in Mathematical Physics* 284 (2008), pp. 281–290.

- [CRS08] T. S. Cubitt, M. B. Ruskai, and G. Smith. “The Structure of Degradable Quantum Channels”. In: *Journal of Mathematical Physics* 49 (10 2008), p. 102104.
- [GB08] M. Grant and S. Boyd. “Graph Implementations for Nonsmooth Convex Programs”. In: *Recent Advances in Learning and Control*. Ed. by V. Blondel, S. Boyd, and H. Kimura. Lecture Notes in Control and Information Sciences. http://stanford.edu/~boyd/graph_dcp.html. Springer-Verlag Limited, 2008, pp. 95–110.
- [HW08] P. Hayden and A. Winter. “Counterexamples to the Maximal p -Norm Multiplicativity Conjecture for all $p > 1$ ”. In: *Communications in Mathematical Physics* 284 (2008), pp. 263–280.
- [SY08] G. Smith and J. Yard. “Quantum Communication with Zero-Capacity Channels”. In: *Science* 321 (5897 2008), pp. 1812–1815.
- [YHD08] J. Yard, P. Hayden, and I. Devetak. “Capacity Theorems for Quantum Multiple-Access Channels: Classical-Quantum and Quantum-Quantum Capacity Regions”. In: *IEEE Transactions on Information Theory* 54 (7 2008), pp. 3091–3113.
- [Has09] M. B. Hastings. “Superadditivity of Communication Capacity Using Entangled Inputs”. In: *Nature Physics* 5 (2009), pp. 255–257.
- [ASW10] G. Aubrun, S. Szarek, and E. Werner. “Hasting’s Additivity Counterexample via Dvoretzky’s Theorem”. In: *Communications in Mathematical Physics* 305 (2010), pp. 85–97.
- [Brá+10] K. Brádler et al. “Conjugate Degradability and the Quantum Capacity of Cloning Channels”. In: *Journal of Mathematical Physics* 51 (7 2010), p. 072201.
- [BH10] F. Brandão and M. Horodecki. “On Hastings’ Counterexamples to the Minimum Output Entropy Additivity Conjecture”. In: *Open Systems & Information Dynamics* 17 (1 2010), pp. 31–52.
- [GHP10] A. Grudka, M. Horodecki, and Ł. Pankowski. “Constructive Counterexamples to the Additivity of the Minimum Output Rényi Entropy of Quantum Channels for all $p > 2$ ”. In: *Journal of Physics A* 43 (2010).
- [NC10] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2010.
- [SSY11] G. Smith, J. A. Smolin, and J. Yard. “Quantum Communication with Gaussian Channels of Zero Quantum Capacity”. In: *Nature Photonics* 5 (11 2011), pp. 624–627.

- [Buf12] A. I. Bufetov. “On the Vershik-Kerov Conjecture Concerning the Shannon-McMillan-Breiman Theorem for the Plancherel Family of Measures on the Space of Young Diagrams”. In: *Geometric and Functional Analysis* 22 (4 2012), pp. 938–975.
- [SS12] G. Smith and J. A. Smolin. “Detecting Incapacity of a Quantum Channel”. In: *Physical Review Letters* 108 (23 2012), p. 230507.
- [Wat12] S. Watanabe. “Private and Quantum Capacities of More Capable and Less Noisy Quantum Channels”. In: *Physical Review A* 85 (1 2012), p. 012326.
- [Cub13] T. Cubitt. *MATLAB Code*. 2013. URL: <https://www.dr-qubit.org/matlab.html>.
- [HS13] A. Holevo and M. Shirokov. “On Classical Capacities of Infinite-Dimensional Quantum Channels”. In: *Problems of Information Transmission* 49 (2013), pp. 15–31.
- [Ana+14] A. Anandkumar et al. “Tensor Decompositions for Learning Latent Variable Models”. In: *Journal of Machine Learning Research* 15 (2014), pp. 2773–2832.
- [GB14] M. Grant and S. Boyd. *CVX: Matlab Software for Disciplined Convex Programming, version 2.1*. <https://cvxr.com/cvx>. 2014.
- [Ouy14] Y. Ouyang. “Channel Covariance, Twirling, Contraction and Some Upper Bounds on the Quantum Capacity”. In: *Quantum Information and Computation* 14 (11 2014), pp. 916–936.
- [Brá15] K. Brádler. “The Pitfalls of Deciding Whether a Quantum Channel is (Conjugate) Degradable and How to Avoid Them”. In: *Open Systems & Information Dynamics* 22 (4 2015), p. 1550026.
- [Cub+15] T. Cubitt et al. “Unbounded Number of Channel Uses May Be Required to Detect Quantum Capacity”. In: *Nature Communications* 6 (2015), p. 6739.
- [GF15] Y. Guo and H. Fan. “A Generalization of Schmidt Number for Multipartite States”. In: *International Journal of Quantum Information* 13 (3 2015), p. 1550025.
- [BCN16] S. T. Belinschi, B. Collins, and I. Nechita. “Almost One Bit Violation for the Additivity of the Minimum Output Entropy”. In: *Communications in Mathematical Physics* 341 (2016), pp. 885–909.
- [Rob16] E. Robeva. “Orthogonal Decomposition of Symmetric Tensors”. In: *SIAM Journal on Matrix Analysis and Applications* 37 (1 2016), pp. 86–102.

- [AS17] G. Aubrun and S. Szarek. *Alice and Bob Meet Banach. The Interface of Asymptotic Geometric Analysis and Quantum Information Theory*. American Mathematical Society, 2017.
- [BŻ17] I. Bengtsson and K. Życzkowski. *Geometry of Quantum States*. Cambridge University Press, 2017.
- [CLS17] A. Cross, K. Li, and G. Smith. “Uniform Additivity in Classical and Quantum Information”. In: *Physical Review Letters* 118 (2017), p. 040501.
- [Sut+17] D. Sutter et al. “Approximate Degradable Quantum Channels”. In: *IEEE Transactions on Information Theory* 63 (12 2017), pp. 7832–7844.
- [Wil17] M. M. Wilde. *Quantum Information Theory*. Cambridge University Press, 2017.
- [BC18] M. Brannan and B. Collins. “Highly Entangled, Non-Random Subspaces of Tensor Products from Quantum Groups”. In: *Communications in Mathematical Physics* 358 (2018), pp. 1007–1025.
- [GJL18] L. Gao, M. Junge, and N. LaRacunte. “Capacity Estimates via Comparison with TRO Channels”. In: *Communications in Mathematical Physics* 364 (1 2018), pp. 83–121.
- [MSH18] M. Mozrzyk, M. Studziński, and M. Horodecki. “A Simplified Formalism of the Algebra of Partially Transposed Permutation Operators with Applications”. In: *Journal of Physics A: Mathematical and Theoretical* 51 (12 2018), p. 125202.
- [WXD18] X. Wang, W. Xie, and R. Duan. “Semidefinite Programming Strong Converse Bounds for Classical Capacity”. In: *IEEE Transactions on Information Theory* 64 (1 2018), pp. 640–653.
- [Lim+19] Y. Lim et al. “Activation and Superactivation of Single-Mode Gaussian Quantum Channels”. In: *Physical Review A* 99 (3 2019), p. 032337.
- [Pas+19] A. Paszke et al. “PyTorch: An Imperative Style, High-Performance Deep Learning Library”. In: *Advances in Neural Information Processing Systems*. 2019, pp. 8024–8035.
- [Vir+20] P. Virtanen et al. “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”. In: *Nature Methods* 17 (2020), pp. 261–272.
- [FF21] K. Fang and H. Fawzi. “Geometric Rényi Divergence and its Applications in Quantum Channel Capacities”. In: *Communications in Mathematical Physics* 384 (2021), pp. 1615–1677.

- [Haa+21] E. Haapasalo et al. “An Operational Characterization of Infinite-Dimensional Quantum Resources”. In: *Physical Review Letters* 127 (2021), p. 250401.
- [Sid21] V. Siddhu. “Entropic Singularities Give Rise to Quantum Transmission”. In: *Nature Communications* 12 (2021), p. 5750.
- [Wan21] X. Wang. “Pursuing the Fundamental Limits for Quantum Computation”. In: *IEEE Transactions on Information Theory* 67 (7 2021), pp. 4524–4532.
- [SD22] S. Singh and N. Datta. “Detecting Positive Quantum Capacities of Quantum Channels”. In: *npj Quantum Information* 8 (1 2022).
- [CG23] S. Chessa and V. Giovannetti. “Resonant Multilevel Amplitude Damping Channels”. In: *Quantum* 7 (2023), p. 902.
- [Led+23] F. Leditzky et al. “The Platypus of the Quantum Channel Zoo”. In: *IEEE Transactions on Information Theory* 69 (2023), pp. 3825–3849.
- [SC23] P. Skrzypczyk and D. Cavalcanti. *Semidefinite Programming in Quantum Information Science*. IOP Publishing, 2023.
- [AK24] M. A. Alhejji and E. Knill. “Towards a Resolution of the Spin Alignment Conjecture”. In: *Communications in Mathematical Physics* 405 (5 2024), p. 119.
- [GOT24] M. Griffin, K. Ono, and W. Tsai. “Distributions of Hook Lengths in Integer Partitions”. In: *Proceedings of the American Mathematical Society Series B* 11 (2024), pp. 422–435.
- [JKM24] V. Jain, M. Kwan, and M. Michelen. *Entangled States Are Typically Incomparable*. 2024. URL: <https://arxiv.org/abs/2406.03335>.
- [SW24] G. Smith and P. Wu. *Additivity of Quantum Capacities in Simple Non-Degradable Quantum Channels*. 2024. URL: <https://arxiv.org/abs/2409.03927>.
- [SS24] K. Szczygielski and M. Studziński. “New Constructive Counterexamples to Additivity of Minimum Output Rényi p -Entropy of Quantum Channels”. In: *IEEE Transactions on Information Theory* 70 (10 2024), pp. 7023–7035.
- [Alh25] M. Alhejji. “Refining Ky Fan’s Majorization Relation with Linear Programming”. In: *Annales Henri Poincaré* (2025).
- [GOY25] S. R. Garcia, R. O’Loughlin, and J. Yu. “Symmetric and Antisymmetric Tensor Products for the Function-Theoretic Operator Theorist”. In: *Canadian Journal of Mathematics* 77 (1 2025), pp. 324–346.
- [Sid25] V. Siddhu. *Leaking Information to Gain Entanglement*. 2025. URL: <https://arxiv.org/abs/2011.15116>.

- [WW25] P. Wu and Y. Wang. “Quantum Capacity Amplification via Privacy”. In: *2025 IEEE International Symposium on Information Theory (ISIT)*. IEEE. 2025.

Appendices

A Symbolic Study of Subchannels and SDPs for Degradability

Listing A.1: Mathematica functions for symbolically analyzing subchannels and degrading maps with example usage for finding antidegradable regions of \mathcal{N}_{12} .

```

1 PartialTrace = ResourceFunction["MatrixPartialTrace"];
2 FullChannel[iso_, state_] := Refine[iso . state . ConjugateTranspose[iso]];
3
4 Deleter[state_, resdim_] := ReplacePart[state, {{resdim, i_} -> 0, {i_,
   resdim} -> 0}];
5 Adder[state_, resdim_] := Transpose[ Insert[ Transpose[
6   Insert[state, ConstantArray[0, Length[state]], resdim]
7   ],
8   ConstantArray[0, Length[state] + 1],
9   resdim]
10  ];
11 FullChannelRes[iso_, state_, resdim_] := Refine[iso . Adder[state, resdim] .
   ConjugateTranspose[iso]];
12
13 Bob[iso_, state_, dims_] := PartialTrace[FullChannel[iso, state], 2, dims];
14 BobRes[iso_, state_, dims_, resdim_] := PartialTrace[FullChannelRes[iso,
   state, resdim], 2, dims];
15 Env[iso_, state_, dims_] := PartialTrace[FullChannel[iso, state], 1, dims];
16 EnvRes[iso_, state_, dims_, resdim_] := PartialTrace[FullChannelRes[iso,
   state, resdim], 1, dims];
17
18 ChoiMatrix[ch_, indim_] := Sum[
19   KroneckerProduct[
20     Array[Function[{x,y}, DiscreteDelta[x-i,y-j]], {indim, indim}],
21     ch[Array[Function[{x,y}, DiscreteDelta[x-i,y-j]], {indim, indim}]]
22   ],
23   {i, 1, indim}, {j, 1, indim}
24 ];
25 KrausOperators[ch_, indim_, outdim_] := Map[Function[kraus, Sqrt[kraus
   [[1]]] * Transpose[ArrayReshape[kraus[[2]], {indim, outdim}]] / Norm[
   kraus[[2]]]], Transpose[Eigensystem[ChoiMatrix[ch, indim]]]];
26 TransferMatrix[ch_, indim_, outdim_] := Refine[Total[Map[Function[kraus,
   KroneckerProduct[kraus, kraus]], KrausOperators[ch, indim, outdim]]]];

```

```

27
28 Involution[matrix_, dimA_, dimB_] := Module[
29   {
30     dims = {dimA, dimB},
31     tensor, permutedTensor, resultMatrix
32   },
33   tensor = ArrayReshape[matrix, Join[dims, dims]];
34   permutedTensor = Transpose[tensor, {4, 2, 3, 1}];
35   resultMatrix = ArrayReshape[permutedTensor, {dimA * dimA, dimB * dimB}];
36   resultMatrix
37 ];
38
39 ChoiToChannel[choi_, dimA_, dimB_] := Function[state, PartialTrace[choi .
    KroneckerProduct[Transpose[state], IdentityMatrix[dimA]], 1, {dimA, dimB
    }]];
40
41 Clear[r,s,t]
42 $Assumptions = (
43 0 <= r <= 1 &&
44 0 <= s <= 1 &&
45 0 <= t <= 1 &&
46 0 <= 1-s-t <= 1
47 );
48
49 V = {
50   {1, 0, 0},
51   {0, Sqrt[r], 0},
52   {0, 0, Sqrt[s]},
53   {0, Sqrt[1-r], 0},
54   {0, 0, Sqrt[t]},
55   {0, 0, 0},
56   {0, 0, Sqrt[1-s-t]},
57   {0, 0, 0},
58   {0, 0, 0}
59 };
60
61 Chan12 := Function[state, BobRes[V, state, {3, 3}, 1]];
62 ChanC12 := Function[state, EnvRes[V, state, {3, 3}, 1]];
63
64 Trans12 = Simplify[TransferMatrix[Chan12, 2, 3]];
65 TransC12 = Simplify[TransferMatrix[ChanC12, 2, 3]];
66
67 AntiDegrading01 = Transpose[LinearSolve[Transpose[TransC01], Transpose[
    Trans01]]];
68 AntiDegradeChoi01 = Involution[AntiDegrading01, 3, 3];

```

```
69 AntiEig12 = FullSimplify[Eigenvalues[AntiDegradeChoi01]];
```

Listing A.2: MATLAB implementation of [COROLLARY 2](#) to find degradable regions of \mathcal{N}_{12} using slack variables to control for diagonal binary partial trace.

```
1 resolution = 101;
2 r = linspace(0, 1, resolution);
3 s = linspace(0, 1, resolution);
4 t = linspace(0, 1, resolution);
5
6 rst = cartesian(r, s, t);
7 rst = rst(1 - rst(:, 2) - rst(:, 3) >= 0, :);
8 rst(:, 4) = false;
9
10 for i = 1:size(rst)
11
12     r = rst(i, 1);
13     s = rst(i, 2);
14     t = rst(i, 3);
15
16     sprintf("r: %.2f, s: %.2f, t: %.2f", [r, s, t])
17
18     TN = [
19         r, 0, 0, s;
20         0, sqrt(r*t), 0, 0;
21         0, 0, 0, 0;
22         0, 0, sqrt(r*t), 0;
23         1-r, 0, 0, 0;
24         0, sqrt((1-r)*(1-s-t)), 0, 0;
25         0, 0, 0, 0;
26         0, 0, sqrt((1-r)*(1-s-t)), 0;
27         0, 0, 0, 1-s-t
28     ];
29
30     TNC = [
31         1-r, 0, 0, 1-s-t;
32         0, sqrt(t*(1-r)), 0, 0;
33         0, 0, 0, 0;
34         0, 0, sqrt(t*(1-r)), 0;
35         r, 0, 0, t;
36         0, sqrt(r*s), 0, 0;
37         0, 0, 0, 0;
38         0, 0, sqrt(r*s), 0;
39         0, 0, 0, s
40     ];
41
```

```

42     [status, coefs, choi] = findValidTransferMatrix(TN, TNC, 3, 3);
43     rst(i, 4) = status == "Solved";
44
45 end
46
47 function C = cartesian(varargin)
48     args = varargin;
49     n = nargin;
50
51     [F{1:n}] = ndgrid(args{:});
52
53     for i=n:-1:1
54         G(:,i) = F{i}(:);
55     end
56
57     C = unique(G , 'rows');
58 end
59
60 function [status, validK, choiMtx] = findValidTransferMatrix(TN, TNC, dimA,
    dimB)
61
62     D0 = TN' \ TNC';
63     D0 = D0';
64     n = dimA * dimB;
65
66     cvx_begin quiet sdp
67         cvx_precision best
68
69         variable K(n, n); % Initialize K
70         variable c(4);
71
72         combined = D0 + K;
73         choi = involution(combined, dimA, dimB);
74
75         K*TN == 0;
76         choi == semidefinite(n);
77
78         partial = TrX(choi, 2, [dimA, dimB]);
79         partial = c(1) * diag([0,1,1]);
80         partial = c(2) * diag([1,0,1]);
81         partial = c(3) * diag([1,1,0]);
82         partial = c(4) * diag([1,1,1]);
83
84         c >= 0;
85         sum(c) == 1;

```

```
86
87     cvx_end
88
89     status = cvx_status;
90     if strcmp(cvx_status, 'Solved')
91         validK = K; % Return the valid coefficients
92         choiMtx = choi;
93     else
94         validK = [];
95         choiMtx = [];
96     end
97 end
```

B Optimization of Coherent Information

Listing B.1: Computation of coherent information, namely of $Q^{(1)}(\mathcal{N}_{r,s,t})$, $Q^{(2)}(\mathcal{N}_{r,s,t})$, and $Q^{(1)}(\mathcal{N}_{r,s,t} \otimes \mathcal{E}^{1/2})$.

```
1 function f = pure_state_density(params)
2     r = length(params);
3     if (-1)^r == 1
4         r = r/2;
5     else
6         disp('Not a valid even length real vector!')
7     end
8     % Extract real and imaginary parts
9     realPart = params(1:r);
10    imagPart = params(r + 1:2*r);
11    % Construct the complex state vector
12    psi = complex(realPart, imagPart);
13    % Normalize the state vector
14    f = psi / norm(psi);
15 end
16
17 function res = compute_vs_ci(V, rho, dims)
18     sigma = V*rho*V';
19     res = real(VNent(TrX(sigma, 2, dims))-VNent(TrX(sigma, 1, dims)));
20 end
21
22 function res = ReMAD_channel(gamma, mode)
23
24     [d, ~] = size(gamma);
25     V = zeros(d^2, d);
26     V(1,1) = 1;
27     disp(d)
28     for j = 1:d
29         V(:, j) = sqrt(gamma(j,j)) * kron(basis_vector(j, d), basis_vector
30             (1, d));
31         for k = 1:(j-1)
32             V(:, j) = V(:, j) + sqrt(gamma(j,j-k)) * kron(basis_vector(j-k,
33                 d), basis_vector(k + 1, d));
34         end
35     end
36 end
```

```

35     if (nargin == 1)
36         mode = 'isom';
37     end
38
39     switch mode
40         case 'isom'
41             res = V;
42         otherwise
43             error('Unknown channel representation!')
44     end
45 end
46
47 function vec = basis_vector(j, n)
48     vec = zeros(n, 1);
49     vec(j, 1) = 1;
50 end
51
52 % number of optimizations
53 M = 50;
54 % optimization options
55 opt = optimoptions('fminunc', 'disp', 'none');
56 opt_ps = optimoptions('particleswarm', 'UseParallel', true, 'Display', 'none');
57
58 resolution = 101;
59 r = linspace(0, 1, resolution);
60 s = linspace(0, 1, resolution);
61 t = linspace(0, 1, resolution);
62
63 rst = cartesian(r, s, t);
64 r = rst(:, 1);
65 s = rst(:, 2);
66 t = rst(:, 3);
67
68 rst(:, 4) = -1;
69 rst(:, 5) = -1;
70 rst(:, 6) = -1;
71
72 p = 0.5;
73 a_eras = 2;
74 b_eras = 3;
75 e_eras = 3;
76 V_eras = erasure(a_eras, p, 'isom');
77
78 a_remad = 3;
79 b_remad = 3;

```

```

80 e_remad = 3;
81
82 dim = [a_remad * a_eras, b_remad * b_eras, e_remad * e_eras, 2 * e_remad *
      a_eras];
83 k = 2 * a_remad * a_eras;
84
85 res_general_two_copy = zeros(1,M);
86 res_general_rho_two_copy = cell(1, M);
87
88 res_general_erasure = zeros(1,M);
89 res_general_rho_erasure = cell(1, M);
90
91 res_general_one_copy = zeros(1,M);
92 res_general_rho_one_copy = cell(1, M);
93
94 for j = 1:length(rst)
95     rr = rst(j, 1);
96     ss = rst(j, 2);
97     tt = rst(j, 3);
98
99     if 1-ss-tt >= 0
100
101         gamma = [
102             1, 0, 0;
103             rr, 1-rr, 0;
104             ss, tt, 1-ss-tt;
105         ];
106         %V_remad = ReMAD_channel(gamma, 'isom');
107         V_remad = Generalized_vs_channel(tt, ss, 'isom');
108
109         disp(['ReMAD params: r=', num2str(rr), ', s=', num2str(ss), ', t=',
110             num2str(tt)])
111
112         VE = kron(V_remad, V_eras);
113         VE_input = a_remad * a_eras;
114         VE_output = b_remad * b_eras;
115         for kk = 1:size(VE, 2)
116             VE(:, kk) = syspermute(VE(:, kk), [1,3,2,4], [b_remad, e_remad,
117                 b_eras, e_eras]);
118         end
119
120         V_2 = kron(V_remad, V_remad);
121         V_2_input = a_remad * a_remad;
122         V_2_output = b_remad * b_remad;

```

```

122     for kk = 1:size(V_2, 2)
123         V_2(:, kk) = syspermute(V_2(:, kk), [1,3,2,4], [b_remad, e_remad
           , b_remad, e_remad]);
124     end
125
126
127     disp('Computing Q^1 of V @ V...')
128     for m=1:M
129         obj = @(x) -compute_ci(V_2, TrX(pure_state_density(x)'*
           pure_state_density(x), 2, [V_2_input, V_2_input]),
           V_2_output);
130         x0 = rand(1,2*V_2_input*V_2_input);
131         % Perform the optimization
132         [x, f] = fminunc(obj, x0, opt);
133         % Desired density:
134         res_general_two_copy(m) = -f;
135         res_general_rho_two_copy{m} = TrX(pure_state_density(x)'*
           pure_state_density(x), 2, [V_2_input, V_2_input]);
136     end
137
138     disp('Computing Q^1 of V @ E...')
139     for m=1:M
140         obj = @(x) -compute_ci(VE, TrX(pure_state_density(x)'*
           pure_state_density(x), 2, [VE_input, VE_input]), VE_output);
141         x0 = rand(1,2*VE_input*VE_input);
142         [x, f] = fminunc(obj, x0, opt);
143         res_general_erasure(m) = -f;
144         res_general_rho_erasure{m} = TrX(pure_state_density(x)'*
           pure_state_density(x), 2, [VE_input, VE_input]);
145     end
146
147     disp('Computing Q^1 of V...')
148     for m=1:M
149         obj = @(x) -compute_ci(V_remad, TrX(pure_state_density(x)'*
           pure_state_density(x), 2, [a_remad, a_remad]), b_remad);
150         x0 = rand(1,2*a_remad*a_remad);
151         [x, f] = fminunc(obj, x0, opt);
152         res_general_one_copy(m) = -f;
153         res_general_rho_one_copy{m} = TrX(pure_state_density(x)'*
           pure_state_density(x), 2, [a_remad, a_remad]);
154     end
155
156     [ci_one_copy, index_one_copy] = max(res_general_one_copy);
157     [ci_two_copy, index_two_copy] = max(res_general_two_copy);
158     [ci_erasure, index_erasure] = max(res_general_erasure);

```

```
159
160     disp(['Q^1 of V @ V = ', num2str(ci_two_copy)])
161     rst(j, 4) = ci_two_copy;
162     disp(['Q^1 of V @ E = ', num2str(ci_erasure)])
163     rst(j, 5) = ci_erasure;
164     disp(['Q^1 of V = ', num2str(ci_one_copy)])
165     rst(j, 6) = ci_one_copy;
166
167     end
168 end
169
170 writematrix(rst, 'remad_quant.csv')
```

C Optimization of Schmidt Coefficients and Numerical Procedures Involving Young Tableaux

Listing C.1: General purpose optimization procedure to search for non-additive p -Rényi entropy given computing upper bounds of Schmidt coefficients. Takes in the orthonormal basis of any space, and searches over specified subspaces and bipartitions.

```
1 import torch
2 import time
3 import numpy as np
4 from itertools import product, combinations
5 import torch.linalg as la
6
7 DEVICE = torch.device("cuda:0" if torch.cuda.is_available() else "cpu")
8
9 # brings subsystems to the front of the tensor
10 # product, to more easily write as B @ E
11 def swap_tensor(dims, tensor, subsystems):
12
13     k, n = dims
14     leftover_subsystems = [i for i in range(n) if i not in subsystems]
15     reordered_subsystems = subsystems + leftover_subsystems
16
17     tensor = tensor.reshape((k,) * n)
18     tensor = torch.permute(tensor, dims = reordered_subsystems)
19
20     return tensor.flatten()
21
22 def create_complex_tensor(real, imaginary):
23
24     real = torch.tensor(real, dtype = torch.float64)
25     imaginary = torch.tensor(imaginary, dtype = torch.float64)
26     complex_tensor = torch.complex(real, imaginary)
27     complex_tensor.requires_grad_()
28     complex_tensor.to(torch.complex128)
29     complex_tensor.to(DEVICE)
```

```

30     return complex_tensor
31
32 # norm of projection of unit tensor on subspace
33 def proj_objective(bob, env, proj, dims):
34
35     bob = bob / la.norm(bob)
36     env = env / la.norm(env)
37
38     k, n = dims
39     bob_size = len(bob)
40
41     num_bob = int(np.log(bob_size) / np.log(k))
42     num_env = n - num_bob
43
44     bob = bob.reshape((k,) * num_bob)
45     env = env.reshape((k,) * num_env)
46     bob_tensor_env = torch.tensordot(bob, env, dims = 0).flatten()
47
48     projected_vec = proj @ bob_tensor_env
49     projected_norm = la.norm(projected_vec)**2
50
51     return -projected_norm, bob_tensor_env
52
53 # optimizes norm of a projection
54 def optim_proj_objective(bob_real, bob_im, env_real, env_im, dims, proj, lr
    = 1e-3, epochs = 1000, eps = 1e-5):
55
56     bob = create_complex_tensor(bob_real, bob_im)
57     env = create_complex_tensor(env_real, env_im)
58
59     optimizer = torch.optim.Adam([bob, env], lr = lr)
60
61     norm_old = torch.inf
62     for i in range(epochs):
63         #torch.autograd.set_detect_anomaly(True)
64         optimizer.zero_grad()
65
66         norm, vec = proj_objective(bob, env, proj, dims)
67         norm.backward()
68         optimizer.step()
69
70         # if i % 10 == 0:
71         #     print(norm, vec)
72
73         # convergence threshold

```

```

74     if torch.abs(norm_old - norm) < eps:
75         break
76     else:
77         norm_old = norm
78
79     return norm, vec
80
81 # finds upper bound on schmidt coefficients for a fixed
82 # number of bob components and subdimension size of a subset
83 def schmidt_bound(onb, dims, num_bob, sub_dim = None, num_subbasis = None,
84                  num_bob_partitions = None, epochs = 1000, batch_size = 50, eps = 1e-5, lr
85                  = 1e-3):
86
87     k, n = dims
88     dim = onb.shape[1]
89     # if not searching over a particular subset size, just take the whole
90     # space
91     if sub_dim is None: sub_dim = dim
92
93     # if num_subset isnt passed, sample over all possible subsets
94     # otherwise, randomly draw the specified amount
95     basis_idx = range(dim)
96     if num_subbasis is None:
97         sub_bases = combinations(basis_idx, sub_dim)
98     else:
99         sub_bases = (np.random.choice(basis_idx, sub_dim) for _ in range(
100             num_subbasis))
101
102     # if num_bob_partitions isnt passed, sample over possible partitions
103     # otherwise, randomly draw the specified amount
104     if num_bob_partitions is None:
105         bob_partitions = combinations(range(n), num_bob)
106     else:
107         bob_partitions = (random.sample(range(n), num_bob) for _ in range(
108             num_bob_partitions))
109
110     subs_and_bobs = list(product(sub_bases, bob_partitions))
111     norms = torch.empty(len(subs_and_bobs))
112
113     for trial, (sub_basis_idx, bob_partition) in enumerate(subs_and_bobs):
114
115         sub_basis = onb[:, sub_basis_idx]
116         # swap the tensors around in our sub basis so that
117         # all of bobs are at the front
118         for i in range(sub_dim):

```

```

114         sub_basis[:, i] = swap_tensor(dims, sub_basis[:, i], list(
115             bob_partition))
116
117     proj = (sub_basis @ sub_basis.T).to(torch.complex128)
118
119     dim_bob = k**num_bob
120     dim_env = k**(n-num_bob)
121
122     bob_real = torch.randn((batch_size, dim_bob)).tolist()
123     bob_im = torch.randn((batch_size, dim_bob)).tolist()
124     env_real = torch.randn((batch_size, dim_env)).tolist()
125     env_im = torch.randn((batch_size, dim_env)).tolist()
126
127     batch_norms = torch.empty(batch_size)
128     for batch, (br, bi, er, ei) in enumerate(zip(bob_real, bob_im,
129         env_real, env_im)):
130         batch_norms[batch], _ = optim_proj_objective(br, bi, er, ei,
131             dims, proj, epochs = epochs, eps = eps, lr = lr)
132         print(batch_norms[batch])
133
134     local_best = torch.min(batch_norms)
135     norms[trial] = -local_best
136
137     return norms, subs_and_bobs
138
139 # dimensional bound to find additivity breaking p
140 def min_required_p(W, B, E, mu, min_p = 1, max_p = 1e4):
141     if torch.is_tensor(mu): mu = mu.detach().numpy()
142
143     p_range = np.linspace(min_p, max_p)
144     syndrome = W/(B*E) - ((1-mu)**p_range + mu**p_range)**(2/p_range)
145     min_p_idx = np.nonzero(syndrome > 0)[0]
146
147     if len(min_p_idx) == 0:
148         min_p = np.inf
149     else:
150         min_p = p_range[min_p_idx[0]]
151
152     return min_p
153
154 def schmidt_bound_loop(onb, dims, bobs_range = None, sub_dims_range = None,
155     num_subbasis = None, num_bob_partitions = None, epochs = 1000, batch_size
156     = 50, lr = 1e-3):

```

```

154 k, n = dims
155 full_dim = onb.shape[1]
156
157 if bobs_range is None:
158     # this will not include n, so at least one
159     # component will be the environment
160     bobs = range(1, n)
161 else:
162     bobs = range(bobs_range[0], bobs_range[1] + 1)
163
164 if sub_dims_range is None:
165     sub_dims = range(1, full_dim + 1)
166 else:
167     sub_dims = range(sub_dims_range[0], sub_dims_range[1] + 1)
168
169 num_searches = len(sub_dims) * len(bobs)
170 best_p = np.empty(num_searches) * np.inf
171 best_mu = np.zeros(num_searches)
172 best_subbases = [0] * (num_searches)
173 best_bobs = [0] * (num_searches)
174 total = 0
175
176 for i, (sub_dim, bob) in enumerate(product(sub_dims, bobs)):
177     print(f"\nsub dim size: {sub_dim}, bob components: {bob}")
178
179     start = time.time()
180     opts, subs_and_bobs = schmidt_bound(onb, dims, bob, sub_dim,
181                                         num_subbasis = num_subbasis,
182                                         num_bob_partitions =
183                                             num_bob_partitions,
184                                         epochs = epochs,
185                                         batch_size = batch_size,
186                                         lr = lr)
187
188     end = time.time()
189     diff = end - start
190     total += diff
191
192     min_p = np.empty(len(opts))
193     for j, opt in enumerate(opts):
194         min_p[j] = min_required_p(sub_dim, k**bob, k**(n - bob), opt)
195
196     # argmin only returns array if more than one matching argmin,
197     # otherwise is integer
198     lowest_p_idx = np.argmin(min_p)
199     if isinstance(lowest_p_idx, np.ndarray): lowest_p_idx = lowest_p_idx

```

```

[0]
198     lowest_p = min_p[lowest_p_idx]
199     if np.isfinite(lowest_p):
200         print(f"found breaking p: {lowest_p:.3f}")
201         best_p[i] = lowest_p
202         best_mu[i] = opts[lowest_p_idx]
203         sub_and_bob = subs_and_bobs[lowest_p_idx]
204         best_subbases[i] = sub_and_bob[0]
205         best_bobs[i] = sub_and_bob[1]
206
207
208     print(f"total: {total:.0f}s")
209
210     return best_p, best_subbases, best_bobs, best_mu

```

Listing C.2: Implementations of Young symmetrizers on tensor product spaces. Includes various other helper functions such as computing hook-lengths of partitions and computing orthonormal bases of the associated irreducible representations W_{nk}^λ .

```

1 import numpy as np
2
3 from sympy.combinatorics.permutations import Cycle, Permutation
4 from itertools import product, combinations, permutations, zip_longest,
   chain
5 from math import factorial
6 from scipy.linalg import orth, svd
7 from scipy.sparse import csr_matrix
8 from scipy.sparse.linalg import svds
9
10 def young_tableau(partition):
11
12     n = partition[-1]
13     m = len(partition)
14     young = np.zeros((m, n))
15
16     idx = 0
17     row = 0
18     # number of entries that need to be filled
19     # in the tableau. our step consists in filling
20     # in entries up to the first element of the partition
21     max_col = -1
22     col = partition[0]
23     while idx < m - 1:
24         young[row, :col] = 1
25         row += 1

```

```

26     idx += 1
27     col = partition[idx] - partition[idx - 1]
28     if col > max_col: max_col = col
29
30     young[row, :col] = 1
31     if max_col != -1:
32         young = young[:, :(max_col + 1)]
33
34     # now reorder so the rows are from longest to shortest,
35     # relevant for hook length calculations
36     lengths = np.sum(young, axis=1)
37     sorted_rows = np.argsort(-lengths)
38     young = young[sorted_rows]
39
40     return young
41
42 def hook_length(partition):
43
44     young = young_tableau(partition)
45     m, n = young.shape
46
47     hook = 1
48     for i in range(m):
49         for j in range(n):
50             if young[i, j] > 0:
51                 row_length = np.count_nonzero(young[i, j:])
52                 col_length = np.count_nonzero(young[i:, j])
53                 # -1 because we double count the starting cell
54                 hook *= row_length + col_length - 1
55
56     return factorial(partition[-1]) // hook
57
58 def cyclic_perms(seq):
59     # cut off one element, get all permutations of the rest,
60     # which will return them in each possible order. do this
61     # for all subsets, excluding empty and singletons
62     subseq = chain.from_iterable(combinations(seq, r) for r in range(2, len(
63         seq) + 1))
64     perms = []
65     for ss in subseq:
66         perms += [list(ss[:1]) + list(perm) for perm in permutations(ss[1:])]
67     return perms
68
69 def partition_intervals(partition):

```

```

69
70     return [range(partition[i-1] + 1 if i > 0 else 1, partition[i] + 1) for
71             i in range(len(partition))]
72 def row_perms(partition):
73
74     rows_perms = []
75     full_perms = []
76     for i in range(len(partition)):
77
78         end = partition[i]
79         start = partition[i-1] if i > 0 else 0
80         # dont want to include trivial permutations
81         if end - 1 == start: continue
82         ran = np.arange(start + 1, end + 1)
83         perms = cyclic_perms(ran)
84         rows_perms.append(perms)
85         full_perms += perms
86         if len(rows_perms) > 1:
87             full_perms += list(product(*rows_perms))
88
89     return full_perms
90
91 def col_perms(partition):
92
93     cols_perms = []
94     full_perms = []
95     rows = partition_intervals(partition)
96     columns = zip_longest(*rows)
97
98     for column in columns:
99
100         column = [c for c in column if c is not None]
101         if len(column) == 1: break
102         perms = cyclic_perms(column)
103         cols_perms.append(perms)
104         full_perms += perms
105         if len(cols_perms) > 1:
106             full_perms += list(product(*cols_perms))
107
108     return full_perms
109
110 def young_symmetrizer(partition):
111
112     rows = row_perms(partition)

```

```

113     cols = col_perms(partition)
114
115     identity = Permutation(range(partition[-1]))
116     rows += [identity]
117     cols += [identity]
118
119     pairs = []
120     for r,c in product(rows, cols):
121         rperm = Permutation([r] if isinstance(r, list) else r)
122         cperm = Permutation([c] if isinstance(c, list) else c)
123         # sympy composition is backwards
124         perm = cperm * rperm
125         sign = cperm.signature()
126
127         # we now convert these to cycles, since they can
128         # handle inputs which are outside of their domain
129         cycle = Cycle()
130         for c in perm.cyclic_form:
131             cycle = cycle(*c)
132         pairs.append((sign, cycle))
133
134     return pairs
135
136 def young_symmetrize_tensor(tensor_indices, dims, symmetrizer):
137
138     # create empty tensor of correct dimension
139     k, n = dims
140     #tensor = torch.zeros(k**n).to(torch.float64)
141     terms = []
142     for sign, perm in symmetrizer:
143         # add and subtract one since our permutations
144         # are 1-indexed but tensors are 0-indexed
145         perm_index_order = [perm[i+1] - 1 for i in range(n)]
146         perm_indices = tuple(tensor_indices[i] for i in perm_index_order)
147         #tensor += sign * basis_tensor(dims, perm_indices)
148         terms.append(sign * basis_tensor(dims, perm_indices))
149
150     tensor = torch.stack(terms).sum(dim = 0)
151     return tensor
152
153 def perm_onb(partition, ret_type = 'onb'):
154
155     n = partition[-1]
156     k = len(partition)
157     vecs = torch.zeros((k**n, k**n)).to(torch.float64)

```

```

158
159     ys = young_symmetrizer(partition)
160
161     tensors = product(* ([range(k)] * n))
162     for i, tensor in enumerate(tensors):
163         symmetrized_tensor = young_symmetrize_tensor(tensor, (k, n), ys)
164         vecs[:, i] = symmetrized_tensor
165
166     match ret_type:
167         case 'onb':
168             basis = torch.tensor(orth(vecs))
169             return basis
170         case 'onb_sparse':
171             dim = hook_length(partition)
172             vecs = vecs[:, torch.sum(torch.abs(vecs), dim = 1) > 0]
173             vecs_sparse = csr_matrix(vecs)
174             U, _, _ = svds(vecs_sparse, dim, which = "lm")
175             basis = torch.tensor(U[:, :dim].copy())
176             return basis
177         case 'proj':
178             Q, _ = la.qr(vecs)
179             return Q @ Q.T
180         case 'full':
181             return vecs
182
183 # https://stackoverflow.com/questions/10035752/elegant-python-code-for-integer-partitioning
184 def partition_gen(n):
185     a = [0 for i in range(n + 1)]
186     k = 1
187     y = n - 1
188     while k != 0:
189         x = a[k - 1] + 1
190         k -= 1
191         while 2 * x <= y:
192             a[k] = x
193             y -= x
194             k += 1
195         l = k + 1
196         while x <= y:
197             a[k] = x
198             a[l] = y
199             yield a[:k + 2]
200             x += 1
201             y -= 1

```

```

202     a[k] = x + y
203     y = x + y - 1
204     yield a[:k + 1]
205
206 # convert a partition of the form n_1 + ... + n_m = n
207 # into the list format (lambda_1,...,lambda_m)
208 def format_partition(partition):
209     partition[0] -= 1
210     new_partition = np.cumsum(partition)
211     return list(new_partition + 1)

```

Listing C.3: Multipurpose function, allowing for computation of an orthonormal of the maximal completely entangled subspace W_{nk}^{\otimes} , as well as removing antisymmetric tensors from its span to generate W_{nk}^{\wedge} . Relies on [LISTING C.2](#) for computing an orthonormal basis of W_{nk}^{\wedge} .

```

1 def custom_orth(A):
2     # use gesvd from LAPACK as it is more numerically stable
3     # for our uses than the default gesdd
4     u, s, vh = svd(A, full_matrices=False, lapack_driver = "gesvd")
5     M, N = u.shape[0], vh.shape[1]
6     rcond = np.finfo(s.dtype).eps * max(M, N)
7     tol = np.amax(s, initial=0.) * rcond
8     num = np.sum(s > tol, dtype=int)
9     Q = u[:, :num]
10    return Q
11
12 def find_minimal_removals(W, v):
13
14     coefs = W.T @ v
15     v_reconstructed = W @ coefs
16     # v is not in the span
17     if not torch.allclose(v, v_reconstructed):
18         return None
19
20     support = torch.where(~torch.isclose(coefs,
21                                         torch.tensor(0., device=DEVICE,
22                                                         dtype = torch.double)
23                                         )
24                          )[0].tolist()
25     # arbitrarily pick the first index
26     vec_to_remove = support[0]
27
28     return vec_to_remove

```

```

29 def anti_onb_explicit(dims):
30
31     k, n = dims
32
33     vecs = torch.zeros((k**n, comb(k, n))).to(torch.float64)
34     anti_partition = list(range(1, n + 1))
35     antisymmetrizer = young_symmetrizer(anti_partition)
36
37     tensors = product(* ([range(k)] * n))
38     for i, tensor in enumerate(filter(strict_inc, tensors)):
39         symmetrized_tensor = young_symmetrize_tensor(tensor, (k, n),
40             antisymmetrizer)
41         vecs[:, i] = symmetrized_tensor / la.norm(symmetrized_tensor)
42
43     return vecs
44
45 def completely_entangled_onb(dims, ret_type = "onb"):
46
47     k, n = dims
48
49     # tensors are tuples of the form (i_1,...,i_n)
50     # with 1 <= i_j <= k (i.e. |i_1 *** i_n>)
51     add = product(* ([range(k)] * n))
52     sub = product(* ([range(k)] * n))
53     vecs = []
54     for a, s in product(add, sub):
55         if a == s: continue
56         if sum(a) == sum(s):
57             vecs.append(basis_tensor(dims, a) - basis_tensor(dims, s))
58
59     vecs = torch.stack(vecs, dim = 1)
60
61     match ret_type:
62         case 'onb':
63             basis = torch.tensor(custom_orth(vecs,))
64             return basis
65         case 'onb_no_anti':
66             basis = torch.tensor(custom_orth(vecs))
67
68             anti_onb = anti_onb_explicit(dims)
69             for anti in anti_onb.T:
70                 idx = find_minimal_removals(basis, anti)
71                 if idx is not None:
72                     basis = basis[:, list(range(0,idx)) + list(range(idx +
73                         1, basis.shape[1]))]

```

```
72
73         return basis
74
75     dims = (10,2)
76     to_add = [2,4]
77
78     onb_ces = completely_entangled_onb(dims, "onb_no_anti")
79     onb_anti = anti_onb_explicit(dims)
80
81     extended_onb_anti = torch.cat([onb_anti, onb_ces[:, to_add]], dim = 1)
82     extended_onb_anti = torch.tensor(orth(extended_onb_anti))
```