Quantum Information at High and Low Energies

Thesis by
Yu Qing (Eugene) Tang

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

Caltech

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2021
Defended May 11, 2021

ACKNOWLEDGEMENTS

A PhD is a difficult yet tremendously rewarding experience, made easier and all the richer by the people in one's life. To that end, I have had the wonderful fortune of having absolutely terrific people in my life. It is my pleasure to acknowledge here the ways they've brightened my days:

My fellow graduate students, with whom I stand in solidarity: Thom Bohdanowicz, Alex Buser, Pouria Dadras, Alex Dalzell, Bailey Gu, Robert Huang, Joe Iverson, Alex Kliesch, Junyu Liu, Chris Pattison, Tian Wang.

The post-docs (current and former), from whom I've learned much: Victor Albert, Charles Cao, Elizabeth Crosson, Hrant Gharibyan, Alex Jahn, Natalie Klco, Richard Kueng, Angelo Lucia, Sepehr Nezami, Grant Salton. I especially thank Burak Şahinoğlu, with whom I've had great fun. And to think that it all began when we got lost together in Argentina.

My friends in Markham, who've offered me much needed escape: Glad, Louie, Jack, Sebastian, James.

It perhaps my greatest fortune to have had a plethora of wonderful mentors:

From my undergraduate days at Waterloo, and continuing into the present, I'd like to thank Achim Kempf and Eduardo Martin-Martinez for not only their guidance, but also their friendship.

Robert König, who has been a teacher (even if he has no memory of such a role), mentor, collaborator, and friend.

My thesis committee, who generously offered their time to listen to me ramble on about black holes and error-correction: Fernando Brandao, Xie Chen, David Hsieh, Alexei Kitaev.

Spiros Michalakis, who has been a constant source of companionship, and encouragement throughout my years at Caltech. Nobody has taught me to dig deep, question everything, and take nothing for granted quite the way Spiros has.

And of course, special thanks to my advisor John Preskill, who has been a truly excellent presence in my life, and a source of unwavering support.

And finally, I thank my parents, to whom I owe everything.

ABSTRACT

In this thesis, we take a look at how quantum information theory can be used to study physical systems at both high and low energies.

In the first part of this thesis, we examine the structure of the low-energy subspaces of quantum many-body systems. Motivated by considerations from topological order and holography, we show that the existence of error-correcting properties in low-energy subspaces is a generic feature of quantum systems. Using the formalism of matrix product states, we construct explicit quantum error-detecting codes formed from the momentum eigenstates of a quantum many-body system. We do so for both generic gapped systems as well as the gapless Heisenberg model.

We also examine how topological order can persist past the ground state space into the low-energy subspace of excited states by studying the No Low-Energy Trivial States (NLTS) conjecture. We prove a version of the NLTS conjecture under the assumption of symmetry protection. Moreover, we show that our symmetric NLTS result has implications for the performance of quantum variational optimization algorithms by using it to prove a bound on the Quantum Approximate Optimization Algorithm (QAOA). Specifically, we show that there exists a family of graphs on which QAOA for the MAXCUT problem can always be outperformed at logarithmic depths by the classical Goemans-Williamson algorithm.

In the second part of this thesis, we examine problems related to bulk reconstruction in holography and the black hole firewall paradox. Using the formalism of the tensor Radon transform, we devise and implement a numerical algorithm for reconstructing (perturbatively in AdS_3/CFT_2) the bulk metric tensor from a given boundary entropy profile. We argue that the fidelity of this reconstruction serves as a useful criterion for when a semi-classical bulk dual should be expected to exist.

We finally examine the black hole firewall problem from the perspective of quantum error-correction and quantum computational complexity. We argue that the state of the Hawking radiation has the special property of being computationally pseudo-random, meaning that it cannot be distinguished from the maximally mixed state by any efficient quantum computation. We show that this implies that each black hole has a natural structure as a quantum error-correcting code. These black hole codes give an operational meaning to some of the *ER=EPR* type proposals for resolving the firewall paradox within the framework of quantum information theory.

PUBLISHED CONTENT AND CONTRIBUTIONS

The following publications each constitute a chapter of this thesis. I was a major contributing author for all of these articles and participated in all parts of the writing and revision process.

[1] S. Bravyi, A. Kliesch, R. Koenig, and E. Tang, "Obstacles to variational quantum optimization from symmetry protection," *Physical Review Letters*, vol. 125, no. 26, Dec. 2020. DOI: `10.1103/physrevlett.125.260505`.

[2] C. Cao, X.-L. Qi, B. Swingle, and E. Tang, "Building bulk geometry from the tensor radon transform," *Journal of High Energy Physics*, no. 12, Dec. 2020. DOI: `10.1007/jhep12(2020)033`.

[3] I. Kim, E. Tang, and J. Preskill, "The ghost in the radiation: Robust encodings of the black hole interior," *Journal of High Energy Physics*, no. 6, Jun. 2020. DOI: `10.1007/jhep06(2020)031`.

[4] M. Gschwendtner, R. König, B. Şahinoğlu, and E. Tang, "Quantum error-detection at low energies," *Journal of High Energy Physics*, no. 9, Sep. 2019. DOI: `10.1007/jhep09(2019)021`.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

C h a p t e r   1

INTRODUCTION

Quantum information theory has become an increasingly versatile tool for studying various physical systems. In recent years, the incorporation of quantum information has been especially fruitful for high-energy and condensed matter physics, leading to advances in our understanding of how black holes process information, and of the topological orders associated with quantum many-body systems. In this thesis, we take a look at how quantum information theory can be used to study physical systems at both high and low energies.

Part I — Low Energies

In the first part of this thesis, we examine the structure of low-energy subspaces of quantum many-body systems, in particular with regard to the error-correcting properties that such subspaces can exhibit.

Quantum Error-Correcting Codes in Low-Energy Subspaces

A primary motivation for studying the error-correcting properties of low-energy subspaces comes from the theory of topological phases. Topological phases are special phases of matter which exhibit long-range entanglement structure. In particular, one necessary requirement for the existence of topological order is that the ground states of a topologically ordered model cannot be distinguished by local operations; this is succinctly summarized by saying that the ground state space of a topologically ordered model constitutes a quantum error-detecting code against local errors [1]. This characterization of topological order leads to the natural follow-up question of whether error-correcting properties can extend beyond ground state spaces into the low-energy subspaces of many-body systems.

A second, independent motivation also comes through holography. The AdS/CFT correspondence states that gravity in $d + 1$ dimensional Anti-de Sitter spacetime (AdS) is in some ways equivalent to a conformal field theory (CFT) defined on the d-dimensional AdS boundary [2]. Under the AdS/CFT correspondence, the bulk geometry of AdS spacetime is generally viewed as an emergent property of the underlying degrees of freedom present in the CFT. One puzzling feature of

the correspondence is that the map relating the CFT degrees of freedom and the bulk geometry must be highly degenerate and non-local. For example, through the holographic dictionary, it appears that an operator located sufficiently deeply in the bulk must be independently reconstructible on multiple, distinct regions of the boundary CFT.

In [3], it was argued that the paradoxical features of the bulk-to-boundary reconstruction map was in fact a reflection of the fact that it is the encoding map of a quantum error-correcting code. More precisely, [3] proposed that the action of bulk operators are most appropriately viewed as logical operations on a quantum error-correction code defined within boundary Hilbert space. This idea was later given concrete form in [4], which constructed an explicit example of a holographic quantum error-correcting code that manifestly exhibits many features of the bulk-to-boundary reconstruction map as a concrete toy model.

The characterization of holographic quantum error-correcting codes remains an important and challenging problem. The AdS/CFT correspondence suggests that any holographic CFT must contain, within its low-energy subspaces, an abundance of quantum error-correcting codes resulting from the encoding of the logical bulk operators. A natural question is to what extent is such a property holographic? Is the presence of quantum error-correcting properties something unique to holographic CFTs or is it something which must be present within a generic quantum system? This again motivates a study of quantum error-correction within the low-energy subspaces of a quantum system.

In Chapter 2, we study the quantum error-detecting[1] properties associated with the low-energy subspaces of a quantum many-body system. Using the formalism of matrix product states, we explicitly construst a quantum error-detecting code from the excited states of a gapped 1D local Hamiltonian. Specifically, it is shown that a generic choice of distinct momentum eigenstates selected from a gapped band is sufficient to form an error-detecting code against sublinearly many local errors. The construction of this code shows that error-correcting behavior is a generic feature of low-energy subspaces of gapped, local Hamiltonians. We also construct a similar code (with the same parameter scaling) consisting of magnon states selected from a gapless 1D integrable system - the isotropic Heisenberg model. The codes constructed in Chapter 2 show that the existence of error-correcting properties in

---

[1]Although we focus on quantum error-detection since it is a natural primitive in the context of the problem that we study, all of the results in Chapter 2 hold equivalently as statements about quantum error-correcting codes, albeit with logarithmic, as opposed to sublinear $\Omega\left(n^{1-\epsilon}\right)$, distance.

low-energy subspaces is a generic feature of quantum many-body systems, even in 1D where the analogous notion of topological order in the ground state space is absent [5], [6], and is not tied to the particular presence or absence of gap or integrability.

The No-Low Energy Trivial States Conjecture

A topologically ordered model has the additional feature that its ground states are non-trivial, i.e., they cannot be prepared from a product state by a constant depth local circuit. As with the aforementioned error-correcting properties associated with topologically ordered ground states, it becomes natural to ask if this non-triviality can also be extended into the low-energy subspaces of many-body systems. It is widely thought to be the case that there *does* exist physically reasonable gapped local Hamiltonians which satisfies the property that all states of sufficiently low energy are non-trivial, a conjecture which is formalized as the No Low-Energy Trivial States (NLTS) conjecture [7].

More precisely, let $\{H_n\}_{n\in\mathbb{N}}$ be a family of local Hamiltonians (with n parametrizing system size) such that:

1. Each H_n has ground state energy 0.

2. Each interaction term has bounded norm.

3. Each interaction term involves a constant number of qubits.

4. Each qubit participates in a constant number of interactions.

We say that the family has the *No Low-Energy Trivial States* (NLTS) property if there exists some $\epsilon > 0$ and a function $f : \mathbb{N} \to \mathbb{N}$ such that, given a depth-d local quantum circuit U, we have

$$\langle 0|U^{\dagger}H_nU|0\rangle > \epsilon n$$

for all $n \geq f(d)$.

Conjecture 1 (No-Low Energy States (NLTS)). *There exists a family of local Hamiltonians which possesses the NLTS property.*

The general NLTS conjecture is, aside from being of independent interest, a precursor to the quantum PCP conjecture, and remains open at this time. Note in particular

that the existence of non-trivial topological order is equivalent to the $\epsilon = 0$ case of the NLTS conjecture. The NLTS conjecture therefore posits that there exists a family of Hamiltonians such that non-trivial topological order persists beyond the ground state space into the low-energy subspace of excited states.

In Chapter 3, we prove a specific case of the NLTS conjecture under the assumption of a local $\mathbb{Z}_2$ symmetry. Specifically, we prove that the NLTS conjecture holds under the assumption that the Hamiltonian, starting product state, and preparation circuit all satisfy a $\mathbb{Z}_2$ symmetry condition. It is interesting to note that the presence of $\mathbb{Z}_2$-symmetry is essential in our proof: the same family of Hamiltonians that we construct in Chapter 3 does not exhibit NLTS without it. This opens up the intriguing possibility that NLTS may be akin to topological order in 1D systems which only exists under symmetry protection.

Examples of systems which satisfy the assumptions of our theorem include quantum variational algorithms, specifically the Quantum Approximate Optimization Algorithm (QAOA), which is one of the most intensely studied and promising algorithms for near-term quantum computation [8]. Using our symmetric NLTS result, we provide a general bound on the performance of QAOA on the maximum cut problem (MAXCUT) for logarithmic depths. Specifically, we show that for logarithmic depth QAOA, there always exists an infinite family of graphs on which QAOA will be outperformed by the best known polynomial time classical algorithm, i.e., the Goemans-Williamson algorithm employing semi-definite relaxation. Our result shows, for the first time, rigorous bounds on the general performance of QAOA beyond depth one[2], and is amongst the first rigorous performance bounds for QAOA against a well-known classical algorithm.

Part II — High Energies

In the second part of this thesis, we study the applications of quantum information theory to high-energy physics, specifically regarding the bulk reconstruction problem in holography and the black hole firewall problem.

Bulk Reconstruction

It could be said that the flurry of activity involving the application of quantum information theory in holography really begins with the discovery of the Ryu-

---

[2]Previously, for MAXCUT QAOA — one of the best studied cases of the algorithm — the only known performance bounds beyond a single iteration were for specific graphs, such as the cycle graph.

Takayanagi (RT) formula back in 2006 [9], [10]. The RT formula relates the von Neumann entropy of a CFT state on a boundary subregion with a corresponding surface area within the bulk. Specifically, the RT formula states that

$$S(A) = \min_{\gamma_A} \frac{\text{Area}\left[\gamma_A\right]}{4G},$$

where $S(A)$ is the von Neumann entropy of a boundary CFT state on the subregion A, and where the minimization on the right runs through all bulk surfaces γ_A cohomologous with A. For the first time, the RT formula gives a direct and undeniable relation between bulk geometry, which is typically considered a part of the low-energy semiclassical description of the theory, and the underlying microscopic degrees of freedom that describe the dual theory.

The problem of bulk reconstruction asks about the recovery of geometric quantities in the bulk AdS spacetime given certain boundary data, and more generally, of quantifying when a well-defined semi-classical geometric bulk should be expected to exist. In Chapter 4, we study a simplified version of the bulk reconstruction problem aimed at recovering the bulk metric tensor perturbatively in $\text{AdS}_3/\text{CFT}_2$.

In particular, we consider the following scenario. Let us be given a boundary entropy profile $\{S(A)\}_{A \in \mathcal{I}}$ where $\mathcal{I}$ consists of all single interval regions A on the boundary, and where $S(A)$ represents the von Neumann algebra of some CFT state on A. The Ryu-Takayanagi formula now translates the entropy profile into the areas of bulk minimal surfaces. In the special case where we work with $\text{AdS}_3/\text{CFT}_2$, a constant time slice is naturally represented by the Poincare disk model of the hyperbolic plane, and the minimal surfaces γ_A are simply geodesics that subtend the boundary region A. Given the boundary entropy profile, or equivalently the lengths of all geodesics between pairs of points on the boundary, we may ask if this data uniquely determines the bulk metric tensor (up to diffeomorphism). In the context of geometric tomography, this problem is known as the *boundary rigidity problem*. The solution to the boundary rigidity problem is unknown for the general class of manifolds, but well-defined and studied for the case where the underlying manifold is of (nearly) constant curvature. We therefore focus our attention to case of small metric perturbations about the vacuum AdS_3 solution, where the induced metric on a constant time spatial slice is given by $g_{ij} = g_{ij}^{(0)} + h_{ij}$, where $g^{(0)}$ is the background metric of the Poincare disk, and where h is a sufficiently small metric perturbation.

Given a boundary region A and a background geodesic γ_A subtending A, the change

in geodesic length through the addition of a metric perturbation is given by

$$\Delta L(A) = \frac{1}{2} \int_{\gamma_A} h\left(\dot{\gamma}_A, \dot{\gamma}_A\right) \, ds.$$

The integral above can be considered as a map from the space of symmetric 2-tensors h to the space of smooth functions on geodesics γ_A. More precisely, we define the tensor Radon transform $R_2[h]$ of the tensor h to be the function which takes in a background geodesic curve γ_A and returns the value

$$R_2\left[h\right]\left(\gamma_A\right) = \int_{\gamma_A} h\left(\dot{\gamma}_A, \dot{\gamma}_A\right) \, ds.$$

It can be shown that the tensor Radon transform $R_2[h]$ is an injective (up to diffeomorphism) map of h, which implies a positive solution to the linearized version of the boundary rigidity problem: we may uniquely (up to diffeomorphism) recover a metric perturbation h given the lengths of all boundary anchored geodesics. Equivalently, through the RT formula, a bulk metric perturbation is uniquely (up to diffeomorphism) determined by the set of vacuum subtracted boundary entanglement entropies.

In Chapter 4, we review the theory of the tensor Radon transform and implement a numerical algorithm to explicitly reconstruct the bulk metric perturbation given a boundary entropy profile. We apply this reconstruction to a few well-motivated examples of holographic boundary data, as well as to the (ostensibly non-geometric) boundary entropy data generated from quenching a free fermion model. We show that the fidelity of the reconstruction can be used as a characterization of whether a given entropy profile has a well-defined geometric bulk dual, providing an empirical answer to the question of when a CFT state has a well-defined semi-classical bulk dual in the special case of linear perturbations about AdS$_3$.

The Black Hole Firewall Problem

In 2012, [11] presented a sharpened version of the black hole information problem called the black hole firewall paradox. We can present the key points of the paradox through a simplified toy-model of the Hawking radiation process. Consider a simple model for black hole evaporation based on the pair creation picture for Hawking radiation [12]:

1. We begin with a pure state $|\psi\rangle_M$ representing the state of the black hole on an initial time slice.

2. We imagine that Hawking radiation is then induced from pair creation. The schematic evolution of our state is given by

$$|\psi\rangle_M \mapsto |\psi'\rangle_M \otimes \frac{1}{\sqrt{2}} \left(|00\rangle_{\tilde{B}_1 B_1} + |11\rangle_{\tilde{B}_1 B_1} \right),$$

where B_1 describes the subsystem of an outgoing Hawking mode, and $\tilde{B}_1$ describes its interior partner. The fact that the state on $B_1 \tilde{B}_1$ is taken to be a maximally entangled state reflects the fact that we are assuming quantum field theory to be a valid semi-classical description of the evaporation process. For a sufficiently large black hole, the event horizon is an ordinary region of spacetime with low curvature. The validity of quantum field theory in such regimes implies that the structure of the vacuum state in the near horizon region is described by a maximally entangled state between outgoing Hawking modes and their interior partners.

3. Continued time evolution repeatedly produce new correlated pairs, with the state at step N being

$$|\Psi_N\rangle = |\psi''\rangle_M \otimes \frac{1}{\sqrt{2^N}} \bigotimes_{k=1}^{N} \left(|00\rangle_{\tilde{B}_k B_k} + |11\rangle_{\tilde{B}_k B_k} \right).$$

If we now trace away the collection of interior modes $\tilde{B}_k$, then the exterior modes B_k are left in a maximally mixed state: this is a reflection of the fact that Hawking radiation is thermal. If the evaporation process continues until the black hole has completely radiated away, then we are left with a maximally mixed thermal state. Since we started with a pure state initially, this is in violation of the unitarity of quantum mechanics.

This is essentially a description of the classic black hole information problem. Note that the contradiction here is a statement about the asymptotic S-matrix being non-unitary. What the authors of [11] did was to sharpen the above argument to finite times using entanglement entropy.

Consider some late outgoing mode B_K for some sufficiently late time step K, and denote the collection of early modes $\{B_k \mid k < K\}$ as E. From before, we know that B_K must be (maximally) entangled with its partner mode $\tilde{B}_K$ in the interior. This is formalized by the entropy condition $S(B_K \tilde{B}_K) = 0$. On the other-hand, if the ultimate process of black hole evaporation is to remain unitary, then B_K must eventually start to purify the early radiation. This means that for sufficiently large K, we must have $S(B_K E) < S(E)$.

Together we now have three conditions on the von Neumann entropy:

1. The first is the strong sub-additivity of von Neumann entropies:

$$S(ABC) + S(B) \leq S(AB) + S(BC),$$

   valid for any tripartite system ABC.

2. The second is the entropy condition

$$S(B_K \tilde{B}_K) = 0.$$

   This condition follows from the maximal entanglement of an outgoing Hawk-ing mode and its interior partner. As argued before, this relation follows from the validity of quantum field theory and the structure of the vacuum state in the near event horizon region. In effect, this equation signifies the smoothness of the black hole event horizon.

3. The final condition is

$$S(B_K E) < S(E),$$

   which signifies the unitary of the black hole evaporation process.

The paradox here is that these three conditions are mutually contradictory. Indeed, these three conditions lead to the following chain of inequalities:

$$\begin{aligned}
S(B_K) + S(E) &= S(B_K) + S(E B_K \tilde{B}_K) \\
&\leq S(B_K \tilde{B}_K) + S(B_K E) \\
&= S(B_K E) \\
&< S(E),
\end{aligned}$$

which implies that $S(B_K) < 0$, in contradiction with the non-negativity of entan-glement entropy. The firewall paradox is so named because, to the original authors of [11], the most acceptable violation is that of the second condition $S(B_K \tilde{B}_K) = 0$, signaling a breakdown of the smoothness of spacetime structure and the presence of an energetic firewall at the black hole event horizon.

Since its very inception, there have been numerous attempts to resolve the firewall paradox. While it is arguable that the firewall paradox has yet to be fully resolved, the

most commonly accepted class of resolutions falls under the collective denomination of *ER=EPR* type solutions. The term *ER=EPR* comes from a postulated equivalence between entanglement (as represented by the maximally entangled EPR pair), and spacetime locality (as represented by the acronym ER, which is short for the Einstein-Rosen bridge). The *ER=EPR* philosophy[3] proposes that instead of considering the black hole interior as being separate from the exterior radiation, the entanglement structure of the system suggests that a part of the black hole interior should instead be identified as a subsystem of the exterior radiation. In short, if the black hole firewall problem stems from a fundamental violation of the monogamy of entanglement, then the most straightforward solution is to simply identify the problematic parties.

Although this solution is at first glance both radical and bewildering, there has been accumulating evidence which suggests that this is truly what happens. Perhaps the most convincing piece of evidence is the recent discovery of the so-called "replica wormhole" geometries of the gravitational path integral. [14], [15] The replica wormhole geometries are additional, and previously unincluded, saddle points to the calculation of the entanglement entropy using the replica trick for the gravitational path integral. The inclusion of these new saddles bring about a modification to the calculation of the radiation entropy at late times, with the striking feature that unitarity is recovered. The inclusion of the replica wormholes produces the so-called *island formula* for the entanglement entropy of the radiation subsystem, given by

$$S(E) = \min_X \left[\frac{\text{Area}(X)}{4G} + S_{\text{bulk}}\left(E \cup I\right) \right].$$

In the formula above, E denotes the exterior radiation subsystem, and $S(E)$ the usual von Neumann entropy. The extermization on the right-hand side is over closed codimension-2 spatial surfaces X, and I is a spatial region such that $\partial I = X$, called the *island*. The extremizing surface $\tilde{X}$ is called a *quantum extremal surface*. The term $S_{\text{bulk}}\left(E \cup I\right)$ denotes the coarse-grained entropy on the union of the radiation and island subsystems, calculated in a semi-classical approximation using the Euclidean path integral.

It can be shown that at early times, there is no non-trivial solution to the extremization problem and the quantum extremal surface $\tilde{X}$ (and hence also the island region I) is

---

[3]The term *ER=EPR* was first coined by Juan Maldacena and Leonard Susskind in [13]. The same term has since been appropriated by various authors for a wide selection of distinct, albeit related, mechanisms and proposals. In this thesis, we will use the term *ER=EPR* to refer to the guiding philosophy that there should be some intimate, even if somewhat ill-defined, relation between spacetime locality and entanglement.

empty. Correspondingly, the island formula simply reproduces Hawking's original entropy calculation at early times, suggesting that

$$S(E) = S_{\text{bulk}}(E)$$

at early times. If this result persists to late times, then we are again confronted with the black hole information problem and the loss of unitarity. At late times, however, there exists a non-trivial solution to the quantum extremal surface which is located near the black hole event horizon, and the corresponding island region is non-empty. The island I in this case contains part of the black hole interior, and so it contains the interior partners of the exterior Hawking radiation. The subsystem E is essentially purified by the inclusion of the island I and $S_{\text{bulk}}(E \cup I)$ is consequently small. The island formula thus reduces at late times to

$$S(E) \approx \frac{\text{Area}(\tilde{X})}{4G} \approx S_{\text{BH}},$$

where S_{BH} is the Beckenstein-Hawking entropy of the remaining black hole. The island formula is therefore seen to accurately reproduce the unitary page curve

$$S(E) = \min\{S_{\text{bulk}}(E), S_{\text{BH}}\}.$$

The island formula suggests that unitary is recovered by identifying an island region in the interior of the black hole with the exterior radiation, and it provides what is perhaps the most concrete implementation of the *ER=EPR* philosophy yet. However, there are also many unresolved issues remaining.

One puzzling feature is how the semi-classical path integral has knowledge of the fine-grained entropy in the first place. In particular, the island formula provides no suggestions as to what is going on microscopically. In Chapter 5, we aim to examine some of these issues from the perspective of quantum error-correction and quantum computational complexity. Taking seriously the suggestion that the black hole interior must be embedded within the exterior radiation, and following a general philsophy first outlined by Harlow and Hayden in [16], we suggest computational pseudorandomness as an underlying mechanism for realizing the black hole embedding. By axiomizing the thermality of Hawking radiation as an assumption that the radiation forms a pseudorandom state (i.e., a state which cannot be distinguished from a thermal state by a computation of low-complexity), we show that each black hole defines a natural quantum error-correcting code which encodes the interior modes of the black hole into the exterior radiation. The presence of the black hole

code gives an operational meaning to the suggested embedding of the black hole interior into the exterior radiation within the framework of quantum information theory.

BIBLIOGRAPHY

[1] S. Bravyi, M. B. Hastings, and S. Michalakis, "Topological quantum order: Stability under local perturbations," *Journal of Mathematical Physics*, vol. 51, no. 9, p. 093 512, 2010. DOI: `10.1063/1.3490195`.

[2] J. M. Maldacena, "The large n limit of superconformal field theories and supergravity," *Advances in Theoretical and Mathematical Physics*, vol. 2, pp. 231–252, 1998. DOI: `10.1023/A:1026654312961`.

[3] A. Almheiri, X. Dong, and D. Harlow, "Bulk locality and quantum error correction in AdS/CFT," *Journal of High Energy Physics*, no. 4, p. 163, Apr. 2015. DOI: `10.1007/JHEP04(2015)163`.

[4] F. Pastawski, B. Yoshida, D. Harlow, and J. Preskill, "Holographic quantum error-correcting codes: Toy models for the bulk/boundary correspondence," *Journal of High Energy Physics*, no. 6, p. 149, Jun. 2015. DOI: `10.1007/JHEP06(2015)149`.

[5] X. Chen, Z.-C. Gu, and X.-G. Wen, "Classification of gapped symmetric phases in one-dimensional spin systems," *Physical Review B*, vol. 83, no. 3, p. 035 107, 2011. DOI: `10.1103/PhysRevB.83.035107`.

[6] N. Schuch, D. Pérez-García, and I. Cirac, "Classifying quantum phases using matrix product states and projected entangled pair states," *Physical Review B*, vol. 84, p. 165 139, 16 Oct. 2011. DOI: `10.1103/PhysRevB.84.165139`.

[7] M. H. Freedman and M. B. Hastings, "Quantum systems on non-k-hyperfinite complexes: A generalization of classical statistical mechanics on expander graphs," *Quantum Information and Computation*, vol. 14, no. 144, 2014.

[8] E. Farhi, J. Goldstone, and S. Gutmann, "A Quantum Approximate Optimization Algorithm," Nov. 2014. arXiv: `1411.4028`.

[9] S. Ryu and T. Takayanagi, "Holographic derivation of entanglement entropy from AdS/CFT," *Physical Review Letters*, vol. 96, p. 181 602, 2006. DOI: `10.1103/PhysRevLett.96.181602`.

[10] ——, "Aspects of Holographic Entanglement Entropy," *Journal of High Energy Physics*, no. 08, p. 045, 2006. DOI: `10.1088/1126-6708/2006/08/045`.

[11] A. Almheiri, D. Marolf, J. Polchinski, and J. Sully, "Black holes: Complementarity or firewalls?" *Journal of High Energy Physics*, no. 2, p. 62, Feb. 2013. DOI: `10.1007/JHEP02(2013)062`.

[12] S. D. Mathur, "The information paradox: A pedagogical introduction," *Classical and Quantum Gravity*, vol. 26, no. 22, p. 224 001, Oct. 2009, ISSN: 1361-6382. DOI: `10.1088/0264-9381/26/22/224001`.

[13] J. Maldacena and L. Susskind, "Cool horizons for entangled black holes," *Fortschritte der Physik*, vol. 61, no. 9, pp. 781–811, 2013. DOI: `10.1002/prop.201300020`.

[14] G. Penington, S. H. Shenker, D. Stanford, and Z. Yang, "Replica wormholes and the black hole interior," 2020. arXiv: `1911.11977`.

[15] A. Almheiri, T. Hartman, J. Maldacena, and E. Shaghoulian, "Replica wormholes and the entropy of hawking radiation," *Journal of High Energy Physics*, no. 13, May 2020. DOI: `10.1007/JHEP05(2020)013`.

[16] D. Harlow and P. Hayden, "Quantum computation vs. firewalls," *Journal of High Energy Physics*, no. 6, p. 85, Jun. 2013. DOI: `10.1007/JHEP06(2013)085`.

Part I

Low Energies

Chapter 2

QUANTUM ERROR-DETECTION AT LOW ENERGIES

Motivated by the close relationship between quantum error-correction, topological order, the holographic AdS/CFT duality, and tensor networks, we initiate the study of approximate quantum error-detecting codes in matrix product states (MPS). We first show that using open-boundary MPS to define boundary to bulk encoding maps yields at most constant distance error-detecting codes. These are degenerate ground spaces of gapped local Hamiltonians. To get around this no-go result, we consider excited states, i.e., we use the excitation ansatz to construct encoding maps: these yield error-detecting codes with distance $\Omega(n^{1-\nu})$ for any $\nu \in (0,1)$ and $\Omega(\log n)$ encoded qubits. This shows that gapped systems contain – within isolated energy bands – error-detecting codes spanned by momentum eigenstates. We also consider the gapless Heisenberg-XXX model, whose energy eigenstates can be described via Bethe ansatz tensor networks. We show that it contains – within its low-energy eigenspace – an error-detecting code with the same parameter scaling. All these codes detect arbitrary d-local (not necessarily geometrically local) errors even though they are not permutation-invariant. This suggests that a wide range of naturally occurring many-body systems possess intrinsic error-detecting features.

This chapter is based on the published article:

M. Gschwendtner, R. König, B. Şahinoğlu, and E. Tang, "Quantum error-detection at low energies," *Journal of High Energy Physics*, no. 9, Sep. 2019. DOI: 10.1007/jhep09(2019)021.

2.1 Introduction

Quantum error-correcting codes are fundamental for achieving robust quantum memories and fault-tolerant quantum computation. Following seminal work by Shor [1] and others [2]–[5], the study of quantum error-correction has seen tremendous progress from both the theoretical and the experimental point of view. Beyond its operational implications for the use of faulty quantum hardware, quantum error-correction is closely connected to fundamental physics, as shown early on by the work of Kitaev [6]: the ground space of a topologically ordered model constitutes a quantum error-correcting code whose dimension depends on the topology of the

underlying surface containing the physical degrees of freedom. In addition to giving rise to a new field called topological quantum computing [7]–[13], this work has had a significant impact on the problem of classifying topologically ordered phases in two spatial dimensions [14], [15]. Motivated by the success of this program, follow-up work has pursued the classification of gapped phases of matter with or without global symmetries, starting from one spatial dimension [16]–[19] up to arbitrarily high dimensions [20], [21].

More recently, concepts from quantum error-correction have helped to resolve conceptual puzzles in AdS/CFT holographic duality. Almheiri, Dong, and Harlow **almheiri15** have proposed that subspaces of holographic conformal field theories (CFTs) which are dual to perturbations around a particular classical bulk AdS geometry constitute a quantum error-correcting code robust against erasure errors. In this proposal, the bulk and boundary degrees of freedom correspond to the logical and the physical degrees of freedom of the code, respectively. Puzzling features such as subregion-subregion duality and radial commutativity can naturally be understood in this language, under the hypothesis that the duality map works as a code which recovers, from erasure, part of the boundary degrees of freedom. Related to this picture, Ryu-Takayanagi type formulas have been shown to hold in any quantum error-correcting code that corrects against erasure [22].

Key to many of these results in the context of topological order and the AdS/CFT holographic duality is the language of tensor networks. The latter, originating in work by Fannes, Nachtergaele, and Werner on finitely correlated states [23] and the density matrix renormalization group [24], [25], has seen a revival in the last 15 years. Major conceptual contributions include the introduction of matrix product states by [26]–[30], the introduction of the multi-scale entanglement renormalization ansatz (MERA) [31] by Vidal, and various projected entangled-pair states (PEPS) techniques [29], [32]–[36] for higher dimensional systems.

It has been shown that tensor network techniques provide exact descriptions of topologically ordered states [37]–[39], and furthermore, tensor networks have been instrumental in the characterization and classification of topological order [40]–[44]. This approach has also been generalized to higher dimensions, clarifying the connections to topological quantum field theories [45].

A similar success story for the use of tensor networks is emerging in the area of AdS/CFT duality. Aspects of holographic duality have been explored in terms of toy models based on tensor networks [46]–[48]. Indeed, many (though not all)

conjectured features of this duality can be recovered in these examples. This field, while still in its infancy, has provided new appealing conjectures which point to a potentially more concrete understanding of the yet to be uncovered physics of quantum gravity [49], [50].

Given the existing close connections between quantum error-correction and a variety of physical systems ranging from topological order to AdS/CFT, it is natural to ask how generic the appearance of error-correcting features is in naturally occurring quantum many-body systems. A first step towards showing the ubiquity of such features is the work of Brandao, et. al. [51]. There, it is shown that quantum chaotic systems satisfying the Eigenstate Thermalization Hypothesis (ETH) have energy eigenstates that form approximate quantum error-correcting codes. Nearby extensive energy eigenstates of 1D translation invariant Hamiltonians, as well as ground spaces of certain gapless systems (including the Heisenberg and Motzkin models), also contain approximate quantum error-correcting codes. Motivated by this work, we ask if one can demonstrate the existence of error-correcting codes within the low-energy eigenspaces of generic Hamiltonians, whether or not they are gapped or gapless. Specifically, we ask this question for 1D systems.

Our work goes beyond earlier work by considering errors (that is, noise) of a more general form: existing studies of error-correction in the context of entanglement renormalization and/or holography have primarily concentrated on qubit loss, modeled by so-called erasure errors (see e.g., [47], [52], [53]). This erasure noise model has several theoretical advantages. In particular, it permits one to argue about the existence of recovery maps in terms of entanglement entropies of the associated erased regions. This can be connected to well-known results on entanglement entropies in critical 1D systems. Furthermore, the appearance of entanglement entropies in these considerations is natural in the context of the AdS/CFT duality, where these quantities are involved in the connection of the boundary field theory to the bulk geometry via the Ryu-Takayangi formula. However, compared to other forms of errors typically studied in the quantum fault-tolerance community, erasure is quite a restricted form of noise: it is, in a certain sense, much easier to correct than, e.g., depolarizing noise. As an example to illustrate this point, we recall that the toric code can recover from loss of half its qubits [54], whereas it can only tolerate depolarizing noise up to a noise rate of 11% even given perfect syndrome measurements [9]. Motivated by this, we aim to analyze error-correcting properties with respect to more generic noise even though this precludes the use of

entanglement entropies. Again, the work [51] provides first results in this direction by considering errors on a fixed, connected subset of sites (that is, geometrically localized errors). The restriction to a connected subset was motivated in part by the consideration of permutation-invariant subspaces (note other previous works on permutation-invariant code spaces [55], [56]). In our work, we lift the restriction to permutation-invariant codes and instead analyze arbitrary weight-d errors with potentially disconnected supports. Furthermore, we study an operational task – that of error-detection – with respect to a noise model where errors can occur on any subset of qubits of a certain size, instead of only a fixed subset.

We find that the language of matrix product states (MPS) and the related excitation ansatz states provides a powerful analytical tool for studying error-detection in 1D systems. In particular, we relate properties of transfer operators to error-detection features: for MPS describing (degenerate) ground spaces of gapped Hamiltonians, injectivity of the transfer operators gives rise to a no-go theorem. For excitation ansatz states describing the low-energy excitations of gapped systems, we use injectivity and a certain normal form to establish error-correction properties. Finally, for a gapless integrable model, we analyze the Jordan structure of (generalized) transfer matrices to find bounds on code parameters. In this way, our work connects locally defined features of tensor networks to global error-correction properties. This can be seen as a first step in an organized program of studying approximate quantum error-correction in tensor network states.

2.2 Our Contribution

We focus on error-detection, a natural primitive in fault-tolerant quantum computation. Contrary to full error-correction, where the goal is to recover the initial encoded state from its corrupted version, error-detection merely permits one to decide whether or not an error has occurred. Errors (such as local observables) detected by an error-detecting code have expectation values independent of the particular logical state. In the context of topological order, where local errors are considered, error-detection has been referred to as TQO-1 (topological quantum order condition 1); see, e.g., [57]. An approximate version of the latter is discussed in [58].

A code, i.e., a subspace of the physical Hilbert space, is said to be error-detecting (for a set of errors) if the projection back onto the code space after the application of an error results in the original encoded state, up to normalization. Operationally, this

means that one can ensure that no error occurred by performing a binary-outcome POVM consisting of the projection onto the code space or its complement. This notion of an error-detecting code is standard, though quite stringent: unless the code is constructed algebraically (e.g., in terms of Pauli operators), it is typically not going to have this property.

Our first contribution is a relaxed, yet still operationally meaningful definition for approximate error-detection. It relaxes the former notion in two directions: first, the post-measurement state is only required to approximate the original encoded state. Second, we only demand that this approximation condition is satisfied if the projection onto the code space occurs with non-negligible probability. This is motivated by the fact that if this projection does not succeed with any significant probability, the error-detection measurement has little effect (by the gentle measurement lemma [59]) and may as well be omitted. More precisely, we consider a CPTP map $\mathcal{N} : \mathcal{B}((\mathbb{C}^{\mathrm{p}})^{\otimes n}) \to \mathcal{B}((\mathbb{C}^{\mathrm{p}})^{\otimes n})$ modeling noise on n physical qudits (of dimension p). Here the Kraus operators of $\mathcal{N}$ take the role of errors (considered in the original definition). We define the following notion:

Definition 2.3.1 (Approximate quantum error-detecting code). *A subspace $C \subset (\mathbb{C}^{\mathrm{p}})^{\otimes n}$ (with associated projection P) is an (ϵ, δ)-approximate error-detecting code for $\mathcal{N}$ if for any state $|\Psi\rangle \in C$ the following holds:*

$$\text{if} \qquad \text{tr}(P\mathcal{N}(|\Psi\rangle\langle\Psi|)) \geq \delta \qquad \text{then} \qquad \langle\Psi|\rho_{\mathcal{N},P}|\Psi\rangle \geq 1 - \epsilon \,,$$

where $\rho_{\mathcal{N},P} = \text{tr}(P\mathcal{N}(|\Psi\rangle\langle\Psi|))^{-1} \cdot P\mathcal{N}(|\Psi\rangle\langle\Psi|)P.$

This definition ensures that the post-measurement state $\rho_{\mathcal{N},P}$ is close (as quantified by ϵ) to the initial code state when the outcome of the POVM is P. Furthermore, we only demand this in the case where $\mathcal{N}(|\Psi\rangle\langle\Psi|)$ has an overlap with the code space of at least δ.

In the following, we often consider families of codes $\{C_n\}_n$ indexed by the number n of physical spins. In this case, we demand that both approximation parameters ϵ_n and δ_n tend to zero as $n \to \infty$. This is how we make sure that we have a working error-detecting code in the asymptotic or thermodynamic limit of the physical Hilbert space.

Of particular interest are errors of weight d, i.e., errors which only act non-trivially on a subset of d of the n subsystems in the product space $(\mathbb{C}^{\mathrm{p}})^{\otimes n}$. We call this subset the *support* of the error, and refer to the error as *d-local*. We emphasize that

throughout this paper, d-local only refers to the weight of the errors: they do not need to be geometrically local, i.e., their support may be disconnected. In contrast, earlier work on approximate error-correction such as [51] only considered errors with support on a (fixed) connected subset of d sites. We then define the following:

Definition 2.3.3 (Error-detection for d-local errors)**.** *A subspace $C \subset (\mathbb{C}^{\mathsf{p}})^{\otimes n}$ is called an $(\epsilon, \delta)[[n, k, d]]$-approximate quantum error-detecting code (AQEDC) if $\dim C = \mathsf{p}^k$ and if C is an (ϵ, δ)-approximate error-detecting code for any CPTP map $\mathcal{N} : \mathcal{B}((\mathbb{C}^{\mathsf{p}})^{\otimes n}) \to \mathcal{B}((\mathbb{C}^{\mathsf{p}})^{\otimes n})$ of the form*

$$\mathcal{N}(\rho) = \sum_{j \in [J]} p_j F_j \rho F_j^{\dagger} \,, \tag{2.1}$$

where each F_j is a d-local operator with $\|F_j\| \leq 1$ and $\{p_j\}_{j \in [J]}$ is a probability distribution. We refer to d as the distance *of the code.*

In other words, an $(\epsilon, \delta)[[n, k, d]]$-AQEDC deals with error channels which are convex combinations of d-local errors. This includes for example the commonly considered case of random Pauli noise (assuming the distribution is supported on errors having weight at most d). However, it does not cover the most general case of (arbitrary) d-local errors/error channels because of the restriction to convex combinations. The consideration of convex combinations of d-local errors greatly facilitates our estimates and allows us to consider settings that go beyond earlier work. We leave it as an open problem to lift this restriction, and only provide some tentative statements in this direction.

To exemplify in what sense our definition of AQEDC for d-local errors extends earlier considerations, consider the case where the distribution over errors in (2.2) is the uniform distribution over all d-qudit Pauli errors on n qubits. In this case, the number of Kraus operators in the representation (2.2) is polynomial in n even for constant distance d. In particular, arguments involving the number of terms in (2.2) cannot be used to establish bounds on the code distance as in [51], where instead only Pauli errors acting on d fixed sites were considered: The number of such operators is only 4^d instead of the number $\binom{n}{d} 4^d$ of all weight-$\leq d$-Paulis, and, in particular, does not depend on the system size n.

We establish the following approximate Knill-Laflamme type conditions which are sufficient for error-detection:

Corollary 2.3.4. *Let $C \subset (\mathbb{C}^p)^{\otimes n}$ be a code with orthonormal basis $\{\psi_\alpha\}_{\alpha \in [p^k]}$ such that (for some $\gamma > 0$),*

$$\left| \langle \psi_\alpha | F | \psi_\beta \rangle - \delta_{\alpha,\beta} \langle \psi_1 | F | \psi_1 \rangle \right| \leq \gamma \cdot \|F\| \qquad \text{for all } \alpha, \beta \in [p^k] \,,$$

for every d-local operator F on $(\mathbb{C}^p)^{\otimes n}$. Let $\delta > p^{5k}\gamma^2$. Then C is an $(\epsilon = p^{5k}\gamma^2\delta^{-1}, \delta)[[n, k, d]]$-AQEDC.

This condition, which is applicable for "small" code space dimension, i.e., $k = O(\log n)$, allows us to reduce the consideration of approximate error-detection to the estimation of matrix elements of local operators. We also establish a partial converse to this statement: if a subspace $C \subset (\mathbb{C}^p)^{\otimes n}$ contains two orthonormal vectors whose reduced d-local density operators (for some subset of d sites) are almost orthogonal, then C cannot be an error-detecting code with distance d (see Lemma 2.3.6 for a precise statement).

Equipped with these notions of approximate error-detection, we study quantum many-body systems in terms of their error-detecting properties using tensor network techniques. More specifically, we consider two types of code families, namely:

(i) codes that are degenerate ground spaces of local Hamiltonians and permits a description in terms of tensor networks, and

(ii) codes defined by low-energy eigenstates of (geometrically) local Hamiltonians, with the property that these can be efficiently described in terms of tensor networks.

As we explain below, (i) and (ii) are closely connected via the parent Hamiltonian construction. For (i), we follow a correspondence between tensor networks and codes which is implicit in many existing constructions: we may think of a tensor as a map from certain virtual to physical degrees of freedom. To define this map, consider a tensor network given by a graph $G = (V, E)$ and a collection of tensors A. Let us say that an edge $e \in E$ is a dangling edge if one of its vertices has degree 1, and let us call the corresponding vertices the dangling vertices of the tensor network. An edge $e \in E$ is an internal edge if it is not a dangling edge; we use an analogous notion for vertices. We assume that each internal edge $e \in E$ is associated a virtual space of fixed bond dimension D, and each dangling edge with a physical degree of dimension p. Then the tensor network associates a tensor T of degree $\deg(v)$

to each internal vertex v of G, where it is understood that indices corresponding to internal edges are contracted. The tensor network is fully specified by the family A of such tensors. We partition the set of dangling vertices into two subsets M and M^c. Then the tensor network defines a map $\Gamma(A, G) : (\mathbb{C}^p)^{\otimes|M|} \to (\mathbb{C}^p)^{\otimes|M^c|}$ as each fixing of the degrees of freedom in M defines an element of the Hilbert space associated with the degrees of freedom in M^c by tensor contraction. That is, the map depends on the graph G specifying the structure of the tensor network, as well as the family A of local tensors. In particular, fixing a subspace of $(\mathbb{C}^p)^{\otimes|M|}$, its image under the map $\Gamma(A, G)$ defines a subspace $C \subset (\mathbb{C}^p)^{\otimes|M^c|}$ which we will think of as an error-correcting code. In the following, we also allow the physical and virtual (bond) dimensions to vary (depending on the location in the tensor network); however, this description captures the essential construction.

This type of construction is successful in two and higher spatial dimensions, yielding error-correcting codes with macroscopic distance: examples are the ground states of the toric code [40], [60] and other topologically ordered models [37], [39], [41]. However, in 1D, it seems a priori unlikely that the very same setup can generate any nontrivial quantum error-detection code, at least for gapped systems. This is because of the exponential decay of correlations [61]–[63] and the lack of topological order without symmetry protection [18], [64]. We make this precise by stating and proving a no-go theorem.

More precisely, we follow the above setup provided by the boundary-to-bulk tensor network map $\Gamma(A) = \Gamma(A, G)$. Here, G is the 1D line graph with dangling edges attached to internal vertices, which is equivalent to considering the ground space of 1D local gapped Hamiltonians with open boundary conditions. The associated tensor network is a matrix product state.

Generic MPS satisfy a condition called injectivity, which is equivalent to saying that the transfer matrix of the MPS is gapped. Exploiting this property allows us to prove a lower bound on the distinguishability of d-local reduced density operators for any two orthogonal states in the code space. This bound is expressed in terms of the virtual bond dimension D of the MPS tensor A. In particular, the bound implies the following no-go theorem for codes generated by MPS as described above.

Theorem 2.5.3. *Let $C \subset (\mathbb{C}^p)^{\otimes n}$ be an approximate quantum error-detecting code generated by $\Gamma(A)$, i.e., a translation-invariant injective MPS of constant bond dimension D by varying boundary conditions. Then the distance of C is constant.*

The physical interpretation of this theorem is as follows: for every injective MPS with periodic boundary conditions, there exists a strictly $\log D$-geometrically local gapped Hamiltonian such that the MPS is the unique ground state [28]. One can further enlarge the ground space by leaving out a few Hamiltonian terms near the boundary. The degeneracy then depends on the number of terms omitted, and the ground states are described by open boundary condition MPS. Then, our no-go theorem implies that the ground space of any such parent Hamiltonian arising from such a constant bond-dimension MPS is a trivial code, i.e., it can have at most a constant distance. This result is equivalent to saying that there is no topological quantum order in the ground space of 1D gapped systems.[1]

To get around this no-go result, we extend our considerations beyond the ground space and include low-energy subspaces in the code space. We show that this indeed leads to error-detecting codes with macroscopic distance. We identify two ways of constructing nontrivial codes by either considering single-particle excitations of varying momenta, or by considering multi-particle excitations above the ground space. Both constructions provide us with codes having distance scaling asymptotically significantly better than what can be achieved in the setup of our no-go theorem. In fact, the code distance is a polynomial arbitrarily close to linear in the system size (i.e., n) in both cases.

Our first approach, using states of different momenta, involves the formalism of the excitation ansatz (see Section 2.6 for a review). This gives a tensor network parametrization of momentum eigenstates associated with a Hamiltonian having quasi-particle excitations. We show the following:

Theorem 2.6.9. *Let $\nu \in (0, 1)$ and let $\kappa, \lambda > 0$ be such that*

$$5\kappa + \lambda < \nu \ .$$

Let A, $B(p)$ be tensors associated with an injective excitation ansatz state $|\Phi_p(B; A)\rangle$, where p is the momentum of the state. Then there is a subspace $C \subset (\mathbb{C}^p)^{\otimes n}$ spanned by excitation ansatz states $\{|\Phi_p(B; A)\rangle\}_p$ with different momenta p such that C is

---

[1]More precisely, this statement holds for systems whose ground states can be approximated by constant bond dimension MPS. It is not clear whether this is sufficient to make a statement about general 1D local gapped Hamiltonians. The identification of ground states of 1D local gapped Hamiltonians with constant bond dimension MPS is sometimes made in the literature, as for example in the context of classifying phases [18], [19], [64].

an $(\epsilon, \delta)[[n, k, d]]$-AQEDC with parameters

$$k = \kappa \log_\mathsf{p} n \, ,$$
$$d = n^{1-\nu} \, ,$$
$$\epsilon = \Theta(n^{-(\nu-(5\kappa+\lambda))}) \, ,$$
$$\delta = n^{-\lambda} \, .$$

The physical interpretation of this result stems from the fact that excitation ansatz states approximate quasi-particle excitations: given a local gapped Hamiltonian, assuming a good MPS approximation to its ground state, we can construct an arbitrarily good approximation to its isolated quasi-particle excitation bands by the excitation ansatz. This approximation guarantee is shown using Lieb-Robinson type bounds [61], [65] based on a previous result [66] which employs the method of energy filtering operators. Thus our result demonstrates that generic low-energy subspaces contain approximate error-detecting codes with the above parameters. Also, note that unlike the codes considered in [51, Theorem 1], the excitation ansatz codes are comprised of finite energy states, and not finite energy *d*ensity states.

We remark that the choice of momenta is irrelevant for this result; it is not necessary to restrict to nearby momenta. Instead, any subset of momentum eigenstates can be used. The only limitation here is that the number of different momenta is bounded by the dimension of the code space. This is related to the fact that localized wave functions (which would lead to a non-extensive code distance) are a superposition of many different momenta, a fact formally expressed by the position-momentum uncertainty relation.

Our second approach for side-stepping the no-go theorem is to consider multi-particle excitations. We consider a specific model, the periodic Heisenberg-XXX spin chain Hamiltonian H on n qubits. We find that there are good error-detecting codes within the low-energy subspace of this system. For this purpose, we consider the state

$$|\Psi\rangle = \sum_{m=1}^{n} \omega^m \mathbf{s}_m^- |1\rangle^{\otimes n} \qquad \text{where} \qquad \omega = e^{2\pi i/n}, \tag{2.2}$$

and where $\mathbf{s}_m^- = |0\rangle\langle 1|$ changes the state of the m-th spin from $|1\rangle$ to $|0\rangle$. This has energy $O(1/n^2)$ above the ground state energy of H. The corresponding eigenspace is degenerate and contains all "descendants" $S_-^r|\Psi\rangle$ for $r \in \{0, \dots, n-2\}$, where $S_- = \sum_{m=1}^{n} \mathbf{s}_m^-$ is the (total) spin lowering operator. We also note that each state $S_-^r|\Psi\rangle$

has fixed momentum $2\pi/n$, and that r directly corresponds to its total magnetization. We emphasize that these states are, in particular, not permutation-invariant. Our main result concerning these states is the following:

Theorem 2.7.9. *Let $\nu \in (0, 1)$ and $\kappa, \lambda > 0$ be such that*

$$6\kappa + \lambda < \nu .$$

Then there is a subspace C spanned by descendant states $\{S_-^r |\Psi\rangle\}_r$ with magnetization r pairwise differing by at least 2 such that C is an $(\epsilon, \delta)[[n, k, d]]$-AQEDC with parameters

$$
\begin{aligned}
k &= \kappa \log_2 n , \\
d &= n^{1-\nu} , \\
\epsilon &= \Theta\big(n^{-(\nu-(6\kappa+\lambda))}\big) , \\
\delta &= n^{-\lambda} .
\end{aligned}
$$

This code, which we call the magnon-code, can also be seen to be realized by tensor networks. The state (2.2) has an MPS description with bond dimension 2 and the descendants $S_-^r |\Psi\rangle$ can be expressed using a matrix-product operator (MPO) description of the operator S_-. More generally, it is known that these states form an example of the algebraic Bethe ansatz, and the latter have a natural tensor network description [67]. This suggests that our results may generalize to other exactly solvable models.

Outline

The paper is organized as follows. We discuss our notion of approximate error-detection and establish sufficient and necessary conditions in Section 2.3. In Section 2.4, we review the basics of matrix product states. We also establish bounds on expectation values in terms of properties of the associated transfer operators. In Section 2.5, we prove our no-go theorem and show the limits of error-detection for code spaces limited to the ground space of a gapped local Hamiltonian. We then consider low-energy eigenstates of local Hamiltonians and show how they perform asymptotically better than the limits of the no-go theorem. We first consider single-particle momentum eigenstates of generic local gapped Hamiltonians in Section 2.6. In Section 2.7, we consider codes defined by many-particle eigenstates of the Heisenberg-XXX model.

2.3 Approximate Quantum Error-Detection

Here we introduce our notion of approximate quantum error-detection. In Section 2.3.1, we give an operational definition of this notion. In Section 2.3.2, we provide sufficient conditions for approximate quantum error-detection which are analogous to the Knill-Laflamme conditions for quantum error-correction [4]. Finally, in Section 2.3.3, we give necessary conditions for a subspace to be an approximate quantum error-detecting code.

2.3.1 Operational Definition of Approximate Error-Detection

Let $\mathcal{N} : \mathcal{B}((\mathbb{C}^p)^{\otimes n}) \to \mathcal{B}((\mathbb{C}^p)^{\otimes n})$ be a CPTP map modeling noise on n physical qubits. We introduce the following notion:

Definition 2.3.1. *A subspace $C \subset (\mathbb{C}^p)^{\otimes n}$ (with associated projection P) is an (ϵ, δ)-approximate error-detection code for $\mathcal{N}$ if for any pure state $|\Psi\rangle \in C$ the following holds:*

$$\text{if} \qquad \text{tr}(P\mathcal{N}(|\Psi\rangle\langle\Psi|)) \geq \delta \qquad \text{then} \qquad \langle\Psi|\rho_{\mathcal{N},P}|\Psi\rangle \geq 1 - \epsilon \,,$$

where $\rho_{\mathcal{N},P} = \text{tr}(P\mathcal{N}(|\Psi\rangle\langle\Psi|))^{-1} \cdot P\mathcal{N}(|\Psi\rangle\langle\Psi|)P.$

In this definition, $\rho_{\mathcal{N},P}$ is the post-measurement state when applying the POVM $\{P, I - P\}$ to $\mathcal{N}(|\Psi\rangle\langle\Psi|)$. Roughly speaking, this definition ensures that the post-measurement state is ϵ-close to the initial code state if the outcome of the POVM is P. Note, however, that we only demand this in the case where $\mathcal{N}(|\Psi\rangle\langle\Psi|)$ has an overlap with the code space of at least δ. The idea behind this definition is that if this overlap is negligible, then the outcome P does not occur with any significant probability and the error-detection measurement may as well be omitted.

Definition 2.3.1 is similar in spirit to operationally defined notions of approximate quantum error-correction considered previously. In [68], approximate error-correction was defined in terms of the "recoverable fidelity" of any encoded pure state affected by noise. The restriction to pure states in the definition is justified by means of an earlier result by Barnum, Knill, and Nielsen [69].

We note that, by definition, an (ϵ, δ)-approximate error-detection code for $\mathcal{N}$ is also an (ϵ', δ')-approximate error-detection code for any $\epsilon \leq \epsilon'$ and $\delta \leq \delta'$. The traditional "exact" notion of a quantum error-detecting code C (see e.g., [70]) demands that for a set $\mathcal{F} \subset \mathcal{B}((\mathbb{C}^p)^{\otimes n})$ of *detectable errors*, we have

$$\langle\Psi|E|\Phi\rangle = \lambda_E \langle\Psi|\Phi\rangle \qquad \text{for all } |\Psi\rangle, |\Phi\rangle \in C$$

for some scalar $\lambda_E \in \mathbb{C}$ depending only on E, for all $E \in \mathcal{F}$. It is straightforward to see that such a code defines a $(0, 0)$-approximate error-detecting code of any CPTP map $\mathcal{N}$ whose Kraus operators belong to $\mathcal{F}$.

2.3.2 Sufficient Conditions for Approximate Quantum Error-Detection

The following theorem shows that certain approximate Knill-Laflamme-type conditions are sufficient for approximate error-detection.

Theorem 2.3.2. *Let $\mathcal{N}(\rho) = \sum_{j\in[J]} R_j \rho R_j^\dagger$ be a CPTP map on $\mathcal{B}((\mathbb{C}^p)^{\otimes n})$. Let $C \subset (\mathbb{C}^p)^{\otimes n}$ be a subspace with orthonormal basis $\{\psi_\alpha\}_{\alpha\in[K]}$. Define*

$$\epsilon_{\mathsf{approx}} := \max_{\alpha,\beta\in[K]} \sum_{j\in[J]} \left| \langle \psi_\alpha | R_j | \psi_\beta \rangle - \delta_{\alpha,\beta} \langle \psi_1 | R_j | \psi_1 \rangle \right|^2 . \tag{2.3}$$

Let $\delta > K^5 \epsilon_{\mathsf{approx}}$ be arbitrary. Then the subspace C is an (ϵ, δ)-approximate quantum error-detection code for $\mathcal{N}$ with $\epsilon = K^5 \epsilon_{\mathsf{approx}} \delta^{-1}$.

This theorem deals with cases where the code dimension K is "small" compared to other quantities. We will later apply this theorem to the case where K is polynomial, and where $\epsilon_{\mathsf{approx}}$ and δ are inverse polynomial in the system size n.

We note that the conditions of Theorem 2.3.2 may appear more involved than e.g., the Knill-Laflamme type conditions (see [4]) for (exact) quantum error-correction: the latter involve one or two error operators (interpreted as Kraus operators of the channel), whereas in expression (2.3.2), we sum over all Kraus operators. It appears that this is, to some extent, unavoidable when going from exact to approximate error-correction/detection in general. We note that (tight) approximate error-correction conditions [71] obtained by considering the decoupling property of the complementary (encoding plus noise) channel similarly depend on the entire noise channel. Nevertheless, we show below that, at least for probabilistic noise, simple sufficient conditions for quantum error-detection involving only individual Kraus operators can be given.

Proof. Let us define

$$\mathsf{err}^\psi(R, \alpha, \beta) := \langle \psi_\alpha | R | \psi_\beta \rangle - \delta_{\alpha,\beta} \langle \psi_1 | R | \psi_1 \rangle .$$

Consider an arbitrary orthonormal basis $\{\varphi_\alpha\}_{\alpha\in[K]} \in C \subset (\mathbb{C}^p)^{\otimes n}$ of C. Let U be a unitary matrix such that

$$\varphi_\alpha = \sum_{\beta\in[K]} U_{\alpha,\beta} \psi_\beta \qquad \text{for all } \alpha \in [K] .$$

Because $\sum_{\gamma\in[K]}(U^{\dagger})_{\alpha,\gamma}U_{\gamma,\beta} = \delta_{\alpha,\beta}$, we obtain by straightforward computation

$$\langle\varphi_\alpha|R|\varphi_\beta\rangle - \delta_{\alpha,\beta}\langle\psi_1|R|\psi_1\rangle = \sum_{\gamma,\delta\in[K]}\overline{U_{\alpha,\gamma}}U_{\beta,\delta}\,\mathrm{err}^\psi(R,\gamma,\delta)\ .$$

We conclude that

$$|\langle\varphi_\alpha|R|\varphi_\beta\rangle| \leq \sum_{\gamma,\delta\in[K]}|\mathrm{err}^\psi(R,\gamma,\delta)| \leq K\cdot\sqrt{\sum_{\gamma,\delta\in[K]}|\mathrm{err}^\psi(R,\gamma,\delta)|^2} \qquad \text{for } \alpha\neq\beta$$

because $\max_{\gamma,\delta}|\overline{U}_{\alpha,\gamma}U_{\beta,\delta}| \leq 1$ for a unitary matrix U and by using the Cauchy-Schwarz inequality. By definition of err and $\epsilon_{\mathsf{approx}}$, this implies that

$$\langle\varphi_\alpha|\mathcal{N}(|\varphi_\beta\rangle\langle\varphi_\beta|)|\varphi_\alpha\rangle \leq K^4\epsilon_{\mathsf{approx}} \qquad \text{for } \alpha\neq\beta \qquad (2.4)$$

for any orthonormal basis $\{\varphi_\alpha\}_{\alpha\in[K]}$ of C.

Let now $\delta > 0$ be given and let $\Psi \in C$ be an arbitrary state in the code space such that

$$\mathrm{tr}(P\mathcal{N}(|\Psi\rangle\langle\Psi|)) \geq \delta\ . \qquad (2.5)$$

Let us pick an orthonormal basis $\{\varphi_\alpha\}_{\alpha\in[K]} \in C \subset (\mathbb{C}^{\mathsf{p}})^{\otimes n}$ of C such that $\varphi_1 = \Psi$. Then

$$\begin{aligned}
1 - \langle\Psi|\rho_{\mathcal{N},P}|\Psi\rangle &= 1 - \frac{\langle\Psi|\mathcal{N}(|\Psi\rangle\langle\Psi|)|\Psi\rangle}{\mathrm{tr}(P\mathcal{N}(|\Psi\rangle\langle\Psi|))} \\
&= \frac{1}{\mathrm{tr}(P\mathcal{N}(|\Psi\rangle\langle\Psi|))}\cdot(\mathrm{tr}(P\mathcal{N}(|\Psi\rangle\langle\Psi|)) - \langle\Psi|\mathcal{N}(|\Psi\rangle\langle\Psi|)|\Psi\rangle) \\
&= \frac{1}{\mathrm{tr}(P\mathcal{N}(|\Psi\rangle\langle\Psi|))}\cdot\sum_{\alpha=2}^{K}\langle\varphi_\alpha|\mathcal{N}(|\varphi_1\rangle\langle\varphi_1|)|\varphi_\alpha\rangle \\
&\leq \frac{1}{\delta}\cdot K^5\epsilon_{\mathsf{approx}}
\end{aligned}$$

because of (2.3.2) and (2.3.2). The claim follows.

$\square$

If there are vectors $\{\eta_{\alpha,\beta}\}_{\alpha,\beta\in[K]}$ such that

$$\left|\langle\psi_\alpha|R_j|\psi_\beta\rangle - \delta_{\alpha,\beta}\langle\psi_1|R_j|\psi_1\rangle\right| \leq \|R_j\eta_{\alpha,\beta}\| \qquad \text{for all } j\in[J]\ , \qquad (2.6)$$

then this implies the bound

$$\epsilon_{\mathsf{approx}} \leq \max_{\alpha,\beta}\mathrm{tr}(\mathcal{N}(|\eta_{\alpha,\beta}\rangle\langle\eta_{\alpha,\beta}|)) = \max_{\alpha,\beta}\|\eta_{\alpha,\beta}\|^2\ .$$

Unfortunately, good bounds of the form (2.3.2) are not straightforward to establish in the cases considered here. Instead, we consider a slightly weaker condition (see Equation (2.3.4)) which still captures many cases of interest. In particular, it provides a simple criterion for establishing that a code can detect probabilistic Pauli errors with a certain maximum weight. Correspondingly, we introduce the following definition:

Definition 2.3.3. *An $(\epsilon, \delta)[[n, k, d]]$-AQEDC C is a p^k-dimensional subspace of $(\mathbb{C}^{\mathsf{p}})^{\otimes n}$ such that C is an (ϵ, δ)-error-detecting code for any CPTP map of the form*

$$\mathcal{N}(\rho) = \sum_{j \in [J]} p_j F_j \rho F_j^\dagger \,, \tag{2.7}$$

where each F_j is a d-local operator with $\|F\| \leq 1$ and $\{p_j\}_{j \in [J]}$ is a probability distribution.

We then have the following sufficient condition:

Corollary 2.3.4. *Let $K = \mathsf{p}^k$ and $C \subset (\mathbb{C}^{\mathsf{p}})^{\otimes n}$ be a code with orthonormal basis $\{\psi_\alpha\}_{\alpha \in [K]}$ satisfying (for some $\gamma > 0$),*

$$\left| \langle \psi_\alpha | F | \psi_\beta \rangle - \delta_{\alpha,\beta} \langle \psi_1 | F | \psi_1 \rangle \right| \leq \gamma \cdot \|F\| \qquad \text{for all } \alpha, \beta \in [K] \,, \tag{2.8}$$

for every d-local operator F on $(\mathbb{C}^{\mathsf{p}})^{\otimes n}$. Let $\delta > K^5 \gamma^2$. Then C is an $(\epsilon = K^5 \gamma^2 \delta^{-1}, \delta)[[n, k, d]]$-AQEDC.

Proof. Defining $R_j = \sqrt{p_j} F_j$, the claim follows immediately from Theorem 2.3.2. $\square$

Note that the exponents in this statement are not optimized, and could presumably be improved. We have instead opted for the presentation of a simple proof, as this ultimately provides the same qualitative statements.

We also note that the setting considered in Corollary 2.3.4, i.e., our notion of $(\epsilon, \delta)[[n, k, d]]$-error-detecting codes, goes beyond existing work on approximate error-detection/correction [51]–[53], where typically only noise channels with Kraus (error) operators acting on a fixed, contiguous (i.e., geometrically local) set of d physical spins are considered. At the same time, our results are limited to convex combinations of the form (2.3.3). It remains an open problem whether these codes also detect noise given by more general (coherent) channels.

2.3.3 Necessary Conditions for Approximate Quantum Error-Detection

Here we give a partial converse to Corollary 2.3.4, which shows that a condition of the form (2.3.4) is indeed necessary for approximate quantum error-detection.

Lemma 2.3.5. *Let $\psi_1, \psi_2 \in (\mathbb{C}^{\mathsf{p}})^{\otimes n}$ be two orthonormal states in the code space C and $F = F_{\mathcal{S}} \otimes I_{[n]\backslash \mathcal{S}} \in \mathcal{B}((\mathbb{C}^{\mathsf{p}})^{\otimes d})$ an orthogonal projection acting on d sites $\mathcal{S} \subset [n]$ such that*

$$|\langle \psi_1 | F | \psi_1 \rangle - \langle \psi_2 | F | \psi_2 \rangle| = \eta$$

for some $\eta \in [0, 1]$, with $1 - \eta \ll 1$. Then any subspace $C \subset (\mathbb{C}^{\mathsf{p}})^{\otimes n}$ of dimension p^k is not an $(\epsilon, \delta)[[n, k, d]]$-code for

$$\epsilon < 1 - 10(1 - \eta), \quad \text{and}$$
$$\delta < \eta^2 .$$

Proof. Let

$$F_{j,k} := \langle \psi_j | F | \psi_k \rangle \qquad \text{for } j, k \in \{1, 2\} .$$

By choosing the phase of $|\psi_1\rangle$ appropriately, we may assume that $F_{1,2} \geq 0$. Note that $F_{1,2} = F_{2,1} \leq \|F\psi_2\| = \sqrt{\langle \psi_2 | F | \psi_2 \rangle}$ by the Cauchy-Schwarz inequality and because F is a projection. Let us denote the entries of F by

$$F = \begin{pmatrix} p & r \\ r & q \end{pmatrix}$$

where $q \in [0, 1 - \eta]$, $p = q + \eta$, and $r \in [0, \sqrt{q}]$. Let us define a CPTP map $\mathcal{N}$ of the form (2.3.3) by

$$\mathcal{N}(\rho) = e^{i\pi F} \rho e^{-i\pi F} \qquad \text{where} \qquad F = F_{\mathcal{S}} \otimes I_{[n]\backslash \mathcal{S}} .$$

Let $\hat{P} = \sum_{j=1}^{2} |\psi_j\rangle\langle\psi_j|$. Consider the normalized vector $|\Psi\rangle = \frac{1}{\sqrt{2}}(|\psi_1\rangle + |\psi_2\rangle)$. Then

$$\hat{P}\mathcal{N}(|\Psi\rangle\langle\Psi|)\hat{P} = \frac{1}{2} \sum_{i,j,k,\ell} W_{k,i} \overline{W_{\ell,j}} |\psi_k\rangle\langle\psi_\ell| , \tag{2.9}$$

where

$$W_{j,k} := \langle \psi_j | e^{i\pi F} | \psi_k \rangle \qquad \text{for } j, k \in \{1, 2\} .$$

Observe that since $F^2 = F$ is a projection, we have $e^{i\pi F} = I - 2F$, thus the entries of W are

$$W_{j,k} = \delta_{j,k} - 2F_{j,k} \qquad \text{for } j, k \in \{1, 2\} .$$

In particular, from (2.3.3) we obtain for the projection P onto C

$$
\begin{aligned}
\operatorname{tr}(P\mathcal{N}(|\Psi\rangle\langle\Psi|)P) &\geq \operatorname{tr}\left(\hat{P}\mathcal{N}(|\Psi\rangle\langle\Psi|)\hat{P}\right) \\
&= \frac{1}{2} \sum_{i,j,k} W_{k,i}\overline{W_{k,j}} \\
&= 1 - 2p + 2p^2 - 2q + 2q^2 + 4r(p + q - 1 + r) \\
&\geq (p - q)^2 = \eta^2 ,
\end{aligned}
\tag{2.10}
$$

where we used that the last expression is minimal (and equal to $(p - q)^2$) for $r = 1/2(1 - p - q)$. We also have

$$
\begin{aligned}
\langle\Psi|\mathcal{N}(|\Psi\rangle\langle\Psi|)|\Psi\rangle &= \frac{1}{4} \sum_{i,j,k,\ell} W_{k,i}\overline{W_{\ell,j}} \\
&= (2r + p + q - 1)^2 \\
&= (2(r + q) - (1 - \eta))^2 .
\end{aligned}
$$

This expression is maximal for (r, q) each maximal (since both are non-negative), hence for $(r, q) = (\sqrt{1 - \eta}, 1 - \eta)$, and we obtain the upper bound

$$\langle\Psi|\mathcal{N}(|\Psi\rangle\langle\Psi|)|\Psi\rangle \leq (1 - \eta + 2\sqrt{1 - \eta})^2 \leq 9(1 - \eta) ,$$

where we used that $x \leq \sqrt{x}$ for $x \in [0, 1]$. This implies with (2.3.3) that for $\rho_{\mathcal{N},P} = \operatorname{tr}(P\mathcal{N}(|\Psi\rangle\langle\Psi|))^{-1} \cdot P\mathcal{N}(|\Psi\rangle\langle\Psi|)P$, we have

$$\langle\Psi|\rho_{\mathcal{N},P}|\Psi\rangle \leq \frac{9(1 - \eta)}{\eta^2} = \frac{9(1 - \eta)}{(1 - (1 - \eta))^2} \leq 10(1 - \eta)$$

for $1 - \eta \ll 1$. Thus

$$1 - \langle\Psi|\rho_{\mathcal{N},P}|\Psi\rangle \geq 1 - 10(1 - \eta) .$$

With (2.3.3), this implies the claim.

$\square$

We reformulate Lemma 2.3.5, by stating it in terms of reduced density matrices, as follows:

Lemma 2.3.6. *Let $\psi_1, \psi_2 \in (\mathbb{C}^p)^{\otimes n}$ be two orthonormal vectors in a subspace $C \subset (\mathbb{C}^p)^{\otimes n}$ of dimension p^k. Fix a region $R \subset [n]$ of size $|R| = d$ and let $\rho_j = \mathrm{tr}_{[n]\backslash R} |\psi_j\rangle\langle\psi_j|$, $j = 1, 2$ be the reduced density matrices on R. Then C is not a $(\epsilon, \delta)[[n, k, d]]$-error-detecting code for*

$$\epsilon < 1 - 10\zeta(\rho_1, \rho_2), \qquad and$$

$$\delta < (1 - \zeta(\rho_1, \rho_2))^2,$$

where $\zeta(\rho_1, \rho_2) := \max\{\mathrm{rank}\,\rho_1, \mathrm{rank}\,\rho_2\}^2 \cdot \mathrm{tr}(\rho_1\rho_2)$.

Proof. By definition of the trace distance, the projection F onto the positive part of $\rho_1 - \rho_2$ satisfies

$$\eta := \frac{1}{2}\|\rho_1 - \rho_2\|_1 = \mathrm{tr}(F(\rho_1 - \rho_2)).$$

With the inequality $\|A\|_1 \leq \sqrt{\mathrm{rank}(A)}\|A\|_F$, we get the bound

$$F(\rho_1, \rho_2) = \|\sqrt{\rho_1}\sqrt{\rho_2}\|_1^2 \leq D^2\|\sqrt{\rho_1}\sqrt{\rho_2}\|_F^2 = D^2\,\mathrm{tr}(\rho_1\rho_2)$$

on the fidelity of ρ_1 and ρ_2, where $D = \mathrm{rank}(\sqrt{\rho_1}\sqrt{\rho_2}) \leq \max\{\mathrm{rank}\,\rho_1, \mathrm{rank}\,\rho_2\}$. Inserting this into the inequality $\frac{1}{2}\|\rho_1 - \rho_2\|_1 \geq 1 - F(\rho_1, \rho_2)$ yields

$$\eta \geq 1 - \max\{\mathrm{rank}\,\rho_1, \mathrm{rank}\,\rho_2\}^2 \cdot \mathrm{tr}(\rho_1\rho_2).$$

The claim then follows from Lemma 2.3.5 and the fact that if C is not an $(\epsilon, \delta)[[n, k, d]]$-code, then it is not an $(\epsilon', \delta')[[n, k, d]]$-code for any $\delta' \leq \delta$ and $\epsilon' \leq \epsilon$. $\qquad\square$

We will use Lemma 2.3.6 below to establish our no-go result for codes based on injective MPS with open boundary conditions.

2.4 On Expectation Values of Local Operators in MPS

Key to our analysis are expectation values of local observables in MPS, and more generally, matrix elements of local operators with respect to different MPS. These directly determine whether or not the considered subspace satisfies the approximate quantum error-detection conditions. To study these quantities, we first review the terminology of transfer operators (and, in particular, injective MPS) in Section 2.4.1. In Section 2.4.2, we establish bounds on the matrix elements and the norms of transfer operators. These will subsequently be applied in all our derivations.
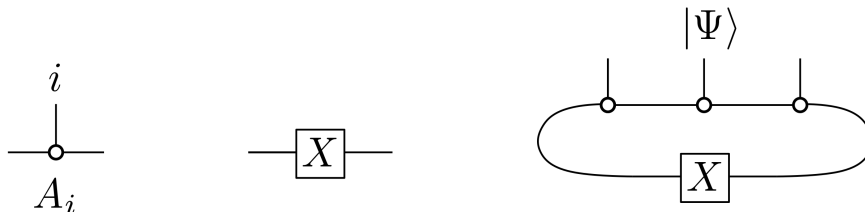
Figure 2.1: This figure illustrates an MPS with $n = 3$ physical spins, defined in terms of a family $\{A_j\}_{j=1}^{\mathsf{p}}$ of matrices and a matrix X.

2.4.1 Review of Matrix Product States

A matrix product state (or MPS) of bond dimension D is a state $|\Psi\rangle$ on $(\mathbb{C}^{\mathsf{p}})^{\otimes n}$ which is parametrized by a collection of $D \times D$ matrices. In this paper, we focus on uniform, site-independent MPS. Such a state is fully specified by a family $A = \{A_j\}_{j=1}^{\mathsf{p}}$ of $D \times D$ matrices describing the "bulk properties" of the state, together with a "boundary condition" matrix $X \in \mathcal{B}(\mathbb{C}^D)$. We write $|\Psi\rangle = |\Psi(A, X, n)\rangle$ for such a state, where we often suppress the defining parameters (A, X, n) for brevity.

Written in the standard computational basis, the state $|\Psi(A, X, n)\rangle$ is expressed as

$$|\Psi\rangle = \sum_{i_1,\ldots,i_n \in [\mathsf{p}]} \mathrm{tr}\left(A_{i_1} \cdots A_{i_n} X\right) |i_1 \cdots i_n\rangle \tag{2.11}$$

for a family $\{A_j\}_{j=1}^{\mathsf{p}} \subset \mathcal{B}(\mathbb{C}^D)$ of matrices. The number of sites $n \in \mathbb{N}$ is called the system size, and each site is of local dimension $\mathsf{p} \in \mathbb{N}$, which is called the physical dimension of the system. The parameter $D \in \mathbb{N}$ is called the bond, or virtual, dimension. This state can be represented graphically as a tensor network as in Figure 2.1.

Note that the family of matrices $A = \{A_j\}_{j=1}^{\mathsf{p}}$ of a site-independent MPS equivalently defines a three-index tensor $(A_j)_{\alpha\beta}$ with one "physical" (j) and two "virtual" (α, β) indices. We call this the local MPS tensor associated to $|\Psi(A, X, n)\rangle$.

The matrices $\{A_j\}_{j=1}^{\mathsf{p}}$ defining a site-independent MPS $|\Psi(A, X, n)\rangle$ give rise to a completely positive (CP) linear map $\mathcal{E} : \mathcal{B}(\mathbb{C}^D) \to \mathcal{B}(\mathbb{C}^D)$ which acts on $Y \in \mathcal{B}(\mathbb{C}^D)$ by

$$\mathcal{E}(Y) = \sum_{i=1}^{\mathsf{p}} A_i Y A_i^{\dagger} . \tag{2.12}$$

Without loss of generality (by suitably normalizing the matrices $\{A_j\}_{j=1}^{\mathsf{p}}$), we assume that $\mathcal{E}$ has spectral radius 1. This implies that $\mathcal{E}$ has a positive semi-definite fixed point $r \in \mathcal{B}(\mathbb{C}^D)$ by the Perron-Frobenius Theorem, see [72, Theorem 2.5]. We

say that the MPS $|\Psi(A, X, n)\rangle$ is *injective*[2] if the associated map $\mathcal{E}$ is *primitive*, i.e., if the fixed-point r is positive definite (and not just positive semi-definite), and if the eigenvalue 1 associated to r is the only eigenvalue on the unit circle, including multiplicity [74, Theorem 6.7].

From expression (2.4.1), we can see that there is a gauge freedom of the form

$$\tilde{A}_j = P^{-1}A_jP, \qquad \tilde{X} = P^{-1}XP, \quad \text{for } j = 1, \ldots, \mathsf{p}, \tag{2.13}$$

for every invertible matrix $P \in GL(\mathbb{C}^D)$, for which $|\Psi(A, X, n)\rangle = |\Psi(\tilde{A}, \tilde{X}, n)\rangle$. Given an injective MPS, the defining tensors can be brought into a *canonical* form by exploiting this gauge freedom in the definition of the MPS.[3]

One proceeds as follows: given an injective MPS, let r denote the unique fixed-point of the transfer operator $\mathcal{E}$. We can apply the gauge freedom (2.4.1) with $P = \sqrt{r}$ to obtain an equivalent MPS description by matrices $\{\tilde{A}_j := r^{-1/2}A_jr^{1/2}\}_{j=1}^{\mathsf{p}}$, where the associated map $\tilde{\mathcal{E}}$ is again primitive with spectral radius 1, but now with the identity operator $\tilde{r} = I_{\mathbb{C}^D}$ as the unique fixed-point.

Similarly, one can show that the adjoint

$$\mathcal{E}^{\dagger}(Y) = \sum_{i=1}^{\mathsf{p}} A_i^{\dagger}YA_i$$

of a primitive map $\mathcal{E}$ is also primitive.[4] Since the spectrum of $\mathcal{E}^{\dagger}$ is given by $\mathsf{spec}(\mathcal{E}^{\dagger}) = \overline{\mathsf{spec}(\mathcal{E})}$, this implies that the map $\mathcal{E}^{\dagger}$ has a unique positive fixed-point ℓ with eigenvalue 1, with all other eigenvalues having magnitude less than 1.

Now, let $\tilde{\ell}$ denote the unique fixed-point of the previously defined $\tilde{\mathcal{E}}$. Since $\tilde{\ell}$ is positive definite, it is unitarily diagonalizable:

$$\tilde{\ell} = U\Lambda U^{\dagger},$$

---

[2]Injective MPS are known to be "generic." More precisely, consider the space $\mathbb{C}^D \otimes \mathbb{C}^D \otimes \mathbb{C}^{\mathsf{p}}$ of all defining tensors with physical dimension p and bond dimension D. Then the set of defining tensors with a primitive transfer operator forms an open, co-measure zero set. The definition of injective that we use here differs from the one commonly used in the literature (cf. [28]), but is ultimately equivalent. For a proof of equivalence, see Definition 8, Lemma 6, and Theorem 18 of [73].

[3]The canonical form holds for non-injective MPS as well, see [28]. We only consider the injective case here.

[4]Note that the adjoint is taken with respect to the Hilbert-Schmidt inner product on $\mathcal{B}(\mathbb{C}^D)$. One way to see that the adjoint of a primitive map is primitive is to note that an equivalent definition for primitivity given in [74, Theorem 6.7(2)] is in terms of irreducible maps. A map is irreducible if and only if its adjoint is irreducible (see the remarks in [74] after Theorem 6.2). This in turn means that a map is primitive if and only if its adjoint is primitive.

with U being a unitary matrix, and Λ being a diagonal matrix with all diagonal entries positive. Using the gauge freedom (2.4.1) in the form

$$\tilde{A}_j \mapsto \tilde{\tilde{A}}_j = U^\dagger \tilde{A}_j U \qquad \text{for } j = 1, \ldots, \mathsf{p},$$

we obtain an equivalent MPS description such that the associated channel $\tilde{\tilde{\mathcal{E}}}^\dagger$ has a fixed-point given by a positive definite diagonal matrix Λ. We may without loss of generality take Λ to be normalized as $\mathrm{tr}(\Lambda) = 1$. It is also easy to check that the identity operator $\mathbb{I}_D$ remains the unique fixed-point of $\tilde{\tilde{\mathcal{E}}}$.

In summary, given an injective MPS with associated map $\mathcal{E}$, one may, by using the gauge freedom, assume without loss of generality that:

(i) The unique fixed-point r of $\mathcal{E}$ is equal to the identity, i.e., $r = I_{\mathbb{C}^D}$.

(ii) The unique fixed-point ℓ of $\mathcal{E}^\dagger$ is given by a positive definite diagonal matrix $\ell = \Lambda$, normalized so that $\mathrm{tr}(\Lambda) = 1$.

An MPS with defining tensors A satisfying these two properties above is said to be in *canonical* form.

In the following, after fixing a standard orthonormal basis $\{|\alpha\rangle\}_{\alpha=1}^D$ of $\mathbb{C}^D$, we identify elements $X \in \mathcal{B}(\mathbb{C}^D)$ with vectors $|X\rangle\!\rangle \in \mathbb{C}^D \otimes \mathbb{C}^D$ via the vectorization isomorphism

$$X \mapsto |X\rangle\!\rangle := (X^T \otimes I) \sum_{\alpha=1}^D |\alpha\rangle \otimes |\alpha\rangle = \sum_{\alpha,\beta=1}^D X_{\alpha,\beta}|\beta\rangle \otimes |\alpha\rangle,$$

where $X = \sum_{\alpha,\beta=1}^D X_{\alpha,\beta}|\alpha\rangle\langle\beta|$. It is easy to verify that $\langle\!\langle X|Y\rangle\!\rangle = \mathrm{tr}(X^\dagger Y)$, i.e., the standard inner product on $\mathbb{C}^D \otimes \mathbb{C}^D$ directly corresponds to the Hilbert-Schmidt inner product of operators in $\mathcal{B}(\mathbb{C}^D)$ under this identification. Furthermore, under this isomorphism, a super-operator $\mathcal{E} : \mathcal{B}(\mathbb{C}^D) \to \mathcal{B}(\mathbb{C}^D)$ becomes a linear map $E : \mathbb{C}^D \otimes \mathbb{C}^D \to \mathbb{C}^D \otimes \mathbb{C}^D$ defined by

$$|\mathcal{E}(X)\rangle\!\rangle = E|X\rangle\!\rangle$$

for all $X \in \mathcal{B}(\mathbb{C}^D)$. The matrix E is simply the matrix representation of $\mathcal{E}$, thus E has the same spectrum as $\mathcal{E}$. Explicitly, for a map of the form (2.4.1), it is given by

$$E = \sum_{i=1}^{\mathsf{p}} \overline{A_i} \otimes A_i . \tag{2.14}$$

The fixed-point equations for a fixed-point r of $\mathcal{E}$ and a fixed-point ℓ of $\mathcal{E}^\dagger$ become

$$\langle\!\langle \ell|E = \langle\!\langle \ell| \qquad \text{and} \qquad E|r\rangle\!\rangle = |r\rangle\!\rangle \,, \tag{2.15}$$

i.e., the corresponding vectors are left and right eigenvectors of E, respectively.

For a site-independent MPS $|\Psi(A, X, n)\rangle$, defined by matrices $\{A_j\}_{j=1}^{\mathsf{p}}$, we call the associated matrix E (cf. (2.4.1)) the *transfer matrix*. Many key properties of a site-independent MPS are captured by its transfer matrix. For example, the normalization of the state is given by

$$\|\Psi\|^2 = \langle\Psi|\Psi\rangle = \text{tr}(E^n(\overline{X} \otimes X)) \,.$$

If the MPS is injective, then, according to (i)–(ii), it has a Jordan decomposition of the form

$$E = |I\rangle\!\rangle\langle\!\langle\Lambda| \oplus \tilde{E}.$$

In this expression, $|I\rangle\!\rangle\langle\!\langle\Lambda|$ is the (1-dimensional) Jordan block corresponding to eigenvalue 1, whereas $\tilde{E}$ is a direct sum of Jordan blocks with eigenvalues of modulus less than 1. The second largest eigenvalue λ_2 of E has a direct interpretation in terms of the correlation length ξ of the state, which determines two-point correlators $|\langle\sigma_j\sigma_{j'}\rangle - \langle\sigma_j\rangle \cdot \langle\sigma_{j'}\rangle| \sim e^{-|j-j'|/\xi}$. The latter is given by $\xi = \log(1/\lambda_2)$.

For an injective MPS, the fact that $|I\rangle\!\rangle$ is the unique right-eigenvector of E with eigenvalue 1 implies the normalization condition

$$\langle\!\langle\Lambda|I\rangle\!\rangle = \text{tr}(\Lambda) = 1 \,. \tag{2.16}$$

We will represent these identities diagrammatically, which is convenient for later reference. The matrix Λ will be shown by a square box, the identity matrix corresponds to a straight line. That is, the normalization condition (2.4.1) takes the form

and the left and right eigenvalue equations (2.4.1)
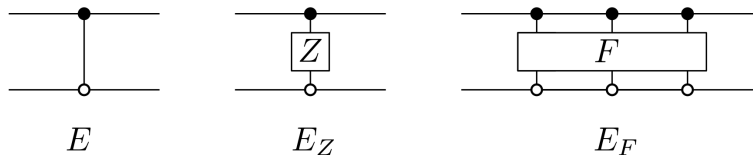
Figure 2.2: The transfer operator E, as well as E_Z for $Z \in \mathcal{B}(\mathbb{C}^\mathsf{p})$, and E_F for $F \in (\mathbb{C}^\mathsf{p})^{\otimes 3}$.

2.4.2 Transfer Matrix Techniques

Here we establish some essential statements for the analysis of transfer operators. In Section 2.4.2, we introduce generalized (non-standard) transfer operators: these can be used to express the matrix elements of the form $\langle \Psi | F | \Psi' \rangle$ of local operators F with respect to pairs of MPS (Ψ, Ψ'). In Section 2.4.2, we establish bounds on the norm of such operators. Relevant quantities appearing in these bounds are the second largest eigenvalue λ_2 of the transfer matrix, as well as the sizes of its Jordan blocks.

More General and Mixed Transfer Operators

Consider a single-site operator $Z \in \mathcal{B}(\mathbb{C}^\mathsf{p})$. The generalized transfer matrix $E_Z \in \mathcal{B}(\mathbb{C}^D \otimes \mathbb{C}^D)$ is defined as

$$E_Z = \sum_{n,m} \langle m | Z | n \rangle \overline{A_m} \otimes A_n \ . \tag{2.17}$$

We further generalize this as follows: if $Z_1, \ldots, Z_d \in \mathcal{B}(\mathbb{C}^\mathsf{p})$, then $E_{Z_1 \otimes \cdots \otimes Z_d} \in \mathcal{B}(\mathbb{C}^D \otimes \mathbb{C}^D)$ is the operator

$$E_{Z_1 \otimes \cdots \otimes Z_d} = E_{Z_1} \cdots E_{Z_d} \ .$$

This definition extends by linearity to any operator $F \in \mathcal{B}((\mathbb{C}^\mathsf{p})^{\otimes d})$, and gives a corresponding operator $E_F \in \mathcal{B}(\mathbb{C}^D \otimes \mathbb{C}^D)$. The tensor network diagrams for these definitions are given in Figure 2.2, and the composition of the corresponding maps is illustrated in Figure 2.3.

In the following, we are interested in inner products $\langle \Psi(A, X, n) | \Psi(B, Y, n) \rangle$ of two MPS, defined by local tensors A and B, with boundary matrices X and Y, which may have different bond dimensions D_1 and D_2, respectively. To analyze these, it is convenient to introduce an "overlap" transfer operator $E = E(A, B)$ which now depends on both MPS tensors A and B. First, we define $E \in \mathcal{B}(\mathbb{C}^{D_1} \otimes \mathbb{C}^{D_2})$ by

$$E = \sum_{m=1}^\mathsf{p} \overline{A_m} \otimes B_m \ .$$
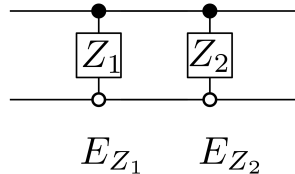
$$E_{Z_1} \qquad E_{Z_2}$$

Figure 2.3: The product $E_{Z_1 \otimes Z_2} = E_{Z_1} E_{Z_2}$ of two transfer operators. Left-multiplication by an operator corresponds to attaching the corresponding diagram on the left.

The definition of E_Z for $Z \in \mathcal{B}(\mathbb{C}^{\mathsf{p}})$ is analogous to Equation (2.4.2), but with appropriate substitutions. We set

$$E_Z = \sum_{n,m} \langle m|Z|n\rangle \overline{A_m} \otimes B_n \ .$$

Starting from this definition, the expression $E_F \in \mathcal{B}(\mathbb{C}^{D_1} \otimes \mathbb{C}^{D_2})$ for $F \in \mathcal{B}((\mathbb{C}^{\mathsf{p}})^{\otimes d})$ is then defined analogously as before.

Norm Bounds on Generalized Transfer Operators

A first key observation is that the (operator) norm of powers of any transfer operator scales (at most) as a polynomial in the number of physical spins, with the degree of the polynomial determined by the size of the largest Jordan block. We need these bounds explicitly and start with the following simple bounds.

Below, we often consider families of parameters depending on the system size n, i.e., the total number of spins. We write $m \gg h$ as a shorthand for a parameter m "being sufficiently large" compared to another parameter h. More precisely, this signifies that we assume that $|h/m| \to 0$ for $n \to \infty$, and that by a corresponding choice of a sufficiently large n, the term $|h/m|$ can be made sufficiently small for a given bound to hold. Oftentimes h will in fact be constant, with $m \to \infty$ as $n \to \infty$.

Lemma 2.4.1. *For $m > h$, the Frobenius norm of the m-th power $(\lambda I + N)^m$ of a Jordan block $\lambda I + N \in \mathcal{B}(\mathbb{C}^h)$ with eigenvalue λ, such that $|\lambda| \leq 1$, and size h is bounded by*

$$\|(\lambda I + N)^m\|_F \leq 3h^{3/2} m^{h-1} |\lambda|^{m-(h-1)} \ . \tag{2.18}$$

Furthermore,

$$\|(\lambda I + N)^m\| \leq 4m^{h-1} \qquad \textit{for } m \gg h \ . \tag{2.19}$$

Proof. For $h = 0$ the claim is trivial. Assume that $h > 1$. Because $N^h = 0$ and N^r has exactly $h - r$ non-zero entries for $r < h$, we have

$$\|(\lambda I_h + N)^m\|_F \le \sum_{r=0}^{h-1} \binom{m}{r} |\lambda|^{m-r} \|N^r\|_F$$

$$\le |\lambda|^m \cdot |\lambda|^{-(h-1)} \sum_{r=0}^{h-1} \binom{m}{r} (h - r)^{1/2}$$

$$\le h^{1/2} |\lambda|^m \cdot |\lambda|^{-(h-1)} \sum_{r=0}^{h-1} \binom{m}{r} .$$

Since the right hand side is maximal for $r = h - 1$, and the binomial coefficient can be bounded from above by $\binom{m}{r} \le \left(\frac{em}{r}\right)^r \le 3 \cdot m^r$, we obtain

$$\sum_{r=0}^{h-1} \binom{m}{r} \le 3h \cdot m^{h-1} ,$$

hence, the first claim follows.

For the second claim, recall that the entries of the m-th power of a Jordan block are

$$((\lambda I_h + N)^m)_{p,q} = \begin{cases} \binom{m}{q-p} \lambda^{m+(p-q)} & \text{if } q \ge p \\ 0 & \text{otherwise} \end{cases} \qquad (2.20)$$

for $p, q \in [h]$. This means that if $|\lambda| = 1$, the maximum matrix element $| ((\lambda I_h + N)^m)_{p,q} | = \binom{m}{h-1}$ is attained for $(p, q) = (1, h)$. Using the Cauchy-Schwarz inequality, it is straightforward to check that

$$\| (\lambda I + N)^m \| \le h \max_{p,q} | (\lambda I + N)^m_{p,q} | = h \cdot \binom{m}{h-1} = \frac{h}{(h-1)!} \frac{m!}{(m - (h-1))!} .$$

Since $\frac{h}{(h-1)!} \le 2$ for $h \in \mathbb{N}$ and

$$\frac{m!}{(m - (h-1))!} = m^{h-1} (1 + O(h/m)) \le 2m^{h-1} \qquad \text{for } m \gg h ,$$

the claim follows. $\qquad\square$

Now, we apply Lemma 2.4.1 to (standard and mixed) transfer operators. It is convenient to state these bounds as follows. The first two statements are about the scaling of the norms of powers of E; the last statement is about the magnitude of matrix elements in powers of E.

Lemma 2.4.2. *Let $\rho(E)$ denote the spectral radius of a matrix $E \in \mathcal{B}(\mathbb{C}^{D_1} \otimes \mathbb{C}^{D_2})$.*

(i) Suppose $\rho(E) \leq 1$. Let h^* be the size of the largest Jordan block(s) of E. Then

$$\|E^m\| \leq 4m^{h-1} \qquad \text{for } m \gg h .$$

(ii) If $\rho(E) < 1$, then

$$\|E^m\|_F \leq \rho(E)^{m/2} \qquad \text{for } m \gg D_1 D_2 .$$

We will often use $\|E^m\|_F \leq 1$ as a coarse bound.

(iii) Suppose that $\rho(E) = 1$. Let h^* denote the size of the largest Jordan block(s) in E. For $p, q \in [D_1 D_2]$, let $(E^m)_{p,q}$ denote the matrix element of E^m with respect to the standard computational basis $\{|p\rangle\}_{p=1}^{D_1 D_2}$. Then the following holds: for all $p, q \in [D_1 D_2]$, there is a constant $c_{p,q}$ with $c_{p,q} = O(1)$ as $m \to \infty$ and some $\ell \in \{1, \ldots, h^*\}$ such that

$$|(E^m)_{p,q}| = cm^{\ell-1}(1 + O(m^{-1})) .$$

Proof. For $\lambda \in \operatorname{spec}(E)$, let us denote by $\lambda I_{h(\lambda)} + N_{h(\lambda)}$ the associated Jordan block, where $h(\lambda)$ is its size. Then

$$\|E^m\| = \max_{\lambda \in \operatorname{spec}(E)} \|(\lambda I_{h(\lambda)} + N_{h(\lambda)})^m\| \leq \max_{\lambda \in \operatorname{spec}(E)} \|(\lambda I_{h(\lambda)} + N_{h(\lambda)})^m\|_F$$

where we assumed that $m \gg h^* \geq h(\lambda)$, $|\lambda| \leq 1$ and (2.4.1). This shows claim (i).

Claim (ii) immediately follows from (2.4.1) and the observation that $m^{D_1 D_2 - 1} \rho(E)^m = O(\rho(E)^{m/2})$.

For the proof of statement (iii), observe that matrix elements (2.4.2) of a Jordan block (matrix) with eigenvalue λ (with $|\lambda| = 1$) of size h scale as

$$\left|((\lambda I + N)^m)_{p,q}\right| = \frac{1}{(q-p)!} \frac{m!}{(m-(q-p))!} = \frac{1}{(q-p)!} m^{q-p} (1 + O(1/m))$$

for $q > p$ and are constant otherwise. Because $q - p \in \{1, \ldots, h-1\}$ when $q > p$, this is of the form $m^\ell (1 + O(1/m))$ for some $\ell \in [h-1]$. Since E^m is similar (as a matrix) to a direct sum of such powers of Jordan blocks, and the form of this scaling does not change under linear combination of matrix coefficients, the claim follows. $\square$

Now let us consider the case where $E = |\ell\rangle\!\rangle\langle\!\langle r| \oplus \tilde{E}$ is the transfer operator of an injective MPS, normalized with maximum eigenvalue 1. Let $\lambda_2 < 1$ denote the second largest eigenvalue. Without loss of generality, we can take the MPS to be in canonical form, so that E has a unique right fixed-point given by the identity matrix I and a unique left fixed-point given by some positive-definite diagonal matrix Λ with unit trace. We can then write the Jordan decomposition of the transfer matrix as

$$E = |I\rangle\!\rangle\langle\!\langle\Lambda| \oplus \tilde{E} \,, \tag{2.21}$$

where $|I\rangle\!\rangle$ and $|\Lambda\rangle\!\rangle$ denotes the vectorization of I and Λ, respectively, and where $\tilde{E}$ denotes the remaining Jordan blocks of E. Note that powers of E can then be expressed as

$$E^m = |I\rangle\!\rangle\langle\!\langle\Lambda| \oplus \tilde{E}^m \,.$$

We can bound the Frobenius norm of the transfer matrix as

$$\|E^m\|_F^2 = \||I\rangle\!\rangle\langle\!\langle\Lambda| \oplus \tilde{E}^m\|_F^2 = \mathrm{tr}(I)\,\mathrm{tr}(\Lambda^2) + \|\tilde{E}^m\|_F^2 \leq D + \|\tilde{E}^m\|_F^2,$$

where $\mathrm{tr}(I) = D$ and $\mathrm{tr}(\Lambda^2) \leq \mathrm{tr}(\Lambda)^2 = 1$. In particular, since $\rho(\tilde{E}) = \lambda_2$, we obtain from Lemma 2.4.2(ii) that

$$\|\tilde{E}^m\|_F \leq \lambda_2^{m/2} \quad \text{for } m \gg D \,. \tag{2.22}$$

This implies the following statement:

Lemma 2.4.3. *The transfer operator E of an injective MPS satisfies*

$$\|E^m\|_F \leq \sqrt{D+1} \quad \text{for } m \gg D. \tag{2.23}$$

We also need a bound on the norm $\|E_F^\dagger(\psi_1 \otimes \psi_2)\|$, where E is a mixed transfer operator, $F \in \mathcal{B}((\mathbb{C}^p)^{\otimes d})$ is an operator acting on d sites, and where $\psi_j \in \mathbb{C}^{D_j}$ for $j = 1, 2$.

Lemma 2.4.4. *Let $E_1 = E(A)$ and $E_2 = E(B)$ be the transfer operators associated with the tensors A and B, respectively, with bond dimensions D_1 and D_2. Let $E = E(A, B) \in \mathcal{B}(\mathbb{C}^{D_1} \otimes \mathbb{C}^{D_2})$ be the combined transfer operator. Let $\psi_1 \in \mathbb{C}^{D_1}$ and $\psi_2 \in \mathbb{C}^{D_2}$ be unit vectors. Then*

$$\|(E_F)^\dagger(\psi_1 \otimes \psi_2)\| \leq \|F\|\sqrt{\|E_1^d\| \cdot \|E_2^d\|} \,, \tag{2.24}$$

$$\|E_F(\psi_1 \otimes \psi_2)\| \leq \|F\|\sqrt{\|E_1^d\| \cdot \|E_2^d\|} \,, \tag{2.25}$$

$$\|E_F\|_F \leq D_1 D_2 \|F\|\sqrt{\|E_1^d\| \cdot \|E_2^d\|} \,, \tag{2.26}$$

for all $F \in \mathcal{B}((\mathbb{C}^p)^{\otimes d})$.

Proof. Writing matrix elements in the computational basis as

$$F_{j_1 \cdots j_d, i_1 \cdots i_d} := \langle j_1 \cdots j_d | F | i_1 \cdots i_d \rangle,$$

we have

$$E_F = \sum_{(i_1,\ldots,i_d),(j_1,\ldots,j_d)} F_{j_1 \cdots j_d, i_1 \cdots i_d} (\overline{A}_{j_1} \otimes B_{j_1})(\overline{A}_{j_2} \otimes B_{j_2}) \cdots (\overline{A}_{j_d} \otimes B_{j_d}).$$

Therefore,

$$(E_F)^\dagger = \sum_{(i_1,\ldots,i_d),(j_1,\ldots,j_d)} \overline{F_{j_1 \cdots j_d, i_1 \cdots i_d}} (\overline{A^\dagger}_{j_d} \otimes B^\dagger_{j_d}) \cdots (\overline{A^\dagger}_{j_2} \otimes B^\dagger_{j_2})(\overline{A^\dagger}_{j_1} \otimes B^\dagger_{j_1})$$

$$= \sum_{(i_1,\ldots,i_d),(j_1,\ldots,j_d)} (\pi \overline{F} \pi^\dagger)_{j_d \cdots j_1, i_d \cdots i_1} (\overline{A^\dagger}_{j_d} \otimes B^\dagger_{j_d}) \cdots (\overline{A^\dagger}_{j_2} \otimes B^\dagger_{j_2})(\overline{A^\dagger}_{j_1} \otimes B^\dagger_{j_1}),$$

where π is the permutation which maps the j-th factor in the tensor product $(\mathbb{C}^p)^{\otimes n}$ to the $(n - j + 1)$-th factor, and where $\overline{F}$ is obtained by complex conjugating the matrix elements in the computational basis. This means that

$$(E_F)^\dagger = E^\dagger_{\pi \overline{F} \pi^\dagger}, \tag{2.27}$$

with $E^\dagger$ being the mixed transfer operator $E^\dagger = E(A^\dagger, B^\dagger)$ obtained by replacing each A_j (respectively B_j) with its adjoint.

Now consider

$$\|(E_F)^\dagger(\psi_1 \otimes \psi_2)\|^2 = (\langle \psi_1 | \otimes \langle \psi_2 |) E_F (E_F)^\dagger (|\psi_1\rangle \otimes |\psi_2\rangle)$$

$$= (\langle \psi_1 | \otimes \langle \psi_2 |) E_F E^\dagger_{\pi \overline{F} \pi^\dagger} (|\psi_1\rangle \otimes |\psi_2\rangle).$$

This can be represented diagrammatically as

In particular, we have

$$\|(E_F)^\dagger(\psi_1 \otimes \psi_2)\|^2 = \langle \chi | (F \otimes I^{\otimes d})(I^{\otimes d} \otimes \pi \overline{F} \pi^\dagger) | \varphi \rangle, \tag{2.28}$$

where $\varphi, \chi \in (\mathbb{C}^{\mathsf{p}})^{\otimes 2d}$ are defined as

$$|\phi\rangle = \quad \langle\psi_1| \quad \underbrace{}_{A \ A} \cdots \underbrace{}_{A^\dagger \ A^\dagger} \quad |\psi_1\rangle \quad ,$$

$$|\chi\rangle = \quad \langle\psi_2| \quad \underbrace{}_{B \ B} \cdots \underbrace{}_{B^\dagger \ B^\dagger} \quad |\psi_2\rangle \quad .$$

It is straightforward to check that

$$\|\chi\|^2 = (\langle\psi_1| \otimes \langle\psi_1|)E_1^d(E_1^\dagger)^d(|\psi_1\rangle \otimes |\psi_1\rangle) \, ,$$
$$\|\varphi\|^2 = (\langle\psi_2| \otimes \langle\psi_2|)E_2^d(E_2^\dagger)^d(|\psi_2\rangle \otimes |\psi_2\rangle) \, .$$

Since $\|(E_j^\dagger)^d\| = \|E_j^d\|$ for $j = 1, 2$, it follows with the submultiplicativity of the operator norm that

$$\|\chi\|^2 \le \|E_1^d\|^2 \, ,$$
$$\|\varphi\|^2 \le \|E_2^d\|^2 \, . \tag{2.29}$$

Applying the Cauchy-Schwarz inequality to (2.4.2) yields

$$\|(E_F)^\dagger(\psi_1 \otimes \psi_2)\|^2 \le \|(F^\dagger \otimes I^{\otimes d})\chi\| \cdot \|(I^{\otimes d} \otimes \pi\overline{F}\pi^\dagger)\varphi\|$$
$$\le \|F\|^2 \cdot \|\chi\| \cdot \|\varphi\| \, ,$$

where we used the fact that the operator norm satisfies $\|F^\dagger\| = \|\overline{F}\| = \|F\|$ and $\|I \otimes A\| = \|A\|$. The claim (2.4.4) follows from this and (2.4.2).

The claim (2.4.4) follows analogously by using Equation (2.4.2). Finally, the claim (2.4.4) follows from (2.4.4) and

$$\|E_F\|^2 = \sum_{\alpha_1,\alpha_2 \in [D_1]} \sum_{\beta_1,\beta_2 \in [D_2]} |(\langle\alpha_1| \otimes \langle\beta_1|)E_F(|\alpha_2\rangle \otimes |\beta_2\rangle)|^2$$
$$\le \sum_{\alpha_1,\alpha_2 \in [D_1]} \sum_{\beta_1,\beta_2 \in [D_2]} \|E_F(|\alpha_2\rangle \otimes |\beta_2\rangle)\|^2$$
$$\le D_1^2 D_2^2 \max_{\alpha,\beta} \|E_F(|\alpha_2\rangle \otimes |\beta_2\rangle)\|^2 \, ,$$

where we employed the orthonormal basis $\{|\alpha\rangle\}_{\alpha \in [D_1]}$ and $\{|\beta\rangle\}_{\beta \in [D_2]}$ for $\mathbb{C}^{D_1}$ and $\mathbb{C}^{D_2}$, respectively, and applied the Cauchy-Schwarz inequality. $\qquad\square$

The main result of this section is the following upper bound on the matrix elements of geometrically d-local operators with respect to two MPS.

Theorem 2.4.5. *Let* $|\Psi_1\rangle = |\Psi(A_1, X_1, n)\rangle, |\Psi_2\rangle = |\Psi(A_2, X_2, n)\rangle \in (\mathbb{C}^p)^{\otimes n}$ *be two MPS with bond dimensions D_1 and D_2, where*

$$X_j = |\varphi_j\rangle\langle\psi_j|, \qquad with \qquad \|\varphi_j\| = \|\psi_j\| = 1 \qquad for\ j = 1, 2\,,$$

are rank-one operators. Let $E = E(A_1, A_2) \in \mathcal{B}(\mathbb{C}^{D_1} \otimes \mathbb{C}^{D_2})$ denote the combined transfer operator defined by the MPS tensors A_1 and A_2, h_j^* the size of the largest Jordan block of $E_j = E(A_j)$ for $j = 1, 2$, and h^* the size of the largest Jordan block of $E = E(A_1, A_2)$. Assume that the spectral radii $\rho(E)$, $\rho(E_1)$, and $\rho(E_2)$ are contained in $[0, 1]$. Then, for any $F \in \mathcal{B}((\mathbb{C}^p)^{\otimes d})$, we have

$$|\langle\Psi_1|(F \otimes I_{(\mathbb{C}^p)^{\otimes n-d}})|\Psi_2\rangle| \leq 16 \cdot \|F\| \cdot d^{(h_1^* + h_2^* - 2)/2}(n - d)^{h^* - 1}$$

for $d \gg D_1, D_2$ and $(n - d) \gg D_1 D_2$.

Proof. The matrix elements $\alpha = \langle\Psi_1|(F \otimes I_{(\mathbb{C}^p)^{\otimes n-d}})|\Psi_2\rangle$ of interest can be written as

$$\alpha = (\langle\psi_1| \otimes \langle\psi_2|)\, E_F E^{n-d} (|\varphi_1\rangle \otimes |\varphi_2\rangle)\ .$$

By the Cauchy-Schwarz inequality, we have

$$|\alpha| \leq \|E_F^\dagger(|\psi_1\rangle \otimes |\psi_2\rangle))\| \cdot \|E^{n-d}(|\varphi_1\rangle \otimes |\varphi_2\rangle\|$$

$$\leq \|F\|\sqrt{\|E_1^d\| \cdot \|E_2^d\|} \cdot \|E^{n-d}\|\,,$$

by the definition of the operator norm and Lemma 2.4.4. Then, the claim follows from Lemma 2.4.2 (i), which provides the bounds

$$\|E_j^d\| \leq 4d^{h_j^* - 1} \quad for\ j = 1, 2\,,$$
$$\|E^{n-d}\| \leq 4(n - d)^{h^* - 1}$$

by our assumptions: $\rho(E_j) \in [0, 1]$, $\rho(E) \in [0, 1]$, and $d \gg D_j \geq h_j^*$ for $j = 1, 2$, as well as $n - d \gg D_1 D_2 \geq h^*$. $\qquad\qquad\square$

2.5 No-Go Theorem: Degenerate Ground Spaces of Gapped Hamiltonians are Constant-Distance AQEDC

In this section, we prove a no-go result regarding the error-detection performance of the ground spaces of local gapped Hamiltonians: their distance can be no more

than constant. We prove this result by employing the necessary condition for approximate error-detection from Lemma 2.3.6 for the code subspaces generated by varying the boundary conditions of an (open-boundary) injective MPS. Note that, given a translation invariant MPS with periodic boundary conditions and bond dimension D, there exists a local gapped Hamiltonian, called the parent Hamiltonian, with a unique ground state being the MPS [28].

We need the following bounds which follow from the orthogonality and normalization of states in such codes.

Lemma 2.5.1. *Let A be the MPS tensor of an injective MPS with bond dimension D, and let $X, Y \in \mathcal{B}(\mathbb{C}^D)$ be such that the states $|\Psi_X\rangle = |\Psi(A, X, n)\rangle$ and $|\Psi_Y\rangle = |\Psi(A, Y, n)\rangle$ are normalized and orthogonal. Let us write the transfer operator as $E = |I\rangle\rangle\langle\langle\Lambda| \oplus \tilde{E}$ (cf. Equation (2.4.2)). Assume $n \gg D$. Then*

 (i) The Frobenius norm of X (and similarly the norm of Y) is bounded by

$$\|X\|_F = O(1) .$$

 (ii) We have

$$|\langle\langle\Lambda|(\overline{X} \otimes Y)|I\rangle\rangle| = O(\lambda_2^{n/2}) ,$$
$$|\langle\langle\Lambda|(\overline{Y} \otimes X)|I\rangle\rangle| = O(\lambda_2^{n/2}) .$$

In the following proofs, we repeatedly use the inequality

$$|\operatorname{tr}(M_1 \ldots M_k)| \leq \|M_1\|_F \cdot \|M_2\|_F \cdots \|M_k\|_F \tag{2.30}$$

for $D \times D$-matrices $\{M_j\}_{j=1}^k$. Note that the inequality (2.5) is simply the Cauchy-Schwarz inequality for $k = 2$. For $k > 2$, the inequality follows from the inequality for $k = 2$ and the submultiplicativity of the Frobenius-norm because

$$|\operatorname{tr}(M_1 \ldots M_k)| \leq \|M_1\|_F \cdot \|M_2 \cdots M_k\|_F \leq \|M_1\|_F \cdot \|M_2\|_F \cdots \|M_k\|_F .$$

Proof. The proof of (i) follows from the fact that the state Ψ_X is normalized, i.e.,

$$\begin{aligned}
1 &= \|\Psi_X\|^2 \\
&= \operatorname{tr}\left(E^n(\overline{X} \otimes X)\right) \\
&= \operatorname{tr}\left(|I\rangle\rangle\langle\langle\Lambda|(\overline{X} \otimes X)\right) + \operatorname{tr}(\tilde{E}^n(\overline{X} \otimes X)) \\
&= \operatorname{tr}(\Lambda X X^\dagger) + \operatorname{tr}(\tilde{E}^n(\overline{X} \otimes X)) \\
&\geq \lambda_{\min}(\Lambda) \cdot \|X\|_F^2 + \operatorname{tr}(\tilde{E}^n(\overline{X} \otimes X)) ,
\end{aligned}$$

where $\lambda_{\min}(\Lambda)$ denotes the smallest eigenvalue of Λ, and we make use of the fact that $XX^\dagger$ is positive with trace $\mathrm{tr}(XX^\dagger) = \|X\|_F^2$. Since

$$| \mathrm{tr}(\tilde{E}^n(\overline{X} \otimes X))| \leq \|\tilde{E}^n\|_F \cdot \|\overline{X} \otimes X\|_F \leq \lambda_2^{n/2}\|X\|_F^2 \quad \text{for } n \gg D$$

by (2.5) and (2.4.2), we conclude

$$\|X\|_F^2 \leq \left(\lambda_{\min}(\Lambda) - \lambda_2^{n/2}\right)^{-1} = \lambda_{\min}(\Lambda)^{-1}(1 + O(\lambda_2^{n/2})) \ .$$

Then the claim (i) follows since $\lambda_{\min}(\Lambda)^{-1}$ is a constant.

Now, consider the first inequality in (ii) (the bound for $|\langle\!\langle\Lambda|(\overline{Y} \otimes X)|I\rangle\!\rangle|$ is shown analogously). Using the orthogonality of the states $|\Psi_X\rangle$ and $|\Psi_Y\rangle$, we obtain

$$\begin{aligned}
0 = \langle\Psi_X|\Psi_Y\rangle &= \mathrm{tr}\left(E^n(\overline{X} \otimes Y)\right) \\
&= \mathrm{tr}\left((|I\rangle\!\rangle\langle\!\langle\Lambda| + \tilde{E}^n)(\overline{X} \otimes Y)\right) \\
&= \langle\!\langle\Lambda|(\overline{X} \otimes Y)|I\rangle\!\rangle + \mathrm{tr}(\tilde{E}^n(\overline{X} \otimes Y)),
\end{aligned}$$

hence

$$\begin{aligned}
|\langle\!\langle\Lambda|(\overline{X} \otimes Y)|I\rangle\!\rangle| = | \mathrm{tr}(\tilde{E}^n(\overline{X} \otimes Y))| &\leq \|\tilde{E}\|_F \cdot \|\overline{X} \otimes Y\|_F \\
&\leq \lambda_2^{n/2}\|X\|_F \cdot \|Y\|_F \ ,
\end{aligned}$$

using (2.4.2). The claim (ii) then follows from (i). $\qquad\square$

With the following lemma, we prove an upper bound on the overlap of the reduced density matrices ρ_X and ρ_Y, supported on 2Δ-sites surrounding the boundary, of the global states $|\Psi_X\rangle$ and $|\Psi_Y\rangle$, respectively.

Lemma 2.5.2. *Let A be an MPS tensor of an injective MPS with bond dimension D, and let $X, Y \in \mathcal{B}((\mathbb{C}^p)^{\otimes n})$ be such that the states $|\Psi_X\rangle = |\Psi(A, X, n)\rangle$ and $|\Psi_Y\rangle = |\Psi(A, Y, n)\rangle$ are normalized and orthogonal. Let $\Delta \gg D$. Let $\mathcal{S} = \{1, 2, \ldots, \Delta\} \cup \{n - \Delta + 1, n - \Delta + 2, \ldots, n\}$ be the subset of 2Δ spins consisting of Δ systems at the left and and Δ systems at the right boundary. Let $\rho_X = \mathrm{tr}_{[n]\setminus\mathcal{S}} |\Psi_X\rangle\langle\Psi_X|$ and $\rho_Y = \mathrm{tr}_{[n]\setminus\mathcal{S}} |\Psi_Y\rangle\langle\Psi_Y|$ be the reduced density operators on these subsystems. Then*

$$\mathrm{tr}(\rho_X\rho_Y) \leq c\lambda_2^{\frac{\Delta}{2}}$$

where λ_2 is the second largest eigenvalue of the transfer operator $E = E(A)$ and where c is a constant depending only on the minimal eigenvalue of E and the bond dimension D.
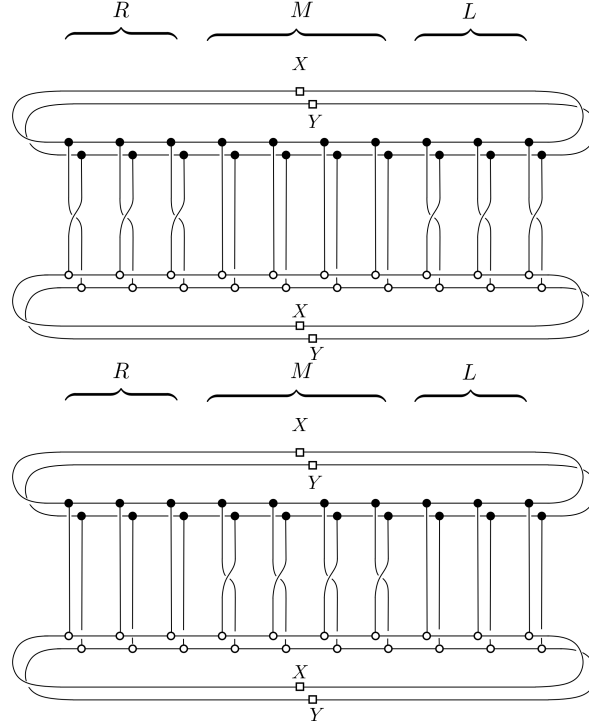
Figure 2.4: The two expressions in Equation (2.5), where L, M, and R are used to denote the sites defined in (2.5).

Proof. For convenience, let us relabel the systems as

$$
\begin{aligned}
(L_1, \ldots, L_\Delta) &= (1, 2, \ldots, \Delta) \\
(M_1, \ldots, M_{n-2\Delta}) &= (\Delta + 1, \Delta + 2, \ldots, n - \Delta) \\
(R_1, \ldots, R_\Delta) &= (n - \Delta + 1, n - \Delta + 2, \ldots, n)
\end{aligned}
\tag{2.31}
$$

indicating their location on the left, in the middle, and on the right, respectively. For the tensor product $\mathcal{H}_A \otimes \mathcal{H}_B$ of two isomorphic Hilbert spaces, we denote by $\mathbb{F}_{AB} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ the flip-operator which swaps the two systems. The following expressions are visualized in Figure 2.4. We have

$$
\mathrm{tr}(\rho_X \rho_Y) = \mathrm{tr}((\rho_X^{L_1 \cdots L_\Delta R_1 \cdots R_\Delta} \otimes \rho_Y^{L'_1 \cdots L'_\Delta R'_1 \cdots R'_\Delta})(\mathbb{F}_{LL'} \otimes \mathbb{F}_{RR'})) , \qquad \text{where}
$$

$$
\mathbb{F}_{LL'} = \mathbb{F}_{L_1 L'_1} \otimes \mathbb{F}_{L_2 L'_2} \otimes \cdots \otimes \mathbb{F}_{L_\Delta L'_\Delta} ,
$$

$$
\mathbb{F}_{RR'} = \mathbb{F}_{R_1 R'_1} \otimes \mathbb{F}_{R_2 R'_2} \otimes \cdots \otimes \mathbb{F}_{R_\Delta R'_\Delta} .
$$

Defining $\mathbb{F}_{MM'}$ analogously, $I_{MM'} = I_{M_1 M'_1} \otimes \cdots \otimes I_{M_{n-2\Delta} M'_{n-2\Delta}}$, and similarly $I_{LL'}$ and $I_{RR'}$, this can be rewritten (by the definition of the partial trace) as

$$
\begin{aligned}
&\mathrm{tr}(\rho_X \rho_Y) \\
&= (\langle \Psi_X^{LMR} | \otimes \langle \Psi_Y^{L'M'R'} |)(\mathbb{F}_{LL'} \otimes I_{MM'} \otimes \mathbb{F}_{RR'})(|\Psi_X^{LMR}\rangle \otimes |\Psi_Y^{L'M'R'}\rangle) \\
&= (\langle \Psi_X^{LMR} | \otimes \langle \Psi_Y^{L'M'R'} |)(I_{LL'} \otimes \mathbb{F}_{MM'} \otimes I_{RR'})(|\Psi_Y^{LMR}\rangle \otimes |\Psi_X^{L'M'R'}\rangle) .
\end{aligned}
\tag{2.32}
$$

In the last identity, we have used that $\mathbb{F}^2 = I$ is the identity.

Reordering and regrouping the systems as

$$(L_1 L_1')(L_2 L_2') \cdots (L_\Delta L_\Delta')(M_1 M_1')(M_2 M_2') \cdots (M_{n-2\Delta} M_{n-2\Delta}')(R_1 R_1')(R_2 R_2') \cdots (R_\Delta R_\Delta') \, ,$$

we observe that $|\Psi_X^{LMR}\rangle \otimes |\Psi_Y^{L'M'R'}\rangle$ is an MPS with MPS tensor $A \otimes A$ and boundary tensor $X \otimes Y$ and $|\Psi_Y^{LMR}\rangle \otimes |\Psi_X^{L'M'R'}\rangle$ is an MPS with MPS tensor $A \otimes A$ and boundary tensor $Y \otimes X$. Let us denote the virtual systems of the first MPS by $V_1 V_2$, and those of the second MPS by $W_1 W_2$, such that the boundary tensors are $X^{V_1} \otimes Y^{V_2}$ and $Y^{W_1} \otimes X^{W_2}$, respectively. Let $\hat{E} = E^{V_1 W_1} \otimes E^{V_2 W_2}$ be the associated transfer operator. Then we have from (2.5)

$$\mathrm{tr}(\rho_X \rho_Y) = \mathrm{tr}\left(\hat{E}^\Delta \hat{E}_{\mathbb{F}^{\otimes n - 2\Delta}} \hat{E}^\Delta \left[(\overline{X}^{V_1} \otimes \overline{Y}^{V_2}) \otimes (Y^{W_1} \otimes X^{W_2}) \right] \right) \, . \qquad (2.33)$$

Recall that $E^\Delta = |I\rangle\!\rangle\langle\!\langle \Lambda| \oplus \tilde{E}^\Delta$, where we have

$$\|\tilde{E}^\Delta\|_F \le \sqrt{D^2} \cdot \|\tilde{E}^\Delta\| \le D \cdot \lambda_2^{\Delta/2} \quad \text{for } \Delta \gg D \, ,$$
$$\||I\rangle\!\rangle\langle\!\langle \Lambda|\|_F = \||I\rangle\!\rangle\|^2 \cdot \||\Lambda\rangle\!\rangle\|^2 \le D^2 \, .$$

In the second line, we use the fact that $\||\Lambda\rangle\!\rangle\|^2 = \mathrm{tr}(\Lambda^\dagger \Lambda) = \sum_i \lambda_i^2 \le 1$ and $\||I\rangle\!\rangle\|^2 = D^2$. Therefore, we have

$$E^\Delta = \sum_{b \in \{0,1\}} H_b \, ,$$

where $H_0 = |I\rangle\!\rangle\langle\!\langle \Lambda|$ and $H_1 = \tilde{E}^\Delta$ satisfy

$$\|H_0\|_F \le D^2 \, , \qquad \text{and} \qquad \|H_1\|_F \le D \cdot \lambda_2^{\Delta/2} \qquad \text{for } \Delta \gg D \, . \qquad (2.34)$$

Note that

$$\hat{E}^\Delta = E^\Delta \otimes E^\Delta = \sum_{b_1, b_2 \in \{0,1\}} H_{b_1} \otimes H_{b_2} \, .$$

Inserting this into (2.5) gives a sum of 16 terms

$$\mathrm{tr}(\rho_X \rho_Y) \le \sum_{b_1,b_2,b_3,b_4 \in \{0,1\}} |\alpha_{b_1,b_2,b_3,b_4}| \, ,$$

where

$$\alpha_{b_1,b_2,b_3,b_4} = \mathrm{tr}\left((H_{b_1}^{V_1 W_1} \otimes H_{b_2}^{V_2 W_2}) \hat{E}_{\mathbb{F}^{\otimes n - 2\Delta}} (H_{b_3}^{V_1 W_1} \otimes H_{b_4}^{V_2 W_2}) \left[(\overline{X}^{V_1} \otimes \overline{Y}^{V_2}) \otimes (Y^{W_1} \otimes X^{W_2}) \right] \right) \, .$$

Consider the term with $b_j = 0$ for all $j \in \{1, \ldots, 4\}$. This is given by

$$\alpha_{0,0,0,0} = \operatorname{tr}\left((|I\rangle\!\rangle\langle\!\langle\Lambda|^{V_1 W_1} \otimes |I\rangle\!\rangle\langle\!\langle\Lambda|^{V_2 W_2}) \hat{E}_{\mathbb{F}^{\otimes n-2\Delta}} (|I\rangle\!\rangle\langle\!\langle\Lambda|^{V_1 W_1} \otimes |I\rangle\!\rangle\langle\!\langle\Lambda|^{V_2 W_2}) \right.$$

$$\left. \cdot \left[(\overline{X}^{V_1} \otimes \overline{Y}^{V_2}) \otimes (Y^{W_1} \otimes X^{W_2}) \right] \right)$$

$$= \langle\!\langle\Lambda|(\overline{X} \otimes Y)|I\rangle\!\rangle \langle\!\langle\Lambda|(\overline{Y} \otimes X)|I\rangle\!\rangle (\langle\!\langle\Lambda| \otimes \langle\!\langle\Lambda|) \hat{E}_{\mathbb{F}^{\otimes n-2\Delta}} (|I\rangle\!\rangle \otimes |I\rangle\!\rangle). \quad (2.35)$$

By inserting this into (2.5), we get with Lemma 2.5.1 (ii) and the Cauchy-Schwarz inequality

$$|\alpha_{0,0,0,0}| = O(\lambda_2^n) \cdot \left| (\langle\!\langle\Lambda| \otimes \langle\!\langle\Lambda|) \hat{E}_{\mathbb{F}^{\otimes n-2\Delta}} (|I\rangle\!\rangle \otimes |I\rangle\!\rangle) \right|$$

$$= O(\lambda_2^n) \cdot \||\Lambda\rangle\!\rangle \otimes |\Lambda\rangle\!\rangle\| \cdot \|\hat{E}_{\mathbb{F}^{\otimes n-2\Delta}} (|I\rangle\!\rangle \otimes |I\rangle\!\rangle)\| .$$

With Lemma 2.4.4, this can further be bounded as

$$|\alpha_{0,0,0,0}| = O(\lambda_2^n) \cdot \||\Lambda\rangle\!\rangle\|^2 \cdot \||I\rangle\!\rangle\|^2 \cdot \|\mathbb{F}^{\otimes n-2\Delta}\| \cdot \|E^{n-2\Delta}\| .$$

Since $\|\mathbb{F}\| = 1$ and $\||\Lambda\rangle\!\rangle\| = O(1)$, $\||I\rangle\!\rangle\| = O(1)$ and $\|E^{n-2\Delta}\| = O(1)$ (cf. (2.4.3)), we conclude that

$$|\alpha_{0,0,0,0}| = O(\lambda_2^n) . \quad (2.36)$$

The remaining terms $|\alpha_{b_1,b_2,b_3,b_4}|$ with $(b_1, b_2, b_3, b_4) \neq (0, 0, 0, 0)$ can be bounded as follows: using inequality (2.5), we have

$$|\alpha_{b_1,b_2,b_3,b_4}| = \left| \operatorname{tr}\left((H_{b_1} \otimes H_{b_2}) \hat{E}_{\mathbb{F}^{\otimes N-2\Delta}} (H_{b_3} \otimes H_{b_4}) \left[(\overline{X} \otimes \overline{Y}) \otimes (Y \otimes X) \right] \right) \right|$$

$$\leq \|H_{b_1} \otimes H_{b_2}\|_F \cdot \|E_{\mathbb{F}^{\otimes N-2n}}\|_F \cdot \|H_{b_3} \otimes H_{b_4}\|_F \cdot \|\overline{X} \otimes \overline{Y} \otimes Y \otimes X\|_F$$

$$= \|X\|_F^2 \cdot \|Y\|_F^2 \cdot \left(\prod_{j=1}^{4} \|H_{b_j}\|_F \right) \cdot \|E_{\mathbb{F}^{\otimes n-2\Delta}}\|_F$$

$$= O(\lambda_2^{\Delta/2}) \cdot \|X\|_F^2 \cdot \|Y\|_F^2 \cdot \|E_{\mathbb{F}^{\otimes n-2\Delta}}\|_F ,$$

where we use (2.5) and the assumption that $(b_1, b_2, b_3, b_4) \neq (0, 0, 0, 0)$. We use Lemma 2.4.4 and (2.4.3) to get the upper bound $\|E_{\mathbb{F}^{\otimes n-2\Delta}}\| \leq D^2 \|F^{\otimes n-2\Delta}\| \cdot \|E^{n-D}\| = O(1)$. Thus

$$|\alpha_{b_1,b_2,b_3,b_4}| = O(\lambda_2^{\Delta/2}) \qquad \text{for } (b_1, b_2, b_3, b_4) \neq (0, 0, 0, 0) . \quad (2.37)$$

Combining (2.5) with (2.5), we conclude that

$$|\operatorname{tr}(\rho_X \rho_Y)| \leq \sum_{b_1,b_2,b_3,b_4 \in \{0,1\}} |\alpha_{b_1,b_2,b_3,b_4}| \leq |\alpha_{0,0,0,0}| + 15 \max_{(b_1,b_2,b_3,b_4) \neq (0,0,0,0)} |\alpha_{b_1,b_2,b_3,b_4}|$$

$$= O(\lambda_2^{\Delta/2}) .$$

The claim follows from this. $\qquad\square$

Recall that we call (a family of subspaces) $C \subset (\mathbb{C}^{\mathsf{p}})^{\otimes n}$ an approximate error-detection code if it is an $(\epsilon, \delta)[[n, k, d]]$-code with $\epsilon \to 0$ and $\delta \to 0$ for $n \to \infty$. Our main result is the following:

Theorem 2.5.3. *Let $C \subset (\mathbb{C}^{\mathsf{p}})^{\otimes n}$ be an approximate error-detecting code generated from a translation-invariant injective MPS of constant bond dimension D by varying boundary conditions. Then the distance of C is constant.*

Proof. Let $C = C_n \subset (\mathbb{C}^{\mathsf{p}})^{\otimes n}$ be a (family of) subspace(s) of dimension p^k defined by an MPS tensor A by choosing different boundary conditions, i.e.,

$$C_n = \{|\Psi(A, X, n)\rangle \mid X \in \mathcal{X}\} \subset (\mathbb{C}^{\mathsf{p}})^{\otimes n}$$

for some (fixed) subspace $\mathcal{X} \subset \mathcal{B}(\mathbb{C}^D)$. For the sake of contradiction, assume that C_n is an $(\epsilon_n, \delta_n)[[n, k, d_n]]$-code with

$$\epsilon_n, \delta_n \to 0 \qquad \text{and} \qquad \text{code distance } d_n \to \infty \qquad \text{for } n \to \infty . \quad (2.38)$$

Let $|\Psi_X\rangle = |\Psi(A, X, n)\rangle, |\Psi_Y\rangle = |\Psi(A, Y, n)\rangle \in C$ be two orthonormal states defined by choosing different boundary conditions $X, Y \in \mathcal{X}$. From Lemma 2.5.2, we may choose Δ sufficiently large such that the reduced density operators ρ_X, ρ_Y on d sites surrounding the boundary satisfies

$$\text{tr}(\rho_X \rho_Y) \leq c\lambda_2^{d/4} \qquad \text{for all} \quad d \geq 2\Delta . \quad (2.39)$$

We note that Δ only depends on the transfer operator and is independent of n. Fix any constant $\epsilon, \delta \in (0, 1)$ and choose some $d \geq 2\Delta$ sufficiently large such that

$$\zeta(\rho_X, \rho_Y) := cD^2 \lambda_2^{d/4}$$

satisfies

$$\epsilon < 1 - 10\zeta \quad \text{and} \quad \delta < (1 - \zeta)^2 . \quad (2.40)$$

Since by assumption $d_n \to \infty$, there exists some $N_0 \in \mathbb{N}$ such that

$$d_n > d \qquad \text{for all } n \geq N_0 . \quad (2.41)$$

Combining (2.5), (2.5), and (2.5) with Lemma 2.3.6, we conclude that C_n is not an $(\epsilon, \delta)[[n, k, d_n]]$-code for any $n \geq N_0$.

By assumption (2.5), there exists some $N_1 \in \mathbb{N}$ such that

$$\epsilon_n < \epsilon \qquad \text{and} \qquad \delta_n < \delta \qquad \text{for all } n \geq N_1 \ .$$

Let us set $N = \max\{N_0, N_1\}$. Then we obtain that C_n is not an $(\epsilon_n, \delta_n)[[n, k, d_n]]$-code for any $n \geq N$, a contradiction. $\qquad\qquad\square$

In terms of the TQO-1 condition (cf. [57]), Theorem 2.5.3 shows the absence of topological order in 1D gapped systems. The theorem also tells us that we should not restrict our attention to the ground space of a local Hamiltonian when looking for quantum error-detecting codes.[5] In the following sections, we bypass this no-go result by extending our search for codes to low-energy states. In particular, we show that single quasi-particle momentum eigenstates of local gapped Hamiltonians and multi-particle excitations of the gapless Heisenberg model constitute error-detecting codes. See Sections 2.6 and 2.7, respectively.

2.6 AQEDC at Low Energies: The Excitation Ansatz

In this section, we employ tangent space methods for the matrix product state formalism, i.e., the excitation ansatz [66], [76], [77], in order to show that quasi-particle momentum eigenstates of local gapped Hamiltonians yield an error-detecting code with distance $\Omega(n^{1-\nu})$ for any $\nu \in (0, 1)$ and $\Omega(\log n)$ encoded qubits.
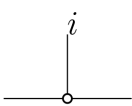
In order to render the formalism accessible to an unfamiliar reader, we review the definition of the excitation ansatz in Section 2.6.1. We then develop the necessary calculational ingredients in order to prove the error-detection properties. In Section 2.6.2, we compute the norm of the excitation ansatz states to lowest order. In Section 2.6.3, we establish (norm) bounds on the transfer operators associated with the excitation ansatz. Then, in Section 2.6.4, we provide estimates on matrix elements of local operators with respect to states appearing in the definition of the excitation ansatz states. Finally, in Section 2.6.5, we combine these results to obtain the parameters of quantum error-detecting codes based on the excitation ansatz.[6]

---

[5]Note that this conclusion is only valid for local gapped Hamiltonians in one dimension. When the spatial dimension $d \geq 2$, there are ground spaces that have topological order, e.g. the toric code. Also for higher dimensions, good quantum LDPC codes are shown to exist in the ground space of frustration-free Hamiltonians [75].

[6]A simple yet illustrative example of the excitation ansatz states is the following: consider the n-fold product state $|0\rangle^{\otimes n}$, the n-body W-state

$$\frac{|10\cdots0\rangle + \cdots + |00\cdots1\rangle}{\sqrt{n}},$$

$$|\Phi_p(B;A)\rangle = \underbrace{B(p) \cdots}_{} + e^{ip} \underbrace{\cdots B(p) \cdots}_{} + \cdots + e^{(n-1)ip} \underbrace{\cdots B(p)}_{}$$
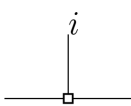
$$i \qquad\qquad i$$
$$A_i \qquad\qquad B_i$$

Figure 2.5: This figure illustrates the excitation ansatz $|\Phi_p(B;A)\rangle$ for n physical spins.

2.6.1 MPS Tangent Space Methods: The Excitation Ansatz

In [76], the MPS ansatz was generalized to a variational class of states which have non-zero momentum. The resulting states are called the *excitation ansatz*. An excitation ansatz state $|\Phi_p(B;A)\rangle \in (\mathbb{C}^p)^{\otimes n}$ is specified by two MPS tensors $\{A_i\}_{i=1}^{p}$ and $\{B_i\}_{i=1}^{p}$ of the same bond and physical dimensions, together with a parameter $p \in \{2\pi k/n \mid k = 0, \dots, n\}$ indicating the momentum. It is defined as

$$|\Phi_p(B;A)\rangle$$
$$= e^{-ip} \sum_{j=1}^{n} e^{ipj} \sum_{i_1,\dots,i_n \in [p]} \mathrm{tr}(A_{i_1} \cdots A_{i_{j-1}} B_{i_j} A_{i_{j+1}} \cdots A_{i_n})|i_1 \dots i_n\rangle. \qquad (2.42)$$

The definition of these states is illustrated in Figure 2.5. Note that we allow the B tensors themselves to depend on the momentum p, so we will sometimes write $B(p)$ when we feel the need to be explicit, and the notation $|\Phi_p(B;A)\rangle$ should really be read as a short-hand for $|\Phi_p(B(p);A)\rangle$.

---

as well as other W-like states with position dependent phase, such as

$$\frac{|10\cdots 0\rangle + e^{ip}|01\cdots 0\rangle + \cdots + e^{ip(n-1)}|00\cdots 1\rangle}{\sqrt{n}}.$$

Here p can be interpreted as the momentum of a single particle excitation. These states are the ground state and first excited states with different momenta of the non-interacting Hamiltonian $H = -\sum_i Z_i$. One can represent them by a bond-dimension $D = 2$ non-injective MPS which is obtained by expressing the excitation ansatz as a single MPS instead of a sum of injective MPS. One can also consider higher (multi-particle) excitations, which can again be treated by using non-injective MPS.

We note that error-detecting properties of various subspaces of the low-energy space of this particular simple non-interacting Hamiltonian can be studied either with or without the formalism of MPS. The tangent space methods serve as a powerful tool that allow us to perform our error-detection analysis, not only for the non-interacting cases, but also for the most general interacting Hamiltonians.

It is also useful to define the constituent "position space" states

$$|\Phi_{j,p}(B;A)\rangle = \sum_{i_1,\ldots,i_n\in[\mathsf{p}]} \mathrm{tr}(A_{i_1}\cdots A_{i_{j-1}}B_{i_j}A_{i_{j+1}}\cdots A_{i_n})|i_1\ldots i_n\rangle,$$

$$, (2.43)$$

which is the state with a "single $B(p)$ excitation" at site j. Note that we retain the p dependence in the definition of these "position space" states since the B tensors themselves are generally p dependent.

We call an excitation ansatz state $|\Phi_p(B;A)\rangle$ *injective* if the transfer operator $\mathcal{E}(A)$ associated with $\{A_j\}_{j=1}^{\mathsf{p}}$ is primitive, which is the only case we consider in this work. Denoting the transfer matrix associated with $\mathcal{E}(A)$ simply as E, it will also be useful to define several other mixed transfer matrices as follows:

$$E_{B(p)} \quad = \quad \sum_{j=1}^{D}\overline{A}_j\otimes B_j(p) \quad = \quad$$

$$,$$

$$E_{\overline{B(p)}} \quad = \quad \sum_{j=1}^{D}\overline{B_j(p)}\otimes A_j \quad = \quad$$

$$,$$

$$E_{\overline{B(p')}B(p)} \quad = \quad \sum_{j=1}^{D}\overline{B_j(p')}\otimes B_j(p) \quad = \quad$$

$$.$$

For brevity, we often suppress the dependence on A and B and simply write $|\Phi_p\rangle \equiv |\Phi_p(B;A)\rangle$ when no confusion is possible.

In addition to the multiplicative gauge freedom (2.4.1), the excitation ansatz admits an additional additive gauge freedom. Exploiting this additive gauge freedom, the following statement can be shown (see [77, Equation (154)]):

Lemma 2.6.1. *Let $|\Phi_p(B;A)\rangle$ be an injective excitation ansatz state and assume that A is normalized such that the transfer operator has spectral radius* 1. *Let ℓ and*

r be the corresponding left- and right-eigenvectors corresponding to eigenvalue 1.
Assume $p \neq 0$.[7] *Then there exists a tensor* $\tilde{B}$ *such that* $|\Phi_p(B; A)\rangle = |\Phi_p(\tilde{B}; A)\rangle$,
and such that

$$\langle\!\langle \ell | E_{\tilde{B}(p)} = 0 \qquad and \qquad \langle\!\langle \ell | E_{\overline{\tilde{B}(p)}} = 0 \,. \tag{2.44}$$

For completeness, we give a proof of this statement in Appendix 2.A. Below, we
assume that all excitation ansatz states satisfy the gauge condition (2.6.1).

2.6.2 The Norm of an Excitation Ansatz State

For a family of excitation ansatz states $\{|\Phi_p(B; A)\rangle\}_p$, we define the constants

$$c_{pp'} = \langle\!\langle \ell | E_{\overline{B(p')}B(p)} | r \rangle\!\rangle = \ell \begin{array}{c} \overline{B(p')} \\ \\ B(p) \end{array} r.$$

We also write $c_p := c_{pp}$. These appear in the norm of the excitation ansatz states as
follows:

Lemma 2.6.2. *The norm of an excitation ansatz state* $|\Phi_p(B; A)\rangle \in (\mathbb{C}^p)^{\otimes n}$ *satisfies*

$$\||\Phi_p(B; A)\rangle\| = \sqrt{nc_p} + O(n^{3/2}\lambda_2^{n/6}),$$

where λ_2 *is the second largest eigenvalue of the transfer matrix* E.

Proof. Using the mixed transfer operators defined in (2.6.1), we can write the norm
of the state $|\Phi_p(B; A)\rangle$ as a sum over pairs $(j, j') \in [n]^2$ satisfying $j < j'$, $j = j'$,
and $j > j'$ respectively, as follows:

$$\||\Phi_p(B; A)\rangle\|^2 = \sum_{j<j'} e^{ip(j-j')} \text{tr} \left(E^{j-1} E_{B(p)} E^{j'-j-1} E_{\overline{B(p)}} E^{n-j'} \right)$$

$$+ \sum_{j>j'} e^{ip(j-j')} \text{tr} \left(E^{j-1} E_{\overline{B(p)}} E^{j'-j-1} E_{B(p)} E^{n-j'} \right)$$

$$+ \sum_{j=1}^{n} \text{tr} \left(E^{j-1} E_{\overline{B(p)}B(p)} E^{n-j} \right). \tag{2.45}$$

---

[7]We have made the $p \neq 0$ assumption here for simplicity. The gauge condition also holds for
$p = 0$ in the form $\langle\!\langle \ell | E_{\tilde{B}(p)} = \langle\!\langle \ell | E_{\overline{\tilde{B}(p)}} = O(\lambda_2^n)$. All of the results presented below for $p \neq 0$ also
hold for $p = 0$ up to an exponentially small error.

Consider an individual term $\text{tr}\left(E^{j-1}E_{B(p)}E^{j'-j-1}E_{\overline{B(p)}}E^{n-j'}\right)$ in the first sum. By the cyclicity of the trace, it can be expressed as

$$\text{tr}\left(E^{j-1}E_{B(p)}E^{j'-j-1}E_{\overline{B(p)}}E^{n-j'}\right) = \text{tr}\left(E_{B(p)}E^{\Delta-1}E_{\overline{B(p)}}E^{n-\Delta-1}\right),$$

where $\Delta = j' - j$. Clearly, one of the terms $\Delta - 1$ or $n - \Delta - 1$ must be lower bounded by $n/3$. Assume that it is the first (the argument for the other case is analogous), i.e., that

$$\Delta - 1 > n/3. \tag{2.46}$$

Then we may substitute the Jordan decomposition of E in the form

$$E^{\Delta-1} = |r\rangle\!\rangle\langle\!\langle\ell| \oplus \tilde{E}^{\Delta-1},$$

which allows us to write

$$\text{tr}\left(E^{j-1}E_{B(p)}E^{j'-j-1}E_{\overline{B(p)}}E^{n-j'}\right) = \text{tr}\left(E_{B(p)}|r\rangle\!\rangle\langle\!\langle\ell|E_{\overline{B(p)}}E^{n-\Delta-1}\right)$$
$$+ \text{tr}\left(E_{B(p)}\tilde{E}^{\Delta-1}E_{\overline{B(p)}}E^{n-\Delta-1}\right).$$

By the gauge condition (2.6.1), the first term vanishes. The magnitude of the second term can be bounded by inequality (2.5), giving

$$\text{tr}\left(E_{B(p)}\tilde{E}^{\Delta-1}E_{\overline{B(p)}}E^{n-\Delta-1}\right) \leq O(1) \cdot \|\tilde{E}^{\Delta-1}\|_F \cdot \|E^{n-\Delta-1}\|_F.$$

Here we used the fact that $\|E_{\overline{B(p)}}\|_F = O(1)$ and $\|E_{B(p)}\|_F = O(1)$. With (2.6.2) and Lemma 2.4.2(ii), we have $\|\tilde{E}^{\Delta-1}\|_F \leq \lambda_2^{n/6}$ and $\|E^{n-\Delta-1}\|_F = O(1)$. We conclude that

$$\left|\text{tr}\left(E^{j-1}E_{B(p)}E^{j'-j-1}E_{\overline{B(p)}}E^{n-j'}\right)\right| = O(\lambda_2^{n/6})$$

for all pairs (j, j') with $j < j'$.

Identical reasoning gives us a bound of the form

$$\text{tr}\left(E^{j-1}E_{\overline{B(p)}}E^{j'-j-1}E_{B(p)}E^{n-j'}\right) = O(\lambda_2^{n/6})$$

for all pairs (j, j') with $j > j'$. Inserting this into the sum (2.6.2), we obtain

$$\|\Phi_p(B;A)\|^2 = \sum_{j=1}^{n} \text{tr}\left(E^{j-1}E_{\overline{B(p)}B(p)}E^{n-j}\right) + O(n^2 \cdot \lambda_2^{n/6}). \tag{2.47}$$

By the cyclicity of the trace and the Jordan decomposition of E, we have

$$\mathrm{tr}\left(E^{j-1}E_{\overline{B(p)}B(p)}E^{n-j}\right) = \mathrm{tr}(E_{\overline{B(p)}B(p)}E^{n-1})$$
$$= \langle\!\langle \ell|E_{\overline{B(p)}B(p)}|r\rangle\!\rangle + \mathrm{tr}(E_{\overline{B(p)}B(p)}\tilde{E}^{n-1})$$
$$= c_p + \mathrm{tr}(E_{\overline{B(p)}B(p)}\tilde{E}^{n-1}).$$

Again using inequality (2.5) and Lemma 2.4.2(ii), we get

$$\left|\mathrm{tr}(E_{\overline{B(p)}B(p)}\tilde{E}^{n-1})\right| \leq \|E_{\overline{B(p)}B(p)}\|_F \cdot \|\tilde{E}^{n-1}\|_F = O\left(\lambda_2^{(n-1)/2}\right).$$

Inserting this into (2.6.2) and noting that $\lambda_2^{(n-1)/2} \leq n \cdot \lambda_2^{n/6}$ gives us

$$\|\Phi_p(B;A)\|^2 = nc_p + O(n^2 \cdot \lambda^{n/6}) = nc_p(1 + O(n \cdot \lambda^{n/6})).$$

Taking the square root yields the desired claim. $\qquad\square$

2.6.3 Bounds on Transfer Operators Associated with the Excitation Ansatz

For an operator $F \in (\mathbb{C}^p)^{\otimes L}$, sites $j, j' \in [L]$, and momenta p, p', let us define operators on $\mathbb{C}^D \otimes \mathbb{C}^D$ by the diagrams

$$E_F(j, p, j', p') = \quad , \quad \text{and} \quad E_F(j, p) = \quad .$$

We also denote by $E(j, p, j', p')$ the expression $E_I(j, p, j', p')$.

We keep the dependence of E_F on L implicit, since none of our computations will explicitly depend on L. Similarly to the bounds discussed in Section 2.4.2, we require bounds on the norm (respectively matrix elements) of these transfer operators. These are given by the following:

Lemma 2.6.3. *Let $F \in (\mathbb{C}^p)^{\otimes L}$, $j, j' \in [L]$, and momenta p, p' be arbitrary. Then we have*

$$\langle\!\langle \ell|E(j, p, j', p')|r\rangle\!\rangle = \delta_{j,j'}c_{pp'}, \tag{2.48}$$

and

$$|\langle\!\langle \ell|E_F(j,p,j',p')|r\rangle\!\rangle| \le \|F\|\sqrt{c_p c_{p'}} \,, \tag{2.49}$$

$$\|E_F(j,p,j',p')\|_F \le D^2\|F\|\sqrt{\|E_{\overline{B(p)}B(p)}\|_F \, \|E_{\overline{B(p')}B(p')}\|_F} \,, \tag{2.50}$$

$$\|E_F(j,p)\|_F \le D^2\|F\|\sqrt{\|E_{\overline{B(p)}B(p)}\|_F} \,, \tag{2.51}$$

$$\|E_F\|_F \le D^2\|F\| \,. \tag{2.52}$$

For the proof of Lemma 2.6.3 (and other arguments below), we make repeated use of the following states. Let $L \in [n]$. Define

$$|\Phi^L_{j,p}\rangle = \ell^{\frac{1}{2}} \quad\text{...}\quad r^{\frac{1}{2}} \tag{2.53}$$

on $\mathbb{C}^D \otimes (\mathbb{C}^{\mathrm{p}})^{\otimes L} \otimes \mathbb{C}^D$. Despite the similar notation, these states are not to be confused with the "position space" states $|\Phi_{j,p}\rangle$ introduced in Equation (2.6.1). The key property of the states $|\Phi^L_{j,p}\rangle$ is the following:

Lemma 2.6.4. *The states* (2.6.3) *have inner product*

$$\langle\Phi^L_{j',p'}|\Phi^L_{j,p}\rangle = \delta_{j,j'}c_{pp'} \,, \tag{2.54}$$

independently of the value of L.

Proof. First, consider the case where $j' = j$. Then we have

$$\langle\Phi^L_{j,p'}|\Phi^L_{j,p}\rangle = \langle\!\langle \ell|E^{j-1}E_{\overline{B(p')}B(p)}E^{L-j}|r\rangle\!\rangle = \langle\!\langle \ell|E_{\overline{B(p')}B(p)}|r\rangle\!\rangle = c_{pp'} \,,$$

where we have used the fixed-point equations (2.4.1). That is, we have

$$\langle\Phi^L_{j,p'}|\Phi^L_{j,p}\rangle = \ell \quad\text{...}\quad r = \ell \quad\text{...}\quad r \,.$$

In a similar fashion, we can compute, for $j < j'$,

$$\langle\Phi^L_{j',p'}|\Phi^L_{j,p}\rangle = \langle\!\langle \ell|E^{j-1}E_{B(p)}E^{j'-j}E_{\overline{B(p')}}E^{L-j'}|r\rangle\!\rangle = \langle\!\langle \ell|E_{B(p)}E^{j'-j}E_{\overline{B(p')}}|r\rangle\!\rangle = 0,$$

where we have used the fixed-point equations (2.4.1) and the gauge condition (2.6.1). The proof for $j > j'$ is analogous. $\square$

Proof of Lemma 2.6.3. We first prove (2.6.3). The expression of interest can be written diagrammatically as

$$\langle\!\langle \ell | E_F(j,p,j',p') | r \rangle\!\rangle = \ell \; \boxed{F} \; r = \langle \Phi^L_{j',p'} | (I \otimes F \otimes I) | \Phi^L_{j,p} \rangle \; .$$

Equation (2.6.3) follows by setting F to be equal to the identity on $(\mathbb{C}^p)^{\otimes L}$ and using the orthogonality relation (2.6.4). Furthermore, we have

$$|\langle\!\langle \ell | E_F(j,p,j',p') | r \rangle\!\rangle| = \left| \langle \Phi^L_{j',p'} | (I \otimes F \otimes I) | \Phi^L_{j,p} \rangle \right| \le \|F\| \cdot \|\Phi^L_{j,p}\| \cdot \|\Phi^L_{j',p'}\|.$$

The claim (2.6.3) then follows from (2.6.4).

Let us next prove (2.6.3). By the definition of the Frobenius norm $\|\cdot\|_F$, we have

$$\|E_F(j,p,j',p')\|^2_F = \sum_{\alpha_1,\alpha_2,\beta_1,\beta_2=1}^{D} \left| (\langle \alpha_1| \otimes \langle \alpha_2|) E_F(j,p,j',p') (|\beta_1\rangle \otimes |\beta_2\rangle) \right|^2$$

where $\{|\alpha\rangle\}_{\alpha=1}^D$ is an orthonormal basis of $\mathbb{C}^D$. The terms in the sum can be written diagrammatically as

$$(\langle \alpha_1| \otimes \langle \alpha_2|) E_F(j,p,j',p') (|\beta_1\rangle \otimes |\beta_2\rangle) = \quad .$$

Defining vectors

$$|\Psi_{j,p}(\alpha,\beta)\rangle = \qquad\qquad (2.55)$$

on $(\mathbb{C}^p)^{\otimes L}$, we have

$$|(\langle \alpha_1| \otimes \langle \alpha_2|) E_F(j,p,j',p') (|\beta_1\rangle \otimes |\beta_2\rangle)|^2 = |\langle \Psi_{j',p'}(\alpha_1,\beta_1) | F | \Psi_{j,p}(\alpha_2,\beta_2) \rangle|^2$$

$$\le \|F\|^2 \cdot \|\Psi_{j,p}(\alpha_2,\beta_2)\|^2 \cdot \|\Psi_{j',p'}(\alpha_1,\beta_1)\|^2 \; .$$

The norm of the vector (2.6.3) can be bounded as

$$\|\Psi_{j,p}(\alpha,\beta)\|^2 = \quad$$

$$= \left| \operatorname{tr}\left(E^{j-1} E_{\overline{B(p)}B(p)} E^{L-j} (|\beta\rangle\langle\alpha| \otimes |\beta\rangle\langle\alpha|) \right) \right|$$

$$\leq \|E_{\overline{B(p)}B(p)}\|_F \cdot \|E^{j-1}\|_F \cdot \|E^{L-j}\|_F$$

$$\leq \|E_{\overline{B(p)}B(p)}\|_F .$$

In the first inequality, we have used (2.5), together with the fact that

$$\||\beta\rangle\langle\alpha| \otimes |\beta\rangle\langle\alpha|\|_F = 1 .$$

In the second inequality, we have used Lemma 2.4.2, along with the fact $\rho(E) = 1$. The claim (2.6.3) follows from this.

With a completely analogous proof, we also have

$$\|E_F(j,p)\|_F \leq D^2 \|F\| \sqrt{\|E_{\overline{B(p)}B(p)}\|_F}, \qquad \text{and} \qquad \|E_F\|_F \leq D^2 \|F\|,$$

which are claims (2.6.3) and (2.6.3). $\qquad\square$

2.6.4 Matrix Elements of Local Operators in the Excitation Ansatz

Overview of the Proof

Let us give a high-level overview of the argument used to establish our main technical result, Lemma 2.6.8. The latter gives estimates on matrix elements $\langle\phi_{p'}|F|\phi_p\rangle$ of a d-local operator F with respect to normalized excitation ansatz states $|\phi_p\rangle$ and $|\phi_{p'}\rangle$, with possibly different momenta p and p'. More precisely, to apply the approximate Knill-Laflamme conditions for approximate error-detection, we need to establish two kinds of bounds:

1. For $p \neq p'$ (i.e., the non-diagonal elements), our aim is to argue that $|\langle\phi_{p'}|F|\phi_p\rangle|$ vanishes as an inverse polynomial in n. This is ultimately a consequence of the fact that in the Jordan decomposition $E = |r\rangle\!\rangle\langle\!\langle\ell| \oplus \tilde{E}$ of the transfer matrix, the sub-dominant term $\tilde{E}$ has norm decaying exponentially with a rate determined by the second largest eigenvalue λ_2.

2. For the diagonal elements, our aim is to argue that $\langle\phi_p|F|\phi_p\rangle$ is almost independent of p, that is, we want to show $|\langle\phi_p|F|\phi_p\rangle - \langle\phi_{p'}|F|\phi_{p'}\rangle|$ is small

for different momenta $p \neq p'$. For this purpose, we need to identify the leading order term in the expression $\langle \phi_p | F | \phi_p \rangle$. Higher order terms are again small by the properties of the transfer operator.

To establish these bounds, first observe that an unnormalized excitation ansatz state $|\Phi_p(B; A)\rangle$ is a superposition of the "position space" states $\{|\Phi_{j,p}\rangle\}_{j=1}^n$, where each state $|\Phi_{j,p}\rangle$ is given by a simple tensor network with an "insertion" of an operator at site j'. Correspondingly, we first study matrix elements of the form $\langle \Phi_{j',p'} | F | \Phi_{j,p} \rangle$. Bounds on these matrix elements are given in Lemma 2.6.5. The idea of the proof of this statement is simple: in the tensor network diagram for the matrix element, subdiagrams associated with powers E^Δ with sufficiently large Δ may be replaced by the diagram associated with the map $|r\rangle\!\rangle\langle\!\langle \ell|$, with an error scaling term scaling as $O(\lambda_2^{\Delta/2})$. This is due to the Jordan decomposition of the transfer operator. Thanks to the gauge condition (2.6.1), the resulting diagrams then simplify, allowing us to identify the leading order term.

To realize this approach, a key step is to identify suitable subdiagrams corresponding to powers E^Δ in the diagram associated with $\langle \Phi_{j',p'} | F | \Phi_{j,p} \rangle$. These are associated with connected regions of size Δ where the operator F acts trivially, and there is no insertion of $B(p)$ (respectively $\overline{B(p')}$), meaning that j and j' do not belong to the region. Lemma 2.6.5 provides a careful case-by-case analysis depending on, at the coarsest level of detail, whether or not j and j' belong to a Δ-neighborhood of the support of F.

Some subleties that arise are the following: to obtain estimates on the leading-order terms for the diagonal matrix elements (see (2) above) as well as related expressions, a bound on the magnitude of the matrix element $\langle \Phi_{j,p'} | F | \Phi_{j,p} \rangle$ only is not sufficient. The lowest-order approximating expression to $\langle \Phi_{j,p'} | F | \Phi_{j,p} \rangle$ obtained by making the above substitutions of the transfer operators a priori seems to depend on the exact site location j. This is awkward because the term $\langle \Phi_{j,p'} | F | \Phi_{j,p} \rangle$ appears as a summand (with sum taken over j) when computing matrix elements of excitation ansatz states. We argue that in fact, the leading order term of $\langle \Phi_{j,p'} | F | \Phi_{j,p} \rangle$ is identical for all values of j not belonging to the support of F. This statement is formalized in Lemma 2.6.6 and allows us to subsequently estimate sums of interest without worry about the explicit dependence on j.

Finally, we require a strengthening of the estimates obtained in Lemma 2.6.5 because we are ultimately interested in excitation ansatz states: these are superpositions of

the states $|\Phi_{j,p}\rangle$, with phases of the form e^{ipj}. Estimating only the magnitude of matrix elements of the form $\langle\Phi_{j',p'}|F|\Phi_{j,p}\rangle$ is not sufficient to establish our results. Instead, we need to treat the phases "coherently," which leads to certain cancellations. The corresponding statement is given in Lemma 2.6.7.

The Proof

We will envision the sites $\{1, \ldots, n\}$ as points on a ring, i.e., using periodic boundary conditions, and measure the distance between sites j, j' by

$$\mathrm{dist}(j, j') := \min_{k \in \mathbb{Z}} |j - j' + k \cdot n|.$$

For $\Delta \in \{0, \ldots, n\}$ and a subset $\mathcal{F} \subset \{1, \ldots, n\}$, let

$$\mathcal{B}^{\Delta}(\mathcal{F}) = \{j \in \{1, \ldots, n\} \mid \exists\, j' \in \mathcal{F} \text{ such that } \mathrm{dist}(j, j') \le \Delta\}$$

be the Δ-*thickening* of $\mathcal{F}$.

We say that $j' \in \{1, \ldots, n\}$ is a *left neighbor of* (or is *left-adjacent to*) $j \in \{1, \ldots, n\}$ if $j' = j - 1$ for $j > 1$, or $j' = n$ for $j = 1$. A connected region $\mathcal{R} \subset \{1, \ldots, n\}$ is said to lie on the left of (or be left-adjacent to) $j \in \{1, \ldots, n\}$ if it is of the form $\mathcal{R} = \{j_1, \ldots, j_r\}$, with $j_{\alpha+1}$ left-adjacent to j_α for $\alpha \in \{0, \ldots, r-1\}$ with the convention that $j_0 = j_r$. Analogous definitions hold for right-adjacency.

For an operator F acting on $(\mathbb{C}^p)^{\otimes n}$, let $\mathrm{supp}(F) \subset \{1, \ldots, n\}$ denote its support, i.e., the sites of the system that the operator acts on non-trivially. We say that F is d-local if $|\mathrm{supp}(F)| = d$. Let us assume that $\mathrm{supp}(F)$ decomposes into κ disjoint connected components

$$\mathrm{supp}(F) = \bigcup_{\alpha=0}^{\kappa-1} \mathcal{F}_\alpha \,. \tag{2.56}$$

We may, without loss of generality, assume that this gives a partition of $\{1, \ldots, n\}$ into disjoint connected sets

$$\{1, \ldots, n\} = \mathcal{A}_0 \cup \mathcal{F}_0 \cup \mathcal{A}_1 \cup \mathcal{F}_1 \cup \cdots \cup \mathcal{A}_{\kappa-1} \cup \mathcal{F}_{\kappa-1}$$

where $\mathcal{A}_\alpha$ is left-adjacent to $\mathcal{F}_\alpha$ for $\alpha \in \{0, \ldots, \kappa-1\}$, $\mathcal{A}_{\alpha+1}$ is right-adjacent to $\mathcal{F}_\alpha$ for $\alpha \in \{0, \ldots, \kappa-2\}$, and $\mathcal{A}_0$ is right-adjacent to $\mathcal{F}_{\kappa-1}$. We may then decompose the operator F as

$$F = \sum_i \bigotimes_{\alpha=0}^{\kappa-1} (I_{\mathcal{A}_\alpha} \otimes F_{i,\alpha}),$$
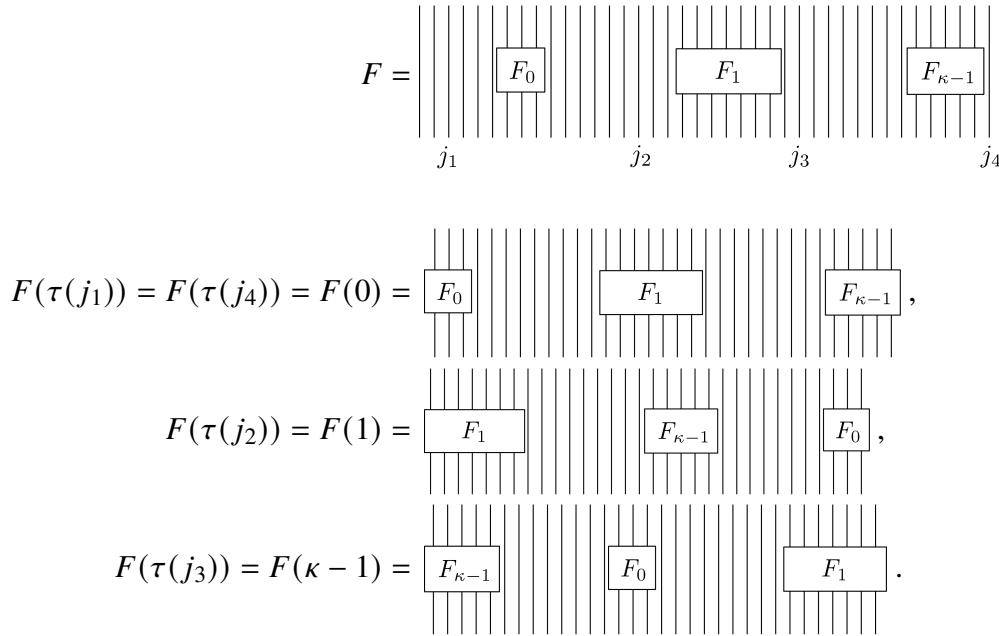
Figure 2.6: Example for F and sites $j_1, j_2, j_3, j_4 \in \{1, \ldots, n\}$ with $\iota(j_1) = \iota(j_4) = 7$, $\iota(j_2) = 19$, and $\iota(j_3) = 35$.

where we write F as a sum of decomposable tensor operators (indexed by i), with each $F_{i,\alpha}$ being an operator acting on the component $\mathcal{F}_\alpha$.

Let us define a function $\tau : \{1, \ldots, n\} \backslash \mathsf{supp}(F) \to \{0, \ldots, \kappa - 1\}$ which associates to every site $j \notin \mathsf{supp}(\mathcal{F})$ the unique index $\tau(j)$ for the component $\mathcal{A}_{\tau(j)}$ of the complement of $\mathsf{supp}(F)$ such that $j \in \mathcal{A}_{\tau(j)}$.

It is also convenient to introduce the following operators $\{F(\tau)\}_{\tau=0}^{\kappa-1}$. The operator $F(\tau)$ is obtained by removing the identity factor on the sites $\mathcal{A}_\tau$ of F, and cyclically permuting the remaining components in such a way that $\mathcal{F}_\tau$ ends up on the sites $\{1, \ldots, |\mathcal{F}_\tau|\}$. More precisely, we define $F(\tau) \in \mathcal{B}((\mathbb{C}^{\mathsf{p}})^{\otimes(n-|\mathcal{A}_\tau|)})$ by

$$F(\tau) = \sum_i F_{i,\tau} \otimes \left(\bigotimes_{\alpha=\tau+1}^{\tau+\kappa-1} I_{\mathbb{C}^{\mathsf{p}}}^{\otimes|\mathcal{A}_{\alpha \ (\mathrm{mod}\ \kappa)}|} \otimes F_{i,\alpha \ (\mathrm{mod}\ \kappa)} \right), \tag{2.57}$$

for $\tau \in \{0, \ldots, \kappa - 1\}$. We note that $j \mapsto F(\tau(j))$ associates a permuted operator to each site j not belonging to the support of F. Let us also define $\iota(j)$ to be the index of the site which gets cyclically shifted to the first site when defining $F_{\tau(j)}$. An example is shown diagrammatically in Figure 2.6.

For two excitation ansatz states $|\Phi_p\rangle$ and $|\Phi_{p'}\rangle$, and an operator F on $(\mathbb{C}^{\mathsf{p}})^{\otimes n}$, we

may write the corresponding matrix element as

$$\langle \Phi_{p'}|F|\Phi_p \rangle = \sum_{j,j'=1}^{n} e^{i(pj-p'j')} \langle \Phi_{j',p'}|F|\Phi_{j,p} \rangle \,, \tag{2.58}$$

where $|\Phi_{j,p}\rangle$ are the "position space" states introduced in Equation (2.6.1). We are interested in bounding the magnitude of this quantity.

We begin by bounding the individual terms in the sum (2.6.4).

Lemma 2.6.5. *Let $j, j' \in \{1, \ldots, n\}$ and let p, p' be arbitrary non-zero momenta. Consider the states $|\Phi_{j,p}\rangle$ and $|\Phi_{j',p'}\rangle$ defined by (2.6.1). Let $\Delta = \Delta(n)$ and $d = d(n)$ be monotonically increasing functions of n. Suppose further that we have*

$$10\Delta d < n \,.$$

Assume F is a d-local operator of unit norm on $(\mathbb{C}^p)^{\otimes n}$ whose support has κ connected components as in (2.6.4). Then we have the following:

(i) *There is some fixed $q \in [n]$ such that for all $j, j' \in \mathcal{B}^{\Delta}(\text{supp}(F))$, we have*

$$\langle \Phi_{j',p'}|F|\Phi_{j,p} \rangle = \langle\!\langle \ell | E_{I^{\otimes\Delta}\otimes F_{\tau(q)}\otimes I^{\otimes\Delta}}(\hat{j}, p, \hat{j}', p')|r \rangle\!\rangle + O(\lambda_2^{\Delta}) \,,$$

where $\hat{j} = j - \iota(q) + \Delta + 1 \pmod n$ and $\hat{j}' = j' - \iota(q) + \Delta + 1 \pmod n$. Furthermore,

$$\langle \Phi_{j',p'}|\Phi_{j,p} \rangle = \Delta_{j,j'} c_{pp'} + O(\lambda_2^{\Delta}) \,. \tag{2.59}$$

(ii) *If $j, j' \notin \mathcal{B}^{\Delta}(\text{supp}(F))$, then*

    (a) $|\langle \Phi_{j',p'}|F|\Phi_{j,p} \rangle| = O(\lambda_2^{\Delta/2})$ *if $j \neq j'$.*

    (b) $\langle \Phi_{j',p'}|F|\Phi_{j,p} \rangle = \langle\!\langle \ell | E_{F(\tau(j))}|r \rangle\!\rangle \cdot c_{pp'} + O(\lambda_2^{\Delta/2})$.
        Here the operator $F(\tau(j))$ is defined by Equation (2.6.4).

(iii) *If $j \in \mathcal{B}^{\Delta}(\text{supp}(F))$ and $j' \notin \mathcal{B}^{\Delta}(\text{supp}(F))$, then*

    (a) $|\langle \Phi_{j',p'}|F|\Phi_{j,p} \rangle| = O(\lambda_2^{\Delta/2})$ *if $j' \notin \mathcal{B}^{2\Delta}(\text{supp}(F))$.*

    (b) *There exists some fixed $q \in [n]$ such that, for all $j \in \mathcal{B}^{\Delta}(\text{supp}(F))$ and $j' \in \mathcal{B}^{2\Delta}(\text{supp}(F))\backslash\mathcal{B}^{\Delta}(\text{supp}(F))$, we have*

$$\langle \Phi_{j',p'}|F|\Phi_{j,p} \rangle = \langle\!\langle \ell | E_{F(\tau(q))}(\hat{j}, p, \hat{j}', p', 2\Delta)|r \rangle\!\rangle + O(\lambda_2^{2\Delta}),$$

        where $\hat{j} = j - \iota(q) + 2\Delta + 1 \pmod n$ and $\hat{j}' = j' - \iota(q) + 2\Delta + 1 \pmod n$.

(iv) If $j' \in \mathcal{B}^\Delta(\text{supp}(F))$ and $j \notin \mathcal{B}^\Delta(\text{supp}(F))$, then

(a) $|\langle \Phi_{j',p'}|F|\Phi_{j,p}\rangle| = O(\lambda_2^{\Delta/2})$ if $j \notin \mathcal{B}^{2\Delta}(\text{supp}(F))$.

(b) There exists some fixed $q \in [n]$ such that, for all $j' \in \mathcal{B}^\Delta(\text{supp}(F))$) and $j \in \mathcal{B}^{2\Delta}(\text{supp}(F)) \backslash \mathcal{B}^\Delta(\text{supp}(F))$, we have

$$\langle \Phi_{j',p'}|F|\Phi_{j,p}\rangle = \langle\!\langle \ell | E_{F(\tau(q))}(\hat{j}, p, \hat{j}', p', 2\Delta)|r\rangle\!\rangle + O(\lambda_2^{2\Delta}),$$

where $\hat{j} = j - \iota(q) + 2\Delta + 1 \pmod{n}$ and $\hat{j}' = j' - \iota(q) + 2\Delta + 1 \pmod{n}$.

Proof. For the proof of (i), suppose that $j, j' \in \mathcal{B}^\Delta(\text{supp}(F))$. Pick any site $q \notin \mathcal{B}^{2\Delta}(\text{supp}(F))$. We note that such a site always exists since

$$|\mathcal{B}^{2\Delta}(\text{supp}(F))| \leq 5\Delta|\text{supp}(F)| = 5\Delta d < 10\Delta d < n$$

by assumption. Let us define the shifted indices

$$\hat{j} = j - \iota(q) + \Delta + 1 \pmod{n}, \quad \text{and} \quad \hat{j}' = j' - \iota(q) + \Delta + 1 \pmod{n}.$$

Then we may write

$$\langle \Phi_{j',p'}|F|\Phi_{j,p}\rangle = \text{tr}\left(E_F(j, p, j', p')\right)$$
$$= \text{tr}\left(E^s E_{I^{\otimes\Delta}\otimes F_{\tau(q)}\otimes I^{\otimes\Delta}}(\hat{j}, p, \hat{j}', p')\right) \tag{2.60}$$

where $s \geq 2\Delta$. This is because by the choice of q, there are at least 2Δ sites not belonging to $\text{supp}(F)$ both on the left and the right of q. Each of these 4Δ sites contributes a factor $E = E_I$ (i.e., a single transfer operator) to the expression within the trace. The term $E_{I^{\otimes\Delta}\otimes F_{\tau(q)}\otimes I^{\otimes\Delta}}(\hat{j}, p, \hat{j}', p')$ incorporates Δ of the associated transfer operators $E = E_I$ on the left- and right of q, respectively, such that at least 2Δ factors of E remain. By the cyclicity of the trace, these can be consolidated into a single term E^s with $s \geq 2\Delta$. The operator $I^{\otimes\Delta} \otimes F_{\tau(q)} \otimes I^{\otimes\Delta}$ (i.e., the additional $I^{\otimes\Delta}$ factors) in the term $E_{I^{\otimes\Delta}\otimes F_{\tau(q)}\otimes I^{\otimes\Delta}}(\hat{j}, p, \hat{j}', p')$ is used to ensure that j and j' are correctly "retained" when going from the first to the second line in (2.6.4). Inserting the Jordan decomposition $E = |r\rangle\!\rangle\langle\!\langle \ell| \oplus \tilde{E}$, we obtain

$$\langle \Phi_{j',p'}|F|\Phi_{j,p}\rangle = \langle\!\langle \ell | E_{I^{\otimes\Delta}\otimes F_{\tau(q)}\otimes I^{\otimes\Delta}}(\hat{j}, p, \hat{j}', p')|r\rangle\!\rangle + \text{tr}\left(\tilde{E}^s E_{I^{\otimes\Delta}\otimes F_{\tau(q)}\otimes I^{\otimes\Delta}}(\hat{j}, p, \hat{j}', p')\right).$$
$$\tag{2.61}$$

By Lemma 2.4.2(ii) and Lemma 2.6.3, we have the bound

$$\left| \text{tr}\left(\tilde{E}^s E_{I^{\otimes \Delta} \otimes F_{\tau(q)} \otimes I^{\otimes \Delta}}(\hat{j}, p, \hat{j}', p') \right) \right| \leq \|\tilde{E}^s\|_F \cdot \|E_{I^{\otimes \Delta} \otimes F_{\tau(q)} \otimes I^{\otimes \Delta}}(\hat{j}, p, \hat{j}', p')\|_F$$

$$\leq \lambda_2^{s/2} \cdot D^2 \|F\| \cdot \sqrt{\|E_{\overline{B(p')}B(p')}\|_F \|E_{\overline{B(p')}B(p)}\|_F}$$

$$= O(\lambda_2^\Delta),$$

where we have used the fact that $\lambda_2^{s/2} \leq \lambda_2^\Delta$ in the last line. We have also absorbed the dependence on the constants D, $\|F\|$, and $\sqrt{\|E_{\overline{B(p')}B(p')}\|_F \|E_{\overline{B(p')}B(p)}\|_F}$ into the big-O notation. Inserting this into (2.6.4) gives the first claim of (i).

Now consider the inner product $\langle \Phi_{j',p'} | \Phi_{j,p} \rangle = \text{tr}(E(j, p, j', p'))$, which corresponds to the case where F is the identity. By the cyclicity of the trace, this can be written as $\langle \Phi_{j',p'} | \Phi_{j,p} \rangle = \text{tr}(E^s E(\hat{j}, p, \hat{j}', p'))$ for some $s \geq 2\Delta$ and suitably defined $\hat{j}, \hat{j}'$. Repeating the same argument as above and using the fact that

$$\langle\!\langle \ell | E(\hat{j}, p, \hat{j}', p') | r \rangle\!\rangle = \Delta_{\hat{j}, \hat{j}'} c_{pp'} = \Delta_{j, j'} c_{pp'}$$

by definition of $E(\hat{j}, p, \hat{j}', p')$, Equation (2.4.1) (i.e., the fact that $|\ell\rangle\!\rangle$ and $|r\rangle\!\rangle$ are left, respectively right, eigenvectors of E), and the gauge identities (2.6.1) of $E_{B(p)}$ and $E_{\overline{B(p)}}$, we obtain the claim (i).

Now consider claim (ii). Suppose that $j, j' \notin \mathcal{B}^\Delta(\text{supp}(F))$. We consider the following two cases:

(iia) If $j \neq j'$, then there is a connected region of at least Δ sites not belonging to $\text{supp}(F)$ to either the left of j' and not containing j, or the left of j and not containing j'. Without loss of generality, we assume the former is the case. By the cyclicity of the trace, we may also assume without loss of generality that $j' = \Delta + 1$, $j > j'$, and that F is supported on the sites $\{2\Delta + 2, \ldots, n\}$. Let $\hat{F}$ denote the restriction of F to the sites $\{\Delta + 2, \ldots, n\}$, and let $\hat{j} := j - (\Delta + 1)$. Then we may write

$$\langle \Phi_{j',p'} | F | \Phi_{j,p} \rangle = \text{tr}\left(E^\Delta E_{\overline{B(p')}} E_{\hat{F}}(\hat{j}, p) \right).$$

Substituting the Jordan decomposition $E^\Delta = |r\rangle\!\rangle\langle\!\langle \ell | \oplus \tilde{E}^\Delta$, we have

$$\langle \Phi_{j',p'} | F | \Phi_{j,p} \rangle = \langle\!\langle \ell | E_{\overline{B(p')}} E_{\hat{F}}(\hat{j}, p) | r \rangle\!\rangle + \text{tr}\left(\tilde{E}^\Delta E_{\overline{B(p')}} E_{\hat{F}}(\hat{j}, p) \right).$$

Since we assume that $p \neq 0$, the gauge condition (2.6.1) states that $\langle\!\langle \ell | E_{\overline{B(p)}} = 0$, hence the first term vanishes and it follows that

$$
\begin{aligned}
|\langle \Phi_{j',p'} | F | \Phi_{j,p} \rangle| &= \left| \mathrm{tr} \left(\tilde{E}^\Delta E_{\overline{B(p')}} E_{\hat{F}}(\hat{j}, p) \right) \right| \\
&\leq \|\tilde{E}^\Delta\|_F \cdot \|E_{\overline{B(p')}}\|_F \cdot \|E_{\hat{F}}(\hat{j}, p)\|_F \\
&\leq \lambda_2^{\Delta/2} \|E_{\overline{B(p')}}\|_F \cdot D^2 \|\hat{F}\| \sqrt{\|E_{\overline{B(p)}B(p)}\|_F} \\
&= O(\lambda_2^{\Delta/2}) \,,
\end{aligned}
$$

as claimed in (iia). In the last line, we have again absorbed the constants into the big-O-expression. This proves part (iia) of Claim (ii).

(iib) If $j = j'$, then there are at least Δ sites to the left and right of j which do not belong to $\mathsf{supp}(F)$. Therefore we may write

$$
\langle \Phi_{j',p'} | F | \Phi_{j,p} \rangle = \mathrm{tr} \left(E^s E_{\overline{B(p')}B(p)} E^t E_{F(\tau(j))} \right) \,,
$$

where s and t are integers greater than Δ, representing the sites surrounding j which are not in the support of F.

Applying the Jordan decomposition $E^\Delta = |r\rangle\!\rangle\langle\!\langle \ell| \oplus \tilde{E}^\Delta$ twice (for E^s and E^t) then gives four terms

$$
\begin{aligned}
\langle \Phi_{j',p'} | F | \Phi_{j,p} \rangle &= \langle\!\langle \ell | E_{\overline{B(p')}B(p)} | r \rangle\!\rangle \langle\!\langle \ell | E_{F(\tau(j))} | r \rangle\!\rangle \\
&\quad + \mathrm{tr} \left(|r\rangle\!\rangle \langle\!\langle \ell | E_{\overline{B(p')}B(p)} \tilde{E}^s E_{F(\tau(j))} \right) \\
&\quad + \mathrm{tr} \left(\tilde{E}^t E_{\overline{B(p')}B(p)} |r\rangle\!\rangle \langle\!\langle \ell | E_{F(\tau(j))} \right) \\
&\quad + \mathrm{tr} \left(\tilde{E}^t E_{\overline{B(p')}B(p)} \tilde{E}^s E_{F(\tau(j))} \right) \,.
\end{aligned}
$$

Since s and t are both larger than Δ, by the same arguments from before, it is clear that the last three terms can each be bounded by $O(\lambda_2^{\Delta/2})$. The claim follows since $\langle\!\langle \ell | E_{\overline{B(p')}B(p)} | r \rangle\!\rangle = c_{pp'}$.

Next, we give the proof of claim (iii). Let us consider the situation where $j \in \mathcal{B}^\Delta(\mathsf{supp}(F))$ and $j' \notin \mathcal{B}^\Delta(\mathsf{supp}(F))$. The proof of the other setting is analogous. We consider two cases:

(iiia) Suppose $j' \notin \mathcal{B}^{2\Delta}(\mathsf{supp}(F))$. Let us define the shifted index $\hat{j} = j - \iota(j') + \Delta + 1 \pmod{n}$. Then we may write

$$
\langle \Phi_{j',p'} | F | \Phi_{j,p} \rangle = \mathrm{tr} \left(E^s E_{\overline{B(p')}} E^t E_{I^{\otimes\Delta} \otimes F(\tau(j')) \otimes I^{\otimes\Delta}}(\hat{j}, p) \right),
$$

where s and t are integers larger than Δ, representing the number of sites adjacent to j' on the left and right which are not in $\mathcal{B}^\Delta(\mathrm{supp}(F))$. We use the Jordan decomposition $E = |r\rangle\!\rangle\langle\!\langle \ell| \oplus \tilde{E}$ on E^s to get

$$\mathrm{tr}\left(E^s E_{\overline{B(p')}}E^t E_{I^{\otimes\Delta}\otimes F(\tau(j'))\otimes I^{\otimes\Delta}}(\hat{j}, p)\right) = \langle\!\langle \ell|E_{\overline{B(p')}}E^t E_{I^{\otimes\Delta}\otimes F(\tau(j'))\otimes I^{\otimes\Delta}}(\hat{j}, p)|r\rangle\!\rangle$$

$$+ \mathrm{tr}\left(\tilde{E}^s E_{\overline{B(p')}}E^t E_{I^{\otimes\Delta}\otimes F(\tau(j'))\otimes I^{\otimes\Delta}}(\hat{j}, p)\right)$$

$$= \mathrm{tr}\left(\tilde{E}^s E_{\overline{B(p')}}E^t E_{I^{\otimes\Delta}\otimes F(\tau(j'))\otimes I^{\otimes\Delta}}(\hat{j}, p)\right),$$

where the first term vanishes due to the gauge condition (2.6.1). From Lemma 2.4.2(ii) we have $\|E^t\|_F \leq 1$, and repeating the same arguments as before, we get the bound

$$|\langle\Phi_{j',p'}|F|\Phi_{j,p}\rangle| = \left|\mathrm{tr}\left(\tilde{E}^s E_{\overline{B(p')}}E^t E_{I^{\otimes\Delta}\otimes F(\tau(j'))\otimes I^{\otimes\Delta}}(\hat{j}, p)\right)\right|$$

$$\leq \|\tilde{E}^s\|_F \cdot \|E_{\overline{B(p')}}\|_F \cdot \|E^t\|_F \cdot D^2\|F\| \cdot \|E_{\overline{B(p)}B(p)}\|_F$$

$$\leq \lambda_2^{s/2}\|E_{\overline{B(p')}}\|_F \cdot D^2\|F\|\sqrt{\|E_{\overline{B(p)}B(p)}\|_F} \; .$$

Since $s \geq \Delta$, we conclude that

$$|\langle\Phi_{j',p'}|F|\Phi_{j,p}\rangle| = O\left(\lambda_2^{\Delta/2}\right) \; .$$

(iiib) Suppose now that $j' \in \mathcal{B}^{2\Delta}(\mathrm{supp}(F))$. Then by repeating the argument for case (i), with Δ replaced by 2Δ, we obtain

$$\langle\Phi_{j',p'}|F|\Phi_{j,p}\rangle = \langle\!\langle \ell|E_{I^{\otimes 2\Delta}\otimes F(\tau(q))\otimes I^{\otimes 2\Delta}}(\hat{j}, p, \hat{j}', p')|r\rangle\!\rangle + O(\lambda_2^{2\Delta}),$$

where we now have $q \notin \mathcal{B}^{4\Delta}(\mathcal{F})$. Again, the existence of such a q is guaranteed by the condition $10\Delta d < n$.

We note that (iv) follows immediately from (iii) by interchanging the roles of (j, p) and (j', p'). Note that we can write

$$\langle\Phi_{j',p'}|F|\Phi_{j,p}\rangle = \overline{\overline{\langle\Phi_{j',p'}|F|\Phi_{j,p}\rangle}} = \overline{\langle\Phi_{j,p}|F^\dagger|\Phi_{j',p'}\rangle} \; .$$

The last expression within the parentheses is precisely what we had calculated in (iii), so this implies the following:

(iva) If $j \notin \mathcal{B}^{2\Delta}(\text{supp}(F))$, then

$$\left|\langle \Phi_{j',p'}|F|\Phi_{j,p}\rangle\right| = \left|\langle \Phi_{j,p}|F^{\dagger}|\Phi_{j',p'}\rangle\right| = O(\lambda_2^{\Delta/2}),$$

where we note that the exact same bound holds for F and $F^{\dagger}$ since $\|F\| = \|F^{\dagger}\|$.

(ivb) If $j \in \mathcal{B}^{2\Delta}(\text{supp}(F))$, then

$$\begin{aligned}
\langle \Phi_{j',p'}|F|\Phi_{j,p}\rangle &= \overline{\langle \Phi_{j,p}|F^{\dagger}|\Phi_{j',p'}\rangle} \\
&= \overline{\langle\!\langle \ell|E_{I^{\otimes 2\Delta}\otimes F^{\dagger}(\tau(q))\otimes I^{\otimes 2\Delta}}(\hat{j}', p', \hat{j}, p)|r\rangle\!\rangle} + O(\lambda_2^{2\Delta}) \\
&= \overline{\langle\!\langle \ell|E_{I^{\otimes 2\Delta}\otimes F^{\dagger}(\tau(q))\otimes I^{\otimes 2\Delta}}(\hat{j}', p', \hat{j}, p)|r\rangle\!\rangle} + O(\lambda_2^{2\Delta}) \\
&= \langle\!\langle \ell|E_{I^{\otimes 2\Delta}\otimes F(\tau(q))\otimes I^{\otimes 2\Delta}}(\hat{j}, p, \hat{j}', p')|r\rangle\!\rangle + O(\lambda_2^{2\Delta}).
\end{aligned}$$

This proves the claim.[8]

$\square$

Note that in the statement (iib), the dependence on j in the expression $\langle\!\langle \ell|E_{F(\tau(j))}|r\rangle\!\rangle$ can be eliminated as follows:

Lemma 2.6.6. *Suppose $j_1, j_2 \notin \mathcal{B}^{\Delta}(\text{supp}(F))$. Then*

$$\left|\langle\!\langle \ell|E_{F(\tau(j_1))}|r\rangle\!\rangle - \langle\!\langle \ell|E_{F(\tau(j_2))}|r\rangle\!\rangle\right| = O(\lambda_2^{\Delta}). \tag{2.62}$$

In particular, for any fixed $j_0 \notin \mathcal{B}^{\Delta}(\text{supp}(F))$ we have

$$\langle \Phi_{j,p'}|F|\Phi_{j,p}\rangle = \langle\!\langle \ell|E_{F(\tau(j_0))}|r\rangle\!\rangle \cdot c_{pp'} + O(\lambda_2^{\Delta/2}), \tag{2.63}$$

for all $j \notin \mathcal{B}^{\Delta}(\text{supp}(F))$.

---

[8]To clarify how the term $\langle\!\langle \ell|E_{I^{\otimes\Delta}\otimes F^{\dagger}(\tau(q))\otimes I^{\otimes\Delta}}(\hat{j}', p', \hat{j}, p)|r\rangle\!\rangle$ is complex conjugated, first write

$$\langle\!\langle \ell|E_{I^{\otimes\Delta}\otimes F^{\dagger}(\tau(q))\otimes I^{\otimes\Delta}}(\hat{j}', p', \hat{j}, p)|r\rangle\!\rangle = \langle \Phi_{\hat{j}',p'}^L|I \otimes I^{\otimes 2\Delta} \otimes F_{\tau(q)}^{\dagger} \otimes I^{\otimes 2\Delta} \otimes I|\Phi_{\hat{j},p}^L\rangle,$$

where $|\Phi_{\hat{j},p}^L\rangle$ are the states defined by (2.6.3), for some appropriate length L. Then we can proceed to conjugate the matrix element, giving us

$$\begin{aligned}
\overline{\langle \Phi_{\hat{j}',p'}^L|I \otimes I^{\otimes 2\Delta} \otimes F_{\tau(q)}^{\dagger} \otimes I^{\otimes 2\Delta} \otimes I|\Phi_{\hat{j},p}^L\rangle} &= \langle \Phi_{\hat{j},p}^L|I \otimes I^{2\Delta} \otimes F_{\tau(q)} \otimes I^{2\Delta} \otimes I|\Phi_{\hat{j}',p'}^L\rangle \\
&= \langle\!\langle \ell|E_{I^{\otimes 2\Delta}\otimes F(\tau(q))\otimes I^{\otimes 2\Delta}}(\hat{j}, p, \hat{j}', p')|r\rangle\!\rangle.
\end{aligned}$$

Proof. The claim (2.6.6) follows immediately from (2.6.6) and claim (iib) of Lemma 2.6.5 since $|c_{pp'}| = O(1)$.

If $\tau(j_1) = \tau(j_2)$, there is nothing to prove. Suppose $\tau(j_1) \neq \tau(j_2)$. Without loss of generality, assume that $\tau(j_1) = 0$ and $\tau(j_2) = \xi$. Then we may write

$$F(\tau(j_1)) = \sum_i F_{i,0} \otimes I^{\otimes a_1} \otimes F_{i,1} \otimes I^{\otimes a_2} \cdots \otimes I^{\otimes a_{\kappa-1}} \otimes F_{i,\kappa-1}, \quad \text{and}$$

$$F(\tau(j_2)) = \sum_i F_{i,\xi} \otimes I^{\otimes a_{\xi+1}} \otimes F_{i,\xi+1} \otimes I^{\otimes a_{\xi+2}} \cdots \otimes I^{\otimes a_{\kappa}} \otimes F_{i,\kappa-1} \otimes I^{\otimes a_0}$$
$$\otimes F_{i,0} \otimes I^{\otimes a_1} \otimes F_{i,1} \otimes I^{\otimes a_2} \otimes \cdots \otimes F_{i,\xi-1},$$

where $a_\alpha = |\mathcal{A}_\alpha|$ for $\alpha \in \{0, \dots, \kappa\}$. Defining the operators

$$\hat{F}_i = F_{i,\xi} \otimes I^{\otimes a_{\xi+1}} \otimes F_{i,\xi+1} \otimes I^{\otimes a_{\xi+2}} \cdots \otimes I^{\otimes a_{\kappa-1}} \otimes F_{i,\kappa-1},$$
$$\hat{G}_i = F_{i,0} \otimes I^{\otimes a_1} \otimes F_{i,1} \otimes I^{\otimes a_2} \otimes \cdots \otimes F_{i,\xi-1},$$

we have

$$F(\tau(j_1)) = \sum_i \hat{G}_i \otimes I^{\otimes a_\xi} \otimes \hat{F}_i, \quad \text{and} \quad F(\tau(j_2)) = \sum_i \hat{F}_i \otimes I^{\otimes a_0} \otimes \hat{G}_i.$$

(We give an example for the operator F, $F(\tau(j_1))$, and $F(\tau(j_2))$ in Figure 2.7.) Therefore we can write

$$\langle\!\langle \ell | E_{F(\tau(j_1))} | r \rangle\!\rangle = \sum_i \langle\!\langle \ell | E_{\hat{G}_i} E^{a_\xi} E_{\hat{F}_i} | r \rangle\!\rangle,$$
$$\langle\!\langle \ell | E_{F(\tau(j_2))} | r \rangle\!\rangle = \sum_i \langle\!\langle \ell | E_{\hat{F}_i} E^{a_0} E_{\hat{G}_i} | r \rangle\!\rangle.$$

Inserting the Jordan decomposition $E = |r\rangle\!\rangle\langle\!\langle \ell| \oplus \tilde{E}$ gives

$$\langle\!\langle \ell | E_{F(\tau(j_1))} | r \rangle\!\rangle = \sum_i \left(\langle\!\langle \ell | E_{\hat{G}_i} | r \rangle\!\rangle \langle\!\langle \ell | E_{\hat{F}_i} | r \rangle\!\rangle + \langle\!\langle \ell | E_{\hat{G}_i} \tilde{E}^{a_\xi} E_{\hat{F}_i} | r \rangle\!\rangle \right),$$
$$\langle\!\langle \ell | E_{F(\tau(j_2))} | r \rangle\!\rangle = \sum_i \left(\langle\!\langle \ell | E_{\hat{F}_i} | r \rangle\!\rangle \langle\!\langle \ell | E_{\hat{G}_i} | r \rangle\!\rangle + \langle\!\langle \ell | E_{\hat{F}_i} \tilde{E}^{a_0} E_{\hat{G}_i} | r \rangle\!\rangle \right).$$

Taking the difference, the first terms of the sums cancel, and we are left with

$$\left| \langle\!\langle \ell | E_{F(\tau(j_1))} | r \rangle\!\rangle - \langle\!\langle \ell | E_{F(\tau(j_2))} | r \rangle\!\rangle \right| = \left| \sum_i \langle\!\langle \ell | E_{\hat{G}_i} \tilde{E}^{a_\xi} E_{\hat{F}_i} | r \rangle\!\rangle - \sum_i \langle\!\langle \ell | E_{\hat{F}_i} \tilde{E}^{a_0} E_{\hat{G}_i} | r \rangle\!\rangle \right|$$
$$\leq \left| \sum_i \langle\!\langle \ell | E_{\hat{G}_i} \tilde{E}^{a_\xi} E_{\hat{F}_i} | r \rangle\!\rangle \right| + \left| \sum_i \langle\!\langle \ell | E_{\hat{F}_i} \tilde{E}^{a_0} E_{\hat{G}_i} | r \rangle\!\rangle \right|.$$

$$(2.64)$$
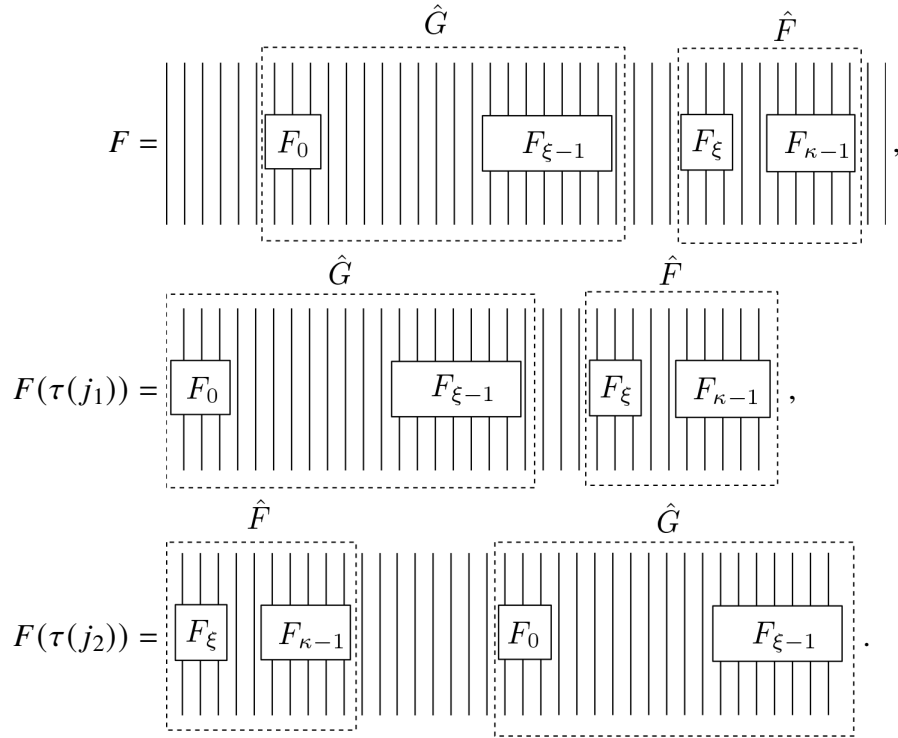
Figure 2.7: Example for the operator F and the corresponding $F(\tau(j_1))$ and $F(\tau(j_2))$.

We can bound the first term $\left| \sum_i \langle\!\langle \ell | E_{\hat{G}_i} \tilde{E}^{a_\xi} E_{\hat{F}_i} | r \rangle\!\rangle \right|$ as follows. First, we write

$$\left| \sum_i \langle\!\langle \ell | E_{\hat{G}_i} \tilde{E}^{a_\xi} E_{\hat{F}_i} | r \rangle\!\rangle \right| = \mathrm{tr}\left(\tilde{E}^{a_\xi} \sum_i E_{\hat{F}_i} | r \rangle\!\rangle \langle\!\langle \ell | E_{\hat{G}_i} \right)$$

$$\leq \| \tilde{E}^{a_\xi} \|_F \left\| \sum_i E_{\hat{F}_i} | r \rangle\!\rangle \langle\!\langle \ell | E_{\hat{G}_i} \right\|_F$$

$$\leq \lambda_2^\Delta \left\| \sum_i E_{\hat{F}_i} | r \rangle\!\rangle \langle\!\langle \ell | E_{\hat{G}_i} \right\|_F ,$$

where the last inequality comes from the fact that $j_2 \notin \mathcal{B}^\Delta(\mathrm{supp}(F))$ and $j_2 \in \mathcal{A}_\xi$ implies that $a_\xi \geq 2\Delta$, so Lemma 2.4.2(ii) gives $\| \tilde{E}^{a_\xi} \|_F \leq \lambda_2^\Delta$. Proceeding as we did in the proof of Lemma 2.6.3, we can write the latter Frobenius norm as

$$\left\| \sum_i E_{\hat{F}_i} | r \rangle\!\rangle \langle\!\langle \ell | E_{\hat{G}_i} \right\|_F^2 = \sum_{\alpha_1,\alpha_2,\beta_1,\beta_2=1}^D \left| \langle \alpha_1 | \langle \alpha_2 | \left(\sum_i E_{\hat{F}_i} | r \rangle\!\rangle \langle\!\langle \ell | E_{\hat{G}_i} \right) | \beta_1 \rangle | \beta_2 \rangle \right|^2 .$$

The individual terms in the sum can be depicted diagrammatically as

$$\langle\alpha_1|\langle\alpha_2| \left(\sum_i E_{\hat{F}_i}|r\rangle\!\rangle\langle\!\langle \ell|E_{\hat{G}_i} \right) |\beta_1\rangle|\beta_2\rangle = \sum_i \quad \cdots$$

Defining the vectors

$$|\Psi(\alpha,\beta)\rangle = \quad \cdots$$

we can then write

$$\langle\alpha_1|\langle\alpha_2| \left(\sum_i E_{\hat{F}_i}|r\rangle\!\rangle\langle\!\langle \ell|E_{\hat{G}_i} \right) |\beta_1\rangle|\beta_2\rangle = \langle\Psi(\alpha_1,\beta_1)| \left(\sum_i \hat{F}_i \otimes I_D \otimes I_D \otimes \hat{G}_i \right) |\Psi(\alpha_2,\beta_2)\rangle.$$

Applying the Cauchy-Schwarz inequality, we get

$$\left| \langle\alpha_1|\langle\alpha_2| \left(\sum_i E_{\hat{F}_i}|r\rangle\!\rangle\langle\!\langle \ell|E_{\hat{G}_i} \right) |\beta_1\rangle|\beta_2\rangle \right|^2 \leq \|\Psi(\alpha_1,\beta_1)\|^2 \cdot \|\Psi(\alpha_2,\beta_2)\|^2 \cdot \left\| \sum_i \hat{F}_i \otimes I_D \otimes I_D \otimes \hat{G}_i \right\|^2.$$

The norm of the vector $|\Psi(\alpha,\beta)\rangle$ is given by

$$\|\Psi(\alpha,\beta)\|^2 = \quad \cdots = \quad \cdots = \langle\alpha|r|\alpha\rangle\langle\beta|\ell|\beta\rangle ,$$

where in the second equality, we have used the fixed-point equations (2.4.1). Therefore we have

$$\left\| \sum_i E_{\hat{F}_i}|r\rangle\!\rangle\langle\!\langle \ell|E_{\hat{G}_i} \right\|_F^2 \leq \left\| \sum_i \hat{F}_i \otimes I_D \otimes I_D \otimes \hat{G}_i \right\|^2 \sum_{\alpha_1,\alpha_2,\beta_1,\beta_2=1}^{D} \langle\alpha_1|r|\alpha_1\rangle\langle\alpha_2|r|\alpha_2\rangle\langle\beta_1|\ell|\beta_1\rangle\langle\beta_2|\ell|\beta_2\rangle$$

$$= \left\| \sum_i \hat{F}_i \otimes I_D \otimes I_D \otimes \hat{G}_i \right\|^2 \cdot |\operatorname{tr}(r)\operatorname{tr}(\ell)|^2 = D^2 \left\| \sum_i \hat{F}_i \otimes I_D \otimes I_D \otimes \hat{G}_i \right\|^2,$$

where the last equality follows from the fact that we gauge-fix the left and right fixed-points such that $r = I_{\mathbb{C}^D}$ and $\operatorname{tr}(\ell) = 1$. Finally, we note that since the operator

norm is multiplicative over tensor products, i.e., $\|A \otimes B\| = \|A\| \cdot \|B\|$, we have

$$\left\| \sum_i \hat{F}_i \otimes I_D \otimes I_D \otimes \hat{G}_i \right\| = \left\| \sum_i \hat{F}_i \otimes \hat{G}_i \right\| = \|F\|.$$

Therefore, we have

$$\left| \sum_i \langle\!\langle \ell | E_{\hat{G}_i} \tilde{E}^{a_\xi} E_{\hat{F}_i} | r \rangle\!\rangle \right| \le D\|F\| \lambda_2^\Delta.$$

The term involving a_0 in (2.6.4) can be bounded identically, and so

$$\left| \langle\!\langle \ell | E_{F(\tau(j_1))} | r \rangle\!\rangle - \langle\!\langle \ell | E_{F(\tau(j_2))} | r \rangle\!\rangle \right| \le 2D\|F\| \lambda_2^\Delta ,$$

which proves (2.6.6). $\qquad \square$

We also need a different version of statement (i), as well as statements (iiib) and (ivb) derived from it.

Lemma 2.6.7. *For $\Omega \subset [n]^2$, let us define*

$$\sigma_{pp'}(\Omega) = \sum_{(j,j') \in \Omega} e^{i(pj - p'j')} \langle \Phi_{j',p'} | F | \Phi_{j,p} \rangle .$$

Let us write $\mathcal{F} := \mathsf{supp}(F)$ and $\mathcal{A}^c = [n] \backslash \mathcal{A}$ for the complement of a subset $\mathcal{A} \subset [n]$. Then:

$$\left| \sigma_{pp'}(\mathcal{B}^\Delta(\mathcal{F}) \times \mathcal{B}^\Delta(\mathcal{F})) \right| \le |\mathcal{B}^\Delta(\mathcal{F})| \cdot \|F\| \sqrt{c_p c_{p'}} + O\left(\sqrt{n} \, \lambda_2^{\Delta/2} \right) , \quad (2.65)$$

$$\left| \sigma_{pp'}(\mathcal{B}^\Delta(\mathcal{F}) \times \mathcal{B}^\Delta(\mathcal{F})^c) \right| \le |\mathcal{B}^{2\Delta}(\mathcal{F})| \cdot \|F\| \sqrt{c_p c_{p'}} + O\left(n^2 \lambda_2^{\Delta/2} \right) , \quad (2.66)$$

$$\left| \sigma_{pp'}(\mathcal{B}^\Delta(\mathcal{F})^c \times \mathcal{B}^\Delta(\mathcal{F})) \right| \le |\mathcal{B}^{2\Delta}(\mathcal{F})| \cdot \|F\| \sqrt{c_p c_{p'}} + O\left(n^2 \lambda_2^{\Delta/2} \right) . \quad (2.67)$$

Finally, we have the following: There exists some fixed $j_0 \in [n]$ such that for $p = p'$, we have

$$\sigma_{pp}(\mathcal{B}^\Delta(\mathcal{F})^c \times \mathcal{B}^\Delta(\mathcal{F})^c) = |\mathcal{B}^\Delta(\mathcal{F})^c| \cdot \langle\!\langle \ell | E_{F_{\tau(j_0)}} | r \rangle\!\rangle c_p + O\left(n^2 \lambda_2^{\Delta/2} \right) . \quad (2.68)$$

For $p \ne p'$, we have

$$\left| \sigma_{pp'}(\mathcal{B}^\Delta(\mathcal{F})^c \times \mathcal{B}^\Delta(\mathcal{F})^c) \right| \le |\mathcal{B}^\Delta(\mathcal{F})| \cdot \|F\| \sqrt{c_p c_{p'}} + O\left(n^2 \lambda_2^{\Delta/2} \right) . \quad (2.69)$$

We observe that the first expression on the right-hand side of the above bound scales linearly with the support size of $\mathcal{F}$ instead of the support size of $\mathcal{F}^c$, as may be naively expected. For (2.6.7), this is due to a cancellation of phases, see (2.6.4) below.

Proof. For the proof of (2.6.7), let us first define the vectors

$$|\Psi(p)\rangle = \sum_{j \in \mathcal{B}^\Delta(\mathcal{F})} e^{ipj} |\Phi_{j,p}\rangle \,.$$

Then we can write

$$|\sigma_{pp'}(\mathcal{B}^\Delta(\mathcal{F}) \times \mathcal{B}^\Delta(\mathcal{F}))| = |\langle\Psi(p')|F|\Psi(p)\rangle|$$
$$\leq \|F\| \cdot \|\Psi(p)\| \cdot \|\Psi(p')\| \,, \qquad (2.70)$$

where the last inequality follows by Cauchy-Schwarz along with the definition of the operator norm $\|F\|$. The vector norm is given by

$$\|\Psi(p)\|^2 = \sum_{j,j' \in \mathcal{B}^\Delta(\mathcal{F})} e^{ip(j-j')} \langle\Phi_{j',p}|\Phi_{j,p}\rangle,$$

and together with Equation (i), we get

$$\|\Psi(p)\|^2 = |\mathcal{B}^\Delta(\mathcal{F})| \cdot c_p + O(\lambda_2^{\Delta/2}) \,.$$

Taking the square root and inserting into Equation (2.6.4), we get

$$|\sigma_{pp'}(\mathcal{B}^\Delta(\mathcal{F}) \times \mathcal{B}^\Delta(\mathcal{F}))| = \|F\| \left(\sqrt{|\mathcal{B}^\Delta(\mathcal{F})| \cdot c_p} + O(\lambda_2^{\Delta/2}) \right) \left(\sqrt{|\mathcal{B}^\Delta(\mathcal{F})| \cdot c_{p'}} + O(\lambda_2^{\Delta/2}) \right)$$

$$= |\mathcal{B}^\Delta(\mathcal{F})| \cdot \|F\| \sqrt{c_p c_{p'}} + O\left(\sqrt{|\mathcal{B}^\Delta(\mathcal{F})|} \cdot \lambda_2^{\Delta/2} \right).$$

Using the bound $\left|\mathcal{B}^\Delta(\mathcal{F})\right| \leq 5d\Delta < n$ gives (2.6.7).

Next, let us look at (2.6.7). We have

$$\sigma_{pp'}(\mathcal{B}^\Delta(\mathcal{F}) \times \mathcal{B}^\Delta(\mathcal{F})^c) = \sum_{j \in \mathcal{B}^\Delta(\mathcal{F})} \sum_{j' \in \mathcal{B}^\Delta(\mathcal{F})^c} e^{i(pj-p'j')} \langle\Phi_{j',p'}|F|\Phi_{j,p}\rangle = \Sigma_1 + \Sigma_2,$$

where we define

$$\Sigma_1 := \sum_{j \in \mathcal{B}^\Delta(\mathcal{F})} \sum_{j' \in \mathcal{B}^{2\Delta}(\mathcal{F}) \backslash \mathcal{B}^\Delta(\mathcal{F})} e^{i(pj-p'j')} \langle\Phi_{j',p'}|F|\Phi_{j,p}\rangle \,,$$

and

$$\Sigma_2 := \sum_{j \in \mathcal{B}^\Delta(\mathcal{F})} \sum_{j' \in \mathcal{B}^{2\Delta}(\mathcal{F})^c} e^{i(pj-p'j')} \langle\Phi_{j',p'}|F|\Phi_{j,p}\rangle \,.$$

The norm of the second sum can be bounded using Lemma 2.6.5(iiia), giving us

$$
\begin{aligned}
|\Sigma_2| &\leq \sum_{j \in \mathcal{B}^\Delta(\mathcal{F})} \sum_{j' \in \mathcal{B}^{2\Delta}(\mathcal{F})^c} |\langle \Phi_{j',p'}|F|\Phi_{j,p}\rangle| \\
&\leq |\mathcal{B}^\Delta(\mathcal{F})| \cdot |\mathcal{B}^{2\Delta}(\mathcal{F})^c| \cdot O(\lambda_2^{\Delta/2}) \\
&= O(n^2 \lambda_2^{\Delta/2}) \,,
\end{aligned} \tag{2.71}
$$

where we again use the trivial bound $\left|\mathcal{B}^\Delta(\mathcal{F})\right|, \left|\mathcal{B}^{2\Delta}(\mathcal{F})^c\right| \leq n$ in the last line. Using Lemma 2.6.5 (iiib), we can express the first sum, with some fixed $q \in [n]$, as

$$
\begin{aligned}
\Sigma_1 &= \sum_{j \in \mathcal{B}^\Delta(\mathcal{F})} \sum_{j' \in \mathcal{B}^{2\Delta}(\mathcal{F}) \setminus \mathcal{B}^\Delta(\mathcal{F})} e^{i(pj - p'j')} \langle\!\langle \ell | E_{F(\tau(q))}(\hat{j}, p, \hat{j}', p') | r \rangle\!\rangle \\
&\quad + |\mathcal{B}^\Delta(\mathcal{F})| \cdot |\mathcal{B}^{2\Delta}(\mathcal{F}) \setminus \mathcal{B}^\Delta(\mathcal{F})| \cdot O(\lambda_2^{\Delta/2}) \\
&= \sum_{j \in \mathcal{B}^\Delta(\mathcal{F})} \sum_{j' \in \mathcal{B}^{2\Delta}(\mathcal{F}) \setminus \mathcal{B}^\Delta(\mathcal{F})} e^{i(pj - p'j')} \langle\!\langle \ell | E_{F(\tau(q))}(\hat{j}, p, \hat{j}', p') | r \rangle\!\rangle + O\left(n^2 \lambda_2^{\Delta/2}\right) \,,
\end{aligned}
$$

where the indices $\hat{j}$ and $\hat{j}'$ are defined as in Lemma 2.6.5. To bound the remaining sum, let us introduce the states

$$
\begin{aligned}
|\Psi_1(p)\rangle &:= \sum_{j \in \mathcal{B}^\Delta(\mathcal{F})} e^{ipj} |\Phi_{\hat{j},p}^L\rangle \,, \qquad \text{and} \\
|\Psi_2(p')\rangle &:= \sum_{j' \in \mathcal{B}^{2\Delta}(\mathcal{F}) \setminus \mathcal{B}^\Delta(\mathcal{F})} e^{ip'j'} |\Phi_{\hat{j}',p'}^L\rangle \,,
\end{aligned}
$$

where we set $L = |\mathsf{supp}(F(\tau(q)))|$. Here, $|\Phi_{j,p}^L\rangle$ are as defined in (2.6.3). Then we can write

$$
\Sigma_1 = \langle \Psi_2(p')|F(\tau(q))|\Psi_1(p)\rangle + O\left(n^2 \lambda_2^{\Delta/2}\right) \,.
$$

By the Cauchy-Schwarz inequality and the orthogonality relations (2.6.4), we have

$$
\begin{aligned}
|\langle \Psi_2(p')|F(\tau(q))|\Psi_1(p)\rangle| &\leq \|F\| \cdot \|\Psi_1(p)\| \cdot \|\Psi_2(p')\| \\
&= \|F\| \sqrt{c_p c_{p'} |\mathcal{B}^\Delta(\mathcal{F})| \cdot |\mathcal{B}^{2\Delta}(\mathcal{F}) \setminus \mathcal{B}^\Delta(\mathcal{F})|} \,,
\end{aligned}
$$

where we bound the states $|\Psi_{1,2}(p)\rangle$ in exactly the same way as we did in the proof of (2.6.7). Using the fact that $|\mathcal{B}^\Delta(\mathcal{F})|, |\mathcal{B}^{2\Delta}(\mathcal{F}) \setminus \mathcal{B}^\Delta(\mathcal{F})| \leq |\mathcal{B}^{2\Delta}(\mathcal{F})|$, we conclude that

$$
|\Sigma_1| \leq |\mathcal{B}^{2\Delta}(\mathcal{F})| \cdot \|F\| \sqrt{c_p c_{p'}} + O\left(n^2 \lambda_2^{\Delta/2}\right) \,.
$$

Combining this with (2.6.4) gives the claim (2.6.7). The proof of (2.6.7) is analogous, using Lemma 2.6.5(iv).

Finally, consider (2.6.7) and (2.6.7). We have

$$\sigma_{pp'}(\mathcal{B}^\Delta(\mathcal{F})^c \times \mathcal{B}^\Delta(\mathcal{F})^c) = \underbrace{\sum_{j \in \mathcal{B}^\Delta(\mathcal{F})^c} e^{ij(p-p')}\langle \Phi_{j,p'}|F|\Phi_{j,p}\rangle}_{=:\Theta_1} + \underbrace{\sum_{\substack{j,j' \in \mathcal{B}^\Delta(\mathcal{F})^c \\ j \neq j'}} e^{i(pj-p'j')}\langle \Phi_{j',p'}|F|\Phi_{j,p}\rangle}_{=:\Theta_2}.$$

Using Lemma 2.6.5(iia), we have

$$|\Theta_2| \leq \|F\| \cdot O(n^2 \lambda_2^{\Delta/2}). \tag{2.72}$$

On the other hand, by Lemma 2.6.5(iib), or more precisely its refinement in the form of Equation (2.6.6) from Lemma 2.6.6, we have

$$\Theta_1 = \left(\sum_{j \in \mathcal{B}^\Delta(\mathcal{F})^c} e^{ij(p-p')} \right) \langle\!\langle \ell | E_{F(\tau(j_0))} | r \rangle\!\rangle c_{pp'} + O(n\lambda_2^{\Delta/2})$$

for some fixed $j_0 \in \mathcal{B}^\Delta(\mathcal{F})^c$. For $p' = p$, the sum above is given trivially by $\sum_{j \in \mathcal{B}^\Delta(\mathcal{F})^c} 1 = |\mathcal{B}^\Delta(\mathcal{F})^c|$. For $p \neq p'$, we have $\sum_{j \in [n]} e^{ij(p-p')} = 0$, and hence

$$\left| \sum_{j \in \mathcal{B}^\Delta(\mathcal{F})^c} e^{ij(p-p')} \right| = \left| \sum_{j \in \mathcal{B}^\Delta(\mathcal{F})} e^{ij(p-p')} \right| \leq |\mathcal{B}^\Delta(\mathcal{F})|. \tag{2.73}$$

Therefore, for $p = p'$, we have

$$\Theta_1 = |\mathcal{B}^\Delta(\mathcal{F})^c| \langle\!\langle \ell | E_{F(\tau(j_0))} | r \rangle\!\rangle c_p + O(n\lambda_2^{\Delta/2}),$$

and for $p \neq p'$, we have

$$|\Theta_1| \leq |\mathcal{B}^\Delta(\mathcal{F})| \langle\!\langle \ell | E_{F(\tau(j_0))} | r \rangle\!\rangle c_{pp'} + O(n\lambda_2^{\Delta/2})$$
$$\leq |\mathcal{B}^\Delta(\mathcal{F})| \cdot \|F\| c_{pp'} + O(n\lambda_2^{\Delta/2}).$$

Note that we also have $c_{pp'} \leq \sqrt{c_p c_{p'}}$ by the Cauchy-Schwarz inequality. Combining these results with (2.6.4) proves claims (2.6.7) and (2.6.7). $\qquad\square$

2.6.5 The Parameters of Codes Based on the Excitation Ansatz

Recall that the normalization of the excitation ansatz states $|\Phi_p\rangle \equiv |\Phi_p(B; A)\rangle$ are given by Lemma 2.6.2 as

$$\|\Phi_p\| = \sqrt{nc_p} + O(n^{3/2}\lambda_2^{n/6}).$$

In the following, we let $|\phi_p\rangle$ denote the normalized versions of $|\Phi_p\rangle$. In terms of matrix elements, we have

$$\langle\phi_p|F|\phi_{p'}\rangle = \frac{\langle\Phi_p|F|\Phi_{p'}\rangle}{n\sqrt{c_p c_{p'}}(1 + O(n^2\lambda_2^{\Delta/6}))} = \frac{\langle\Phi_p|F|\Phi_{p'}\rangle}{n\sqrt{c_p c_{p'}}} + O(n\lambda_2^{\Delta/6}). \quad (2.74)$$

Our main technical result for the excitation ansatz consists of the following estimates:

Lemma 2.6.8. *Let* $\nu \in (0,1)$ *and* $d = n^{1-\nu}$. *Let* $F \in \mathcal{B}((\mathbb{C}^p)^{\otimes n})$ *be a d-local operator with unit norm. Consider the normalized versions* $|\phi_p\rangle$ *and* $|\phi_{p'}\rangle$ *of the excitation ansatz state* (2.6.1). *Then we have*

$$|\langle\phi_{p'}|F|\phi_p\rangle| = O(n^{-\nu/2}) \qquad \text{for } p \neq p', \quad (2.75)$$

and

$$|\langle\phi_p|F|\phi_p\rangle - \langle\phi_{p'}|F|\phi_{p'}\rangle| = O(n^{-\nu/2}) \qquad \text{for all } p, p'. \quad (2.76)$$

Proof. By definition of the excitation ansatz states, we have

$$\langle\Phi_{p'}|F|\Phi_p\rangle = \sum_{j,j'=1}^{n} e^{i(pj-p'j')}\langle\Phi_{j',p'}|F|\Phi_{j,p}\rangle = \sum_{\alpha=1}^{4} \sigma_{pp'}(\Omega_\alpha), \quad (2.77)$$

where

$$\Omega_1 = \mathcal{B}^\Delta(\mathrm{supp}(F)) \times \mathcal{B}^\Delta(\mathrm{supp}(F)),$$
$$\Omega_2 = \mathcal{B}^\Delta(\mathrm{supp}(F)) \times \mathcal{B}^\Delta(\mathrm{supp}(F))^c,$$
$$\Omega_3 = \mathcal{B}^\Delta(\mathrm{supp}(F))^c \times \mathcal{B}^\Delta(\mathrm{supp}(F)),$$
$$\Omega_4 = \mathcal{B}^\Delta(\mathrm{supp}(F))^c \times \mathcal{B}^\Delta(\mathrm{supp}(F))^c,$$

is the partition of $[n]^2$ considered in Lemma 2.6.7. Thus, for $p \neq p'$, we obtain

$$|\langle\Phi_{p'}|F|\Phi_p\rangle| \leq 4|\mathcal{B}^{2\Delta}(\mathcal{F})|\sqrt{c_p c_{p'}} + O(n^2\lambda_2^{\Delta/2}).$$

Inserting the expression (2.6.5) for the normalized matrix element, we get

$$\begin{aligned}
|\langle\phi_{p'}|F|\phi_p\rangle| &\leq \frac{|\langle\Phi_p|F|\Phi_{p'}\rangle|}{n\sqrt{c_p c_{p'}}} + O(n\lambda_2^{\Delta/6}) \\
&\leq \frac{4|\mathcal{B}^{2\Delta}(\mathcal{F})|}{n} + O(n\lambda_2^{\Delta/2}) + O(n\lambda^{\Delta/6}) \\
&= \frac{4|\mathcal{B}^{2\Delta}(\mathcal{F})|}{n} + O(n\lambda_2^{\Delta/6}).
\end{aligned}$$

Assume that $\mathsf{supp}(F)$ consists of κ disjoint connected components. By definition, we have

$$|\mathcal{B}^\Delta(\mathsf{supp}(F))| = \left|\mathsf{supp}(F) \cup \left(\mathcal{B}^\Delta(\mathsf{supp}(F))\backslash\mathsf{supp}(F)\right)\right| \le d + 2\kappa\Delta \le d(1 + 2\Delta),$$

where we use the fact that $\kappa \le d$ in the last inequality. Hence, we have

$$|\langle\phi_{p'}|F|\phi_p\rangle| \le \frac{4d(1 + 4\Delta)}{n} + O(n\lambda_2^{\Delta/6}) .$$

Let $1 > \nu > 0$ be arbitrary. Choosing $d = n^{1-\nu}$ and $\Delta = 6n^{\nu/2}$ gives[9]

$$\frac{|\mathcal{B}^{2\Delta}(\mathsf{supp}(F))|}{n} \le \frac{d(1 + 4\Delta)}{n} = O(n^{-\nu/2}), \tag{2.78}$$

and therefore

$$|\langle\phi_{p'}|F|\phi_p\rangle| = O(n^{-\nu/2}) + O(n\lambda_2^{n^{\nu/2}}) = O(n^{-\nu/2}) .$$

Note that the last equality follows since, for all $\lambda_2 < 1$ and $a, b > 0$, we have $\lim_{n\to\infty} n^a \lambda_2^{n^b} = 0$. This proves claim (2.6.8).

Next, we prove (2.6.8). Making use of Equation (2.6.5) and the decomposition (2.6.5), we have

$$|\langle\phi_p|F|\phi_p\rangle - \langle\phi_{p'}|F|\phi_{p'}\rangle| \le \left|\frac{\langle\Phi_p|F|\Phi_p\rangle}{nc_p} - \frac{\langle\Phi_{p'}|F|\Phi_{p'}\rangle}{nc_{p'}}\right| + O(n\lambda_2^{\Delta/6})$$

$$\le \frac{1}{n}\sum_{\alpha=1}^{4}\left|\left(c_p^{-1}\sigma_{pp}(\Omega_\alpha) - c_{p'}^{-1}\sigma_{p'p'}(\Omega_\alpha)\right)\right| + O(n\lambda_2^{\Delta/6}).$$

By Lemma 2.6.7, we have

$$|\sigma_{pp}(\Omega_\alpha)| \le |\mathcal{B}^{2\Delta}(\mathsf{supp}(F))|c_p + O(n^2\lambda_2^{\Delta/2}) \quad \text{for } \alpha \in \{1, 2, 3\} ,$$

and so we can write

$$\frac{1}{n}\sum_{\alpha=1}^{3}\left|c_p^{-1}\sigma_{pp}(\Omega_\alpha) - c_{p'}^{-1}\sigma_{p'p'}(\Omega_\alpha)\right| \le \frac{1}{n}\sum_{\alpha=1}^{3}\left|c_p^{-1}\sigma_{pp}(\Omega_\alpha)\right| + \frac{1}{n}\sum_{\alpha=1}^{3}\left|c_{p'}^{-1}\sigma_{p'p'}(\Omega_\alpha)\right|$$

$$\le \frac{6|\mathcal{B}^{2\Delta}(\mathsf{supp}(F))|}{n} + O(n\lambda^{\Delta/2}).$$

---

[9]Note that this choice of d and Δ satisfies the requirement in Lemma 2.6.5 for sufficiently large n.

It remains to consider the terms involving Ω_4, whereby using Equation (2.6.7), we get

$$\left| c_p^{-1}\sigma_{pp}(\Omega_4) - c_{p'}^{-1}\sigma_{p'p'}(\Omega_4) \right| = O(n^2\lambda_2^{\Delta/2}).$$

Putting everything together, we have

$$|\langle\phi_p|F|\phi_p\rangle - \langle\phi_{p'}|F|\phi_{p'}\rangle| \leq \frac{6|\mathcal{B}^{2\Delta}(\text{supp}(F))|}{n} + O(n\lambda_2^{\Delta/6})$$
$$= O(n^{-\nu/2}),$$

where we again use the bound (2.6.5) in the last line. This proves claim (2.6.8). $\square$

With Lemma 2.6.8, it is straightforward to check the condition for approximate quantum error-detection from Section 2.3.2. This leads to the following:

Theorem 2.6.9. *Let $\nu \in (0,1)$ and let $\kappa, \Delta > 0$ be such that*

$$5\kappa + \lambda < \nu .$$

Let A, B be tensors associated with an injective excitation ansatz state $|\Phi_p(B;A)\rangle$, where p is the momentum of the state. Then there is a subspace $C \subset (\mathbb{C}^p)^{\otimes n}$ spanned by excitation ansatz states $\{|\Phi_p(B;A)\rangle\}_p$ with different momenta p such that C is an $(\epsilon, \delta)[[n,k,d]]$-AQEDC with parameters

$$k = \kappa \log_p n ,$$
$$d = n^{1-\nu} ,$$
$$\epsilon = \Theta(n^{-(\nu-(5\kappa+\lambda))}) ,$$
$$\delta = n^{-\lambda} .$$

Proof. Let us choose an arbitrary set $\{p_1, \ldots, p_{p^k}\}$ of $p^k = n^\kappa$ distinct, non-zero momenta, and define the space C by

$$C = \text{span}\{|\Phi_{p_j}(B;A)\rangle\}_{j=1}^{p^k} .$$

Since momentum eigenstates to different momenta are orthogonal, the states $\{|\phi_{p_j}\rangle\}_{j=1}^{p^k}$ form an orthonormal basis of C. By Lemma 2.6.8, we have

$$|\langle\phi_{p_r}|F|\phi_{p_s}\rangle - \delta_{r,s}\langle\phi_{p_1}|F|\phi_{p_1}\rangle| = O(n^{-\nu/2})$$

for any d-local unit norm operator $F \in \mathcal{B}((\mathbb{C}^p)^{\otimes n})$ and all $r, s \in [p^k]$. The sufficient conditions of Corollary 2.3.4 for approximate error-detection applied with $\gamma = \Theta(n^{-\nu/2})$ show that C is a $(\Theta(p^{5k}n^{-\nu}/\delta), \delta)[[n,k,d]]$-AQEDC for any δ satisfying $\delta > p^{5k}n^{-\nu}$. This implies the claim for the given choice of parameters. $\square$

From [66], we know that isolated energy bands in gapped systems are well approximated, under mild physical conditions, by the Fourier transforms of local operators. In particular, this means that, possibly after blocking, isolated momentum eigenstates of gapped systems are well approximated by some excitation ansatz state, as one would expect.[10] One consequence of this is that the excitation ansatz codes considered in this section are *generic* among physical systems: essentially any selection of momentum eigenstates from an isolated energy band of a gapped system can be expected to form an error-detecting code with the above parameters.

---

[10]In fact, we expect excitation ansatz states to be even better approximations of momentum eigenstates than the constructions considered in [66]. In [66], the local operators O act on the physical level, whereas the defining tensors B of excitation ansatz states act on the virtual level, and are hence more general.

2.7 AQEDC at Low Energies: An Integrable Model

In this section, we consider the Heisenberg-XXX spin chain. In Section 2.7.1, we introduce the model. The approximate error-detection codes we consider are spanned by eigenstates that we call magnon-states. The latter are particular instances of the algebraic Bethe ansatz, for which a general framework of MPS/MPO descriptions has been introduced in prior work [67]. We review the necessary notation for matrix product operators (MPOs) in Section 2.7.2. In Section 2.7.3, we give an MPS/MPO description of magnon-states. In Section 2.7.4, we provide a second MPS/MPO description with smaller bond dimension. In Section 2.7.5, we consider matrix elements of operators with respect to the magnon-state basis. We show how to relate matrix elements of operators with arbitrary support to matrix elements of operators with connected support. In Section 2.7.6 we analyze the Jordan structure of the transfer operators. In Section 2.7.7, we bound matrix elements of local operators in magnon states. Finally, in Section 2.7.8, we determine the parameters of the magnon code.

2.7.1 The XXX-Model and the Magnon Code

Consider the periodic Heisenberg-XXX spin chain, with Hamiltonian

$$H = -\frac{1}{4} \sum_{m=1}^{n} \left(\sigma_m^x \sigma_{m+1}^x + \sigma_m^y \sigma_{m+1}^y + \sigma_m^z \sigma_{m+1}^z \right) \tag{2.79}$$

on $(\mathbb{C}^2)^{\otimes n}$, where we apply periodic boundary conditions, and where $\sigma_m^x, \sigma_m^y, \sigma_m^z$ are the Pauli matrices acting on the m-th qubit. The model (2.7.1) is gapless and can be solved exactly using the algebraic Bethe ansatz. Our goal here is to argue that (2.7.1) contains error-detecting codes in its low-energy subspace. More precisely, we consider subspaces spanned by non-zero momentum eigenstates.

The Hamiltonian (2.7.1) may alternatively be expressed as

$$H = \frac{n}{4}I - \frac{1}{2} \sum_{m=1}^{n} \mathbb{F}_{m,m+1} , \tag{2.80}$$

where $\mathbb{F}_{m,m+1}$ is the flip-operator acting on the m-th and $(m + 1)$-th qubit. Equation (2.7.1) shows that H commutes with the tensor product representation of the special unitary group $SU(2)$ on $(\mathbb{C}^2)^{\otimes n}$, hence we may restrict to irreducible subspaces (with fixed angular momentum) to diagonalize H. More precisely, let us define, for each qubit m, the operators

$$\mathbf{s}_m^- = |0\rangle\langle 1|, \qquad \mathbf{s}_m^+ = (\mathbf{s}_m^-)^\dagger, \qquad \text{and} \qquad \mathbf{s}_m^3 = \frac{1}{2}(-|0\rangle\langle 0| + |1\rangle\langle 1|).$$

These satisfy the canonical $\mathfrak{su}(2)$ commutation relations, with $\mathbf{s}^+$ and $\mathbf{s}^-$ being the raising and lowering operators of the spin-$1/2$ representation, and the basis states $|0\rangle$ and $|1\rangle$ corresponding to $|j, m\rangle = |1/2, -1/2\rangle$ and $|1/2, 1/2\rangle$, respectively. The total z-angular momentum and raising/lowering-operators for the tensor product representation on $(\mathbb{C}^2)^{\otimes n}$ are given by

$$S_3 = \sum_{m=1}^{n} \mathbf{s}_m^3 \qquad \text{and} \qquad S_\pm = \sum_{m=1}^{n} \mathbf{s}_m^\pm \ .$$

These operators commute with H, and therefore the total Hilbert space splits into a direct sum of spin representations:

$$(\mathbb{C}^2)^{\otimes n} \cong \bigoplus_j \mathcal{H}_j \otimes \mathbb{C}^{m_j} \ ,$$

where the direct sum is taken over all irreducible spin representations (with multiplicity m_j) present in the decomposition of the tensor representation. Each $\mathcal{H}_j$ defines an irreducible $2j + 1$-dimensional angular momentum-j representation, and $H|_{\mathcal{H}_j} = E_j I_{\mathcal{H}_j}$ is proportional to the identity on each of these spaces. For instance, the subspace $\mathcal{H}_{n/2}$ with maximal angular momentum has highest weight vector $|1\rangle^{\otimes n}$ and is spanned by "descendants" obtained by applying the lowering operator, that is,

$$\mathcal{H}_{n/2} = \text{span} \left\{ S_-^r |1\rangle^{\otimes n} \mid r = 0, \ldots, n \right\} \ .$$

It is associated with energy $E_{n/2} = -n/4$, which is the ground state energy of H. Clearly, this is the symmetric subspace, containing only permutation-invariant (i.e., zero-momentum) states. Error-correction within this subspace has been considered in [51]. Indeed, all the examples constructed there consist of subspaces of $\mathcal{H}_{n/2}$.

Here we go beyond permutation-invariance. Specifically, we consider the vector

$$|\Psi\rangle = \overline{\omega} \sum_{r=1}^{n} \omega^r \mathbf{s}_r^- |1\rangle^{\otimes n} \qquad \text{where} \qquad \omega = e^{2\pi i/n} \ . \tag{2.81}$$

The factor $\overline{\omega}$ in front is introduced for convenience. A straightforward calculation shows that $S_+|\Psi\rangle = 0$ and $S_3|\Psi\rangle = (n/2 - 1)|\Psi\rangle$, hence this is a highest weight vector for angular momentum $j = n/2 - 1$ and

$$\mathcal{H}_{n/2-1} = \text{span} \left\{ S_-^r |\Psi\rangle \mid r = 0, \ldots, n - 2 \right\} \ . \tag{2.82}$$

The energy of states in this subspace can be computed to be $E_{n/2-1} = -n/4 + 1 - \cos(2\pi/n) = E_{n/2} + O(1/n^2)$. This shows that these states are associated with

low-lying excitations, and the system is gapless. Observe also that (2.7.1) is an eigenvector of the cyclic shift with eigenvalue ω, that is, it has fixed momentum $p = 2\pi/n$. As S_-^r commutes with the cyclic shift, the same is true for all states in $\mathcal{H}_{n/2-1}$: this is a subspace of fixed momentum and energy. We will argue that $\mathcal{H}_{n/2-1}$ contains error-detecting codes. Specifically, we consider subspaces spanned by states of the form $\{S_-^r|\Psi\rangle\}_r$ for appropriate choices of magnetization r. The state (2.7.1) is sometimes referred to as a one-magnon state. Correspondingly, we call the corresponding code(s) the *magnon-code*. We also refer to the vectors $\{S_-^r|\Psi\rangle\}_r$ (respectively, their normalized versions) as *magnon-states*. For brevity, let us denote the r-th descendant by

$$|\Psi_r\rangle := S_-^r|\Psi\rangle \qquad \text{for} \qquad r = 0, \ldots, n-2 . \tag{2.83}$$

It is clear that the states $|\Psi_r\rangle$ and $|\Psi_s\rangle$ are orthogonal for $r \neq s$ as they have different magnetization, hence they form a basis of the magnon code. It is also convenient to introduce their normalized versions which are given by

$$|\psi_r\rangle = \left(\frac{(n-2-r)!}{n(n-2)!r!}\right)^{1/2} S_-^r|\Psi\rangle \qquad \text{for} \qquad r = 0, \ldots, n-2 , \tag{2.84}$$

as follows from the fact that for a normalized highest-weight vector $|j, j\rangle$ of a spin-j-representation, the vectors

$$|j, j - k\rangle = \left(\frac{(2j-k)!}{(2j)!k!}\right)^{1/2} S_-^k|j, j\rangle,$$

with $k = 0, \ldots, 2j$ form an orthonormal basis.

2.7.2 Matrix Product Operators

Here we briefly review the formalism of matrix product operators (MPO) and introduce the corresponding notation. We only require site-independent MPO. Such an MPO $O \in \mathcal{B}((\mathbb{C}^{\mathsf{p}})^{\otimes n})$ with bond dimension D is given by

$$O = \sum_{\substack{i_1,\ldots,i_n\in[\mathsf{p}] \\ j_1,\ldots,j_n\in[\mathsf{p}]}} \text{tr}(O_{i_1,j_1}\cdots O_{i_n,j_n}X)|i_1\rangle\langle j_1| \otimes \cdots \otimes |i_n\rangle\langle j_n|$$

for a family of local tensors $\{O_{i,j}\}_{i,j\in[\mathsf{p}]} \subset \mathcal{B}(\mathbb{C}^D)$, and a boundary operator $X \in \mathcal{B}(\mathbb{C}^D)$. Alternatively, the MPO O can also be parametrized by the operator $X \in \mathcal{B}(\mathbb{C}^D)$ together with family $\{O^{\alpha,\beta}\}_{\alpha,\beta\in[D]}$ of $\mathsf{p}\times\mathsf{p}$-matrices. In this parametrization (illustrated in Figure 2.8), the MPO is written as

$$O = \sum_{\alpha_0,\ldots,\alpha_n\in[D]} X_{\alpha_n,\alpha_0}O^{\alpha_0,\alpha_1} \otimes O^{\alpha_1,\alpha_2} \otimes \cdots \otimes O^{\alpha_{n-1},\alpha_n} . \tag{2.85}$$
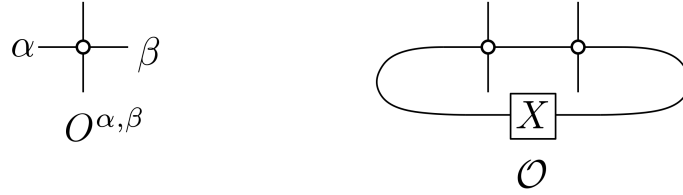
Figure 2.8: Alternative parametrization of an MPO O.
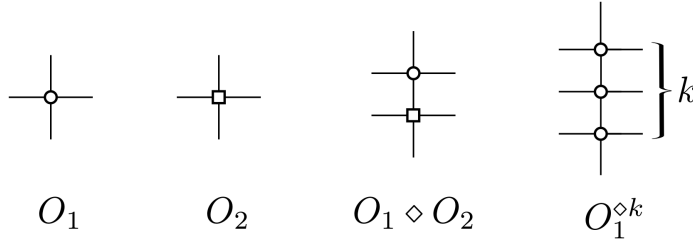
Figure 2.9: The product of two MPO tensors O_1 and O_2, as well as the power $O_1^{\diamond k}$.

Equation (2.7.2) shows that the MPO $O = O(O, X, n) \in \mathcal{B}((\mathbb{C}^p)^{\otimes n})$ is fully specified by three objects:

(i) a four-index tensor O, defined in terms of the collection $\{O_{i,j}\}_{i,j \in [p]}$ of matrices acting on the so-called virtual space $\mathbb{C}^D$ (alternatively, the collection of matrices $\{O^{\alpha,\beta}\}_{\alpha,\beta \in [D]}$ acting on the physical space $\mathbb{C}^p$),

(ii) a matrix $X \in \mathcal{B}(\mathbb{C}^D)$ acting on the virtual space, and

(iii) an integer $n \in \mathbb{N}$ specifying the number of physical spins.

We refer to the tensor O as a local MPO tensor, and to X as a boundary operator.

It is convenient to introduce the following product on MPO tensors. Suppose O_1 and O_2 are MPO tensors associated with MPOs having physical dimension p, and bond dimensions D_1 and D_2, respectively. Then $O_1 \diamond O_2$ is the MPO tensor of an MPO with physical dimension p and bond dimension $D_1 \cdot D_2$. Its tensor network description is given in Figure 2.9. More precisely, if O_α is defined by $\{O_{i,j}^{(x)}\}_{i,j \in [p]}$ for $x = 1, 2$, then $O_1 \diamond O_2$ is defined in terms of the matrices

$$O_{i,j} = \sum_{k=1}^{r} (O^{(1)})_{i,k} \otimes (O^{(2)})_{k,j} \in \mathcal{B}(\mathbb{C}^{D_1} \otimes \mathbb{C}^{D_2}) \qquad \text{for } i, j \in [p] \ .$$

This is clearly associative, and allows us to define $O^{\diamond k} := O \diamond O^{\diamond(k-1)}$ recursively.

Suppose now that an MPO $O = O(O, X, n)$ is given. Observe that for $k \in \mathbb{N}$, the operator O^k is an MPO whose virtual bond space is $(\mathbb{C}^D)^{\otimes k}$ and whose local tensors
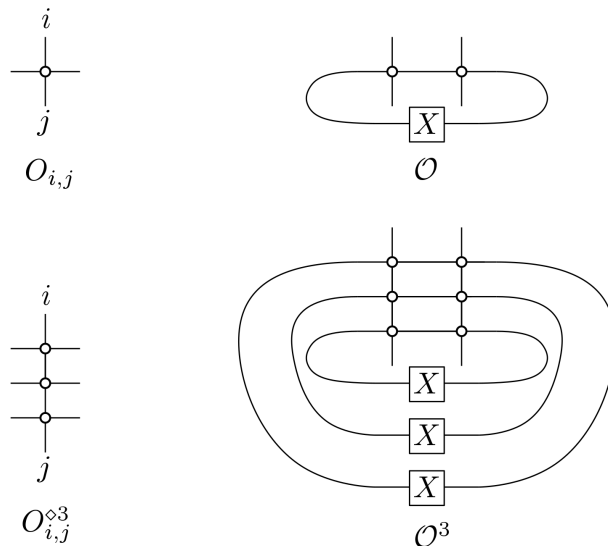
Figure 2.10: This figure shows an MPO $O = O(O, X, 2)$ defined in terms of matrices $\{O_{i,j}\}_{i,j}$ and the matrices $\{O^{\diamond 3}_{i,j}\}_{i,j}$ defining the MPO O^3. Left-multiplication by an operator corresponds to stacking a diagram on top.
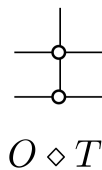
Figure 2.11: Definition of the MPS tensor $O \diamond T$.

take the form

$$\langle \alpha_1 \cdots \alpha_k | (O^{\diamond k})_{i,j} | \beta_1 \cdots \beta_k \rangle = \sum_{s_1,\ldots,s_{k-1} \in [\mathsf{p}]} \begin{array}{c} \langle \alpha_1 | O_{i,s_1} | \beta_1 \rangle \cdot \langle \alpha_2 | O_{s_1,s_2} | \beta_2 \rangle \\ \cdots \langle \alpha_{k-1} | O_{s_{k-2},s_{k-1}} | \beta_{k-1} \rangle \cdot \langle \alpha_k | O_{s_{k-1},j} | \beta_k \rangle \end{array},$$

with boundary tensor $X^{\otimes k}$. Thus the MPO $O^k = O(O^{\diamond k}, X^{\otimes k}, n)$ is defined by the MPO tensor $O^{\diamond k}$ and the boundary operator $X^{\otimes k}$. These are visualized in Figure 2.10, for $k = 3$.

Consider an MPS $|\Psi\rangle = |\Psi(A, X, n)\rangle \in (\mathbb{C}^{\mathsf{p}})^{\otimes n}$ of bond dimension D_1 and an MPO $O = O(O, Y, n) \in \mathcal{B}((\mathbb{C}^{\mathsf{p}})^{\otimes n})$ of bond dimension D_2. Then clearly $O|\Psi\rangle$ is an MPS with bond dimension $D_1 D_2$. We write

$$O|\Psi\rangle = |\Psi(O \diamond A, Y \otimes X, n)\rangle, \tag{2.86}$$

see Figure 2.11 for the definition of the MPS tensor $O \diamond T$.

$E(O)$ $E(O^3)$
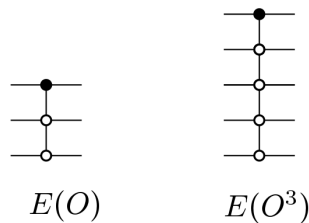
Figure 2.12: This figure shows the tensor network representations of E_O and $E_{O^{\circ 3}}$.

O $O \odot |\varphi\rangle$

$|\varphi\rangle$ $O^{\odot 2}|\varphi\rangle$
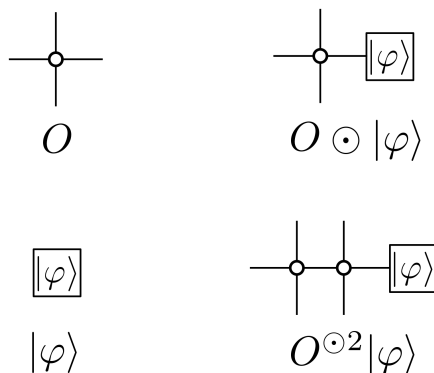
Figure 2.13: This figure illustrates the definition of the product $O \odot |\varphi\rangle$.

In the following, we are interested in matrix elements of the form $\langle \Psi|O|\Psi\rangle$. A central object of study is the generalized transfer operator E_O. If O is specified by matrices $\{O_{i,j}\}_{i,j\in[\mathsf{p}]} \subset \mathcal{B}(\mathbb{C}^{D_2})$, this is given by

$$E_O = \sum_{s,t\in[D_1]} \sum_{j,k\in[D_2]} \langle s|O_{j,k}|t\rangle \overline{A_s} \otimes |j\rangle\langle k| \otimes A_t \in \mathcal{B}(\mathbb{C}^{\mathsf{p}} \otimes \mathbb{C}^{D_2} \otimes \mathbb{C}^{\mathsf{p}}) .$$

This operator, as well as $E_{O^{\circ k}} \in \mathcal{B}(\mathbb{C}^{\mathsf{p}} \otimes (\mathbb{C}^{D_2})^{\otimes k} \otimes \mathbb{C}^{\mathsf{p}})$ for $k = 3$ are illustrated in Figure 2.12.

Consider an MPO tensor O with physical space $\mathcal{H}_p$ and virtual space $\mathcal{H}_v$. Given a vector $|\varphi\rangle \in \mathcal{H}_v$, we can define an element $O \odot |\varphi\rangle \in \mathcal{H}_v \otimes \mathcal{B}(\mathcal{H}_p)$ by attaching O from the left, see Figure 2.13. The map $(O, \varphi) \mapsto O \odot |\varphi\rangle$ is bilinear. Hence we can define

$$O \odot (O \odot |\varphi\rangle) := (O \otimes I_{\mathcal{B}(\mathcal{H}_p)})(O \odot |\varphi\rangle) \in \mathcal{H}_v \otimes \mathcal{B}(\mathcal{H}_p) \otimes \mathcal{B}(\mathcal{H}_p) .$$

This is clearly associative. Correspondingly, we also define $O^{\odot n}|\varphi\rangle \in \mathcal{H}_v \otimes \mathcal{B}(\mathcal{H}_p)^{\otimes n}$ as the result applying this map n times. Note that an MPO defined by $(O, X = |\varphi\rangle\langle\chi|)$ can be written as $\left(\langle\chi| \otimes I_{\mathcal{B}(\mathcal{H}_p)}^{\otimes n}\right) O^{\odot n}|\varphi\rangle$.

Conversely, observe that a bilinear map $\Gamma : \mathcal{H}_v \to \mathcal{H}_v \otimes \mathcal{B}(\mathcal{H}_p)$, together with two states $|\varphi\rangle, |\chi\rangle \in \mathcal{H}_v$, defines a site-independent MPO in this fashion.

$$0$$

$$A_0 \quad = \quad \boxed{\sigma_-} \qquad \sigma_- = |0\rangle\langle 1|$$

$$1$$

$$A_1 \quad = \quad \boxed{Z} \qquad Z = |0\rangle\langle 0| + \omega|1\rangle\langle 1|$$

$$X = \sigma_-$$

Figure 2.14: An MPS description of the one-magnon state $|\Psi\rangle$ (cf. (2.7.1)).

2.7.3 MPS/MPO Representation of the Magnon States

Here we give an MPS/MPO representation of the magnon states that we use throughout our analysis below. We note that more generally, [67] discusses such representations for the Bethe ansatz states.

Consider the one-magnon state $|\Psi\rangle \in (\mathbb{C}^2)^{\otimes n}$ defined by (2.7.1). It is straightforward to check that an MPS representation of $|\Psi\rangle = |\Psi(\{A_0, A_1, X\})\rangle$ with bond dimension $D = 2$ is given by

$$\begin{align}
A_0 &= |1\rangle\langle 0| \, , \\
A_1 &= |0\rangle\langle 0| + \omega|1\rangle\langle 1| \, , \tag{2.87} \\
X &= |0\rangle\langle 1| \, ,
\end{align}$$

where $\omega = e^{2\pi i/n}$, see Figure 2.14. Next, we consider the descendants (2.7.1). The operator $S_- = \sum_{m=1}^{n} \mathbf{s}_m^-$ can be expressed as a bond dimension $D = 2$ MPO, given by

$$S_- = O(O^{0,0}, O^{0,1}, 0^{1,0}, O^{1,1}, X) \in \mathcal{B}((\mathbb{C}^2)^{\otimes n}) \, ,$$

where the boundary tensor is $X = \sigma_- := |0\rangle\langle 1|$, and where the local tensors are defined as

$$\begin{align}
O^{0,0} = O^{1,1} &= I_{\mathbb{C}^2} \, , \\
O^{1,0} &= |0\rangle\langle 1| \, , \tag{2.88} \\
O^{0,1} &= 0 \, .
\end{align}$$

This definition is illustrated in Figure 2.15. The adjoint operator S_+ has an MPO representation described as in Figure 2.16.

$$0 \,-\!\!\diamond\!\!-\, 0 \;=\; \Big| \qquad\qquad 1 \,-\!\!\diamond\!\!-\, 1 \;=\; \Big|$$

$$1 \,-\!\!\diamond\!\!-\, 0 \;=\; \boxed{\sigma^-} \qquad\qquad 0 \,-\!\!\diamond\!\!-\, 1 \;=\; 0$$

$$\langle 1| \,-\!\!\diamond\!\diamond\!\diamond\!\diamond\!-\, |0\rangle$$

$$S_-$$

Figure 2.15: An MPO description of the lowering operator. The MPO $S_- = O(O, X, n)$ is defined with O as given in the figure and with $X = \sigma_-$.

$$0 \,-\!\!\bullet\!\!-\, 0 \;=\; \Big| \qquad\qquad 1 \,-\!\!\bullet\!\!-\, 1 \;=\; \Big|$$

$$1 \,-\!\!\bullet\!\!-\, 0 \;=\; \boxed{\sigma^-} \qquad\qquad 0 \,-\!\!\bullet\!\!-\, 1 \;=\; 0$$

$$\langle 1| \,-\!\!\bullet\!\bullet\!\bullet\!\bullet\!-\, |0\rangle$$

$$S_+$$

Figure 2.16: An MPO description of the adjoint MPO S_+.

$$\left.\begin{array}{l} \langle 1| \;-\!\diamond\!\diamond\!\diamond\!\diamond\!\diamond\!-\; |0\rangle \\ \langle 1| \;-\!\diamond\!\diamond\!\diamond\!\diamond\!\diamond\!-\; |0\rangle \\ \langle 1| \;-\!\diamond\!\diamond\!\diamond\!\diamond\!\diamond\!-\; |0\rangle \\ \langle 1| \;-\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!-\; |0\rangle \end{array}\right\} s \text{ layers}$$

$$|\Psi_s\rangle$$

Figure 2.17: An MPS/MPO representation of the vector $|\Psi_s\rangle = S_-^s |\Psi\rangle$. Seen as an MPS, this has rank-1-boundary tensor $X = (|0\rangle\langle 1|)^{\otimes s+1}$.

It follows that the "descendants" $|\Psi_s\rangle = S_-^s |\Psi\rangle$ can be represented as in Figure 2.17, i.e., they are MPS of the form

$$|\Psi_s\rangle = |\Psi(O^{\diamond s} \diamond A, X^{\otimes(s+1)}, n)\rangle \qquad \text{for } s = 0, \ldots, n-2,$$

where the MPS tensor $O \diamond T$ is defined as in Equation (2.7.2).

2.7.4 A Compressed MPS/MPO Representation of the Magnon States

Consider the MPO representation (2.7.3) of S_-. For $s \in [n]$, it implies the MPO representation

$$S_-^s = O(O^{\diamond s}, \sigma_-^{\otimes s}, n) \tag{2.89}$$

for the s-th power of S_-, which has bond dimension $D = 2^s$. Below we argue that the MPO (2.7.4) can also be expressed as an MPO with bond dimension $s + 1$. We call this the compressed representation:

Lemma 2.7.1. *Let $s \in [n]$ and consider the operator S_-^s, where $S_- = \sum_{m=1}^n s_m^-$. This has the bond dimension $D = s + 1$-MPO representation*

$$S_-^s = O(\tilde{O}_s, \tilde{X}_s, n) \ .$$

Here the virtual space $\mathbb{C}^{s+1}$ is that of a spin-$s/2$ with orthonormal angular momentum eigenstate basis $\{|\frac{s}{2}, m\rangle \mid m = -\frac{s}{2}, -\frac{s}{2} + 1, \ldots, \frac{s}{2}\}$. The boundary tensor is

$$\tilde{X}_s = |\frac{s}{2}, -\frac{s}{2}\rangle\langle\frac{s}{2}, \frac{s}{2}|$$

and the MPO tensor $\tilde{O}_s$ is defined by the matrices

$$
\begin{aligned}
(\tilde{O}_s)_{0,0} = (\tilde{O}_s)_{1,1} &= I \ , \\
(\tilde{O}_s)_{1,0} &= 0 \ , \\
(\tilde{O}_s)_{0,1} &= J_+ \ ,
\end{aligned}
\tag{2.90}
$$

where J_+ is the usual spin-raising operator.[11] In particular, the states $|\Psi_s\rangle = S_-^s|\Psi\rangle$ have an MPS representation of the form

$$|\Psi_s\rangle = |\Psi(\tilde{O}_s \diamond A, \tilde{X}_s \otimes X, n)\rangle \ ,$$

with bond dimension $2(s + 1)$.

Proof. Consider the MPO tensor $O^{\diamond s}$ associated with the MPO representation (2.7.4) of S_-^s. We express it in terms of matrices $\{O_{i,j}\}_{i,j \in \{0,1\}} \subset \mathcal{B}((\mathbb{C}^2)^{\otimes s})$ acting on the

---

[11]With respect to a distinguished orthonormal basis $\{|j, m\rangle\}_{m=-j,-j+1,\cdots,j}$, we have

$$J_+|j, m\rangle = \sqrt{j(j + 1) - m(m + 1)}|j, m + 1\rangle$$

for all $m = -j, \cdots, j - 1$ and $J_+|j, j\rangle = 0$.

virtual space of dimension $D = 2^s$. The latter has orthonormal basis $\{|\alpha\rangle = |\alpha_1\rangle \otimes \cdots \otimes |\alpha_s\rangle\}_{\alpha=(\alpha_1,\ldots,\alpha_s)\in\{0,1\}^s}$. By definition (2.7.3) of O and the fact that $(O^{1,0})^2 = \sigma_-^2 = 0$, it is easy to see that

$$
\begin{aligned}
\langle\alpha|O_{0,0}|\beta\rangle &= \delta_{\alpha,\beta} \\
\langle\alpha|O_{1,1}|\beta\rangle &= \delta_{\alpha,\beta}, \qquad \text{and} \qquad \langle\alpha|O_{0,1}|\beta\rangle = \begin{cases} 1 & \text{if } \beta \leq \alpha \\ 0 & \text{otherwise} \end{cases}, \\
\langle\alpha|O_{1,0}|\beta\rangle &= 0
\end{aligned}
$$

where we write $\beta \leq \alpha$ for $\alpha, \beta \in \{0,1\}^s$ if and only if there is exactly one $k \in [s]$ such that $\beta_k = 0$ and $\alpha_k = 1$, and $\alpha_\ell = \beta_\ell$ for all $\ell \neq k$. Let us define $\mathbf{j}_k^-$ as the operator $|0\rangle\langle 1|$ acting on the k-factor in $(\mathbb{C}^2)^{\otimes s}$, and $\mathbf{j}_k^+ = (\mathbf{j}_k^-)^\dagger$ for $k \in [s]$. Then it is easy to check that $\mathbf{j}_+ := \sum_{k=1}^s \mathbf{j}_k^+$ has the same matrix elements $\langle\alpha|O_{0,1}|\beta\rangle$ as $O_{0,1}$. It follows that

$$
\begin{aligned}
O_{0,0} &= I_{(\mathbb{C}^2)^{\otimes s}} \\
O_{1,1} &= I_{(\mathbb{C}^2)^{\otimes s}}
\end{aligned}, \qquad \text{and} \qquad
\begin{aligned}
O_{1,0} &= 0 \\
O_{0,1} &= \mathbf{j}_+
\end{aligned}. \tag{2.91}
$$

According to the MPO representation (2.7.4) of S_-^s, the matrix elements of this operator can be expressed as

$$
\langle i_1 \cdots i_n|S_-^s|j_1 \cdots j_n\rangle = \langle 1|^{\otimes s} O_{i_1 j_1} \cdots O_{i_n j_n}|0\rangle^{\otimes s} \qquad \text{for all } (i_1,\ldots,i_n), (j_1,\ldots,j_n) \in \{0,1\}^n .
$$

Combining this expression with (2.7.4), it follows that

$$
\langle i_1 \cdots i_n|S_-^s|j_1 \cdots j_n\rangle = \langle 1|^{\otimes s} (\tilde{O}_s)_{i_1 j_1} \cdots (\tilde{O}_s)_{i_n j_n}|0\rangle^{\otimes s} \qquad \text{for all } (i_1,\ldots,i_n), (j_1,\ldots,j_n) \in \{0,1\}^n ,
$$

where $(\tilde{O}_s)_{i,j}$ is the restriction of $(O^{\diamond s})_{i,j}$ to the subspace $\mathsf{span}\{(\mathbf{j}^+)^r|0\rangle^{\otimes s} \mid r = 0,\ldots,s\}$. This implies the claim. $\qquad\square$

2.7.5 Action of the Symmetric Group on the Magnon States

The symmetric group S_n acts on $(\mathbb{C}^2)^{\otimes n}$ by permuting the factors, i.e., we have for an orthonormal basis $\{|e_1\rangle, |e_2\rangle\} \in \mathbb{C}^2$ that

$$
\pi(|e_{i_1}\rangle \otimes \cdots \otimes |e_{i_n}\rangle) = |e_{i_{\pi^{-1}(1)}}\rangle \otimes \cdots \otimes |e_{i_{\pi^{-1}(n)}}\rangle \qquad \text{for all } \pi \in S_n ,
$$

and this is linearly extended to all of $(\mathbb{C}^2)^{\otimes n}$. Since

$$
[\pi, S_-] = 0 \qquad \text{for all } \pi \in S_n , \tag{2.92}
$$

the space $\mathsf{span}\{S_-^k|\Psi\rangle \mid k \in \mathbb{N}_0\}$ is invariant under permutations. In the following, we will show that the restriction of the group action to this space has a particularly simple form: every permutation acts as a tensor product of diagonal unitaries. Our main claim (Theorem 2.7.3 below) follows from (2.7.5) and the following statement.

Lemma 2.7.2. *Let $A_0, A_1 \in \mathcal{B}(\mathbb{C}^2)$ be the matrices defining the MPS $|\Psi\rangle$, cf. Equation (2.7.3). Then*

$$A_c A_b = \omega^c \overline{\omega}^b A_b A_c \qquad \text{for all} \qquad b, c \in \{0, 1\} . \tag{2.93}$$

Consider the MPO tensor O defined by Equation (2.7.3) and set $O_{a,b} = O(O, |b\rangle\langle a|, 2) \in \mathcal{B}((\mathbb{C}^2)^{\otimes 2})$ for $a, b \in \{0, 1\}$. Then

$$O_{a,b}(Z^\dagger \otimes Z) = (Z^\dagger \otimes Z)O_{a,b} \qquad \text{for all } a, b \in \{0, 1\} , \tag{2.94}$$

where $Z = \mathsf{diag}(1, \omega)$.

It is convenient to express the corresponding statements diagramatically. First, observe that specializing (2.7.5) to a neighboring transposition and inserting the MPO description of S_- introduced in Section 2.7.3, we obtain the diagrammatic identity

$$\tag{2.95}$$

Claim (2.7.2) describes the action of a neighboring transposition and can be written as

$$\tag{2.96}$$

Claim (2.7.2) can be written as

$$\tag{2.97}$$

Proof. Equation (2.7.2) can be shown by checking each case:

Similarly, (2.7.2) is shown by direct computation. $\qquad\square$

The main feature we need in what follows is the following statement:

Lemma 2.7.3. *Consider the spin $j = n/2 - 1$ subspace $\mathcal{H}_{n/2-1} \subset (\mathbb{C}^2)^{\otimes n}$ introduced in Equation (2.7.1). Let $\tau = (k\ k+1) \in S_n$ be an arbitrary transposition of nearest neighbors. Then the restriction of τ to $\mathcal{H}_{n/2-1}$ is given by the operator*

$$\tau|_{\mathcal{H}_{n/2-1}} = I^{\otimes k-1} \otimes Z^\dagger \otimes Z \otimes I^{\otimes n-k-1} \, ,$$

where $I = I_{\mathbb{C}^2}$.

Proof. It suffices to check that $\tau S_-^s |\Psi\rangle = (I^{\otimes k-1} \otimes Z^\dagger \otimes Z \otimes I^{\otimes n-k-1}) S_-^s |\Psi\rangle$. This follows immediately from Lemma 2.7.2. A diagrammatic proof of the steps involved

can be given as follows (illustrated for $s = 3$):

Here we used (2.7.5) s times in the first identity, Equation (2.7.5) in the second identity, and Equation (2.7.5) (applied s times) in the last step. □

An immediate and crucial consequence of Lemma 2.7.3 and the unitarity of Z is the fact that matrix elements of an operator acting on d arbitrary sites can be related to matrix elements of a local operator on the d first sites. To express this concisely, we use the following notation: suppose $F = F_1 \otimes \cdots \otimes F_d \in \mathcal{B}((\mathbb{C}^2)^{\otimes d})$ is a tensor product operator and $A = \{a_1 < \cdots < a_d\} \subset [n]$ a subset of $d = |A|$ (ordered) sites. Then we write $F_A \otimes I_{[n]\backslash A} \in \mathcal{B}((\mathbb{C}^2)^{\otimes n})$ for the operator acting as F_k on site a_k, for $k \in [d]$. By linearity, this definition extends to general (not necessarily product) operators $F \in \mathcal{B}((\mathbb{C}^2)^{\otimes d})$. Note that if $A = [d]$ are the first d sites, then $F_A \otimes I_{[n]\backslash A} = F \otimes I^{\otimes n-d}$.

Lemma 2.7.4. *Consider the magnon states $|\Psi_\ell\rangle = S_-^\ell|\Psi\rangle$ and let $r, s \in \{0, \ldots, n-2\}$ be arbitrary. Suppose $F \in \mathcal{B}((\mathbb{C}^2)^{\otimes d})$ acts on a subset $A \subset [n]$ of $d = |A|$ sites. Then*

$$\langle\Psi_r|(F_A \otimes I_{[n]\backslash A})|\Psi_s\rangle = \langle\Psi_r|(\tilde{F}_{[d]} \otimes I^{\otimes n-d})|\Psi_s\rangle$$

where $\tilde{F} \in \mathcal{B}((\mathbb{C}^2)^{\otimes d})$ is given by

$$\tilde{F} = ((Z^\dagger)^a \otimes (Z^\dagger)^{a+1} \cdots \otimes (Z^\dagger)^{a+d})F(Z^a \otimes Z^{a+1} \cdots \otimes Z^{a+d})$$

where $a = (\min A) - 1$.

More generally, if $B \subset [n]$ is a subset of size $b = |B|$ located "to the right of A" (i.e., if $\min B > \max A$) and $G \in \mathcal{B}((\mathbb{C}^2)^{\otimes b})$, then

$$\langle \Psi_r | (F_A \otimes G_B \otimes I_{[n]\backslash(A\cup B)}) | \Psi_s \rangle = \langle \Psi_r | (\tilde{F}_{[d]} \otimes G_B \otimes I_{[n]\backslash([d]\cup B)}) | \Psi_s \rangle \quad (2.98)$$

Furthermore, the analogous statement holds when G is permuted to the right, but with Z replaced by $Z^\dagger$.

Succinctly, Equation (2.7.4) can be represented as follows in the case where A consists of a connected set of sites (and $r = s = 0$):
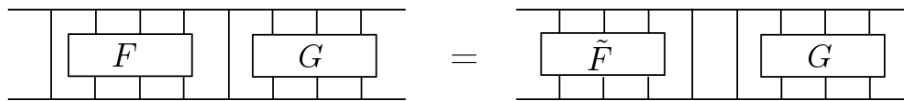
We emphasize, however, that the analogous statement is true for the more general case where A is a union of disconnected components.

Proof. The proof of Equation (2.7.4) for a single-site operator F is immediate. We have (illustrated for $r = s = 0$):
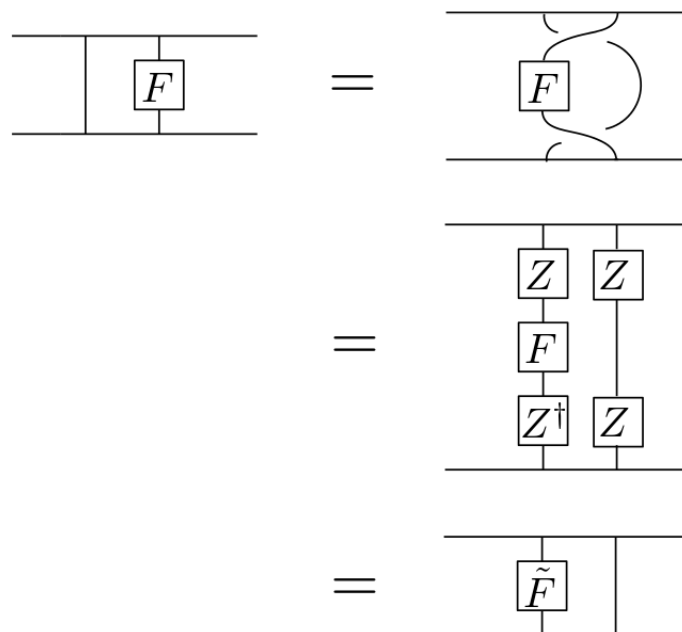
Applying (2.7.4) (with $d = 1$) iteratively then shows that the claim (2.7.4) also holds for any tensor product operator $F = F_1 \otimes \cdots \otimes F_d$. The general claim then follows by decomposing an arbitrary operator F into tensor products and using linearity. $\qquad\square$

2.7.6 The Transfer Operator of the Magnon States and its Jordan Structure

We are ultimately interested in matrix elements $\langle \Psi_r | (F \otimes I^{\otimes(n-d)}) | \Psi_s \rangle$ where $F \in \mathcal{B}((\mathbb{C}^2)^{\otimes d})$ acts on d sites. Using the compressed representation from Lemma 2.7.1, we may write these as

$$\langle \Psi_r | (F \otimes I^{\otimes(n-d)}) | \Psi_s \rangle = \langle \Psi(\tilde{O}_r \diamond A, \tilde{X}_r \otimes X, n) | (F \otimes I^{\otimes(n-d)}) | \Psi(\tilde{O}_s \diamond A, \tilde{X}_s \otimes X, n) \rangle \ .$$

We are thus interested in the ("overlap") transfer operator

$$E_{r,s} = E(\tilde{O}_r \diamond A, \tilde{O}_s \diamond A) \qquad \text{for } r, s \in [n] \ .$$

For convenience, let us also set

$$\begin{aligned}
E_{0,s} &= E(A, \tilde{O}_s \diamond A) && \text{for } s \in [n] \ , \\
E_{r,0} &= E(\tilde{O}_r \diamond A, A) && \text{for } r \in [n] \ , \\
E_{0,0} &= E(A, A) \ .
\end{aligned}$$

Observe that $E_{0,0}$ is the transfer operator E of the MPS $|\Psi\rangle$, whereas $E_{0,s}$ is the transfer operator of $S_-^s | \Psi \rangle$. We will order the tensor factors such that the virtual spaces of the original MPS are the first two factors. Then $E_{r,s} \in \mathcal{B}(\mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^{r+1} \otimes \mathbb{C}^{s+1})$. Our main goal in this section is to show the following:

Theorem 2.7.5. *Let $r, s \in \{0, \ldots, n\}$ be arbitrary. Then the operator $E_{r,s} \in \mathcal{B}(\mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^{r+1} \otimes \mathbb{C}^{s+1})$ has spectrum $\mathsf{spec}(E_{r,s}) = \{1, \omega, \overline{\omega}\}$ (where 1 has multiplicity $2 \cdot (r+1)(s+1)$ and $\omega, \overline{\omega}$ each have multiplicity $(r+1)(s+1)$). The size h^* of the largest Jordan block in $E_{r,s}$ is bounded by*

$$h^* \leq \min\{r, s\} + 2 \ .$$

To prove this theorem, we first rewrite the operator $E_{r,s}$. We have

$$E_{r,s} = E(\tilde{O}_r \diamond A, \tilde{O}_s \diamond A) = E(A, \tilde{O}_r^\dagger \diamond \tilde{O}_s \diamond A) \ ,$$

where $\tilde{O}_r^\dagger$ is obtained from the defining matrices $\{\tilde{O}^{\alpha,\beta}\}_{\alpha,\beta}$ of $\tilde{O}$ by replacing $\tilde{O}^{\alpha,\beta}$ with its adjoint $(\tilde{O}^\dagger)^{\alpha,\beta}$. This amounts to replacing σ_- by σ_+, or alternatively,

swapping the indices in the defining matrices $\{(\tilde{O}_s)_{i,j}\}$ (cf. (2.7.1)). That is,

$$(\tilde{O}_s)_{0,0} = (\tilde{O}_s)_{1,1} \;=\; I_{\mathbb{C}^{s+1}} \qquad\qquad (\tilde{O}^\dagger_r)_{0,0} = (\tilde{O}^\dagger_r)_{1,1} \;=\; I_{\mathbb{C}^{r+1}}$$
$$(\tilde{O}_s)_{1,0} \;=\; 0 \;, \qquad \text{and} \qquad (\tilde{O}^\dagger_r)_{1,0} \;=\; J_{+,r} \;.$$
$$(\tilde{O}_s)_{0,1} \;=\; J_{+,s} \qquad\qquad\qquad\qquad (\tilde{O}^\dagger_r)_{0,1} \;=\; 0$$

where $J_{+,s}$ and $J_{+,r}$ are the raising operators in the spin-$s/2$ and the spin-$r/2$ representation, respectively. We conclude that

$$E_{r,s} = E \otimes I_{\mathbb{C}^{r+1}} \otimes I_{\mathbb{C}^{s+1}} + E_{\sigma_+} \otimes J_{+,r} \otimes I_{\mathbb{C}^{s+1}} + E_{\sigma_-} \otimes I_{\mathbb{C}^{r+1}} \otimes J_{+,s} + E_{\sigma_+\sigma_-} \otimes J_{+,r} \otimes J_{+,s} \;,$$

where S_+ are the raising operator S_+ of the spin-j representation with $j = r/2$ and $j = s/2$, respectively. Here

$$E = |00\rangle\langle 00| + \omega|01\rangle\langle 01| + \overline{\omega}|10\rangle\langle 10| + |11\rangle\langle 11| + |11\rangle\langle 00| \;,$$
$$E_{\sigma_-} = |10\rangle\langle 00| + \omega|11\rangle\langle 01| \;,$$
$$E_{\sigma_+} = |01\rangle\langle 00| + \overline{\omega}|11\rangle\langle 10| \;,$$
$$E_{\sigma_+\sigma_-} = |00\rangle\langle 00| + \overline{\omega}|10\rangle\langle 10| + \omega|01\rangle\langle 01| + |11\rangle\langle 11| \;,$$

are the transfer operators of the MPS $|\Psi\rangle$. We can write down the transfer matrix $E_{r,s}$ more explicitly as

$$E_{r,s} = \overline{A_0} \otimes A_0 \otimes I \otimes I + \overline{A_1} \otimes A_1 \otimes I \otimes I + \overline{A_0} \otimes A_1 \otimes I \otimes J_{+,s}$$
$$+ \overline{A_1} \otimes A_0 \otimes J_{+,r} \otimes I + \overline{A_1} \otimes A_1 \otimes J_{+,r} \otimes J_{+,s} \;, \tag{2.99}$$

where

$$A_0 = \sigma_+ = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \qquad \text{and} \qquad A_1 = \begin{pmatrix} 1 & 0 \\ 0 & \omega \end{pmatrix}$$

are the defining tensors of the original state $|\Psi\rangle$ (cf. (2.7.3)). With this, we can give the proof of the above theorem as follows.

Proof of Theorem 2.7.5. Observe that in the standard basis of the spin-j-representation, the raising operator J_+ is strictly lower diagonal. From (2.7.6) and the definition of A_0 and A_1, it follows that the transfer operator $E_{r,s}$ is lower diagonal in the tensor product basis (consisting of these standard bases and the computational basis of $\mathbb{C}^2$) since each term in the sum is a tensor product of lower diagonal matrices. In fact, every term except

$$D \equiv \overline{A_1} \otimes A_1 \otimes I \otimes I$$

is strictly lower diagonal. Therefore, we see that the eigenvalues of $E_{r,s}$ are given by the diagonal entries of D, and consist of the eigenvalue 1 with multiplicity $2(r+1)(s+1)$, and the eigenvalues ω and $\overline{\omega}$, both with multiplicity $(r+1)(s+1)$. Observe that A_0 and A_1 commute up to a factor of ω, that is,

$$A_1 A_0 = \omega A_0 A_1 . \tag{2.100}$$

To shorten some of the expressions, let us define

$$N_1 = \overline{A_0} \otimes A_0 , \qquad N_2 = \overline{A_0} \otimes A_1 , \qquad N_3 = \overline{A_1} \otimes A_0 , \qquad \text{and} \qquad A = \overline{A_1} \otimes A_1 .$$

We note that each N_i is a nilpotent matrix of order 2, i.e.,

$$N_i^2 = 0 \qquad \text{for } i = 1, 2, 3 \tag{2.101}$$

since $A_0^2 = 0$. Moreover, for the same reason and (2.7.6), we have

$$N_2 N_1 = N_1 N_2 = N_3 N_1 = N_1 N_3 = 0 \qquad \text{and} \qquad N_2 N_3 = \omega^2 N_3 N_2 .$$

Equation (2.7.6) also implies that

$$N_i A = q_i(\omega) A N_i \qquad \text{for} \qquad i = 1, 2, 3 , \tag{2.102}$$

where $q_i(\omega) \in \{1, \omega, \overline{\omega}\}$. Now consider the transfer operator with its diagonal term removed, i.e.,

$$E_{r,s} - D = N_1 \otimes I \otimes I + N_2 \otimes I \otimes J_{+,s} + N_3 \otimes J_{+,r} \otimes I + A \otimes J_{+,r} \otimes J_{+,s} .$$

Let $\mathbb{C}[\omega, \overline{\omega}]$ be the set of polynomials in ω and $\overline{\omega}$. Let us define the set

$$\mathcal{X} = \left\{ p_1(\omega, \overline{\omega}) \, N_1 \otimes I \otimes I + p_2(\omega, \overline{\omega}) \, N_2 \otimes I \otimes J_{+,s} \right.$$
$$\left. + p_3(\omega, \overline{\omega}) \, N_3 \otimes J_{+,r} \otimes I + p_4(\omega, \overline{\omega}) \, A \otimes J_{+,r} \otimes J_{+,s} \;\middle|\; p_i \in \mathbb{C}[\omega, \overline{\omega}] \right\}$$

such that $E_{r,s} - D \in \mathcal{X}$. The key properties of $\mathcal{X}$ which we need are the following:

(i) If $X_1 \in \mathcal{X}$, then $D X_1 = X_2 D$ and $X_1 D = D X_3$ for some $X_2, X_3 \in \mathcal{X}$.

(ii) The product of any $\min\{r, s\} + 2$ operators in $\mathcal{X}$ is equal to zero.

Property (i) follows immediately from the commutation relation (2.7.6) because $D = A \otimes I \otimes I$. Similarly, property (ii) follows from the nilpotency relation (2.7.6), the commutation relation (2.7.6), and the fact that

$$J_{+,r}^{\min\{r,s\}+2} = J_{+,s}^{\min\{r,s\}+2} = 0 \ .$$

We can write these two properties succintly as equalities of sets, that is,

$$D\mathcal{X} = \mathcal{X}D \ , \qquad \text{and} \qquad \qquad \qquad (2.103)$$

$$\mathcal{X}^m = \{0\} \qquad \text{for all } m \geq \min\{r, s\} + 2 \ , \qquad (2.104)$$

where e.g., $\mathcal{X}^2 = \{X_1 X_2 \,|\, X_1, X_2 \in \mathcal{X}\}$. Let us write $D_\lambda = D - \lambda I$. Then

$$E_{r,s} - \lambda I = D_\lambda + (E_{r,s} - D) \in D_\lambda + \mathcal{X} \ .$$

In particular, for $\ell, m, n \in \mathbb{N}_0$ we have

$$(E_{r,s} - I)^\ell (E_{r,s} - \omega I)^m (E_{r,s} - \overline{\omega} I)^n \in (D_1 + \mathcal{X})^\ell (D_\omega + \mathcal{X})^m (D_{\overline{\omega}} + \mathcal{X})^n$$
$$\subseteq \sum_{\substack{a \in \{0,\ldots,\ell\} \\ b \in \{0,\ldots,m\} \\ c \in \{0,\ldots,n\}}} D_1^a D_\omega^b D_{\overline{\omega}}^c \mathcal{X}^{(\ell-a)+(m-b)+(n-c)} \ ,$$

where in the last step, we used the binomial expansion, the pairwise commutativity of the matrices D_1, D_ω, and $D_{\overline{\omega}}$, and (2.7.6). Since $D_1 D_\omega D_{\overline{\omega}} = 0$, the non-zero terms in the expansion must have at least one of a, b, c equal to zero. Choosing

$$\ell = m = n = \min\{r, s\} + 2 \ ,$$

the exponent $(\ell - a) + (m - b) + (n - c)$ is lower bounded by $\min\{r, s\} + 2$ for any such triple (a, b, c). We conclude with (2.7.6) that

$$\sum_{\substack{a \in \{0,\ldots,\ell\} \\ b \in \{0,\ldots,m\} \\ c \in \{0,\ldots,n\}}} D_1^a D_\omega^b D_{\overline{\omega}}^c \mathcal{X}^{(\ell-a)+(m-b)+(n-c)} = \{0\} \ ,$$

and thus

$$(E_{r,s} - I)^{\min\{r,s\}+2} (E_{r,s} - \omega I)^{\min\{r,s\}+2} (E_{r,s} - \overline{\omega} I)^{\min\{r,s\}+2} = 0 \ .$$

Therefore the minimal polynomial of $E_{r,s}$ must divide $p(x) = [(x - 1)(x - \omega)(x - \overline{\omega})]^{\min\{r,s\}+2}$. Thus the Jordan blocks of $E_{r,s}$ are bounded above in size by $\min\{r, s\} + 2$, as claimed. $\qquad \square$

2.7.7 Matrix Elements of Magnon States

With the established bounds on the Jordan structure of $E_{r,s}$, we can derive upper bounds on overlaps of magnon states. Recall that $|\Psi_r\rangle = S_-^r |\Psi\rangle$ for $r = 0, \ldots, n-2$ and $|\psi_r\rangle$ is its normalized version (cf. (2.7.1)) .

Theorem 2.7.6. *Let $F \in \mathcal{B}((\mathbb{C}^2)^{\otimes d})$ be such that $\|F\| \leq 1$. Let $r \neq s$. Then*

$$|\langle \psi_r |(F \otimes I^{\otimes(n-d)})|\psi_s\rangle| = O\left(\frac{d}{n^{|s-r|/2}}\right) .$$

Proof. We can take the complex conjugate, effectively interchanging r and s. Thus we can, without loss of generality, assume that $r < s$. Recall that $|\Psi_r\rangle$ and $|\Psi_s\rangle$ can be represented as MPS using bond dimensions $D_r = 2(r+1)$, $D_s = 2(s+1)$ such that the associated transfer operators E_r, E_s and the combined transfer operator $E_{r,s}$ all have spectrum $\{1, \omega, \omega\}$ and Jordan blocks bounded by 2, 2, and $\min\{r, s\}+2 = r+2$, respectively; see Theorem 2.7.5. Applying Theorem 2.4.5 (with $h_1^* = 2$, $h_2^* = 2$, $h^* = r + 2$), we get

$$|\langle \Psi_r |(F \otimes I^{\otimes n-d})|\Psi_s\rangle| \leq 16 \cdot d(n-d)^{r+1} = O(d \cdot n^{r+1}) .$$

Inserting the normalization (2.7.1)

$$\|\Psi_s\|^2 = \frac{n(n-2)!s!}{(n-2-s)!} \geq s! \cdot n^{s+1}(1 - O(s^2/n))$$

gives

$$|\langle \psi_r |(F \otimes I^{\otimes n-d})|\psi_s\rangle| = \frac{dn^{r+1}}{(r!s!)^{1/2}n^{(r+s)/2+1}} \cdot (1 + O(s^2/n)) = O(d \cdot n^{-(s-r)/2})$$

as claimed. $\qquad\qquad\square$

We also need estimates for the difference of expectation values of magnon states. Let us first show that the reduced d-local operators are all essentially the same.

Lemma 2.7.7. *Let $\{|\psi_s\rangle\}_{s=0}^{n-2}$ be the normalized magnon-states defined in Equation (2.7.1). Then*

$$\langle 1|^{\otimes d} \left(\mathrm{tr}_{n-d} |\psi_s\rangle\langle\psi_s|\right) |1\rangle^{\otimes d} \geq 1 - O(ds/n)$$

for all $s \in [n]$.

Proof. Let us define

$$|\Psi_s^k\rangle = S_-^s \left(\sum_{j=1}^{k} \omega^j \sigma_j^- |1\rangle^{\otimes k} \right).$$

Observe that for $d < n$

$$|\Psi_0^n\rangle = |\Psi_0^d\rangle \otimes |1\rangle^{\otimes n-d} + \omega^d |1\rangle^{\otimes d} \otimes |\Psi_0^{n-d}\rangle.$$

Writing $S_- = S_-^A + S_-^B$ with $S_-^A = \sum_{j=1}^{d} \sigma_j^-$ and $S_-^B = \sum_{j=d+1}^{n} \sigma_j^-$, we get

$$\begin{aligned} |\Psi_s^n\rangle &= \sum_{\ell=0}^{s} \binom{s}{\ell} (S_-^A)^\ell (S_-^B)^{s-\ell} |\Psi_0^n\rangle \\ &= \omega^d |1\rangle^{\otimes d} \otimes (S_-^B)^s |\Psi_0^{n-d}\rangle + |\Phi\rangle \\ &= \omega^d |1\rangle^{\otimes d} \otimes |\Psi_s^{n-d}\rangle + |\Phi\rangle, \end{aligned}$$

for a vector $|\Phi\rangle \in (\mathbb{C}^p)^{\otimes n}$ satisfying $(|1\rangle\langle 1|^{\otimes d} \otimes I^{\otimes n-d})|\Phi\rangle = 0$. In particular, we have

$$(|1\rangle\langle 1|^{\otimes d} \otimes I^{\otimes(n-d)})|\Psi_s^n\rangle = \omega^d |1\rangle^{\otimes d} |\Psi_s^{n-d}\rangle.$$

Tracing out the $(n-d)$ qubits, it follows that

$$\langle 1|^{\otimes d} \operatorname{tr}_{n-d} \left(|\Psi_s^n\rangle\langle\Psi_s^n| \right) |1\rangle^{\otimes d} = \|\Psi_s^{n-d}\|^2.$$

Rewriting the term using the normalized vector $\psi_s^n = \Psi_s^n / \|\Psi_s^n\|$, we get

$$\langle 1|^{\otimes d} \left(\operatorname{tr}_{n-d} |\psi_s^n\rangle\langle\psi_s^n| \right) |1\rangle^{\otimes d} = \frac{\|\Psi_s^{n-d}\|^2}{\|\Psi_s^n\|^2}.$$

Observe that the norm $\|\Psi_s^k\|^2$ is a matrix element of the operator $E_{s,s}^k$. With Lemma 2.4.2 (iii), we obtain

$$\begin{aligned} \langle 1|^{\otimes d} \left(\operatorname{tr}_{n-d} |\psi_s^n\rangle\langle\psi_s^n| \right) |1\rangle^{\otimes d} &= \frac{c \cdot (n-d)^{\ell-1}(1 + O((n-d)^{-1}))}{c \cdot n^{\ell-1}(1 + O(n^{-1}))} \\ &= (1 - d/n)^{\ell-1}(1 + O((n-d)^{-1})) \\ &\geq (1 - d(\ell-1)/n) \cdot (1 + O((n-d)^{-1})) \\ &\geq 1 - \left(\frac{d(\ell-1)}{n} + O(1/n) \right) \end{aligned}$$

for some $\ell \in \{1, \ldots, h^*\}$, where h^* is the size of the largest Jordan block of the transfer operator $E_{s,s}$. Since $h^* \leq s + 2$ by Theorem 2.7.5, the claim follows.

$\square$

Theorem 2.7.8. *Let $F \in \mathcal{B}((\mathbb{C}^2)^{\otimes d})$ be such that $\|F\| \leq 1$. Fix some $s_0 \leq n - 2$. Then*

$$\left|\langle\psi_s|(F \otimes I^{\otimes n-d})|\psi_s\rangle - \langle\psi_r|(F \otimes I^{\otimes n-d})|\psi_r\rangle\right| = O\left(\sqrt{\frac{ds_0}{n}}\right) \qquad \text{for all} \qquad r, s \leq s_0 \ .$$

Proof. For any $F \in \mathcal{B}((\mathbb{C}^2)^{\otimes d})$ with $\|F\| \leq 1$, we have

$$\left|\langle\psi_s|(F \otimes I^{\otimes n-d})|\psi_s\rangle - \langle 1|^{\otimes d} F|1\rangle^{\otimes d}\right| \leq \|\operatorname{tr}_{n-d}|\psi_s\rangle\langle\psi_s| - |1\rangle\langle 1|^{\otimes n}\|$$

$$\leq \sqrt{1 - \langle 1|^{\otimes d} (\operatorname{tr}_{n-d}|\psi_s\rangle\langle\psi_s|) |1\rangle^{\otimes d}} \ ,$$

using the Fuchs — van de Graaf inequality $\frac{1}{2}\|\rho - |\varphi\rangle\langle\varphi|\|_1 \leq \sqrt{1 - \langle\varphi|\rho|\varphi\rangle}$ [78]. With Lemma 2.7.7 we get

$$\left|\langle\psi_s|(F \otimes I^{\otimes n-d})|\psi_s\rangle - \langle 1|^{\otimes d} F|1\rangle^{\otimes d}\right| \leq O\left(\sqrt{\frac{ds_0}{n}}\right) \ .$$

Using the triangle inequality, the claim follows. $\qquad\square$

2.7.8 The Parameters of the Magnon code

Our main result is the following:

Theorem 2.7.9 (Parameters of the magnon-code)**.** *Let $\nu \in (0, 1)$ and $\lambda, \kappa > 0$ be such that*

$$6\kappa + \lambda < \nu \ .$$

Then there is a subspace C spanned by descendant states $\{S_-^r|\Psi\rangle\}_r$ with magnetization r pairwise differing by at least 2 such that C is an $(\epsilon, \delta)[[n, k, d]]$-AQEDC with parameters

$$k = \kappa \log_2 n \ ,$$
$$d = n^{1-\nu} \ ,$$
$$\epsilon = \Theta(n^{-(\nu-(6\kappa+\lambda))}) \ ,$$
$$\delta = n^{-\lambda} \ .$$

Proof. We claim that the subspace

$$C = \operatorname{span}\{\psi_s \mid s \text{ even and } s \leq 2n^\kappa\}$$

spanned by a subset of magnon-states has the claimed property. Clearly, $\dim C = n^\kappa = 2^k$.

Let F be an arbitrary d-local operator on $(\mathbb{C}^2)^{\otimes n}$ of unit norm. According to Lemma 2.7.4, the following considerations concerning matrix elements $\langle \psi_q | F | \psi_p \rangle$ of magnon states do not depend on the location of the support of F as we are interested in the supremum over d-local operators F and unitary conjugation does not change the locality or the norm. Thus, we can assume that $F = \tilde{F} \otimes I^{\otimes n-d}$. That is, we have

$$\sup_{\substack{F\ d\text{-local} \\ \|F\| \leq 1}} |\langle \psi_r | F | \psi_s \rangle| = \sup_{\substack{\tilde{F} \in \mathcal{B}((\mathbb{C}^2)^{\otimes d}) \\ \|\tilde{F}\| \leq 1}} |\langle \psi_r |(\tilde{F} \otimes I^{\otimes n-d})| \psi_s \rangle| = O\left(\frac{d}{n^{|r-s|/2}}\right) \qquad \text{for } r, s \leq 2n^\kappa .$$

by Theorem 2.7.6. In particular, if $|r - s| \geq 2$, then this is bounded by $O(d/n)$. Similarly,

$$\sup_{\substack{F\ d\text{-local} \\ \|F\| \leq 1}} \left|\langle \psi_s | F | \psi_s \rangle - \langle \psi_r | F | \psi_r \rangle\right| = \sup_{\substack{\tilde{F} \in \mathcal{B}((\mathbb{C}^2)^{\otimes d}) \\ \|\tilde{F}\| \leq 1}} \left|\langle \psi_s |(\tilde{F} \otimes I^{\otimes(n-d)})| \psi_s \rangle - \langle \psi_r |(\tilde{F} \otimes I^{\otimes(n-d)})| \psi_r \rangle\right|$$

$$= O\left(\sqrt{\frac{dn^\kappa}{n}}\right) = O(\sqrt{dn^{\kappa-1}}) \qquad \text{for } r, s \leq 2n^\kappa$$

by Theorem 2.7.8. Since $d/n = O(\sqrt{dn^{\kappa-1}})$, we conclude that for all d-local operators F of unit norm, we have

$$\left|\langle \psi_r | F | \psi_s \rangle - \delta_{r,s} \langle \psi_0 | F | \psi_0 \rangle\right| = O(d^{1/2} n^{(\kappa-1)/2}) \qquad \text{for all } r, s \text{ even with } r, s \leq 2n^\kappa .$$

The sufficient conditions of Corollary 2.3.4 for approximate error-detection, applied with $\gamma = \Theta(d^{1/2} n^{(\kappa-1)/2})$, thus imply that C is an $(\Theta(2^{5k} dn^{\kappa-1})/\delta, \delta)[[n, k, d]]$-AQEDC for any δ satisfying

$$\delta > \Theta(2^{5k} dn^{\kappa-1}) = \Theta(n^{6\kappa-\nu})$$

for the choice $d = n^{1-\nu}$. With $\delta = n^{-\lambda}$, the claim follows. $\qquad \square$

Acknowledgements

2.A   Canonical Form of Excitation Ansatz States

For the reader's convenience, we include here a proof of the Lemma 2.6.1 following [77].

Lemma 2.6.1. *Let $|\Phi_p(B; A)\rangle$ be an injective excitation ansatz state and assume that A is normalized such that the transfer operator has spectral radius 1. Let ℓ and r be the corresponding left- and right- eigenvectors corresponding to eigenvalue 1. Assume $p \neq 0$. Then there exists a tensor $\tilde{B}$ such that $|\Phi_p(B; A)\rangle = |\Phi_p(\tilde{B}; A)\rangle$, and such that*

$$\langle\!\langle \ell | E_{\tilde{B}(p)} = 0 \qquad and \qquad \langle\!\langle \ell | E_{\overline{\tilde{B}(p)}} = 0 \, . \qquad ((2.6.1))$$

Proof. We note that the equations (2.6.1) can be written as

$$\sum_{i\in[\mathsf{p}]} A_i^\dagger \ell \tilde{B}_i = 0 \, , \qquad \text{and} \qquad \sum_{i\in[\mathsf{p}]} \tilde{B}_i^\dagger \ell A_i = 0 \, . \qquad (2.105)$$

Diagrammatically, they take the form

where square and round boxes correspond to $\tilde{B}$ and A, respectively.

Let the original MPS tensors be $A = \{A_j\}_{j=1}^{\mathsf{p}}$ and $B = \{B_j\}_{j=1}^{\mathsf{p}} \subset \mathcal{B}(\mathbb{C}^D \otimes \mathbb{C}^D)$. Suppose $X \in \mathcal{B}(\mathbb{C}^D)$ is invertible. Define the MPS tensor $C = \{\tilde{B}_j\}_{j=1}^{\mathsf{p}} \subset \mathcal{B}(\mathbb{C}^D)$ by

$$C_j = A_j X - e^{-ip} X A_j \qquad \text{for } j = 1, \ldots, \mathsf{p}.$$

that is,

$$(2.106)$$

It is then easy to check that

$$|\Phi_p(B; A)\rangle = |\Phi_p(B + C; A)\rangle,$$

where $B + C$ is the MPS tensor obtained by setting $(B + C)_j = B_j + C_j$ for each $j = 1, \ldots, \mathsf{p}$. Indeed, the difference of these two vectors is

$$|\Phi_p(B + C; A)\rangle - |\Phi_p(B; A)\rangle$$

$$= \sum_{i_1, \ldots, i_n \in [\mathsf{p}]} \sum_{k=1}^{n} e^{ipk} \operatorname{tr}(A_{i_1} \cdots A_{i_{k-1}} C_{i_k} A_{i_{k+1}} \cdots A_{i_n}) |i_1 \cdots i_n\rangle$$

$$= \sum_{i_1, \ldots, i_n \in [\mathsf{p}]} \left(\sum_{k=1}^{n} e^{ipk} \operatorname{tr}\left[A_{i_1} \cdots A_{i_{k-1}} (A_{i_k} X - e^{-ip} X A_{i_k}) A_{i_{k+1}} \cdots A_{i_n} \right] \right) |i_1 \cdots i_n\rangle$$

$$= 0,$$

since the terms in the square brackets vanish because of the cyclicity of the trace (alternatively, this can be seen by substituting each square box (corresponding to B) in Figure 2.5 by a formal linear combination of a square box (B) and diagram (2.A)).

Observe that the second equation in (2.A) can be obtained from the first by taking the adjoint since ℓ is a self-adjoint operator. It thus suffices to show that there is an MPS tensor $\tilde{B}$ with the desired property $|\Phi_p(\tilde{B}; A)\rangle = |\Phi_p(B; A)\rangle$ such that

$$\sum_{i \in [\mathsf{p}]} A_i^\dagger \ell \tilde{B}_i = 0. \qquad (2.107)$$

It turns out that setting $\tilde{B} = B + C$ for an appropriate choice of X (and thus C) suffices. Equation (2.A) then amounts to the identity

$$\sum_{j \in [\mathsf{p}]} A_j^\dagger \ell (B_j + A_j X - e^{-ip} X A_j) = 0, \qquad (2.108)$$

or diagrammatically,

$$\ell \,\left[\!\!\begin{array}{c} \bar{A} \\ \\ \tilde{B} \end{array}\!\!\right] = \ell \,\left[\!\!\begin{array}{c} \bar{A} \\ \\ B \end{array}\!\!\right] + \ell \,\left[\!\!\begin{array}{c} \bar{A} \\ \\ A\ \ X \end{array}\!\!\right] - e^{-ip}\,\ell \,\left[\!\!\begin{array}{c} \bar{A} \\ \\ X\ \ A \end{array}\!\!\right] = 0 \ .$$

Because ℓ is the unique eigenvector of $\mathcal{E}^\dagger(\rho) = \sum_{j\in[\mathsf{p}]} A_j^\dagger \rho A_j$ to eigenvalue 1, Equation (2.A) simplifies to

$$\sum_{j\in[\mathsf{p}]} A_j^\dagger \ell B_j + \ell X - e^{-ip} \sum_{j\in[\mathsf{p}]} A_j^\dagger \ell X A_j = 0\,,$$

or

$$\ell \,\left[\!\!\begin{array}{c} \bar{A} \\ \\ B \end{array}\!\!\right] + \ell \,\left[\!\!\begin{array}{c} \\ \\ X \end{array}\!\!\right] - e^{-ip}\,\ell \,\left[\!\!\begin{array}{c} \bar{A} \\ \\ X\ \ A \end{array}\!\!\right] = 0 \ .$$

Since ℓ is full rank, we may substitute $X = \ell^{-1} Y$. Then (2.A) is satisfied if

$$\ell \,\left[\!\!\begin{array}{c} \bar{A} \\ \\ B \end{array}\!\!\right] + \left[\!\!\begin{array}{c} \\ Y \end{array}\!\!\right] - e^{-ip} \left[\!\!\begin{array}{c} \bar{A} \\ Y \\ A \end{array}\!\!\right] = 0 \ ,$$

or

$$\sum_{j\in[\mathsf{p}]} A_j^\dagger \ell B_j + (I - e^{-ip}\mathcal{E})(Y) = 0\,.$$

Because 1 is the unique eigenvalue of magnitude 1 of $\mathcal{E}$, the map $(\lambda I - e^{-ip}\mathcal{E})$ is invertible under the assumption that $p \neq 0$, and we obtain the solution

$$X = \ell^{-1} Y$$

$$= -\ell^{-1} \left(I - e^{-ip}\mathcal{E}\right)^{-1} \left(\sum_{j\in[\mathsf{p}]} A_j^\dagger \ell B_j \right)$$

to Equation (2.A), proving the claim for $p \neq 0$. $\qquad\square$

BIBLIOGRAPHY

[1] P. W. Shor, "Scheme for reducing decoherence in quantum computer memory," *Physical Review A*, vol. 52, no. 4, R2493, 1995. DOI: `10.1103/PhysRevA.52.R2493`.

[2] A. R. Calderbank and P. W. Shor, "Good quantum error-correcting codes exist," *Physical Review A*, vol. 54, no. 2, p. 1098, 1996. DOI: `10.1103/PhysRevA.54.1098`.

[3] A. Y. Kitaev, "Quantum computations: Algorithms and error correction," *Russian Mathematical Surveys*, vol. 52, no. 6, pp. 1191–1249, 1997. DOI: `10.1070/RM1997v052n06ABEH002155`.

[4] E. Knill and R. Laflamme, "Theory of quantum error-correcting codes," *Physical Review A*, vol. 55, no. 2, p. 900, 1997. DOI: `10.1103/PhysRevLett.84.2525`.

[5] E. Knill, R. Laflamme, and W. H. Zurek, "Resilient quantum computation," *Science*, vol. 279, no. 5349, pp. 342–345, 1998. DOI: `10.1126/science.279.5349.342`.

[6] A. Y. Kitaev, "Fault-tolerant quantum computation by anyons," *Annals of Physics*, vol. 303, no. 1, pp. 2–30, 2003. DOI: `10.1016/S0003-4916(02)00018-0`.

[7] M. Freedman, A. Kitaev, M. Larsen, and Z. Wang, "Topological quantum computation," *Bulletin of the American Mathematical Society*, vol. 40, no. 1, pp. 31–38, 2003. arXiv: `quant-ph/0101025`.

[8] R. W. Ogburn and J. Preskill, "Topological quantum computation," in *Quantum computing and quantum communications*, Springer, 1999, pp. 341–356. DOI: `10.1007/3-540-49208-9_31`.

[9] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, "Topological quantum memory," *Journal of Mathematical Physics*, vol. 43, no. 9, pp. 4452–4505, 2002. DOI: `10.1063/1.1499754`.

[10] C. Nayak, S. H. Simon, A. Stern, M. Freedman, and S. D. Sarma, "Non-abelian anyons and topological quantum computation," *Reviews of Modern Physics*, vol. 80, no. 3, p. 1083, 2008. DOI: `10.1103/RevModPhys.80.1083`.

[11] R. Raussendorf, J. Harrington, and K. Goyal, "A fault-tolerant one-way quantum computer," *Annals of physics*, vol. 321, no. 9, pp. 2242–2270, 2006. DOI: `10.1016/j.aop.2006.01.012`.

[12] A. Stern and N. H. Lindner, "Topological quantum computation - from basic concepts to first experiments," *Science*, vol. 339, no. 6124, pp. 1179–1184, 2013. DOI: `10.1126/science.1231473`.

[13] B. M. Terhal, "Quantum error correction for quantum memories," *Reviews of Modern Physics*, vol. 87, no. 2, p. 307, 2015. DOI: `10.1103/RevModPhys.87.307`.

[14] A. Kitaev, "Anyons in an exactly solved model and beyond," *Annals of Physics*, vol. 321, no. 1, pp. 2–111, 2006. DOI: `10.1016/j.aop.2005.10.005`.

[15] M. A. Levin and X.-G. Wen, "String-net condensation: A physical mechanism for topological phases," *Physical Review B*, vol. 71, no. 4, p. 045 110, 2005. DOI: `10.1103/PhysRevB.71.045110`.

[16] A. Kitaev, "Periodic table for topological insulators and superconductors," in *AIP Conference Proceedings*, AIP, vol. 1134, 2009, pp. 22–30. DOI: `10.1063/1.3149495`.

[17] L. Fidkowski and A. Kitaev, "Topological phases of fermions in one dimension," *Physical Review B*, vol. 83, no. 7, p. 075 103, 2011. DOI: `10.1103/PhysRevB.83.075103`.

[18] X. Chen, Z.-C. Gu, and X.-G. Wen, "Classification of gapped symmetric phases in one-dimensional spin systems," *Physical Review B*, vol. 83, no. 3, p. 035 107, 2011. DOI: `10.1103/PhysRevB.83.035107`.

[19] ——, "Complete classification of one-dimensional gapped quantum phases in interacting spin systems," *Physical Review B*, vol. 84, no. 23, p. 235 128, 2011. DOI: `10.1103/PhysRevB.84.235128`.

[20] X. Chen, Z.-X. Liu, and X.-G. Wen, "Two-dimensional symmetry-protected topological orders and their protected gapless edge excitations," *Physical Review B*, vol. 84, no. 23, p. 235 141, 2011. DOI: `10.1103/PhysRevB.84.235141`.

[21] X. Chen, Z.-C. Gu, Z.-X. Liu, and X.-G. Wen, "Symmetry protected topological orders and the group cohomology of their symmetry group," *Physical Review B*, vol. 87, no. 15, p. 155 114, 2013. DOI: `10.1103/PhysRevB.87.155114`.

[22] D. Harlow, "The Ryu–Takayanagi formula from quantum error correction," *Communications in Mathematical Physics*, vol. 354, no. 3, pp. 865–912, 2017. DOI: `10.1007/s00220-017-2904-z`.

[23] M. Fannes, B. Nachtergaele, and R. F. Werner, "Finitely correlated states on quantum spin chains," *Communications in Mathematical Physics*, vol. 144, no. 3, pp. 443–490, 1992. DOI: `10.1007/BF02099178`.

[24] S. R. White, "Density matrix formulation for quantum renormalization groups," *Physical Review Letters*, vol. 69, no. 19, p. 2863, 1992. DOI: `10.1103/PhysRevLett.69.2863`.

[25] ——, "Density-matrix algorithms for quantum renormalization groups," *Physical Review B*, vol. 48, no. 14, p. 10 345, 1993. DOI: `10.1103/PhysRevB.48.10345`.

[26] G. Vidal, "Efficient classical simulation of slightly entangled quantum computations," *Physical Review Letters*, vol. 91, no. 14, p. 147 902, 2003. DOI: `10.1103/PhysRevLett.91.147902`.

[27] ——, "Efficient simulation of one-dimensional quantum many-body systems," *Physical Review Letters*, vol. 93, no. 4, p. 040 502, 2004. DOI: `10.1103/PhysRevLett.93.040502`.

[28] D. Pérez-García, F. Verstraete, M. M. Wolf, and J. I. Cirac, "Matrix product state representations," *Quantum Information and Computation*, vol. 7, no. 401, 2007. DOI: `10.26421/QIC7.5-6-1`.

[29] F. Verstraete and J. I. Cirac, "Renormalization algorithms for quantum-many body systems in two and higher dimensions," 2004. arXiv: `cond-mat/0407066`.

[30] F. Verstraete, V. Murg, and J. I. Cirac, "Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems," *Advances in Physics*, vol. 57, no. 2, pp. 143–224, 2008. DOI: `10.1080/14789940801912366`.

[31] G. Vidal, "Class of quantum many-body states that can be efficiently simulated," *Physical Review Letters*, vol. 101, no. 11, p. 110 501, 2008. DOI: `10.1103/PhysRevLett.101.110501`.

[32] F. Verstraete, M. Wolf, D. Pérez-García, and J. I. Cirac, "Projected entangled states: Properties and applications," *International Journal of Modern Physics B*, vol. 20, no. 30n31, pp. 5142–5153, 2006. DOI: `10.1142/S021797920603620X`.

[33] D. Pérez-García, F. Verstraete, J. I. Cirac, and M. M. Wolf, "PEPS as unique ground states of local hamiltonians," 2007. arXiv: `0707.2260`.

[34] C. V. Kraus, N. Schuch, F. Verstraete, and J. I. Cirac, "Fermionic projected entangled pair states," *Physical Review A*, vol. 81, no. 5, p. 052 338, 2010. DOI: `10.1103/PhysRevA.81.052338`.

[35] M. Schwarz, K. Temme, and F. Verstraete, "Preparing projected entangled pair states on a quantum computer," *Physical Review Letters*, vol. 108, no. 11, p. 110 502, 2012. DOI: `10.1103/PhysRevLett.108.110502`.

[36] M. Fishman, L. Vanderstraeten, V. Zauner-Stauber, J. Haegeman, and F. Verstraete, "Faster methods for contracting infinite two-dimensional tensor networks," *Physical Review B*, vol. 98, no. 23, p. 235 148, 2018. DOI: `10.1103/PhysRevB.98.235148`.

[37] O. Buerschaper, M. Aguado, and G. Vidal, "Explicit tensor network representation for the ground states of string-net models," *Physical Review B*, vol. 79, no. 8, p. 085 119, 2009. DOI: `10.1103/PhysRevB.79.085119`.

[38] Z.-C. Gu, M. Levin, B. Swingle, and X.-G. Wen, "Tensor-product representations for string-net condensed states," *Physical Review B*, vol. 79, no. 8, p. 085 118, 2009. DOI: `10.1103/PhysRevB.79.085118`.

[39] R. König, B. W. Reichardt, and G. Vidal, "Exact entanglement renormalization for string-net models," *Physical Review B*, vol. 79, no. 19, p. 195 123, 2009. DOI: `10.1103/PhysRevB.79.195123`.

[40] N. Schuch, I. Cirac, and D. Pérez-García, "PEPS as ground states: Degeneracy and topology," *Annals of Physics*, vol. 325, no. 10, pp. 2153–2192, 2010. DOI: `10.1016/j.aop.2010.05.008`.

[41] O. Buerschaper, "Twisted injectivity in projected entangled pair states and the classification of quantum phases," *Annals of Physics*, vol. 351, pp. 447–476, 2014. DOI: `10.1016/j.aop.2014.09.007`.

[42] M. B. Şahinoğlu, D. Williamson, N. Bultinck, M. Mariën, J. Haegeman, N. Schuch, and F. Verstraete, "Characterizing topological order with matrix product operators," 2014. arXiv: `1409.2150`.

[43] D. J. Williamson, N. Bultinck, M. Mariën, M. B. Şahinoğlu, J. Haegeman, and F. Verstraete, "Matrix product operators for symmetry-protected topological phases: Gauging and edge theories," *Physical Review B*, vol. 94, no. 20, p. 205 150, 2016. DOI: `10.1103/PhysRevB.94.205150`.

[44] N. Bultinck, M. Mariën, D. J. Williamson, M. B. Şahinoğlu, J. Haegeman, and F. Verstraete, "Anyons and matrix product operator algebras," *Annals of Physics*, vol. 378, pp. 183–233, 2017. DOI: `10.1016/j.aop.2017.01.004`.

[45] M. B. Şahinoğlu, M. Walter, and D. J. Williamson, "A tensor network framework for topological order in higher dimensions," *Bulletin of the American Physical Society*, 2016.

[46] B. Swingle, "Entanglement renormalization and holography," *Physical Review D*, vol. 86, no. 6, p. 065 007, 2012. DOI: `10.1103/PhysRevD.86.065007`.

[47] F. Pastawski, B. Yoshida, D. Harlow, and J. Preskill, "Holographic quantum error-correcting codes: Toy models for the bulk/boundary correspondence," *Journal of High Energy Physics*, no. 6, p. 149, Jun. 2015. DOI: `10.1007/JHEP06(2015)149`.

[48] P. Hayden, S. Nezami, X.-L. Qi, N. Thomas, M. Walter, and Z. Yang, "Holographic duality from random tensor networks," *Journal of High Energy Physics*, no. 11, p. 009, 2016. DOI: `10.1007/JHEP11(2016)009`.

[49] C. Akers and P. Rath, "Holographic Rényi entropy from quantum error correction," 2018. arXiv: `1811.05171`.

[50] X. Dong, D. Harlow, and D. Marolf, "Flat entanglement spectra in fixed-area states of quantum gravity," 2018. arXiv: `1811.05382`.

[51] F. G. Brandao, E. Crosson, M. B. Şahinoğlu, and J. Bowen, "Quantum error correcting codes in eigenstates of translation-invariant spin chains," 2017. arXiv: 1710.04631.

[52] I. H. Kim and M. J. Kastoryano, "Entanglement renormalization, quantum error correction, and bulk causality," *Journal of High Energy Physics*, no. 4, p. 40, Apr. 2017, ISSN: 1029-8479. DOI: 10.1007/JHEP04(2017)040.

[53] F. Pastawski, J. Eisert, and H. Wilming, "Towards holography via quantum source-channel codes," *Physical Review Letters*, vol. 119, no. 2, p. 020 501, 2017. DOI: 10.1103/PhysRevLett.119.020501.

[54] T. M. Stace, S. D. Barrett, and A. C. Doherty, "Thresholds for topological codes in the presence of loss," *Physical Review Letters*, vol. 102, p. 200 501, 20 May 2009. DOI: 10.1103/PhysRevLett.102.200501.

[55] H. Pollatsek and M. B. Ruskai, "Permutationally invariant codes for quantum error correction," *Linear Algebra and its Applications*, vol. 392, pp. 255–288, 2004. DOI: 10.1016/j.laa.2004.06.014.

[56] Y. Ouyang, "Permutation-invariant quantum codes," *Physical Review A*, vol. 90, no. 6, p. 062 317, 2014. DOI: 10.1103/PhysRevA.90.062317.

[57] S. Bravyi, M. B. Hastings, and S. Michalakis, "Topological quantum order: Stability under local perturbations," *Journal of Mathematical Physics*, vol. 51, no. 9, p. 093 512, 2010. DOI: 10.1063/1.3490195.

[58] M. B. Hastings, "Topological order at nonzero temperature," *Physical Review Letters*, vol. 107, p. 210 501, 21 Nov. 2011. DOI: 10.1103/PhysRevLett.107.210501.

[59] T. Ogawa and H. Nagaoka, "A new proof of the channel coding theorem via hypothesis testing in quantum information theory," in *Proceedings IEEE International Symposium on Information Theory,*, Jun. 2002, pp. 73–. DOI: 10.1109/ISIT.2002.1023345.

[60] M. Aguado and G. Vidal, "Entanglement renormalization and topological order," *Physical Review Letters*, vol. 100, p. 070 404, 7 Feb. 2008. DOI: 10.1103/PhysRevLett.100.070404.

[61] M. B. Hastings and T. Koma, "Spectral gap and exponential decay of correlations," *Communications in Mathematical Physics*, vol. 265, no. 3, pp. 781–804, 2006. DOI: 10.1007/s00220-006-0030-4.

[62] F. G. Brandão and M. Horodecki, "An area law for entanglement from exponential decay of correlations," *Nature Physics*, vol. 9, no. 11, p. 721, 2013. DOI: 10.1038/nphys2747.

[63] ——, "Exponential decay of correlations implies area law," *Communications in Mathematical Physics*, vol. 333, no. 2, pp. 761–798, 2015. DOI: 10.1007/s00220-014-2213-8.

[64]   N. Schuch, D. Pérez-García, and I. Cirac, "Classifying quantum phases using matrix product states and projected entangled pair states," *Physical Review B*, vol. 84, p. 165 139, 16 Oct. 2011. DOI: `10.1103/PhysRevB.84.165139`.

[65]   M. B. Hastings, "Solving gapped Hamiltonians locally," *Physical Review B*, vol. 73, no. 8, p. 085 115, 2006. DOI: `10.1103/PhysRevB.73.085115`.

[66]   J. Haegeman, S. Michalakis, B. Nachtergaele, T. J. Osborne, N. Schuch, and F. Verstraete, "Elementary excitations in gapped quantum spin systems," *Physical Review Letters*, vol. 111, no. 8, p. 080 401, 2013. DOI: `10.1103/PhysRevLett.111.080401`.

[67]   V. Murg, V. E. Korepin, and F. Verstraete, "Algebraic Bethe ansatz and tensor networks," *Physical Review B*, vol. 86, no. 4, 2012. DOI: `10.1103/PhysRevB.86.045125`.

[68]   C. Crépeau, D. Gottesman, and A. Smith, "Approximate quantum error-correcting codes and secret sharing schemes," in *Advances in Cryptology – EUROCRYPT 2005*, R. Cramer, Ed., Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 285–301, ISBN: 978-3-540-32055-5.

[69]   H. Barnum, E. Knill, and M. A. Nielsen, "On quantum fidelities and channel capacities," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1317–1329, Jul. 2000, ISSN: 0018-9448. DOI: `10.1109/18.850671`.

[70]   E. Knill, R. Laflamme, A. Ashikhmin, H. Barnum, L. Viola, and W. Zurek, "Introduction to quantum error correction," 2002. arXiv: `quant-ph/0207170`.

[71]   C. Bény and O. Oreshkov, "General conditions for approximate quantum error correction and near-optimal recovery channels," *Physical Review Letters*, vol. 104, p. 120 501, 12 Mar. 2010. DOI: `10.1103/PhysRevLett.104.120501`.

[72]   D. E. Evans and R. Hoegh-Krohn, "Spectral properties of positive maps on C*-algebras," *Journal of the London Mathematical Society*, vol. s2-17, no. 2, pp. 345–355, 1978. DOI: `10.1112/jlms/s2-17.2.345`.

[73]   M. S. Ruiz, "Tensor networks in condensed matter," Ph.D. dissertation, Technische Universität München, 2011.

[74]   M. M. Wolf, "Quantum channels & operations," 2012. [Online]. Available: `https://www-m5.ma.tum.de/foswiki/pub/M5/Allgemeines/MichaelWolf/QChannelLecture.pdf`.

[75]   T. C. Bohdanowicz, E. Crosson, C. Nirkhe, and H. Yuen, "Good approximate quantum LDPC codes from spacetime circuit hamiltonians," 2018. arXiv: `1811.00277`.

[76]   J. Haegeman, T. Osborne, and F. Verstraete, "Post-matrix product state methods: To tangent space and beyond," *Physical Review B*, vol. 88, p. 075 133, 2013. DOI: `10.1103/PhysRevB.88.075133`.

[77] J. Haegeman, M. Mariën, T.Osborne, and F. Verstraete, "Geometry of matrix product states: Metric, parallel transport, and curvature," *Journal of Mathematical Physics*, vol. 55, no. 2, p. 021 902, 2014. DOI: `10.1063/1.4862851`.

[78] C. A. Fuchs and J. van de Graaf, "Cryptographic distinguishability measures for quantum mechanical states," 1997. arXiv: `quant-ph/9712042`.

Chapter 3

OBSTACLES TO STATE PREPARATION AND VARIATIONAL OPTIMIZATION FROM SYMMETRY PROTECTION

Local Hamiltonians with topological quantum order exhibit highly entangled ground states that cannot be prepared by shallow quantum circuits. Here, we show that this property may extend to all low-energy states in the presence of an on-site $\mathbb{Z}_2$ symmetry. This proves a version of the No Low-Energy Trivial States (NLTS) conjecture for a family of local Hamiltonians with symmetry protected topological order. A surprising consequence of this result is that the Goemans-Williamson algorithm outperforms the Quantum Approximate Optimization Algorithm (QAOA) for certain instances of MaxCut, at any constant level. We argue that the locality and symmetry of QAOA severely limits its performance. To overcome these limitations, we propose a non-local version of QAOA, and give numerical evidence that it significantly outperforms standard QAOA for frustrated Ising models on random 3-regular graphs.

This chapter is based on the published article:

3.1 Introduction

Classifying topological phases of matter is amongst the main objectives of modern condensed matter physics [1]. Central to this program is the characterization of entanglement structures that can emerge in ground states of many-body systems. Of particular interest are topologically non-trivial ground states [2]. Such states cannot be generated by a constant-depth quantum circuit starting from a product state. Non-trivial states exhibit complex, non-local entanglement properties and are thus expected to have highly non-classical features. Remarkably, certain gapped local Hamiltonians have non-trivial ground states. For example, preparing a ground state of Kitaev's toric code [3] from a product state requires a circuit depth growing at least polynomially in the system size using nearest-neighbor gates [4], and logarithmically using non-local gates [5].

Going beyond ground states, a natural next question is whether there are local Hamiltonians with the property that any low-energy state is non-trivial. Formalized by Freedman and Hastings [6], this is known as the No Low-energy Trivial States (NLTS) conjecture. To state it in detail, consider many-body systems composed of n finite-dimensional subsystems – assumed here to be qubits for simplicity. A local Hamiltonian is a sum of interaction terms acting non-trivially on $O(1)$ qubits each. We require that each term has operator norm $O(1)$ and each qubit is involved in $O(1)$ terms. The interaction terms may be long-range (no geometric locality is needed). It will be assumed that a local Hamiltonian H_n as above is defined for each $n \in \mathcal{I}$, where $\mathcal{I}$ is some infinite set of system sizes. A family of local Hamiltonians $\{H_n\}_{n \in \mathcal{I}}$ is said to have the NLTS property if there exists a constant $\epsilon > 0$ and a function $f : \mathbb{Z}_+ \to \mathbb{Z}_+$ such that:

1. H_n has ground state energy 0 for any $n \in \mathcal{I}$,

2. $\langle 0^n | U^\dagger H_n U | 0^n \rangle > \epsilon n$ for any depth-d circuit U composed of two-qubit gates and for any $n \geq f(d), n \in \mathcal{I}$.

Here the circuit depth d can be arbitrary. The conjecture is that the NLTS property holds for some family of local Hamiltonians.

The validity of the NLTS conjecture is a necessary condition for the stronger quantum PCP conjecture to hold [7]: the latter posits that there are local Hamiltonians whose ground state energy is QMA-hard to approximate with an extensive error ϵn for some constant $\epsilon > 0$.

A proof of the NLTS conjecture is still outstanding. Although many natural families of Hamiltonians provably do not have the NLTS property (see [8] for a comprehensive list), evidence for its validity has been provided by a number of related results. Ref. [6] constructs Hamiltonians satisfying a certain one-sided NLTS property: these have excitations of two kinds (similar to the toric code), and low-energy states with no excitations of the first kind are non-trivial. The construction crucially relies on expander graphs as even the one-sided NLTS property does not hold for similar constructions on regular lattices [6].

Eldar and Harrow [8] construct families of local Hamiltonians (also based on expander graphs) such that any state whose reduced density operators on a constant fraction of sites coincides with that of a ground state is non-trivial. This feature,

called the No Low-Error Trivial States (NLETS) property, is clearly related to robustness of entanglement in the ground state with respect to erasure errors [9], [10]. The existence of Hamiltonians with the NLETS property is a necessary condition [8] for the existence of good quantum LDPC codes, another central conjecture in quantum information.

Here we pursue a different approach to the NLTS conjecture by imposing an additional symmetry in the initial state as well as the preparation circuit. This mirrors similar considerations in the classification of topological phases, where the concept of symmetry-protected topological (SPT) phases [11] has been extremely fruitful. Indeed, the study of SPT equivalence classes of states, pioneered in [12], [13], has led to a complete classification of 1D phases [11], [14], [15], and also plays an essential role in measurement-based quantum computation [16], [17].

3.2 Implications for the Quantum Approximate Optimization Algorithm

For concreteness, we focus on the simplest case of onsite $\mathbb{Z}_2$-symmetry. A local Hamiltonian is said to be $\mathbb{Z}_2$-symmetric if all interaction terms commute with $X^{\otimes n}$, where X is the single-qubit Pauli-X operator. Likewise, a quantum circuit U acting on n qubits is said to be $\mathbb{Z}_2$-symmetric if it obeys

$$UX^{\otimes n} = X^{\otimes n}U \ .$$

We do not impose the symmetry on the individual gates of U, though this will naturally be the case in many interesting examples, such as the QAOA circuits considered below. Finally, let us say that a state Ψ of n qubits is $\mathbb{Z}_2$-symmetric if $X^{\otimes n}\Psi = \pm\Psi$. Our first result is a proof of the NLTS conjecture in the presence of onsite $\mathbb{Z}_2$-symmetry:

Theorem 3.2.1. *There exist constants $\epsilon, c > 0$ and a family of $\mathbb{Z}_2$-symmetric local Hamiltonians $\{H_n\}_{n \in \mathcal{I}}$ such that H_n has ground state energy 0 for any $n \in \mathcal{I}$ while*

$$\langle\varphi|U^\dagger H_n U|\varphi\rangle > \epsilon n \tag{3.1}$$

for any $\mathbb{Z}_2$-symmetric depth-d circuit U composed of two-qubit gates, any $\mathbb{Z}_2$-symmetric product state φ, and any $n \geq 2^{cd}$, $n \in \mathcal{I}$.

Our starting point to establish Theorem 3.2.1 is a fascinating result by Eldar and Harrow stated as Corollary 43 in [8]. It shows that the output distribution of a shallow quantum circuit cannot assign a non-negligible probability to subsets of bit strings

that are separated far apart w.r.t. the Hamming distance. More precisely, define the distribution $p(x) = |\langle x|U|\varphi\rangle|^2$, where $x \in \{0, 1\}^n$. Given a subset $S \subseteq \{0, 1\}^n$, let $p(S) = \sum_{x \in S} p(x)$.

Fact 1 ([8]). *For all subsets $S, S' \subseteq \{0, 1\}^n$, one has*

$$\text{dist}(S, S') \leq \frac{4n^{1/2}2^{3d/2}}{\min\{p(S), p(S')\}}.$$

Here $\text{dist}(S, S')$ is the Hamming distance, i.e. the minimum number of bit flips required to get from S to S'. We emphasize that Eq. (3.1) holds for all depth-d circuits U and all product states φ ($\mathbb{Z}_2$-symmetry is not needed).

Given a bit string x, let $\overline{x}$ be the bit-wise negation of x. Note that $p(x) = p(\overline{x})$ since $U\varphi$ is $\mathbb{Z}_2$-symmetric. Choose S and S' as the sets of all n-bit strings with the Hamming weight $\leq n/3$ and $\geq 2n/3$, respectively. Then $p(S') = p(S)$ and $\text{dist}(S, S') = n/3$. Eq. (3.1) gives

$$p(S) \leq 12n^{-1/2}2^{3d/2}.$$

Our strategy is to choose the Hamiltonian H_n such that low-energy states of H_n are concentrated on bit strings with the Hamming weight close to 0 or n such that $p(S)$ is non-negligible. Then Eq. (3.2) provides a logarithmic lower bound on the depth d for symmetric low-energy states.

Suppose $G = (V, E)$ is a graph with n vertices. The Cheeger constant of G is defined as

$$h(G) = \min_{\substack{S \subseteq V \\ 0 < |S| \leq n/2}} \frac{|\partial S|}{|S|}, \tag{3.4}$$

where $\partial S \subseteq E$ is the subset of edges that have exactly one endpoint in S. Families of expander graphs are infinite collections of bounded degree graphs $\{G_n\}_{n \in \mathcal{I}}$ whose Cheeger constant is lower bounded by a constant, i.e. $h(G_n) \geq h > 0$ for all $n \in \mathcal{I}$. Explicit constructions of degree-3 expanders can be found in [18]. Fix a family of degree-3 expanders $\{G_n\}_{n \in \mathcal{I}}$ and define H_n as the ferromagnetic Ising model on the graph G_n, i.e.

$$H_n = \frac{1}{2} \sum_{(u,v) \in E} (I - Z_u Z_v).$$

Here Z_u is the Pauli-Z applied to a site u and I is the identity. Clearly, H_n is $\mathbb{Z}_2$-symmetric, each term in H_n acts on two qubits and each qubit is involved

in three terms. Thus $\{H_n\}_{n\in I}$ is a family of local Hamiltonians. H_n has $\mathbb{Z}_2$-symmetric ground states $\frac{1}{\sqrt{2}}(|0^n\rangle \pm |1^n\rangle)$ with zero energy. Given a bit string x, let $\mathsf{supp}(x) = \{j \in [n] : x_j = 1\}$ be the support of x. From equations (3.2) and (3.3), one gets

$$\langle x|H_n|x\rangle = |\partial\, \mathsf{supp}(x)| \geq h \cdot \min\{|x|, n - |x|\},$$

where $|x|$ is the Hamming weight of x. Assume Eq. (3.2.1) is false. Then $p(x)$ is a low-energy distribution such that $\sum_x p(x)\langle x|H_n|x\rangle \leq \epsilon n$. By Markov's inequality, $p(x)$ has a non-negligible weight on low-energy basis states,

$$p(S_{\text{low}}) \geq 1/2, \qquad S_{\text{low}} = \{x : \langle x|H_n|x\rangle \leq 2\epsilon n\}.$$

By Eq. (3.4), $\min\{|x|, n - |x|\} \leq 2n\epsilon h^{-1}$ for all $x \in S_{\text{low}}$. Choose $\epsilon = h/6$. Then $S_{\text{low}} \subseteq S \cup S'$ and $p(S_{\text{low}}) \leq p(S) + p(S') = 2p(S)$. Here the last equality uses the symmetry of $p(x)$. By Eq. (3.5), $p(S) \geq 1/4$. Combining this and Eq. (3.2), one gets $n^{1/2} \leq 48 \cdot 2^{3d/2}$. We conclude that Eq. (3.2.1) holds whenever $n > 48^2 \cdot 8^d$. This proves Theorem 3.2.1.

The Hamiltonians Eq. (3.3) are diagonal in the computational basis and have product ground states $|0^n\rangle$ and $|1^n\rangle$. The presence of the $\mathbb{Z}_2$-symmetry is therefore essential: the same family of Hamiltonians do not exhibit NLTS without it. In this sense, the NLTS property here behaves similarly to topological order in 1D systems which only exists under symmetry protection [11], [14].

Theorem 3.2.1 implies restrictions on the performance of variational quantum algorithms for combinatorial optimization. Recall that the Quantum Approximate Optimization Algorithm (QAOA) [19] seeks to approximate the maximum of a cost function $C : \{0, 1\}^n \to \mathbb{R}$ by encoding it into a Hamiltonian $H_n = \sum_{x\in\{0,1\}^n} C(x)|x\rangle\langle x|$. It variationally optimizes the expected energy of H_n over quantum states of the form $U(\beta, \gamma)|+^n\rangle$, where

$$U(\beta, \gamma) = \prod_{k=1}^{p} e^{i\beta_k B} e^{i\gamma_k H_n},$$

and where $B = \sum_{j=1}^{n} X_j$. The integer $p \geq 1$ is called the QAOA *level*. It controls non-locality of the variational circuit.

A paradigmatic test case for QAOA is the MaxCut problem. Given a graph $G_n = (V, E)$ with n vertices, the corresponding MaxCut Hamiltonian is defined by Eq. (3.3). The maximum energy of H_n coincides with the number of edges in the

maximum cut of G_n. Crucially, the QAOA circuit $U(\beta, \gamma)$ with the Hamiltonian H_n as well as the initial QAOA state $|+^n\rangle$ obey the $\mathbb{Z}_2$-symmetry property. Furthermore, the circuit $U(\beta, \gamma)$ has depth $O(Dp)$, where D is the maximum vertex degree of G_n. Specializing Theorem 3.2.1 to bipartite graphs, we obtain an upper bound on the approximation ratio achieved by the level-p QAOA circuits for the MaxCut cost function (see Appendix 3.A for a proof):

Corollary 3.2.2. *For every integer $D \geq 3$, there exists an infinite family of bipartite D-regular graphs $\{G_n\}_{n \in \mathcal{I}}$ such that the Hamiltonians H_n defined in Eq.* (3.3) *obey*

$$\frac{1}{|E|}\langle +^n|U^{-1}H_nU|+^n\rangle \leq \frac{5}{6} + \frac{\sqrt{D-1}}{3D} \tag{3.8}$$

for any level-p QAOA circuit $U \equiv U(\beta, \gamma)$ as long as $p < (1/3 \log_2 n - 4)D^{-1}$.

Note that any bipartite graph with a set of edges E has maximum cut size $|E|$. In this case, the left-hand side of Eq. (3.2.2) coincides with the approximation ratio, i.e., the ratio between the expected value of the MaxCut cost function on the (optimal) level-p variational state and the maximum cut size. Thus Corollary 3.2.2 provides an explicit upper bound on the approximation ratio achieved by level-p QAOA. Such bounds were previously known only for $p = 1$ [19]. Statement (3.2.2) severely limits the performance of QAOA at any constant level p, rigorously establishing a widely believed conjecture [20]: constant-level QAOA is inferior to the classical Goemans-Williamson algorithm for MaxCut, which achieves an approximation ratio of approximately 0.878 on an arbitrary graph [21]. Indeed, the right-hand side of Eq. (3.2.2) is approximately $5/6 \approx 0.833$ for large vertex degree D.

QAOA circuits $U(\beta, \gamma)$ possess a form of locality which is stronger than the one assumed in Theorem 3.2.1. Indeed, if p and D are constants, the unitary $U(\beta, \gamma)$ can be realized by a constant-depth circuit composed of *nearest-neighbor* gates, i.e., the circuit is geometrically local.

A natural question is whether more general bounds on the variational energy can be established for states generated by geometrically local $\mathbb{Z}_2$-symmetric circuits. Of particular interest are graphs that lack the expansion property, such as regular lattices, where the arguments used in the proof of Theorem 3.2.1 no longer apply. A simple model of this type is the *ring of disagrees* [19]. It describes the MaxCut problem on the cycle graph $\mathbb{Z}_n$.

Quite recently, Ref. [22] proved that the optimal approximation ratio achieved by level-p QAOA for the ring of disagrees is bounded above by $(2p+1)/(2p+2)$ for all

p and conjectured that this bound is tight. Here we prove a version of this conjecture for arbitrary geometrically local $\mathbb{Z}_2$-symmetric circuits. To quantify the notion of geometric locality, let us say that a unitary U acting on n qubits located at vertices of the cycle graph has range R if the operator $U^\dagger Z_j U$ has support on the interval $[j - R, j + R]$ for any qubit j. For example, the level-p QAOA circuit associated with the ring of disagrees has range $R = p$.

Theorem 3.2.3. *Let H_n be the ring of disagrees Hamiltonian,*

$$H_n = \frac{1}{2} \sum_{p \in \mathbb{Z}_n} (I - Z_p Z_{p+1}),$$

where n is even. Let U be a $\mathbb{Z}_2$-symmetric unitary with range $R < n/4$. Then

$$\frac{1}{n} \langle +^n | U^\dagger H_n U | +^n \rangle \leq \frac{2R + 1/2}{2R + 1}.$$

This bound is tight whenever n is a multiple of $2R + 1$.

Since one can always round n to the nearest multiple of $2R + 1$, the bound Eq. (3.6) is tight for all n up to corrections $O(1/n)$, assuming that $R = O(1)$.

Let us first prove the upper bound Eq. (3.6). Define $\overline{X} = (XI)^{\otimes n/2}$. Then

$$\overline{X} H_n \overline{X} + H_n = nI.$$

Let $V = \overline{X} U$. Note that V is a $\mathbb{Z}_2$ symmetric circuit with range R. Taking the expected value of Eq. (3.7) on the state $U|+^n\rangle$, one infers that Eq. (3.6) holds whenever

$$\frac{1}{n} \langle +^n | V^\dagger H_n V | +^n \rangle \geq 1 - \frac{2R + 1/2}{2R + 1} = \frac{1}{2(2R + 1)}.$$

Thus it suffices to prove that Eq. (3.8) holds for any $\mathbb{Z}_2$-symmetric range-R circuit V. For each $j, k \in \mathbb{Z}_n$, define

$$\epsilon_{j,k} = \frac{1}{2} \langle +^n | V^\dagger (I - Z_j Z_k) V | +^n \rangle.$$

Let $\mathrm{dist}(j, k)$ be the distance between j and k with respect to the cycle graph $\mathbb{Z}_n$. We claim that

$$\epsilon_{j,k} = 1/2 \quad \text{if} \quad \mathrm{dist}(j, k) > 2R.$$

Indeed, $\langle +^n | V^\dagger Z_i V | +^n \rangle = 0$ for any qubit i since $V|+^n\rangle$ and $Z_i V|+^n\rangle$ are eigenvectors of $X^{\otimes n}$ with eigenvalues 1 and -1. Such eigenvectors have to be orthogonal. From

dist$(j, k) > 2R$, one infers that $V^\dagger Z_j V$ and $V^\dagger Z_k V$ have disjoint support. Thus

$$\langle +^n | V^\dagger Z_j Z_k V | +^n \rangle = \langle +^n | (V^\dagger Z_j V)(V^\dagger Z_k V) | +^n \rangle$$
$$= \langle +^n | V^\dagger Z_j V | +^n \rangle \cdot \langle +^n | V^\dagger Z_k V | +^n \rangle = 0.$$

This proves Eq. (3.9). Suppose one prepares the state $V|+^n\rangle$ and measures a pair of qubits $j < k$ in the standard basis. Then $\epsilon_{j,k}$ is the probability that the measured values on qubits j and k disagree. By the union bound,

$$\epsilon_{j,k} \le \sum_{i=j}^{k-1} \epsilon_{i,i+1}.$$

Indeed, if qubits j and k disagree, at least one pair of consecutive qubits located in the interval $[j, k]$ must disagree. Set $k = j + 2R + 1$. Then $\epsilon_{j,k} = 1/2$ by Eq. (3.9). Take the expected value of Eq. (3.10) with respect to random uniform $j \in \mathbb{Z}_n$. This gives

$$\frac{1}{2} \le \frac{2R+1}{n} \sum_{i \in \mathbb{Z}_n} \epsilon_{i,i+1} = \frac{2R+1}{n} \langle +^n | V^\dagger H_n V | +^n \rangle$$

proving Eq. (3.8). In Appendix 3.B, we construct a $\mathbb{Z}_2$-symmetric range-R circuit U such that $U|+^n\rangle$ is a tensor product of GHZ-like states on consecutive segments of $2R + 1$ qubits. We show that such circuit saturates the upper bound Eq. (3.6). This completes the proof of Theorem 3.2.3.

3.3 The Recursive Quantum Approximate Optimization Algorithm

Concerns about limitations of QAOA have previously been voiced by Hastings [20] who showed analytically that certain local classical algorithms match the performance of level-1 QAOA for Ising-like cost functions with multi-spin interactions. Hastings also gave numerical evidence for the same phenomenon for MaxCut with $p = 1$, and argued that this should extend to $p > 1$ [20].

Motivated by these limitations, we propose a non-local modification of QAOA which we call the recursive quantum approximate optimization algorithm (RQAOA). To sketch the main ideas behind RQAOA, consider an Ising-like Hamiltonian

$$H_n = \sum_{(p,q) \in E} J_{p,q} Z_p Z_q \tag{3.14}$$

defined on a graph $G_n = (V, E)$ with n vertices. Here $J_{p,q}$ are arbitrary real coefficients. RQAOA aims to approximate the maximum energy $\max_z \langle z | H_n | z \rangle$, where $z \in \{1, -1\}^n$. It consists of the following steps.

First, run the standard QAOA to maximize the expected value of H_n on the state $|\psi\rangle = U(\beta, \gamma)|+^n\rangle$. For every edge $(j, k) \in E$, compute $M_{j,k} = \langle \psi^*|Z_j Z_k|\psi^*\rangle$, where ψ^* is the optimal variational state.

Next, find a pair of qubits $(i, j) \in E$ with the largest magnitude of $M_{i,j}$ (breaking ties arbitrarily). The corresponding variables Z_i and Z_j are correlated if $M_{i,j} > 0$ and anti-correlated if $M_{i,j} < 0$. Impose the constraint

$$Z_j = \text{sgn}(M_{i,j})Z_i$$

and substitute it into the Hamiltonian H_n to eliminate the variable Z_j. For example, a term $Z_j Z_k$ with $k \notin \{i, j\}$ gets mapped to $\text{sgn}(M_{i,j})Z_i Z_k$. The term $J_{i,j}Z_i Z_j$ gets mapped to a constant energy shift $J_{i,j}\text{sgn}(M_{i,j})$. All other terms remain unchanged. This yields a new Ising Hamiltonian H_{n-1} that depends on $n - 1$ variables. By construction, the maximum energy of H_{n-1} coincides with the maximum energy of H_n over the subset of assignments satisfying the constraint Eq. (3.11).

Finally, call RQAOA recursively to maximize the expected value of H_{n-1}. Each recursion step eliminates one variable from the cost function. The recursion stops when the number of variables reaches some specified threshold value $n_c \ll n$. The remaining instance of the problem with n_c variables is then solved by a purely classical algorithm (for example, by a brute force method). Thus the value of n_c controls how the workload is distributed between quantum and classical computers. We describe a generalization of RQAOA applicable to Ising-like cost functions with multi-spin interactions in Appendix 3.C.

Imposing a constraint of the form (3.11) can be viewed as rounding correlations among the variables Z_i and Z_j. Indeed, the constraint demands that these variables must be perfectly correlated or anti-correlated. This is analogous to rounding fractional solutions obtained by solving linear programming relaxations of combinatorial optimization problems. We note that reducing the size of a problem to the point that it can be solved optimally by brute force is a widely used and effective approach in combinatorial optimization.

We compare the performance of the standard QAOA, RQAOA, and local classical algorithms by considering the Ising Hamiltonians in Eq. (3.3) with couplings $J_{p,q} = \pm 1$ defined on the cycle graph. In Appendix 3.D, we prove:

Theorem 3.3.1. *For each integer n divisible by 6, there exists a family of $2^{n/3}$ Ising Hamiltonians of the form $H_n = \sum_{k \in \mathbb{Z}_n} J_k Z_k Z_{k+1}$ with $J_k \in \{1, -1\}$ such that the*
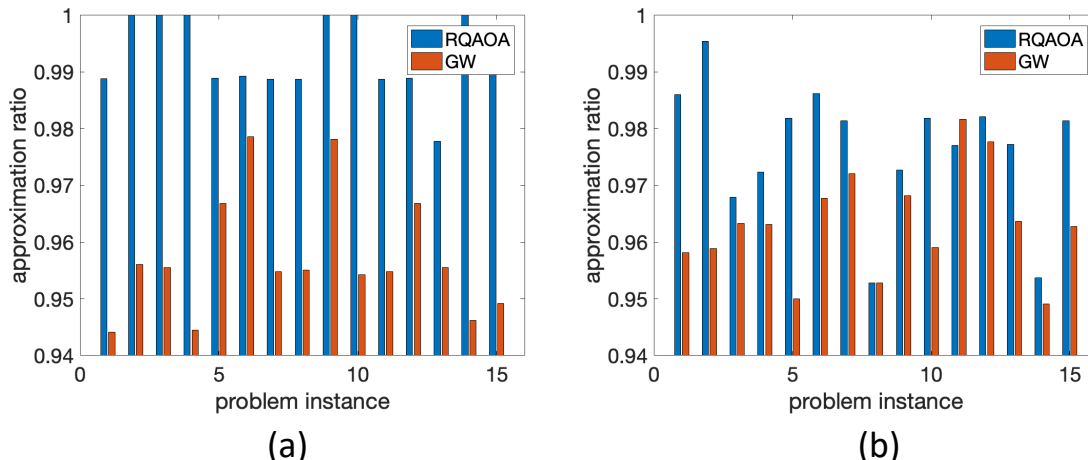
Figure 3.1: Comparisons of level-1 RQAOA and the Goemans-Williamson Algorithm. **(a)** Approximation ratios achieved by level-1 RQAOA (blue) and the Goemans-Williamson (GW) algorithm [21] (red) for 15 instances of the Ising cost function with random ±1 couplings defined on the 2D toric grid of size 16×16. In case **(b)** the Ising Hamiltonian also includes random ±1 external fields. The RQAOA threshold value is $n_c = 20$. We found that the standard level-1 QAOA achieves approximation ratios below $1/2$ for all considered instances (not shown). The GW algorithm was implemented with $n = 256$ rounding attempts and the best found solution was selected. The exact maximum energy was computed using integer linear programming.

following holds for all Hamiltonians in the family:

(i) There is a local classical algorithm which achieves the approximation ratio 1.

(ii) Level-p QAOA achieves an approximation ratio of at most $p/(p+1)$.

(iii) Level-1 RQAOA achieves the approximation ratio 1.

Our definition of local classical algorithms follows [20]. We also show that the level-1 RQAOA achieves the optimal approximation ratio for any 1D Ising model with coupling coefficients $J_k \in \{1, -1\}$. This proves that, in certain cases, RQAOA is strictly more powerful than QAOA.

Finally, we report a numerical comparison between the level-1 RQAOA and the Goemans-Williamson algorithm [21] for the Ising cost function Eq. (3.3) with random coefficients $J_{j,k} = \pm 1$. Two graphs are considered: (a) the 2D grid, and (b) the 2D grid with one extra vertex connected to all grid points. The latter is equivalent to the 2D Ising model with random ±1 external fields. As shown in [23], the problem of maximizing the energy $C(x)$ admits an efficient algorithm in case (a) while case (b) is NP-hard. To compute the mean values $\langle \psi(\beta, \gamma) | Z_j Z_k | \psi(\beta, \gamma) \rangle$,

we used a version of the algorithm by Wang et al [24], as detailed in Appendix 3.C. Figure 3.1 shows approximation ratios achieved by each algorithm for 15 problem instances with the grid size 16×16. It can be seen that RQAOA outperforms the Goemans-Williamson algorithm for all except for one instance. We found that the standard level-1 QAOA achieves approximation ratio below 1/2 for all considered instances.

Note added: After submission of this work, analogous limitations were established for random regular graphs by exploiting the locality and spatial uniformity of QAOA [25], [26]. We focus on $\mathbb{Z}_2$-symmetry and locality, and our statements also apply to non-uniform local algorithms.

Acknowledgements

3.A   Proof of Corollary 3.2.2

In this appendix, we give a proof of Corollary 3.2.2 in the main text. Here and below, we will denote the expected approximation ratio achieved by the QAOA with Hamiltonian H as

$$\mathsf{QAOA}_p(H) = \left(\max_{\beta, \gamma \in \mathbb{R}^p} \langle \Psi_H(\beta, \gamma) | H | \Psi_H(\beta, \gamma) \rangle \right) \cdot \left(\max_{x \in \{0,1\}^n} \langle x | H | x \rangle \right)^{-1},$$

where

$$|\Psi_H(\beta, \gamma)\rangle = U_H(\beta, \gamma)|+^n\rangle \qquad \text{and} \qquad U_H(\beta, \gamma) = \prod_{m=1}^{p} \left(e^{i\beta_m B} e^{i\gamma_m H} \right) \tag{3.16}$$

for $\beta, \gamma \in \mathbb{R}^p$ and where $B = \sum_{j=1}^{n} X_j$. Let us first record a few general features of the QAOA for later use.

Let $G = (V, E)$ be a graph, $n = |V|$, $m = |E|$, and let $J = (J_e)_{e \in E} \in \mathbb{R}^E$ be an assignment of edge weights on G. Let us define the Hamiltonian $H_G(J)$ as

$$H_G(J) = \sum_{\{u,v\} \in E} J_{\{u,v\}} Z_u Z_v . \tag{3.17}$$

It will be useful for later to also define

$$H_G = \sum_{\{u,v\} \in E} Z_u Z_v, \qquad \text{and} \qquad H_G^{\text{MaxCut}} = \frac{1}{2}(mI - H_G),$$

where H_G^{MaxCut} is the Hamiltonian used in QAOA for the Maximum Cut problem on the graph G. We will use the following bound on the circuit depth of a QAOA unitary.

Lemma 3.A.1. *Let $U = U_H(\beta, \gamma)$ with $\beta, \gamma \in \mathbb{R}^p$ be a level-p QAOA unitary (cf. Eq. (3.A)) for a Hamiltonian $H = H_G(J)$ on a graph G (cf. (3.A)). Let D be the maximum degree of G. Then U can be realized by a circuit of depth $d \leq p(D + 1)$ consisting of 2-qubit gates.*

If G is D-regular and bipartite, then the circuit depth of U can be bounded by $d \leq pD$.

Proof. By Vizing's theorem [27], there is an edge coloring of G with at most $D + 1$ colors. Taking such a coloring $E = E_1 \cup \cdots \cup E_{D+1}$, we may apply each level $e^{i\beta B} e^{i\gamma H}$ of U in depth $D + 1$ by applying $(\prod_{v \in V} e^{i\beta X_v}) \prod_{c=1}^{D+1} V_c(\gamma)$, where each $V_c(\gamma) = \left(\prod_{\{u,v\} \in E_c} e^{i\gamma J_{\{u,v\}} Z_u Z_v} \right)$ is a depth-1-circuit of two-local gates.

If G is D-regular and bipartite, we may reduce the chromatic number upper bound from $D + 1$ to D since all bipartite graphs are D-edge-colorable by Kőnig's line coloring theorem. We illustrate the construction of the circuit on Figure 3.2 for the case $D = 3$ and $p = 1$. □

The expected QAOA approximation ratios of suitably related instances are identical:

Lemma 3.A.2. *Let $\mathcal{L} \subset V$ be an arbitrary subset of vertices and $\partial \mathcal{L}$ be the set of edges that have exactly one endpoint in $\mathcal{L}$. Let $J = (J_e)_{e \in E} \in \mathbb{R}^E$ be arbitrary edge weights. Define $\tilde{J} = (\tilde{J}_e)_{e \in E} \in \mathbb{R}^E$ by*

$$\tilde{J}_e = \begin{cases} -J_e & \text{if } e \in \partial \mathcal{L} \\ J_e & \text{otherwise} . \end{cases}$$
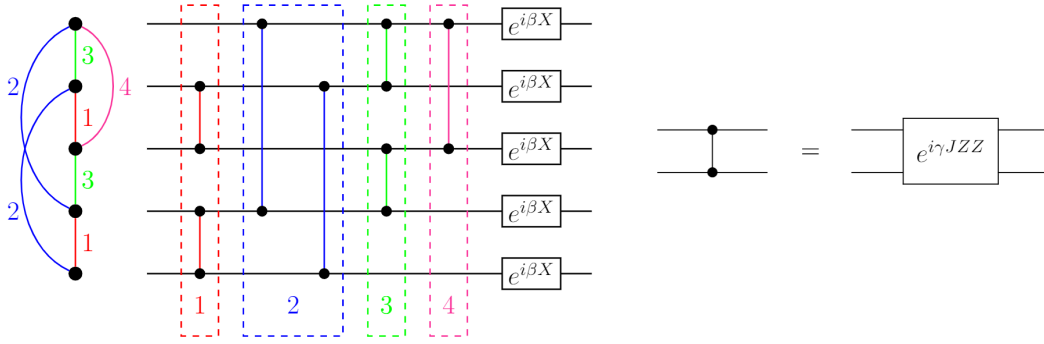
Figure 3.2: Example for the construction of the circuit given in Lemma 3.A.1: a 4-colorable graph with maximum degree 3 alongside its associated depth-5 quantum circuit for the level-1 QAOA unitary.

Then expected QAOA approximation ratios satisfy

$$\mathsf{QAOA}_p(H_G(J)) = \mathsf{QAOA}_p(H_G(\tilde{J})).$$

Proof. Let us write $H = H_G(J)$ and $\tilde{H} = H_G(\tilde{J})$ for brevity. Let $\overline{X} = \overline{X}[\mathcal{L}]$ be a tensor product of Pauli-X operators acting on every qubit in $\mathcal{L} \subset V$. Then $\tilde{H} = \overline{X}H\overline{X}$, which implies that

$$\max_{x \in \{0,1\}^n} \langle x|H|x \rangle = \max_{x \in \{0,1\}^n} \langle x|\tilde{H}|x \rangle . \tag{3.18}$$

Let $\beta, \gamma \in \mathbb{R}^p$ be arbitrary. Then we also have

$$\overline{X}|\Psi_{\tilde{H}}(\beta,\gamma)\rangle = \prod_{m=1}^{p}(\overline{X}e^{i\beta_m B}e^{i\gamma_m \tilde{H}}\overline{X})|+^n\rangle = \prod_{m=1}^{p}(e^{i\beta_m B}e^{i\gamma_m H})|+^n\rangle = |\Psi_H(\beta,\gamma)\rangle,$$

where identities in the middle follow since $|+^n\rangle$ is stabilized by $\overline{X}$, and since $[\overline{X}, B] = 0$. Therefore we have

$$\langle\Psi_{\tilde{H}}(\beta,\gamma)|\tilde{H}|\Psi_{\tilde{H}}(\beta,\gamma)\rangle = \langle\Psi_{\tilde{H}}(\beta,\gamma)|\overline{X}H\overline{X}|\Psi_{\tilde{H}}(\beta,\gamma)\rangle = \langle\Psi_H(\beta,\gamma)|H|\Psi_H(\beta,\gamma)\rangle.$$

Combined with (3.A), this implies the claim. □

In particular, if $G = (V, E)$ is a bipartite graph, then Lemma 3.A.2 implies that

$$\mathsf{QAOA}_p(H_G) = \mathsf{QAOA}_p(-H_G)$$

and

$$\mathrm{QAOA}_p(H_G^{\mathrm{MaxCut}}) = \frac{1}{2}(1 + \mathrm{QAOA}_p(H_G)). \tag{3.19}$$

We now prove Corollary 1. It is a direct consequence of Theorem 1, which we restate here for convenience in the notation of this appendix:

Theorem 3.A.1. *Consider a family $\{G_n = ([n], E_n)\}_{n \in \mathcal{I}}$ of graphs with Cheeger constant lower bounded as $h(G_n) \geq h > 0$ for all $n \in \mathcal{I}$. Then*

$$\langle \varphi | U^\dagger H_{G_n} U | \varphi \rangle < |E_n| - \frac{hn}{3}$$

for any $\mathbb{Z}_2$-symmetric depth-d circuit U composed of two-qubit gates, any $\mathbb{Z}_2$-symmetric product state φ, and any $n > 48^2 8^d$, $n \in \mathcal{I}$.

Then we have the following:

Corollary 3.A.1. *For every integer $D \geq 3$, there exists an infinite family of bipartite D-regular graphs $\{G_n\}_{n \in \mathcal{I}}$ such that*

$$\mathrm{QAOA}_p(H_{G_n}^{\mathrm{MaxCut}}) \leq \frac{5}{6} + \frac{\sqrt{D-1}}{3D}$$

as long as

$$p < (1/3 \log_2 n - 4)D^{-1} .$$

Proof. Fix some $D \geq 3$. By the results of [28], [29], there exists an infinite family $\{G_n\}_{n \in \mathcal{I}}$ of bipartite D-regular Ramanujan graph with n vertices for every $n \in \mathcal{I}$. Consider a fixed $n \in \mathcal{I}$ and let $p = p(n)$ be the associated QAOA level. Let $U_n = U_{H_{G_n}}(\beta^*, \gamma^*)$ be a level-p QAOA unitary for the Hamiltonian H_{G_n} on G_n, and assume that $\beta^*, \gamma^* \in \mathbb{R}^p$ are such that the expectation of H_{G_n} is maximized. Because G_n is D-regular, the circuit depth of U_n can be bounded from above by pD according to Lemma 3.A.1. Condition (3.12) implies that $n > 48^2 8^{pD}$, thus

$$\mathrm{QAOA}_p(H_{G_n}) = \frac{1}{|E_n|}\langle +^n | U_n^\dagger H_{G_n} U_n | +^n \rangle < 1 - \frac{h}{3|E_n|}n = 1 - \frac{2h}{3D}$$

by Theorem 1, where we have used that $|E_n| = nD/2$. With (3.A) (using that G_n is bipartite), we conclude that

$$\mathrm{QAOA}_p(H_{G_n}^{\mathrm{MaxCut}}) < 1 - \frac{h}{3D} .$$

The claim then follows from the bound $h/D \geq (D - 2\sqrt{D-1})/(2D)$, valid for all Ramanujan graphs. $\qquad\square$

3.B   Optimal Variational Circuit for the Ring of Disagrees

In this section we prove that the upper bound of Theorem 2 in the main text is tight whenever n is a multiple of $2R + 1$. Let

$$|\text{GHZ}_n\rangle = 2^{-1/2}(|0^n\rangle + |1^n\rangle)$$

be the GHZ state of n qubits.

Lemma 3.B.1. *Suppose $n = 2p + 1$ for some integer p. There exists a $\mathbb{Z}_2$-symmetric range-p quantum circuit V such that*

$$|\text{GHZ}_n\rangle = V|+^n\rangle.$$

Proof. We shall write $\text{CX}_{c,t}$ for the CNOT gate with a control qubit c and a target qubit t. Let $c = p + 1$ be the central qubit. One can easily check that

$$|\text{GHZ}_n\rangle = \left(\prod_{j=1}^{p} \text{CX}_{c,c-j}\text{CX}_{c,c+j}\right)H_c|0^n\rangle.$$

All CX gates in the product pairwise commute, so the order does not matter. Inserting a pair of Hadamards on every qubit $j \in [n] \setminus \{c\}$ before and after the respective CX gate and using the identity $(I \otimes H)\text{CX}(I \otimes H) = \text{CZ}$, one gets

$$|\text{GHZ}_n\rangle = \left(\prod_{j\in[n]\setminus\{c\}} H_j\right)\left(\prod_{j=1}^{p} \text{CZ}_{c,c-j}\text{CZ}_{c,c+j}\right)|+^n\rangle.$$

Let $S = \exp\left[i(\pi/4)Z\right]$ be the phase-shift gate. Define the two-qubit Clifford gate

$$\text{RZ} = (S \otimes S)^{-1}\text{CZ} = \exp(-i\pi/4)\exp\left[-i(\pi/4)(Z \otimes Z)\right].$$

Expressing CZ in terms of RZ and S in Eq. (3.14), one gets

$$|\text{GHZ}_n\rangle = S_c^{2p}\left(\prod_{j\in[n]\setminus\{c\}} H_jS_j\right)\left(\prod_{j=1}^{p} \text{RZ}_{c,c-j}\text{RZ}_{c,c+j}\right)|+^n\rangle.$$

Multiply both sides of Eq. (3.15) on the left by a product of S gates over qubits $j \in [n] \setminus \{c\}$. Noting that

$$SHS = i\exp\left[-i(\pi/4)X\right],$$

one gets (ignoring an overall phase factor)

$$\prod_{j\in[n]\setminus\{c\}} S_j|\text{GHZ}_n\rangle = S_c^{2p}\left(\prod_{j\in[n]\setminus\{c\}} \exp\left[-i(\pi/4)X_j\right]\right)\left(\prod_{j=1}^{p} \text{RZ}_{c,c-j}\text{RZ}_{c,c+j}\right)|+^n\rangle.$$

Using the identity

$$\prod_{j\in[n]\setminus\{c\}} S_j|\mathsf{GHZ_n}\rangle = S_c^{2p}|\mathsf{GHZ_n}\rangle \,,$$

one can cancel S_c^{2p} that appears in both sides of Eq. (3.16). We arrive at Eq. (3.13) with

$$V = \left(\prod_{j\in[n]\setminus\{c\}} \exp\left[-i(\pi/4)X_j\right]\right)\left(\prod_{j=1}^{p} \mathsf{RZ}_{c,c-j}\mathsf{RZ}_{c,c+j}\right) \,.$$

The circuit diagram of V in the case $n = 7$ is shown in Figure 3.3. Obviously, V is $\mathbb{Z}_2$-symmetric since any individual gate commutes with $X^{\otimes n}$. Let us check that V has range-p. Consider any single-qubit observable O_q acting on the q-th qubit. Consider three cases:

Case 1: $q = c$. Then $V^\dagger O_q V$ may be supported on all n qubits. However, $[c - p, c + p] = [1, n]$, so the p-range condition is satisfied trivially.

Case 2: $1 \leq q < c$. Then all gates $\mathsf{RZ}_{c,c+j}$ in V cancel the corresponding gates in $V^\dagger$, so that $V^\dagger O_q V$ has support in the interval $[1, c] \subseteq [q - p, q + p]$. Thus the p-range condition is satisfied.

Case 3: $c < q \leq n$. This case is equivalent to Case 2 by symmetry. $\qquad\square$

Recall that we consider the ring of disagrees Hamiltonian

$$H_n = \frac{1}{2} \sum_{p\in\mathbb{Z}_n} (I - Z_p Z_{p+1}).$$

Lemma 3.B.2. *Consider any integers n, p such that n is even and n is a multiple of $2p + 1$. Then there exists a $\mathbb{Z}_2$-symmetric range-p circuit U such that*

$$\langle +^n|U^\dagger H_n U|+^n\rangle = \frac{2p + 1/2}{2p + 1}.$$

Proof. Let V be the $\mathbb{Z}_2$-symmetric range-p unitary operator preparing the GHZ state on $2p + 1$ qubits starting from $|+^{2p+1}\rangle$, see Lemma 3.B.1. Suppose $n = m(2p + 1)$ for some integer m. Define

$$U = U_1 U_2,$$
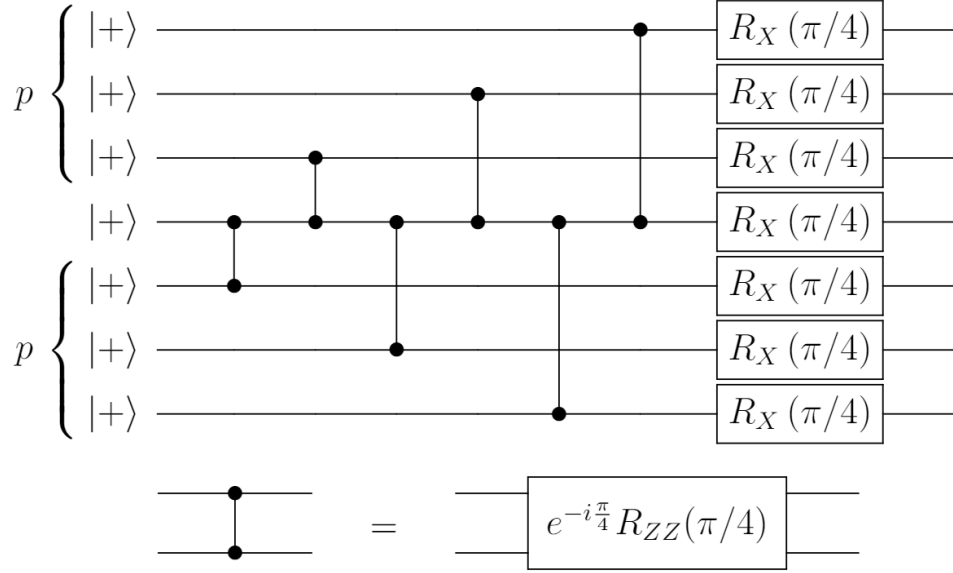
Figure 3.3: The $\mathbb{Z}_2$-symmetric range-3 quantum circuit to prepare the GHZ state $|\text{GHZ}_{2p+1}\rangle$ of $2p + 1 = 7$ qubits ($p = 3$). Here, $R_O(\theta) = \exp(-i\theta O)$.

where
$$U_1 = (X \otimes I)^{\otimes n/2} \quad \text{and} \quad U_2 = V^{\otimes m}.$$

Since each copy of V acts on a consecutive interval of qubits and has range p, one infers that U has range p. We have

$$U_1^\dagger H_n U_1 = \sum_{p \in \mathbb{Z}_n} G_p, \quad G_p = \frac{1}{2}(I + Z_p Z_{p+1}).$$

The state $U_2|+^n\rangle$ is a tensor product of GHZ states supported on consecutive tuples of $2p + 1$ qubits. The expected value of G_p on the state $U_2|+^n\rangle$ equals 1 if G_p is supported on one of the GHZ states. Otherwise, if G_p crosses the boundary between two GHZ states, the expected value of G_p on the state $U_2|+^n\rangle$ equals $1/2$. Thus

$$\langle +^n|U^\dagger H_n U|+^n\rangle = \sum_{p \in \mathbb{Z}_n} \langle +^n|U_2^\dagger G_p U_2|+^n\rangle = m(2p + 1/2) = n\frac{2p + 1/2}{2p + 1}.$$

$\square$

3.C   Recursive QAOA

In this appendix, we outline the Recursive QAOA algorithm (RQAOA) for general cost functions.

C.1 Variable Elimination

Let $G = (V, E)$ be a hypergraph with $|V| = n$ vertices. Suppose a variable $x_v \in \{1, -1\}$ is associated with each vertex $v \in V$. Let $\{1, -1\}^V = \{1, -1\}^n$ be the set of all possible variable assignments. Let $J : E \to \mathbb{R}$ be a function which assigns a real weight J_e to every hyperedge e. Given a subset $f \subset V$ and an assignment $x \in \{1, -1\}^V$, let us write

$$x(f) = \prod_{v \in f} x_v.$$

Let us agree that $x(\emptyset) = 1$. We consider the problem of maximizing cost functions of the form

$$C(x) = \sum_{e \in E} J_e x(e)$$

over $x \in \{1, -1\}^V$.

Fix some vertex $v \in V$. As a motivation, we first describe how a single variable x_v can be eliminated when a suitably constrained problem is considered. Namely, suppose that instead of trying to approximate $\max_{x \in \{1,-1\}^V} C(x)$, we restrict to $x \in \{1, -1\}^V$ satisfying

$$x(f) = \sigma, \tag{3.27}$$

where $f \subset V$ is some fixed subset of vertices containing v, and $\sigma \in \{1, -1\}$ is a constant. If $x \in \{1, -1\}^V$ satisfies the constraint (C.1), then

$$J_e x(e) = J_e x(e) x(f) \sigma = \sigma J_e x(e \triangle f).$$

Here and below, $A \triangle B$ denotes the symmetric difference of sets A and B. We arrive at

$$C(x) = \sum_{\substack{e \in E \\ v \notin e}} J_e x(e) + \sum_{\substack{e \in E: \\ v \in e}} \sigma J_e x(e \triangle f) . \tag{3.28}$$

Note that $C(x)$ does not depend on x_v. Expression (C.1) can be written as a sum over the hyperedges of a hypergraph $G' = (V', E')$ with vertex set $V' = V \backslash \{v\}$ and hyperedges

$$E' = E'_0 \cup E'_1,$$

where

$$E'_0 = \{e \in E : v \notin e\} \quad \text{and} \quad E'_1 = \{e \triangle f : e \in E, \quad v \in e\}.$$

Note that G' no longer contains the vertex v. Define a function $J' : E' \rightarrow \mathbb{R}$ such that

$$J'_e = J_e \quad \text{if } e \in E'_0,$$

and

$$J'_e = \sigma J_{e \triangle f} \quad \text{if } e \in E'_1.$$

By construction, the maximum of

$$C'(x) = \sum_{e \in E'} J'_e x(e)$$

over all assignments $x \in \{1, -1\}^{V'}$ coincides with the maximum of $C(x)$ over all $x \in \{1, -1\}^V$ satisfying the constraint Eq. (C.1). Furthermore, any maximum x^* of C' can directly be translated to a corresponding maximum of C over the restricted set defined by the constraint (C.1) by setting $x_v^* = \sigma \cdot x(f \backslash \{v\})$. That is, we have $x^* = \xi(x)$ for the function $\xi : \{1, -1\}^{V'} \rightarrow \{1, -1\}^V$ defined by

$$\xi(x)_w = \begin{cases} \sigma \cdot x(f \backslash \{v\}) & \text{for } v = w \\ x_w & \text{otherwise} \end{cases} \tag{3.33}$$

for all $w \in V$.

In summary, we have reduced the problem of maximizing $C(x)$ over n variables $x \in \{1, -1\}^V$ satisfying (C.1) to the problem of maximizing $C'(x)$ over $n - 1$ variables $x \in \{1, -1\}^{V'}$. If a global maximum x of $C(x)$ happens to satisfy (C.1), the new reduced problem yields a solution to the original problem.

C.2 Correlation Rounding

To construct an approximation algorithm, we simply *impose* a constraint of the form (C.1) by choosing $f \subset V$, $v \in f$, and $\sigma \in \{1, -1\}$ appropriately. To make the latter choice, we use the standard QAOA_p algorithm with $p = O(1)$. That is, let us set

$$H_G(J) = \sum_{e \in E} J_e Z(e) \qquad \text{where} \qquad Z(e) = \prod_{v \in e} Z_v \tag{3.34}$$

and write $H = H_G(J)$. We first use the standard $\mathsf{QAOA}_p(H)$ algorithm to find an optimal state

$$\Psi = \Psi_{H_G}(\beta_*, \gamma_*) \in (\mathbb{C}^2)^{\otimes |V|}$$

maximing the energy of H_G. The expected value

$$M_e = \langle \Psi | Z(e) | \Psi \rangle$$

can be efficiently approximated on a quantum computer for any $e \in E$.

Suppose that the state Ψ is measured in the computational basis giving a string $x \in \{1, -1\}^V$. Clearly, if $|M_f|$ is close to 1, then the variables $\{x_v\}_{v \in f}$ satisfy a constraint of the form (C.1) with high probability with $\sigma = \mathsf{sign}(M_f) \in \{1, -1\}$ and any $v \in f$. Thus it is natural to choose f such that $|M_f|$ is maximal. Combined with the procedure for eliminating the corresponding variable x_v described in Section C.1, we obtain a subroutine for reducing the problem size by one variable. Pseudocode for this routine is given below.

Imposing a constraint of the form (C.1) can be viewed as rounding correlations among the variables $\{x_w\}_{w \in f}$: indeed, the constraint demands that for $v \in f$, the variable x_v and $x(f \backslash \{v\})$ must be perfectly correlated or anti-correlated.

1: **function** ELIMINATEVARIABLE$(G = (V, E), J)$
2: ────────────────────────────────
3: **Input:** A hypergraph $G = (V, E)$ and a weight function $J : E \to \mathbb{R}$
4: **Output:** A hypergraph $G' = (V', E')$, $J' : E' \to \mathbb{R}$ and a function $\xi : \{1, -1\}^{V'} \to \{1, -1\}^V$.
5: ────────────────────────────────
6:     Run $\mathsf{QAOA}_p(H_G(J))$ to find a state Ψ which maximizes $\langle \Psi | H_G(J) | \Psi \rangle$.
7:     Compute $M_e = \langle \Psi | Z(e) | \Psi \rangle$ for every $e \in E$.
8:     Set $f = \mathrm{argmax}_{f \in E} |M_f|$ (breaking ties arbitrarily).
9:     Pick $v \in f$ arbitrarily.
10:     Set $\sigma = \mathsf{sign}(M_f)$.
11:     Define $V' = V \backslash \{v\}$.   Also define E', J' and ξ by Eqs. (3.19),(3.20),(3.21),(3.22),(C.1).
12:     **return** $(G' = (V', E'), J', \xi)$.
13: **end function**

C.3   The Recursive QAOA (RQAOA) Algorithm

The recursive QAOA algorithm (RQAOA) we propose here proceeds simply by iterating the process of eliminating one variable at a time until the number of variables reaches some specified threshold value $n_c \ll n$. The remaining instance of the problem with n_c variables is solved by a purely classical algorithm (for example, by the brute force method). Thus the value of n_c controls how the workload is distributed between the quantum and the classical computers. Pseudocode for the RQAOA algorithm is given in Fig. 3.4.

```
 1: function RQAOA(G = (V, E), J)
 2: ─────────────────────────────────────────────────
 3: Input: A hypergraph G = (V, E) with n = |V| and a weight function J : E → ℝ
         defining a Hamiltonian H_G(J), see Eq. (C.2).
 4: Output: A variable assignment x ∈ {−1, 1}^V
 5: ─────────────────────────────────────────────────
 6:       Let ξ^(0) : {1, −1}^V → {1, −1}^V be the identity map.
 7:       for k = 1 to n − n_c do
 8:           (G, J, ξ) ← eliminateVARIABLE(G, J).
 9:           ξ^(k) ← ξ^(k−1) ∘ ξ.
10:       end for
11:       Let G = (V, E) be the final hypergraph with |V| = n_c vertices.
12:       Find x* = argmax_{x∈{1,−1}^V} ⟨x|H_G(J)|x⟩.
13:       return ξ(x*)
14: end function
```

Figure 3.4: Pseudocode for the recursive QAOA algorithm.

C.4 Classical Simulability of Level-1 RQAOA for Ising Models

Suppose J is a real symmetric matrix of size n. Here we consider Ising-like cost functions such that the corresponding Hamiltonian is

$$H = \sum_{1 \le p < q \le n} J_{p,q} Z_p Z_q .$$

The mean values of a Pauli operator $Z_p Z_q$ on the level-1 QAOA state

$$|\Psi_H(\beta, \gamma)\rangle = e^{i\beta B} e^{i\gamma H} |+^n\rangle$$

can be computed in time $O(n)$ using an explicit analytic formula. Such a formula was derived for the Max-Cut cost function by Wang et al. [24, Theorem 1]. Below we provide a generalization to general Ising Hamiltonians. Since the total number of terms in the cost function is $O(n^2)$, simulating each step of RQAOA takes time at most $O(n^3)$. Assuming that $n_c = O(1)$, the number of steps is roughly n so that the full simulation cost is $O(n^4)$. Crucially, the simulation cost of this method does not depend on the depth of the variational circuit. This is important because RQAOA may potentially increase the depth from $O(1)$ to $O(n)$ since it adds many new terms to the cost function.

Lemma 3.C.1. *Fix a pair of qubits $1 \le u < v \le n$. Let $c = \cos(2\beta)$ and*

$s = \sin(2\beta)$. *Then*

$$\langle\Psi_H(\beta,1)|Z_uZ_v|\Psi_H(\beta,1)\rangle = (s^2/2)\prod_{p\neq u,v}\cos[2J_{u,p}-2J_{v,p}] - (s^2/2)\prod_{p\neq u,v}\cos[2J_{u,p}+2J_{v,p}]$$

$$+cs\cdot\sin(2J_{u,v})\left[\prod_{p\neq u,v}\cos(2J_{u,p}) + \prod_{p\neq u,v}\cos(2J_{v,p})\right].$$

Here we only consider the case $\gamma = 1$ since γ can be absorbed into the definition of J.

Proof. Given a 2-qubit observable O, define the mean value

$$\mu(O) = \langle\Psi_H(\beta,1)|O_{u,v}|\Psi_H(\beta,1)\rangle.$$

We are interested in the observable $O = ZZ \equiv Z\otimes Z$.

We note that all terms in H and B that act trivially on $\{u,v\}$ do not contribute to $\mu(O)$. Such terms can be set to zero. Given a 2-qubit observable O, define a mean value

$$\mu'(O) = \langle +^n|e^{iH'}O_{u,v}e^{-iH'}|+^n\rangle, \quad \text{where} \quad H' = \sum_{p\neq u,v}(J_{u,p}Z_u + J_{v,p}Z_v)Z_p.$$

Using the identities

$$e^{i\beta(X_u+X_v)}Z_uZ_v e^{-i\beta(X_u+X_v)} = c^2Z_uZ_v + s^2Y_uY_v + cs(Z_uY_v + Y_uZ_v),$$

$$e^{iJ_{u,v}Z_uZ_v}Z_uZ_v e^{-iJ_{u,v}Z_uZ_v} = Z_uZ_v,$$

$$e^{iJ_{u,v}Z_uZ_v}Y_uY_v e^{-iJ_{u,v}Z_uZ_v} = Y_uY_v$$

$$e^{iJ_{u,v}Z_uZ_v}Z_uY_v e^{-iJ_{u,v}Z_uZ_v} = \cos(2J_{u,v})Z_uY_v + \sin(2J_{u,v})X_v,$$

$$e^{iJ_{u,v}Z_uZ_v}Y_uZ_v e^{-iJ_{u,v}Z_uZ_v} = \cos(2J_{u,v})Y_uZ_v + \sin(2J_{u,v})X_u,$$

and noting that $\mu'(ZZ) = 0$, one easily gets

$$\mu(ZZ) = s^2\cdot\mu'(YY) + cs\cdot\cos(2J_{u,v})[\mu'(ZY)+\mu'(YZ)] + cs\cdot\sin(2J_{u,v})[\mu'(XI)+\mu'(IX)].$$

Using the explicit form of H', one gets

$$e^{-iH'}|+^n\rangle = \frac{1}{2}\sum_{a,b=0,1}|a,b\rangle_{u,v}\otimes|\Phi(a,b)\rangle_{\text{else}},$$

where $|\Phi(a,b)\rangle$ is a tensor product state of $n-2$ qubits defined by

$$|\Phi(a,b)\rangle = \bigotimes_{p\neq u,v}|J_{u,p}(-1)^a + J_{v,p}(-1)^b\rangle_p \quad \text{where} \quad |\theta\rangle \equiv e^{-i\theta Z}|+\rangle.$$

Combining Eqs. (3.25) and (3.28), one gets

$$\mu'(O) = (1/4) \sum_{a,b,a',b'=0,1} \langle a',b'|O|a,b \rangle \cdot \langle \Phi(a',b')|\Phi(a,b) \rangle.$$

Using the tensor product form of the states $|\Phi(a,b)\rangle$ and the identity $\langle \theta'|\theta \rangle = \cos(\theta - \theta')$ gives

$$\langle \Phi(a',b')|\Phi(a,b) \rangle = \prod_{p \neq u,v} \cos \left[J_{u,p}(-1)^a - J_{u,p}(-1)^{a'} + J_{v,p}(-1)^b - J_{v,p}(-1)^{b'} \right].$$

From Eqs. (3.30) and (3.31), one can easily compute the mean value $\mu'(O)$ for any 2-qubit observable.

Consider first the case $O = YY$. Then the only terms contributing to Eq. (3.30) are those with $a' = a \oplus 1$ and $b' = b \oplus 1$. The identity $\langle a \oplus 1|Y|a \rangle = -i(-1)^a$ gives

$$\mu'(YY) = -(1/4) \sum_{a,b=0,1} (-1)^{a+b} \prod_{p \neq u,v} \cos \left[2J_{u,p}(-1)^a + 2J_{v,p}(-1)^b \right],$$

that is,

$$\mu'(YY) = (1/2) \prod_{p \neq u,v} \cos \left[2J_{u,p} - 2J_{v,p} \right] - (1/2) \prod_{p \neq u,v} \cos \left[2J_{u,p} + 2J_{v,p} \right].$$

Next, consider the case $O = YZ$. Note that the matrix elements $\langle a',b'|O|a,b \rangle$ have zero real part. From Eqs. (3.30) and (3.31), one infers that $\mu'(YZ)$ has zero real part. This implies

$$\mu'(YZ) = \mu'(ZY) = 0.$$

Finally, consider the case $O = XI$. Then the only terms that contribute to Eq. (3.30) are those with $a' = a \oplus 1$ and $b' = b$. We get

$$\mu'(XI) = \prod_{p \neq u,v} \cos (2J_{u,p}).$$

Here we noted that the inner product Eq. (3.31) with $a' = a \oplus 1$ and $b' = b$ does not depend on a, b. By the same argument,

$$\mu'(IX) = \prod_{p \neq u,v} \cos (2J_{v,p}).$$

Combining Eq. (3.27) and Eqs. (3.33),(3.34),(3.35),(3.36), one arrives at Eq. (3.23).

$\square$

For more general cost functions that include interactions among three or more variables, there are two complications: First, unlike in the Ising case, the variable elimination process will typically increase the degree of non-locality of interactions. Second, mean values of Pauli operators on the QAOA state $\Psi_H(\beta, \gamma)$ lack a simple analytic formula (as far as we know). However, one can approximately compute the mean values using the Monte Carlo method due to Van den Nest [30]. A specialization of this method to simulation of the level-1 QAOA is described in [31]. The Monte Carlo simulator has runtime scaling polynomially with the number of qubits, number of terms in the cost function, and the inverse error tolerance, see [31] for details. This method also requires no restrictions on the depth of the variational circuit.

An important distinction between QAOA and RQAOA lies in the measurement step. QAOA requires few-qubit measurements to estimate the variational energy as well as the final n-qubit measurement that assigns a value to each individual variable. This last step is what makes QAOA hard to simulate classically and may lead to a quantum advantage [32]. In contrast, RQAOA only needs few-qubit measurements to estimate mean values of individual terms in the cost function. The n-qubit measurement step is replaced by the correlation rounding that eliminates variables one by one. One may ask whether the lack of multi-qubit measurements also precludes a quantum advantage. Indeed, in the special case of level-1 variational circuits and the Ising-like cost function RQAOA can be efficiently simulated classically, see above. However, level-p RQAOA with $p > 1$ as well as level-1 RQAOA with more general cost functions are not known to be classically simulable in polynomial time, leaving room for a quantum advantage.

3.D Comparison of QAOA, RQAOA, and Classical Algorithms

D.1 QAOA versus Classical Local Algorithms

In this section, we discuss another limitation of QAOA which results from its locality and the covariance condition discussed in Lemma 3.A.2: we compare QAOA to a certain very simple classical local algorithm (see Lemma 3.D.1 below). We show that there is an exponential number of problem instances for which the classical local algorithm outperforms QAOA.

Let us briefly sketch the notion of a local classical algorithm. We envision that the tuple $(J_e)_{e \in E}$ is given as input. Here we are interested in algorithms which are local

with respect to the underlying graph G. For $r \in \mathbb{N}$ and $v \in V$, define

$$E_r(v) = \bigcup_{\ell=1}^{r} \bigcup_{\substack{(e_1,\ldots,e_\ell) \\ \text{path with } v \in e_1}} \{e_1, \ldots, e_\ell\}$$

to be the set of edges that belong to a path starting at v of length bounded by r. Consider a classical algorithm $\mathcal{A}$ which on input $\{J_e\}_{e\in E}$ outputs $x = (x_1, \ldots, x_n) \in \{0,1\}^n$. We say that $\mathcal{A}$ is r-local if there is a family of functions $\{g_v : \mathbb{R}^{E_r(v)} \to \{0,1\}\}_{v\in V}$ such that the following holds for every problem instance $(J_e)_{e\in E} \in \mathbb{R}^E$: We have

$$x_v = g_v \left(\{J_e\}_{e\in E_r(v)} \right) \qquad \text{for every } v \in V .$$

In other words, in an r-local classical algorithm, every output bit x_v only depends on edge weights J_e belonging to paths of length bounded by r starting at v. We note that this definition can easily be generalized to the probabilistic case (e.g., by including local random bits). For the purposes of this section, deterministic functions turn out to be sufficient.

The (choice of) family $\{g_v\}_{v\in V}$ can be considered as a set of variational parameters for the classical algorithm. To keep the number of variational parameters constant, we consider vertex-transitive graphs G. Fix $v_* \in V$. For every $v \in V$, fix an automorphism π_v of G such that $\pi_v(v_*) = v$. Then the sets $E_r(v)$ for different $v \in V$ can be identified via $E_r(v) = \pi_v(E_r(v^*))$. We say that an r-local classical algorithm is *uniform* if (after this identification) $g_v \equiv g$ for all $v \in V$, i.e., if there is a single function $g : \mathbb{R}^{E_r(v_*)} \to \{0,1\}$ specifying the behavior of the algorithm. To obtain general-purpose algorithms (applicable to any instance), the function $g : \mathbb{R}^{E_r(v_*)} \to \{0,1\}$ should be chosen adapatively (i.e., potentially depending on the instance). The definition of local classical algorithm sketched here includes e.g., the algorithms considered in Ref. [20], though it is slightly more general as the local functions can be arbitrary.

Let $n = 6r$ be a multiple of 6. Consider n-qubit Hamiltonians (cf. (3.A)) of the form

$$H(J) = \sum_{k\in\mathbb{Z}_n} J_k Z_k Z_{k+1} \qquad \text{where} \qquad J = (J_0, \ldots, J_{n-1}) \in \{1, -1\}^n .$$

To define locality and uniformity for the cycle graph $\mathbb{Z}_n$, let $\pi_v(w) = v + w \pmod{n}$ be chosen as translation modulo n for $v \in \mathbb{Z}_n$. We show the following:

Lemma 3.D.1. *There is a subset $S \subset \{1, -1\}^n$ of $2^{n/3}$ problem instances such that the following holds:*

(i) $\mathsf{QAOA}_p(H(J)) \le p/(p+1)$ *for every* $p \in \mathbb{N}$ *and every* $J \in \mathcal{S}$.

(ii) *There is a* 1-*local uniform classical algorithm such that for every* $J \in \mathcal{S}$, *the algorithm outputs* $x \in \{0, 1\}^n$ *such that* $\langle x|H(J)|x \rangle = 1$.

(iii) *Level*-1 RQAOA *achieves the approximation ratio* 1.

Proof. For every $s = (s_0, \ldots, s_{2r-1}) \in \{0, 1\}^{2r}$, define $J = J(s) \in \{1, -1\}^n$ by

$$J_{3a} = J_{3a+1} = (-1)^{s_a}, \qquad \text{and} \qquad J_{3a+2} = 1,$$

for all $a = 0, 1, \ldots, 2r - 1$. We claim that the set $\mathcal{S} = \{J(s) \mid s \in \{0, 1\}^{2r}\}$ has the required properties. Consider an instance $H(J(s))$ with $s \in \mathcal{S}$. Define

$$\overline{X}(s) = \prod_{a=0}^{2r-1} X_{3a+1} \ .$$

Then $H(J(s))$ is related to $H_{\mathbb{Z}_n} = \sum_{j \in \mathbb{Z}_n} Z_j Z_{j+1}$ by the gauge transformation

$$H(J(s)) = \overline{X}(s) H_{\mathbb{Z}_n} \overline{X}(s)^{-1} \ .$$

Since the QAOA algorithm is invariant under such gauge transformation (see Lemma 3.A.2), we obtain

$$\mathsf{QAOA}_p(H(J(s))) = \mathsf{QAOA}_p(H_{\mathbb{Z}_n}) \le \frac{p}{p+1}$$

where we use the bound

$$\mathsf{QAOA}_p(H_{\mathbb{Z}_n}^{\mathrm{MaxCut}}) \le \frac{2p+1}{2p+2},$$

proven in [22] for even n, in combination with Lemma 3.A.2. This shows (i).

For the proof of (ii), consider the classical algorithm $\mathcal{A}$ which on input $J = (J_0, \ldots, J_{n-1})$ outputs

$$x_v = g(J_{v-1}, J_v) \qquad \text{for every } v \in \mathbb{Z}_n \ ,$$

where

$$g(J, J') = \begin{cases} 1 & \text{if } (J, J') = (-1, -1) \\ 0 & \text{otherwise} \ . \end{cases}$$

Clearly, the algorithm $\mathcal{A}$ is uniform and 1-local, and it is easy to check that the output satisfies $\langle x|H(J)|x \rangle = 1$.

The proof of (iii) is given as a part of Lemma 3.D.2. $\qquad \square$

D.2 RQAOA **on the Ising Ring**

Here we prove that level-1 RQAOA achieves approximation ratio 1 on the ring of disagrees, in sharp contrast to (arbitrary) level-p QAOA (see Lemma 3.D.1(i)). More generally, level-1 RQAOA produces $x \in \{0, 1\}^n$ which maximizes the cost function (that is, achieves approximation ratio 1) for any $1D$ Ising model where the coupling coefficients are either $+1$ or -1.

Lemma 3.D.2. *Consider a cost function of the form*

$$C_n(x) = \sum_{k \in \mathbb{Z}_n} J_k x_k x_{k+1} \qquad \text{for } x \in \{1, -1\}^n \,,$$

where $J_k \in \{1, -1\}$ for all $k \in \mathbb{Z}_n$. Then the level-1 RQAOA produces $x^* \in \{1, -1\}^n$ such that $C_n(x^*) = \max_{x \in \{1, -1\}^n} C_n(x)$.

It would be interesting to see additional, more general families of examples where approximation ratios achieved by RQAOA can be computed or lower bounded analytically.

Proof. Let

$$H = \sum_{k \in \mathbb{Z}_n} J_k Z_k Z_{k+1} \,. \tag{3.50}$$

Observe first that $\langle \Psi_H(\beta, \gamma) | Z_i Z_j | \Psi_H(\beta, \gamma) \rangle = 0$ if $|i - j| > 2$ since in this case the operators $U^{-1} Z_i U$ and $U^{-1} Z_j U$ have disjoint support. Lemma 3.C.1 shows that a QAOA$_1$-state $\Psi_H(\beta, \gamma)$ has expectation values

$$\langle \Psi_H(\beta, \gamma) | Z_i Z_j | \Psi_H(\beta, \gamma) \rangle = \begin{cases} \frac{1}{2} J_i \sin(4\beta) \sin(4\gamma) & \text{if } j = i + 1 \\ \frac{1}{4} J_i J_{i+1} \sin^2(2\beta) \sin^2(4\gamma) & \text{if } j = i + 2 \\ 0 & \text{otherwise} \end{cases} \tag{3.51}$$

when $J_k \in \{1, -1\}$ for every $k \in \mathbb{Z}_n$. Thus

$$|\langle \Psi(\beta, \gamma) | Z_i Z_{i+2} | \Psi(\beta, \gamma) \rangle| \leq 1/4 \qquad \text{for all } (\beta, \gamma) \,. \tag{3.52}$$

Assume (β^*, γ^*) are such that $(\beta^*, \gamma^*) = \operatorname{argmax}_{(\beta, \gamma)} \langle \Psi_H(\beta^*, \gamma^*) | H | \Psi_H(\beta^*, \gamma^*) \rangle$. Then we can infer from (D.2) that

$$\langle \Psi(\beta^*, \gamma^*) | Z_i Z_{i+1} | \Psi(\beta^*, \gamma^*) \rangle = J_i/2 \,. \tag{3.53}$$

Combined with (D.2) and (D.2), we conclude that

$$\mathrm{argmax}_{(i,j):i<j} \, |\langle \Psi_H(\beta^*, \gamma^*)|Z_i Z_j|\Psi_H(\beta^*, \gamma^*)\rangle| = (i^*, i^* + 1) \qquad (3.54)$$

for some $i^* \in \mathbb{Z}_n$. Without loss of generality, assume that $i^* = n - 2$. Then, according to (D.2), the RQAOA algorithm eliminates the variable x_{n-1} (i.e., $v = n-1$, $f = \{n - 2, n - 1\}$). By (D.2), this is achieved by imposing the constraint

$$x_{n-1} = x_{n-2} J_{n-2} \qquad (3.55)$$

i.e., $\sigma = J_{n-2}$. The resulting reduced graph $G' = (V', E')$ has vertex set $V' = V \backslash \{n - 1\} = \mathbb{Z}_{n-1}$ and edges

$$E' = \{\{i, i + 1\} \mid i \in \mathbb{Z}_n \backslash \{n - 2\}\} \cup \{\{n - 2, 0\}\}$$
$$= \{\{i, i + 1\} \mid i \in \mathbb{Z}_{n-2}\} \ ,$$

and it is easy to check that the new cost function takes the form

$$C'(x) = 1 + \sum_{k \in \mathbb{Z}_{n-1}} J'_k x_k x_{k+1} \qquad (3.56)$$

with

$$J'_i = \begin{cases} J_i & \text{when } i \neq n - 2 \\ J_{n-2} J_{n-1} & \text{when } i = n - 2 \end{cases}. \qquad (3.57)$$

We note that the transformation (D.2) preserves the parity of the couplings in the sense that

$$\prod_{k \in \mathbb{Z}_n} J_k = \prod_{k \in \mathbb{Z}_{n-1}} J'_k \ . \qquad (3.58)$$

Inductively, the RQAOA thus eliminates variables $x_{n-1}, x_{n-2}, \ldots, x_{n_c}$ while imposing the constraints (cf. (D.2))

$$x_{n-1} = x_{n-2} J_{n-2}$$
$$x_{n-2} = x_{n-3} J'_{n-3}$$
$$\vdots$$

arriving at the cost function $C_{n_c}(x)$ associated with an Ising chain of length n_c having couplings belonging to $\{1, -1\}$. Because of (D.2) and because (D.2) is frustrated

if and only if $\prod_{k \in \mathbb{Z}_n} J_k = -1$, we conclude that any maximum $x^* \in \{1, -1\}^{n_c}$ of $C_{n_c}(x)$ satisfies

$$C_{n_c}(x^*) = \begin{cases} n_c + 1 & \text{if } \prod_{k \in \mathbb{Z}_n} J_k = 1 \\ n_c - 2 & \text{otherwise .} \end{cases}$$

Because the cost function acquires a constant energy shift in every variable elimination step (cf. (D.2)), the output $x = \xi(x^*)$ of the RQAOA algorithm satisfies

$$C(x) = n - n_c + C_{n_c}(x^*) = \begin{cases} n + 1 & \text{if } \prod_{k \in \mathbb{Z}_n} J_k = 1 \\ n - 2 & \text{otherwise .} \end{cases}$$

This implies the claim. □

BIBLIOGRAPHY

[1] C.-K. Chiu, J. C. Y. Teo, A. P. Schnyder, and S. Ryu, "Classification of topological quantum matter with symmetries," *Reviews of Modern Physics*, vol. 88, p. 035 005, 3 Aug. 2016. DOI: `10.1103/RevModPhys.88.035005`.

[2] X. Chen, Z.-C. Gu, and X.-G. Wen, "Local unitary transformation, long-range quantum entanglement, wave function renormalization, and topological order," *Physical Review B*, vol. 82, p. 155 138, 15 Oct. 2010. DOI: `10.1103/PhysRevB.82.155138`.

[3] A. Y. Kitaev, "Fault-tolerant quantum computation by anyons," *Annals of Physics*, vol. 303, no. 1, pp. 2–30, 2003. DOI: `10.1016/S0003-4916(02)00018-0`.

[4] S. Bravyi, M. B. Hastings, and F. Verstraete, "Lieb-Robinson bounds and the generation of correlations and topological quantum order," *Physical Review Letters*, vol. 97, p. 050 401, 5 Jul. 2006. DOI: `10.1103/PhysRevLett.97.050401`.

[5] D. Aharonov and Y. Touati, "Quantum circuit depth lower bounds for homological codes," Oct. 2018. arXiv: `1810.03912`.

[6] M. H. Freedman and M. B. Hastings, "Quantum systems on non-k-hyperfinite complexes: A generalization of classical statistical mechanics on expander graphs," *Quantum Information and Computation*, vol. 14, no. 144, 2014.

[7] M. B. Hastings, "Trivial low energy states for commuting Hamiltonians, and the quantum pcp conjecture," *Quantum Information and Computation*, vol. 13, pp. 393–429, 2013.

[8] L. Eldar and A. W. Harrow, "Local Hamiltonians whose ground states are hard to approximate," in *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, Oct. 2017, pp. 427–438. DOI: `10.1109/FOCS.2017.46`.

[9] D. Aharonov, I. Arad, and T. Vidick, "The Quantum PCP Conjecture," *SIGACT News*, vol. 44, no. 2, pp. 47–79, Jun. 2013, ISSN: 0163-5700. DOI: `10.1145/2491533.2491549`.

[10] C. Nirkhe, U. Vazirani, and H. Yuen, "Approximate Low-Weight Check Codes and Circuit Lower Bounds for Noisy Ground States," in *Proceedings of ICALP 2018*, I. C. et al., Ed., ser. Leibniz International Proceedings in Informatics (LIPIcs), vol. 107, Dagstuhl, Germany: Dagstuhl–Leibniz-Zentrum fuer Informatik, 2018, 91:1–91:11, ISBN: 978-3-95977-076-7. DOI: `10.4230/LIPIcs.ICALP.2018.91`.

[11] Z.-C. Gu and X.-G. Wen, "Tensor-entanglement-filtering renormalization approach and symmetry-protected topological order," *Physical Review B*, vol. 80, p. 155 131, 15 Oct. 2009. DOI: 10.1103/PhysRevB.80.155131.

[12] F. D. M. Haldane, "Nonlinear field theory of large-spin Heisenberg antiferromagnets: Semiclassically quantized solitons of the one-dimensional easy-axis néel state," *Physical Review Letters*, vol. 50, pp. 1153–1156, 15 Apr. 1983. DOI: 10.1103/PhysRevLett.50.1153.

[13] I. Affleck, T. Kennedy, E. H. Lieb, and H. Tasaki, "Rigorous results on valence-bond ground states in antiferromagnets," *Physical Review Letters*, vol. 59, pp. 799–802, 7 Aug. 1987. DOI: 10.1103/PhysRevLett.59.799.

[14] N. Schuch, D. Pérez-García, and I. Cirac, "Classifying quantum phases using matrix product states and projected entangled pair states," *Physical Review B*, vol. 84, p. 165 139, 16 Oct. 2011. DOI: 10.1103/PhysRevB.84.165139.

[15] F. Pollmann, A. M. Turner, E. Berg, and M. Oshikawa, "Entanglement spectrum of a topological phase in one dimension," *Physical Review B*, vol. 81, p. 064 439, 6 Feb. 2010. DOI: 10.1103/PhysRevB.81.064439.

[16] T.-C. Wei, I. Affleck, and R. Raussendorf, "Two-dimensional Affleck-Kennedy-Lieb-Tasaki state on the honeycomb lattice is a universal resource for quantum computation," *Physical Review A*, vol. 86, p. 032 328, 3 Sep. 2012. DOI: 10.1103/PhysRevA.86.032328.

[17] J. Miller and A. Miyake, "Resource quality of a symmetry-protected topologically ordered phase for quantum computation," *Physical Review Letters*, vol. 114, p. 120 506, 12 Mar. 2015. DOI: 10.1103/PhysRevLett.114.120506.

[18] M. Morgenstern, "Existence and explicit constructions of q + 1 regular Ramanujan graphs for every prime power q," *Journal of Combinatorial Theory, Series B*, vol. 62, no. 1, pp. 44–62, 1994, ISSN: 0095-8956. DOI: 10.1006/jctb.1994.1054.

[19] E. Farhi, J. Goldstone, and S. Gutmann, "A Quantum Approximate Optimization Algorithm," Nov. 2014. arXiv: 1411.4028.

[20] M. B. Hastings, "Classical and Quantum Bounded Depth Approximation Algorithms," May 2019. arXiv: 1905.07047.

[21] M. X. Goemans and D. P. Williamson, "Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming," *J. ACM*, vol. 42, no. 6, pp. 1115–1145, Nov. 1995, ISSN: 0004-5411. DOI: 10.1145/227683.227684.

[22] G. Mbeng, R. Fazio, and G. Santoro, "Quantum Annealing: A journey through Digitalization, Control, and hybrid Quantum Variational schemes," Jun. 2019. arXiv: 1906.08948.

[23] F. Barahona, "On the computational complexity of Ising spin glass models," *Journal of Physics A*, vol. 15, no. 10, p. 3241, 1982. DOI: `10.1088/0305-4470/15/10/028`.

[24] Z. Wang, S. Hadfield, Z. Jiang, and E. G. Rieffel, "Quantum approximate optimization algorithm for MaxCut: A fermionic view," *Physical Review A*, vol. 97, p. 022 304, 2 Feb. 2018. DOI: `10.1103/PhysRevA.97.022304`.

[25] E. Farhi, D. Gamarnik, and S. Gutmann, "The quantum approximate optimization algorithm needs to see the whole graph: A typical case," 2020. arXiv: `2004.09002`.

[26] ——, "The quantum approximate optimization algorithm needs to see the whole graph: Worst case examples," 2020. arXiv: `2005.08747`.

[27] V. G. Vizing, "On an estimate of the chromatic class of a p-graph," *Diskret. Analiz*, vol. 3, pp. 25–30, 1964.

[28] A. W. Marcus, D. A. Spielman, and N. Srivastava, "Interlacing families IV: Bipartite Ramanujan graphs of all sizes," in *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, Oct. 2015, pp. 1358–1377. DOI: `10.1109/FOCS.2015.87`.

[29] A. W. Marcus, D. A. Spielman, and N. Srivastava, "Interlacing families I: Bipartite Ramanujan graphs of all degrees," *Annals of Mathematics*, vol. 182, no. 1, pp. 307–325, 2015, ISSN: 0003486X.

[30] M. Van Den Nest, "Simulating quantum computers with probabilistic methods," *Quantum Info. Comput.*, vol. 11, no. 9-10, pp. 784–812, Sep. 2011, ISSN: 1533-7146.

[31] S. Bravyi, D. Browne, P. Calpin, E. Campbell, D. Gosset, and M. Howard, "Simulation of quantum circuits by low-rank stabilizer decompositions," *Quantum*, vol. 3, p. 181, Sep. 2019, ISSN: 2521-327X. DOI: `10.22331/q-2019-09-02-181`.

[32] E. Farhi and A. W. Harrow, "Quantum Supremacy through the Quantum Approximate Optimization Algorithm," Feb. 2016. arXiv: `1602.07674`.

Part II

High Energies

Chapter 4

BUILDING BULK GEOMETRY FROM THE TENSOR RADON TRANSFORM

Using the tensor Radon transform and related numerical methods, we study how bulk geometries can be explicitly reconstructed from boundary entanglement entropies in the specific case of AdS_3/CFT_2. We find that, given the boundary entanglement entropies of a 2d CFT, this framework provides a quantitative measure that detects whether the bulk dual is geometric in the perturbative (near AdS) limit. In the case where a well-defined bulk geometry exists, we explicitly reconstruct the unique bulk metric tensor once a gauge choice is made. We then examine the emergent bulk geometries for static and dynamical scenarios in holography and in many-body systems. Apart from the physics results, our work demonstrates that numerical methods are feasible and effective in the study of bulk reconstruction in AdS/CFT.

This chapter is based on the published article:

4.1 Introduction

Recent progress [1]–[7] in quantum gravity has shown that spacetime geometry can emerge from quantum entanglement. This emergence provides appealing explanations for many intuitive properties of the physical world, including the existence of gravity [8]–[11], conditions on the allowed distribution of energy and matter [12], [13], and the unitarity of black hole dynamics [14], [15]. Most of these developments have taken place in the context of the Anti-de Sitter/Conformal Field Theory (AdS/CFT) correspondence [16], [17], although some proposals relating geometry and entanglement also apply to flat or de Sitter geometries [5], [7], [18]–[20]. As an example of the holographic principle [21], [22], AdS/CFT describes a duality between a $(d+1)$-dimensional bulk theory with the presence of gravity in asymptotic AdS spacetime, and a d-dimensional conformal field theory (CFT) without gravity on the boundary. For the purposes of studying quantum gravity, the duality is especially powerful because the CFT is an object whose basic rules we understand.

However, much remains to be understood about the relationships between the two sides of the duality.

One such challenge in the AdS/CFT duality is to understand how the boundary degrees of freedom without gravity can reorganize themselves into a higher dimensional bulk configuration with gravity. This is called the problem of bulk reconstruction, and this paper reports two results on this topic. First, we describe a perturbative procedure to reconstruct the bulk geometry given an appropriate set of boundary entanglement data. Second, we show that this reconstruction procedure can detect whether the putative bulk dual is semi-classical in the sense of having only weak fluctuations about an average value.

Our first result builds on a number of works that study bulk metric reconstruction using geodesic lengths [23], [24] or entanglement entropy [25][1]. Early efforts in this area often utilized bulk symmetries to simplify the problem of recovering the bulk metric from minimal geodesic data [33]–[35]. Later work on differential entropy and hole-ography [25], [36] furthered our understanding using information theoretic quantities and suggested a method to recover the bulk, although no explicit reconstruction formula was given for generic cases. More recently, it was shown that in certain cases, static geometries [37], or even the full dynamical metric [38], can be fixed non-perturbatively by boundary entanglement data without any prior knowledge of the bulk symmetries. However, barring a few well-known examples with symmetries, there does not exist a reconstruction procedure that directly and explicitly converts entropy data into bulk metrics. Our work addresses this missing element.

Our second result arises from the basic issue that we have a far from complete understanding of what kind of boundary states correspond to semi-classical bulk geometries. Some necessary conditions are expected from holographic entropy inequalities [39] and from consistency relations for any putative metric reconstruction procedure[2]. In general, we do not expect all quantum states from the boundary CFT to correspond to well-defined semi-classical geometries in the bulk. On the contrary, we expect an abundance of non-geometrical states obtained, for instance, by superposing states with macroscopically distinct dual geometries such as the

---

[1]More generally, but less explicitly, there are also proposals for bulk reconstruction using tensor networks [4], [26]–[29], modular Hamiltonians [30], and light-cone cuts [31], [32]. These rely on different boundary data and methodologies which we will not discuss here.

[2]One typically proceeds by assuming the boundary state has a bulk geometry and running the reconstruction procedure. If the boundary state actually does not have a bulk geometry, then the reconstruction procedure will lead to internal inconsistencies.

AdS vacuum and an AdS black-hole geometry. Therefore, to better understand holographic duality, we must also address the necessary and sufficient conditions under which a bulk geometry can emerge from the boundary state. Our work also addresses this open question.

In this paper, we study the above issues in the context of AdS_3/CFT_2 using an approach based on the tensor Radon transform. The method is for metric solutions that are close to AdS_3, hence it is restricted at present to perturbative problems. To linear order in the perturbation, each quantum state on the boundary corresponds to a constant time slice in the bulk, so we provide a numerical reconstruction algorithm that takes the entanglement entropies of intervals in the boundary state as input, and outputs the best-fit bulk metric tensor of the spatial slice in the linearized regime. This solution is unique up to gauge transformations. The algorithm also provides a quantitative indicator of whether the boundary data admits a bulk geometric description near AdS_3. This is measured by the quality of the fit, which intuitively quantifies how far the boundary entanglement data of a given state is from being geometric. A poor quality of the fit indicates the boundary data lack consistency with a semi-classical geometry.

As a proof of principle, we explore several reconstructions numerically in holography and with a 1d free fermion CFT. We find that free fermion ground states in the presence of disorder and a mass deformation do not correspond to a well-defined bulk geometries. Likewise, mixtures of states where each is dual to a distinct classical geometry can also fail to have a well-defined bulk geometry. In addition to these static examples, the method is applied to several dynamical scenarios including global and local quenches in the free fermion model and entanglement dynamics in a toy model scrambling system. In the case where the dynamics is scrambling, we find that the bulk description is qualitatively consistent with an in-falling spherical shell of bulk matter experiencing gravitational attraction. These results further demonstrate that it is both feasible and interesting to study AdS/CFT using tensor Radon transform techniques coupled with modest (laptop-scale) computational resources.

The remainder of the paper is organized as follows. In Section 4.2, we briefly review the basic assumptions, especially how entanglement entropy can be tied to metric tensors via the tensor Radon transform. We give a general review on the tensor Radon transform in Appendix 4.A. Its adaptation to a hyperbolic geometry and the gauge fixing procedure are found in Appendix 4.B and Appendix 4.C, respectively. In Section 4.3, we introduce the numerical reconstruction procedure, which we

elaborate in detail in Appendix 4.D. In Section 4.4, we apply the reconstruction algorithm to static and dynamical boundary entanglement data. In Section 4.5, we discuss these reconstructions and how geometrical and non-geometrical can be distinguished using the relative reconstruction error. Finally we conclude with some remarks and directions for future work in Section 4.6.

4.2 Boundary Rigidity and Bulk Metric Reconstruction

We begin with the vacuum state $|0\rangle_{\text{CFT}}$ of a hologaphic CFT. Because this state has conformal symmetry at all scales, it must be dual to empty AdS [16]. When the bulk theory is Einstein gravity coupled to matter, the bulk geometry controls the leading entanglement structure of the CFT state via the Ryu-Takayanagi (RT) formula [2], [3]. Given a boundary region A, the RT formula computes the von Neumann entropy $S(A)$ in terms of a minimal area surface,

$$S(A) = \frac{1}{4G} \min_{\gamma_A} \text{Area}[\gamma_A],$$

where $\text{Area}[\gamma]$ is the area of γ_A, G is Newton's constant, and the minimum runs over bulk surfaces γ_A that are homologous to A. In the time symmetric case where RT applies, all of these surfaces can be taken to lie in a time-symmetric spacelike surface Σ. In other words, the Ryu-Takayanagi formula says that the von Neumann entropy of a state on a boundary subregion is given by area of the minimal area bulk surface that subtends the region.

If we have access to a boundary state, in the sense of knowing its von Neumann entropies on all (connected) subregions, then the RT formula translates these entropic quantities into a set of boundary anchored minimal surface areas. Because these minimal surfaces all lie on the spatial slice Σ, recovering the bulk geometry from entanglement reduces to a pure geometry problem where we try to find the interior metric g_{ij} of a Riemannian manifold $\mathcal{M}$ while knowing only the areas of minimal surfaces that are anchored to its boundary $\partial\mathcal{M}$. This is precisely the statement of the boundary rigidity problem [40], which is well-studied in the field of integral geometry [41].

For this work, we focus exclusively on the case of $\text{AdS}_3/\text{CFT}_2$, where minimal surfaces are simply geodesics and spatial slices are 2d Riemannian manifolds. For a class of 2d Riemanian manifolds (called *simple* manifolds), it is known that the lengths of all boundary-anchored geodesics indeed fixes the bulk metric uniquely up to gauge equivalence [42]. Because the single interval von Neumann entropies in the

ground state of a 2d CFT are universally determined by the central charge [43], the RT formula combined with boundary rigidity completely fixes the bulk geometry to be that of hyperbolic space [44], [45]. Of course, this is precisely the induced geometry on a time-symmetric slice of AdS, as it had to be based on the grounds of symmetry[3].

It is natural to ask whether we can exploit the power of the RT formula and the results from boundary rigidity theory to reconstruct dual geometries from boundary states other than the vacuum. For instance, given a generic non-vacuum state $|\psi\rangle_{\text{CFT}}$, can we apply the same principles to reconstruct the metric tensor for the bulk geometry from the set of boundary-anchored geodesic lengths? There are several obstacles that prevent us from recovering the bulk metric tensor exactly using the above methods, even if $|\psi\rangle_{\text{CFT}}$ has a well-defined dual geometry. First, since it may not be possible for minimal surfaces (or extremal surfaces in the dynamical case [46]) to foliate entire space(time) manifold, thus we can only reconstruct regions where there is at least a local foliation with minimal or extremal surfaces. Second, even for the regions whose geometries are fixed by the entanglement data [38], there is no explicit reconstruction formula for the general boundary rigidity problem.

Although such problems are difficult to solve in general, it is typically easier to reconstruct the difference between the dual geometry and a known reference, or background, geometry. This is known as the linearized boundary rigidity problem [40], [41]. In this work, instead of a direct reconstruction of the of the dual geometry for $|\psi\rangle_{\text{CFT}}$, we reconstruct the differences in the entanglement patterns as linearized metric perturbations around the AdS background.

To do so, we first fix the background geometry to be vacuum AdS$_3$. Working with a given constant-time slice, let us suppose that $|\psi\rangle_{\text{CFT}}$ has a slightly different entanglement structure compared to $|0\rangle_{\text{CFT}}$, and is dual to a bulk geometry with a metric that is close, but not equal, to that of pure hyperbolic space on our time-slice. Then by the RT formula, the boundary-anchored geodesic lengths now differ slightly from those of pure hyperbolic space. For a given boundary sub-region A, the change in the geodesic length anchored at A is related to the vacuum subtracted entropy of

---

[3]Since the lengths of geodesics on an asymptotically AdS spacetime are divergent when extended to the boundary, it is understood that the geodesics are actually regularized to be anchored on a cutoff surface that is some finite distance away from an arbitrarily chosen coordinate in the bulk. For the rest of this work, we will always impose a UV cutoff for the CFT, or equivalently, an IR cut off in the bulk geometry so that the geodesic lengths stay finite.

the state by

$$\Delta L(A) = L_\psi(A) - L_0(A) = \frac{S_\psi(A) - S_0(A)}{4G},$$

where $L_\psi(A)$ and $L_0(A)$ denotes the lengths of geodesics anchored at the end points of A for states $|\psi\rangle_{\text{CFT}}$ and $|0\rangle_{\text{CFT}}$, respectively.

This change in geodesic length corresponds to a change in the bulk metric

$$g_{ij}^{(0)} \mapsto g_{ij} = g_{ij}^{(0)} + h_{ij},$$

where $g_{ij}^{(0)}$ is the pure hyperbolic metric and h_{ij} is the perturbation. For the linearized problem, the goal is to find h_{ij} to leading order, with h_{ij} viewed as a rank two symmetric tensor field on the hyperbolic background. Note that geodesics of the background metric remain geodesics of the perturbed metric to first order in h since geodesics satisfy an extremality condition. Changes in entanglement due to variations in the minimal surface itself are of order h^2 [8]. The leading order change in geodesic length can then be written as

$$\begin{aligned}
\Delta L(A) &= L_\psi(A) - L_0(A) \\
&= \int_{\gamma_A} \sqrt{(g_{ij}^{(0)} + h_{ij})\dot{\gamma}_A^i \dot{\gamma}_A^j}\, ds - \int_{\gamma_A} \sqrt{g_{ij}^{(0)} \dot{\gamma}_A^i \dot{\gamma}_A^j}\, ds \\
&= \frac{1}{2} \int_{\gamma_A} h_{ij} \dot{\gamma}_A^i \dot{\gamma}_A^j\, ds + O(h^2),
\end{aligned} \tag{4.1}$$

where γ_A is the geodesic of the hyperbolic background anchored at the end points of A and $\dot{\gamma}_A^i$ denotes unit tangent vectors along γ_A. The tangent vectors are normalized such that

$$g_{ij}^{(0)} \dot{\gamma}_A^i \dot{\gamma}_A^j = 1.$$

For simplicity, we will use $L(\gamma_A)$ and $L(A)$ interchangeably, often dropping the explicit dependence on A and simply writing $L(\gamma)$ when there is no confusion, with the understanding that γ is a boundary-anchored geodesic.

To make concrete progress in this work, we simplify the full problem by taking the linearized bulk result for changes in geodesic length to be equal to the full change in boundary entanglement entropy,

$$\Delta L(\gamma) \approx \frac{1}{2} \int_\gamma \dot{\gamma}^i \dot{\gamma}^j h_{ij}\, ds.$$

This approximation enables calculations; going beyond it may be technically non-trivial, but it is likely not a fundamental obstacle.

With this simplification, the length perturbation becomes the integrated longitudinal projection of the metric perturbation along a geodesic γ. This is precisely the *tensor Radon transform* $R_2[h]$ of the metric perturbation h_{ij} [47]. Given a symmetric 2-tensor field h_{ij}, the tensor Radon transform $R_2[h]$ defines a map from the space of boundary geodesics to the complex numbers given by

$$R_2[h_{ij}](\gamma_A) \equiv \int_{\gamma_A} h_{ij} \dot{\gamma}_A^i(s) \dot{\gamma}_A^j(s) \, ds,$$

where γ_A is the background geodesic anchored at the boundary of A.

As an aside, we note that there exist several related notions of Radon transform. A standard Radon transform on a Riemannian manifold is defined by integrating some quantity on a minimal co-dimension one surface, whereas an X-ray Radon transform is defined similarly, but for a dimension one surface, i.e., a geodesic. For two spatial dimensions, as is the case we consider here, the two definitions coincide. For more details on the Radon transform, see Appendix 4.A.

Formally, the bulk metric deformation can be recovered by inverting the tensor Radon transform. Schematically, we can write

$$R_2^{-1}[2\Delta L] = R_2^{-1}[R_2[h_{ij}]] = h_{ij},$$

where R_2^{-1} denotes some (not yet properly defined) inverse Radon transform, and where ΔL denotes the collection of boundary anchored geodesic length deviations.

Throughout, we work exclusively in the perturbative regime, to leading order in h, which allows us to relate geodesic length deformations to the Radon transform through (4.2). It also ensures that the resulting geometric solution, when it exists, is uniquely determined by the boundary entropy data [42], [45].[4] We wish to comment here that despite restrictions to the perturbative regime, h_{ij} can still capture highly non-trivial physics. Indeed, standard calculations of gravitational waves and the dynamics of typical stars, planets, and galaxies are all done in the weak-field regime.

Before we can proceed with the inversion process, we must give meaning to the inverse Radon transform. For this purpose, it is important to note that the Radon

---

[4]In general, we are not guaranteed a unique solution for the bulk metric, even if one exists, because the assumption that the manifold is simple breaks down for sufficiently large deviations from a constant curvature background [23]. Working in the perturbative limit ensures that the Radon transform remains well-defined. See Appendix 4.A.

transform has a non-trivial kernel: given any vector field ξ on M such that $\xi|_{\partial M} = 0$, we necessarily have

$$R_2[\nabla_i \xi_j + \nabla_j \xi_i] = 0.$$

Physically, any Radon transform of a pure gauge deformation that reduces to the identity at the boundary is zero. The presence of this kernel is natural because the transform relates geodesic lengths (which are gauge invariant) to metric tensors (which are not), so the Radon transform can only be injective up to gauge.

In the presence of such a kernel, we must fix a gauge prescription in order to recover a metric tensor uniquely. We will use a prescription which we call the *holomorphic gauge* [48]. In a crude sense, the holomorphic gauge preferentially reconstructs the trace part of the metric at the cost of diminishing non-zero contributions to the off-diagonal.[5] This provides two independent gauge constraints in two spatial dimensions, which would allow us to proceed with the reconstruction. For a more detailed description of the gauge constraints and Radon transform on a hyperbolic background, see Appendix 4.B.

As a final comment, in the above discussion we have explicitly assumed that the boundary state corresponded to a well-defined bulk geometry. Below we will construct algorithmic machinery which can actually carry out the reconstruction in this case. However, we will also see that the algorithm can be applied to a more general class of states, with interesting results.

4.3   Numerical Methods for Reconstruction

The inversion formula for the flat-space scalar Radon transform is a well-known classical result in integral geometry [41]. Explicit reconstruction formulas for scalar and vector Radon transforms are also available for constant negative curvature backgrounds [42], [49]. However, there are currently no explicit reconstruction formulas available for higher rank tensors on curved backgrounds, although several results in the literature come close to a solution in various regimes [42], [48]–[52]. In the absence of an exact analytic reconstruction formula, we instead draw inspiration from the general principles employed in seismology to study the Earth's interior [53] in developing our numerical method.

In this section, we give a brief overview of the method. The full details of the

---

[5]There exists other gauge fixing prescriptions as well. The most commonly considered prescription is known as the *solenoidal gauge*. See Appendix 4.A. We choose an alternative gauge prescription for various reasons of convenience. For more details on gauge fixing, see Appendix 4.C.

discretization, gauge fixing, and solutions for the constrained least square problem can be found in Appendix 4.D.

4.3.1 Discretization and Optimization Procedures

The basic idea behind our numerical reconstruction is straightforward. We first discretize the bulk and boundary regions into a finite number of tiles. To each tile $\mathcal{T}$ in the bulk, we associate a tensor $h_{ij}(\mathcal{T})$, and for each interval A on the boundary, we associate a geodesic (of the background metric) γ_A anchored on the endpoints of the interval. In two spatial dimensions, a rank-2 symmetric tensor has 3 independent degrees of freedom. Each geodesic anchored at the end points of an interval generates a linear equation via the discretized version of the Radon transform (4.2), defined by

$$2\Delta L(\gamma_A) = R_2[h_{ij}] \approx \sum_{\mathcal{T}} W^{ij}(\mathcal{T}, \gamma_A) h_{ij}(\mathcal{T}), \tag{4.2}$$

where the tensor W^{ij} contains information about the direction of the tangent vectors $\dot{\gamma}_A$, as well as the arc length $\Delta s(\gamma_A, \mathcal{T})$ of the geodesic segment that passes through each tile $\mathcal{T}$. In Equation (4.3.1), we sum over all bulk tiles $\mathcal{T}$ and over repeated indices. Naturally, one has $W^{ij}(\mathcal{T}, \gamma_A) = 0$ if the geodesic does not pass through a tile $\mathcal{T}$.

We can abbreviate Equation (4.3.1) in matrix form as

$$\mathbf{b} = \mathbf{Wh},$$

where $\mathbf{W}$ and $\mathbf{h}$ are vectorized representations of W^{ij} and h_{ij}, and where $\mathbf{b}$ denotes the corresponding geodesic length deformations. As a result, given a specific discretization, the discretized forward Radon transform can be written as a linear map $W : V_B \to V_\gamma$ from the space of tile-wise constant bulk tensor valued functions V_B to the space of boundary anchored geodesic lengths V_γ. Both spaces are finite dimensional due to the discretization.

Since the forward Radon transform has a non-trivial kernel in the continuum limit, we must impose a gauge fixing condition to recover a unique solution. We give the full detail of the gauge fixing conditions and the accompanying partial differential equations in Appendix 4.C. To ensure the problem is well-posed, we set the holomorphic gauge constraints as discretized partial differential equations, which we formally write as

$$\mathbf{Ch} = \mathbf{0}. \tag{4.3}$$

Here, **C** denotes the constraint matrix representing the partial differential operator associated with the gauge constraint. Reconstruction of the metric perturbation then corresponds to finding the solution to the linear equations (4.3.1) above, subject to the linear constraints (4.3.1).

In practice, there does not always exist exact solutions h_{ij} which satisfies the constrained system. This can be due to a variety of reasons, such as the presence of discretization errors, or if the boundary data is simply inconsistent with a geometric bulk, i.e., if the boundary entropy function fails to lie within the range of the forward Radon transform [41], [42], [48]. Instead of trying to look for an exact solution, it is more natural to look for the best-fit solution $\mathbf{h}_*$ which solves the constrained minimization problem

$$\min_{\mathbf{h}} \|\mathbf{W}\mathbf{h} - \mathbf{b}\|,$$

$$\text{subject to } \mathbf{C}\mathbf{h} = \mathbf{0}. \tag{4.4}$$

The objective function is linear and we are guaranteed a unique global minimum. Thus we will say that $\mathbf{h}_*$ is the optimal geometric solution corresponding to boundary data **b**. We will also write $\mathbf{h}_*(\mathbf{b})$ when we need to denote the dependence of $\mathbf{h}_*$ on the initial boundary data.

Even with the existence of an optimal reconstruction $\mathbf{h}_*$, we do not expect generic boundary data to correspond to a geometric dual in general. A useful quantity is the relative error of reconstruction, which measures the tension between the best-fit solution and the actual data. We can consider various relative errors. If we know the exact bulk solution, say $\mathbf{h}_0$, then we can denote the bulk relative error as

$$\mathcal{E}_{\text{bulk}} = \frac{\|\mathbf{h}_* - \mathbf{h}_0\|}{\|\mathbf{h}_0\|},$$

where $\mathbf{h}_*$ is the bulk metric tensor reconstructed from the forward transform of **h**.

More commonly however, we do not have access to an a priori geometric state. Instead, we have a CFT state $|\psi\rangle_{\text{CFT}}$ from which we can extract discretized boundary data $\mathbf{b}_\psi$. In this case, we can likewise consider the boundary relative error, defined by

$$\mathcal{E}_{\text{bdy}} = \frac{\|\mathbf{W}[\mathbf{h}_*(\mathbf{b}_\psi)] - \mathbf{b}_\psi\|}{\|\mathbf{b}_\psi\|}. \tag{4.5}$$

The boundary relative error is simply the normalized distance from $\mathbf{b}_\psi$ to the subspace of boundary data vectors with geometric duals, which is the same quantity

minimized by the constrained least squares problem (4.3.1). The boundary relative error therefore serves to quantify the degree to which a state is geometric or non-geometric. We will discuss in greater detail in Section 4.5 how the relative errors can be used to distinguish geometric data from non-geometric data on the boundary.

To ensure the reliability of the reconstruction algorithm for the inverse tensor Radon transform, we also perform the numerical inversion of boundary data $\mathbf{b}$ whose bulk tensor field is known. We produce such boundary data by preparing various known bulk tensor valued functions $\mathbf{h}_0$ in the holomorphic gauge, and then generating its corresponding geodesic data $\mathbf{b_0}$ through a forward tensor Radon transform. Subsequently, we apply the numerical reconstruction to the geodesic data and compare the reconstructed $\mathbf{h}_*(\mathbf{b_0})$ to the original test function $\mathbf{h}_0$. We find remarkable agreement in our reconstructions. Absent rigorous analytic convergence guarantees, the successful benchmarking of the algorithm on known cases serve to provide confidence in the fidelity of the reconstruction. Details of this benchmarking process are given in Appendix 4.E.

4.4 Reconstructed Geometries

The forward tensor Radon transform is generally neither injective nor surjective. Therefore, not all boundary data can be interpreted as the Radon transform of a bulk tensor field. However, it may be possible to derive necessary and sufficient characterization of what boundary data corresponds to a bulk tensor field. Such a criterion is known as a *range characterization* of the tensor Radon transform. Ranges characterizations of various transforms have been discussed extensively for scalar and vector cases on curved backgrounds and tensor cases on flat background [41], [42], [48]. However, it remains an active topic of research for transforms on curved background for higher rank tensor fields.

Although we lack a rigorous analytic characterization for the tensor Radon transform, our numerical methods can still effectively capture the parts of the entanglement data that do not lie within the range of the tensor Radon transform. Since the Radon transform is linear, any contribution that cannot be fitted to a bulk tensor field in the global best-fit reconstruction effectively captures the non-geometric contribution for the discrete reconstruction. More specifically, the relative boundary error (4.3.1) serves as a indicator for the fraction of the boundary entropy data which does not lie within the range, i.e., the fraction which can be considered non-geometric.

In the upcoming sections, we reconstruct geometries from boundary entanglement

entropies generated by holographic systems as well as those generated numerically from a 1d free fermion system. We then discuss how their relative boundary errors $\mathcal{E}_{\mathrm{bdy}}$ can be used as a standard to distinguish states that have a well-defined classical dual geometry from the ones that do not. This provides a direct quantitative condition for whether a state is geometric.

For entanglement entropies generated by holographic systems, we expect a weakly coupled gravity dual, and therefore spacetime geometries that are generically classical at low energies. The same cannot be expected for a generic many-body quantum system at criticality [54], although they may still capture interesting features using a dual spacetime prescription. For instance, in a free fermion system, the conformal field theory has neither strong coupling nor a large central charge. Although it is unclear what the dual bulk description would be, it is generally expected that any geometric description must be one where gravity is strongly coupled, the leading order RT formula does not apply, and the spacetime is dominated by quantum effects. Therefore, the reconstruction is likely poor in the absence of a large number of symmetries. We will find that numerical evidence support these basic intuitions.

It is difficult to present certain reconstructions from dynamical systems in a paper. A list of the animated reconstructions for various dynamical processes are linked here.[6]

4.4.1 Holographic Reconstructions

As a benchmark for holographic geometry reconstructions, we first reconstruct the metric for the thermal AdS geometry in the linearized regime. The entropy data we use are generated by the minimal geodesic lengths in a BTZ geometry [55] using the Ryu-Takayanagi formula. Since the data corresponds to a bulk geometry by construction, the reconstruction should show good agreement with said geometry at linear order. This includes the correct qualitative behaviour in the bulk, and a positive metric perturbation present deeper into the bulk. It must also afford relatively small reconstruction errors, which can be attributed to factors such as discretization errors, approximations made in linearization, and working on a fixed hyperbolic background despite the changes in bulk geometry.

Indeed, as shown in Figure 4.1, we find good agreement with our expectations in the reconstruction. By probing deeper into the bulk, the geodesics for a thermal geometry travel through larger distances, resulting in a net positive change in the

---

[6]Link: https://www.youtube.com/playlist?list=PLCjJ3kjqxOfw1aIa5c0X6KSpox1-AjM5b
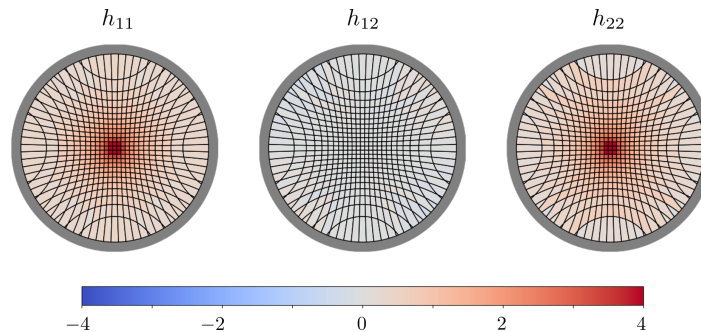
Figure 4.1: Thermal state reconstruction. Plot shows the individual tensor components h_{11}, h_{12}, h_{22} of the metric perturbation from left to right, respectively. The boundary relative errors are $\mathcal{E}_{\text{bdy}} \lesssim 10^{-2}$ across a wide range of temperatures.

metric perturbation. Note that at linear order, we cannot detect a change in bulk topology, which the entanglement data dual to a BTZ geometry should predict. This is because of the fixed background metric.[7]

Motivated by entanglement dynamics in holography, we can also construct heuristics for the entanglement growth of a boundary theory. For instance, under a global quench, we expect qualitative growth of entanglement to be captured by

$$S_A(t) \approx \min\{svt, s|A|\} \tag{4.6}$$

for $t \geq 0$, where entanglement for any region will grow linearly in time after the initial stage, until the entropy satisfies a volume law [56], [57]. Here s denotes the entropy density, v the speed of entropy growth, and t the time that has elapsed since the quench. The size of the boundary interval is denoted by $|A|$. The corresponding metric and curvature perturbations are shown in Figure 4.2.

As the system thermalizes, larger subregions have a volume law entropy. The wavefront of the entanglement spread is reflected in the bulk as a spherically symmetric perturbation moving from the boundary to the center. Assuming Einstein gravity, which holds for holographic CFTs, the curvature perturbation δR also reflects the bulk matter distribution through the linearized Hamiltonian constraint for each instance of time [47]. Therefore, this thermalization process is consistent with the collapse of a spherical shell of matter [56].

---

[7]Suppose the background metric is updated using the new found metric perturbation. Then we should recover the fact that the corrected geodesics avoid the central region of the bulk. However, geodesic avoidance alone does not necessarily indicate the formation of a horizon. For instance, adding a single massive particle in AdS will lead to a back-reacted geometry where geodesics avoid the region near the inserted mass.

(a) Component h_{11}.

(b) Linearized scalar curvature perturbation.
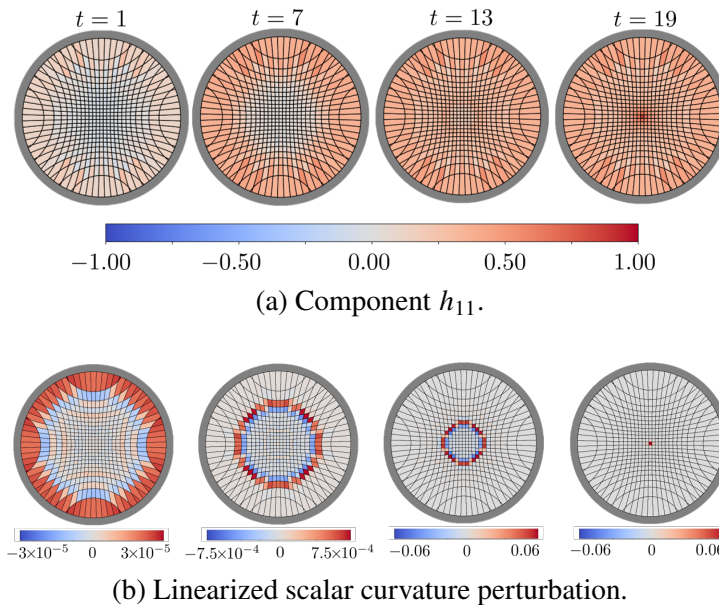
Figure 4.2: Metric perturbation from volume law entropy growth (4.4.1). Plots are ordered from left to right as time increases. We only give h_{11} for the sake of clarity, because $h_{12} \approx 0$ and $h_{22} \approx h_{11}$. The boundary relative error is $\mathcal{E}_{\text{bdy}} \approx 0.03$.

Mixture of Thermal States

There are also instances where we do not expect a well-defined geometry to exist. For example, when the state is taken from a theory where the bulk is strongly coupled and/or the state is a macroscopic superposition of certain classical geometries, quantum gravitational effects can dominate, leading to a breakdown of the classical geometric description.

In this section, as an example of a potentially non-geometric holographic state, we look at mixtures of thermal AdS geometries at various temperatures. We will consider states of the form

$$\rho = p\rho(T_1) + (1 - p)\rho(T_2),$$

where $\rho(T_i)$ are thermal states of a CFT with distinct temperatures T_1 and T_2. From [58], the von Neumann entropy of the mixture is estimated as

$$S(\rho_A) = S(\rho_{1,A}) + S(\rho_{2,A}) + H(p),$$

where we write $\rho_{i,A}$ to denote the reduced state of $\rho(T_i)$ on a boundary region A, and where

$$H(p) = -p \log p - (1 - p) \log(1 - p)$$

is a Shannon-like term that corresponds to the entropy of mixing [59]. Physically, such a state can be created by superposing two thermofield double states at different temperatures, and then tracing out one side of the wormhole. For CFTs with strong bulk gravity where $G_N \approx 1$, we find the superposition incurs a large error, indicating non-geometric configurations in the bulk. However, for a CFT with a weakly coupled dual where $G_N \ll 1$, the geometry smoothly interpolates between the two temperatures at the linearized level, consistent with our expectation that the entropy operator is proportional to the area operator at leading order in N. In the language of Radon transform, this is caused by the entropy of mixing $H(p)$ being a non-geometrical contribution.
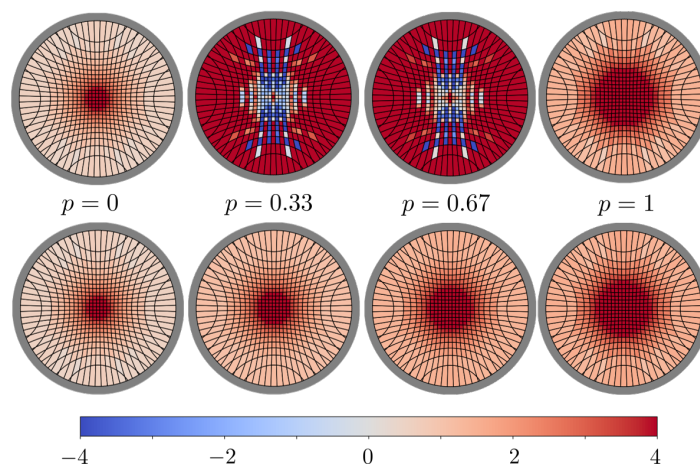
Figure 4.3: Component h_{11} of a mixture of two thermal geometries at two distinct temperatures. Top diagram for $G_N \approx 1$ and bottom for $G_N \ll 1$, both with $L_{\mathrm{AdS}} = 1$. From left to right, we increase the mixing ratio p and the geometry transitions smoothly from one at lower temperature to the other at higher temperature when the gravitational coupling is weak. The reconstruction is dominated by artifacts and have large error if the coupling is strong, leading to a contribution $G_N H(p)$ in the leading order.

4.4.2 1D Free Fermion

Given that the reconstruction procedure relies only on the von Neumann entropy of a state, we may also apply it to various quantum many-body systems where we do not necessarily expect the conformal field theory to be dual to a semi-classical bulk theory with weakly coupled gravity. The lack of a semi-classical geometric dual will generally be reflected in the presence of a large boundary reconstruction error. One example where the separation between geometric and non-geometric states is particularly manifest is the contrast between the reconstruction of holographically motivated data and that of a 1d free fermion.

Generic low-energy states for free fermion systems are generally believed to have a highly quantum bulk because the conformal field theory is non-interacting, and has a small central charge. Heuristically, the strong-weak nature of the holographic duality suggests that such systems should result in a strongly coupled bulk geometry, if such geometries even exist. In this section, we give a few examples indicating that generic excited states in the free fermion system are indeed non-geometric. However, we shall also see that certain large-scale geometric properties may nevertheless be present, even if the overall state is ostensibly non-geometric.

For our reconstruction, we consider a 100-site massless free fermion system with periodic boundary conditions, and with Hamiltonian

$$\hat{H}_0 = -\sum_i \hat{a}_i^\dagger \hat{a}_{i+1} + \text{h.c.},$$

where, as usual, the creation and annihilation operators satisfy the canonical anti-commutation relations

$$\{\hat{a}_i, \hat{a}_j^\dagger\} = \delta_{ij}, \quad \{\hat{a}_i, \hat{a}_j\} = 0, \quad \{\hat{a}_i^\dagger, \hat{a}_j^\dagger\} = 0.$$

We shall also consider various quench dynamics and deformations of the free fermion system.

The free fermion system is described by a $(1 + 1)$-d conformal field theory in the thermodynamic limit, with central charge $c = 1/2$ [60]. Here we study directly the regulated lattice model. Any such model of non-interacting fermionic modes can be studied efficiently numerically with a cost scaling polynomially with the number of fermion modes N. In particular, we can compute all of the single interval entanglement entropies via [61]. The key point is that the state of the system is Gaussian. In particular, given a subset A of the fermions, the reduced density matrix is $\rho_A \propto e^{-K_A}$ with K_A quadratic in $\hat{a}$ and $\hat{a}^\dagger$,

$$K_A = \hat{a}^\dagger k_A \hat{a}.$$

Here k_A is an $|A| \times |A|$ Hermitian matrix that determines all correlation functions of fermions in A. It is related to the 2-point function via

$$G_{ij}^A = \langle \hat{a}_i^\dagger \hat{a}_j \rangle \big|_{i,j \in A} = \left(\frac{1}{e^{k_A^{\mathrm{T}}} + 1} \right)_{ij}.$$

The entropy of A is then determined by the eigenvalues of k_A, but there is also a direct formula in G^A:

$$S(A) = -\text{tr}\left[G^A \ln G^A + (1 - G^A) \ln(1 - G^A) \right].$$

To fix our background geometry, we numerically compute the ground state entanglement of the critical Hamiltonian $\hat{H}_0$. To consider non-vacuum emergent geometries, we consider states $|\psi\rangle \neq |0\rangle$, which are not necessarily energy eigenstates of $\hat{H}_0$. For instance, these excited states can be generated by first deforming the Hamiltonian away from criticality through a mass deformation

$$\hat{H}_0 \mapsto \hat{H}_0 + m\hat{H}_1,$$

and then finding the ground state $|\psi_m\rangle$ of the perturbed Hamiltonian. For small m, the new ground state $|\psi_m\rangle$ is generically an excited state with relatively low energy in the unperturbed system.

We will study reconstructions in both the static and dynamical cases. For the static case, we reconstruct the emergent geometry of the new ground states $|\psi_m\rangle$ by finding the best-fit geometry using the discrete Radon transform. We do this for a number of distinct values of m. For the dynamical case, we consider the quench dynamics corresponding to the deformation $\hat{H}_1$. We start with some fixed deformed ground state $|\psi_m\rangle$ and then time evolve the state with the free Hamiltonian $\hat{H}_0$ and reconstruct the geometry that corresponds to each time step: $|\psi_m(t)\rangle = \exp(-it\hat{H}_0)|\psi_m\rangle$.

Local Deformations

In this section and the next, we will consider deformations of the form

$$\hat{H}_1 = \sum_{i \in S} \hat{a}_i^\dagger \hat{a}_{i+1} + \text{h.c.}, \tag{4.7}$$

where m is a small positive parameter, and where S is a set of sites where we introduce such deformations. The perturbation serves to deform the ground state wavefunction around the sites near S. We label the sites counter-clockwise from 0 to 99, starting with site 0 aligned with the positive x-axis (i.e., site i sits on the unit circle at angle $\theta_i = 2\pi i/100$). We will consider local perturbations where S consists of one or two sites in this section, and more global perturbations in the next section.

Figure 4.4 shows the best-fit geometry arising from the ground state $|\psi_m\rangle$ corresponding to a local mass deformation at a single site $S = \{0\}$, located along the positive x-axis. We can see that, in contrast to the holographic reconstructions in Section 4.4.1, the geometries shown in Figure 4.4 are heavily dominated by noise and local artifacts. The corresponding relative error of reconstruction is also larger by more than an order of magnitude as compared to the holographic cases (see

Figure 4.13a in Section 4.5 for a more comprehensive comparison of relative errors). The large error of reconstruction can be seen as a hallmark of the fact that the underlying state is non-geometric.

$$h_{11} \qquad\qquad h_{12} \qquad\qquad h_{22}$$

Figure 4.4: Components of the reconstructed metric tensor perturbation corresponding to the ground state of a mass deformed Hamiltonian, with the deformation located at a single site located along the positive x-axis. The boundary relative error is $\mathcal{E}_{\mathrm{bdy}} \approx 0.62$.

Note that the lack of geometric features in the reconstruction does not necessarily indicate that the reconstruction procedure is flawed. In fact, non-geometric features are expected because not all boundary data should produce geometric reconstructions. Our reconstruction procedure has equipped us with information of telling apart when that will happen. We will discuss this further in Section 4.5.

Nevertheless, there are cases where some large-scale geometric features can be extracted from the plot. This is especially clear in the dynamical case, where we consider the quench dynamics obtained by evolving the deformed ground state $|\psi\rangle$ using the free fermion Hamiltonian $\hat{H}_0$. The reconstruction, performed timeslice by timeslice, is shown in Figure 4.5 (only the dominant component h_{11} is shown). While it may not be entirely clear, Figure 4.5 reveals a large scale shockwave that originates from the deformation site and then travels across the bulk before being reflected at the left boundary.

Figure 4.5: Quenched time dynamics of the reconstructed geometry corresponding to a single mass deformation along the x-axis (see Figure 4.4 for the $t = 0$ geometries). Only the dominant component h_{11} is shown here. The average reconstruction error over time is $\mathcal{E}_{\text{bdy}} \approx 0.52$.

The large-scale geometry is heavily obscured by non-geometric noise in Figure 4.5. To better extract the large-scale features that we consider to be relevant, we must filter out the small-scale artifacts what we consider to be "noise."

To more clearly extract the underlying large-scale features of what appears to be waves generated by the perturbations, we perform some pre-processing of the boundary data to smooth out the small-scale noise. This is done by averaging the entanglement entropies of two neighbouring intervals of the same size. More precisely, suppose $S(i, j)$ is the von Neumann entropy on the interval from site i to site j. Then the smoothing procedure we are applying is similar to a filter by convolving with a 2-site window function such that

$$S(i, j) \rightarrow \frac{S(i, j) + S(i + 1, j + 1)}{2}.$$

This produces a less noisy reconstruction and significantly reduces the amount of non-geometric contributions to boundary relative errors. Intuitively, the smoothing procedure acts as a low-pass filter in the space of boundary intervals. This serves to remove the short distance artifacts normally associated with non-geometricality. While we used a specific filter in this example, one may apply more general filter constructions for other purposes.[8]

---

[8]It is reasonable to suspect that the two site averaging of entanglement entropy works well as a filter here because UV details in the free fermion Hamiltonian (4.4.2) with Fermi momentum $\pm\pi/2$ contributed to non-geometric noise. We might therefore expect that we can remove such

We wish to emphasize that this filtering procedure is not a part of our reconstruction process. It is merely here to assist us in gaining some intuition regarding the qualitative physical behaviour of this type of dynamical process. One could imagine that the filtering reveals what the bulk physical process might have looked like, had the state actually been geometric.

The corresponding smoothed geometries are shown in Figure 4.6. It can be seen that smoothing significantly reduces local noise, and clearly reveals large-scale time dynamics associated with the quench that looks like a (shock)wave.
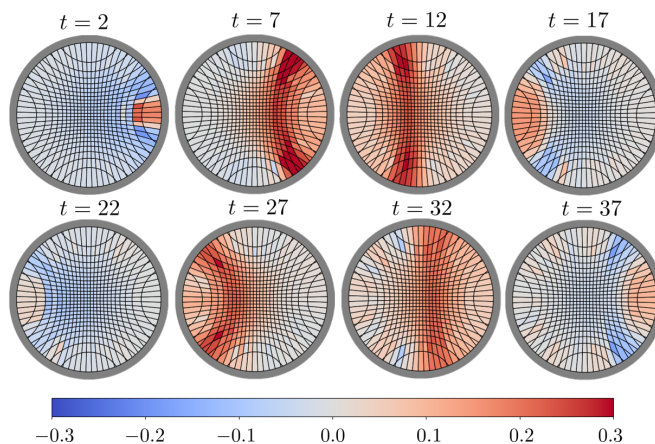
Figure 4.6: Quenched time dynamics of the reconstructed geometry corresponding to a single mass deformation along the x-axis. The boundary data is pre-processed by smoothing to remove small-scale details (see Figure 4.5 for the corresponding unsmoothed reconstruction). Only the dominant component h_{11} is shown here. The relative error is $0.03 \lesssim \mathcal{E}_{\text{bdy}} \lesssim 0.1$.

As another example, the quench dynamics of a state with two distinct deformations at sites $S = \{0, 30\}$ is shown in Figure 4.7 (unsmoothed) and Figure 4.8 (smoothed).

non-geometric contributions through coarse-graining by grouping together adjacent sites in pairs. While this grouping does indeed remove some non-geometric noise, it does not remove it entirely: the reconstruction error after grouping is $0.1 \lesssim \mathcal{E} \lesssim 0.3$, which is in between what is shown in Figure 4.5 and Figure 4.6. Given that the smoothed data reconstruction error is still significantly larger than that of completely geometric data, it is not clear if the non-geometric contribution completely reside in the UV.
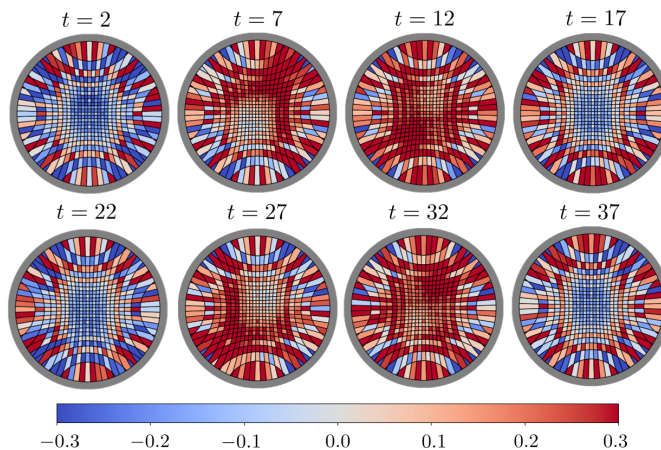
Figure 4.7: Quenched time dynamics of the reconstructed geometry corresponding to a mass deformation at two distinct sites ($S = \{0, 30\}$). The trace of h_{ij} is shown here. The boundary relative error is $\mathcal{E}_{\text{bdy}} \approx 0.41$.
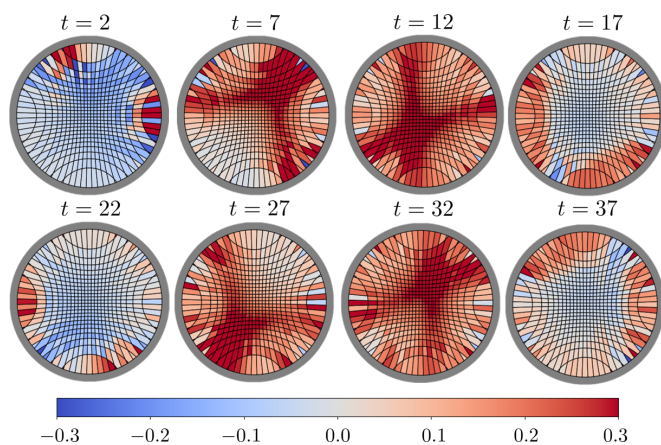
Figure 4.8: Quenched time dynamics of the reconstructed geometry corresponding to a mass deformation at two distinct sites ($S = \{0, 30\}$). The boundary data is pre-processed by smoothing to remove small-scale details. The trace of h_{ij} is shown here. The boundary relative error is $\mathcal{E}_{\text{bdy}} \approx 0.05$.

In all cases, we see that the free fermion Hamiltonian will generate seemingly non-interacting waves that traverse through the hypothetical bulk spacetime. Since the underlying time dynamics is integrable, the same entanglement feature recurs after the waves traverse the entire system; we show one such iteration in our figures.

Global Deformations

Similarly to the previous analysis for local deformations, we can also consider global deformations in the same vein. In this section, we will consider Hamiltonian

perturbations of the same form as (4.4.2), but now with deformations located at every other site, i.e., $S = \{0, 2, 4, \cdots, 98\}$.

In Figure 4.9, we show the geometries reconstructed from the ground state $|\psi_m\rangle$ of a Hamiltonian with the aforementioned global deformations, plotted across a range of m values. Again, we find the overall geometry to be highly dominated by noise, with a large relative error of reconstruction indicating that the underlying state is non-geometric. The relative error of reconstruction generally becomes worse with increasing values of m (see Figure 4.13b). Qualitatively, the reconstructions for the different values of m appear similar, the main difference being the overall magnitude of the metric perturbation.
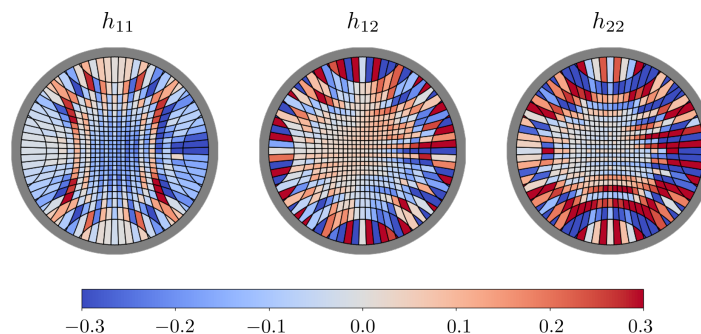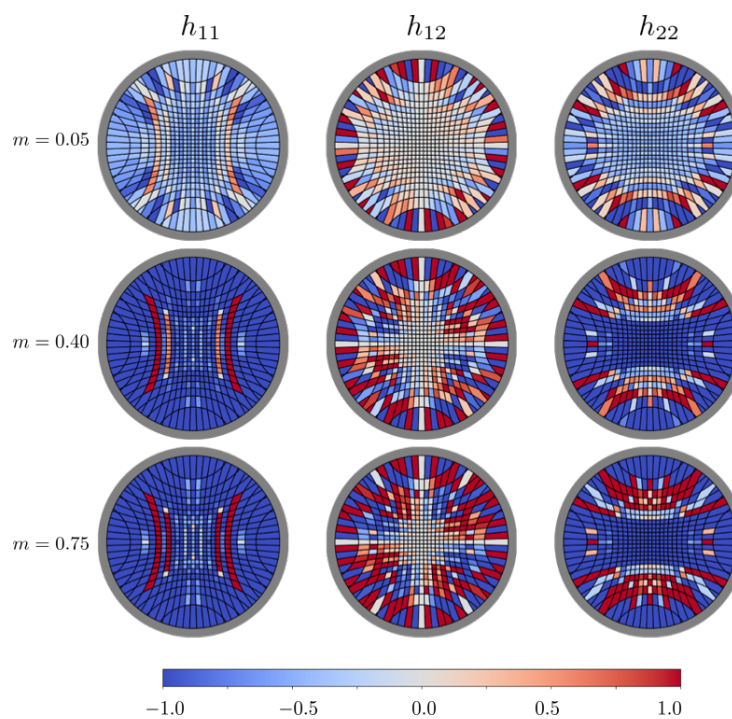

Figure 4.9: Components of the reconstructed metric tensor perturbation corresponding to the ground state of a globally mass deformed Hamiltonian (with m values as shown), with deformations located at every other site. The relative errors are approximately $\mathcal{E}_{\text{bdy}} \approx 0.25$ for this set of plots (see Figure 4.13b).

Looking at the corresponding quench dynamics (see Figure 4.10 for the unsmoothed reconstruction and Figure 4.11 for the smoothed version), we see that the large-scale geometry involves a spread of entanglement that is qualitatively similar to the configuration obtained in the thermalization scenario considered in section 4.4.1. However, in this case the bulk "matter" is non-interacting. The integrability of free

fermion Hamiltonian ensures that the falling shockwave returns to the boundary after some finite time instead of collapsing into a steady configuration.
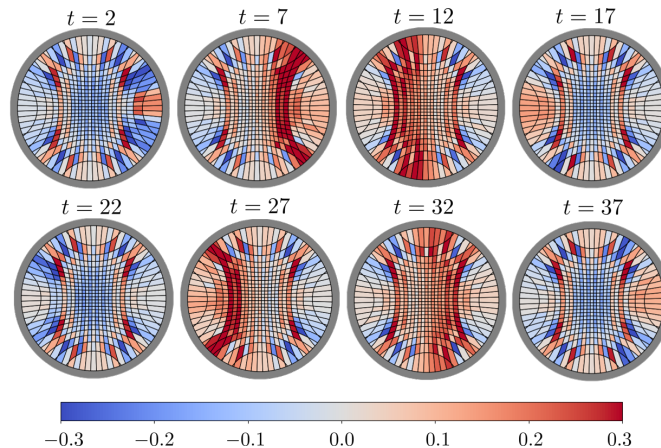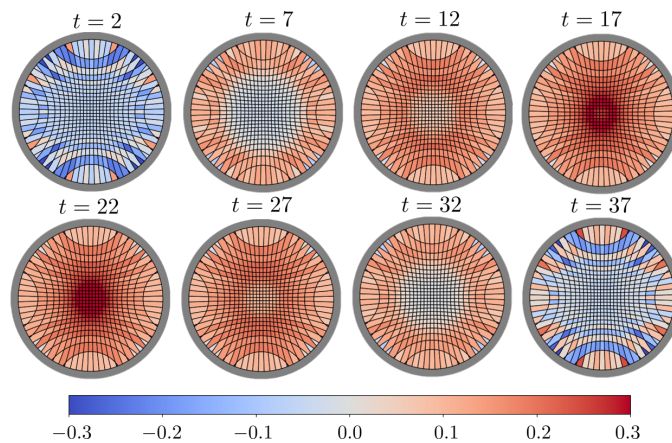

Figure 4.10: Quenched time dynamics of the reconstructed geometry corresponding to a global mass deformation (see Figure 4.9 for the $t = 0$ geometries). The component h_{22} is shown here. The boundary relative error is maximal $\mathcal{E}_{bdy} \approx 0.22$ near $t = 0$ and is minimal $\mathcal{E}_{bdy} \approx 0.006$ around $t = 20$.
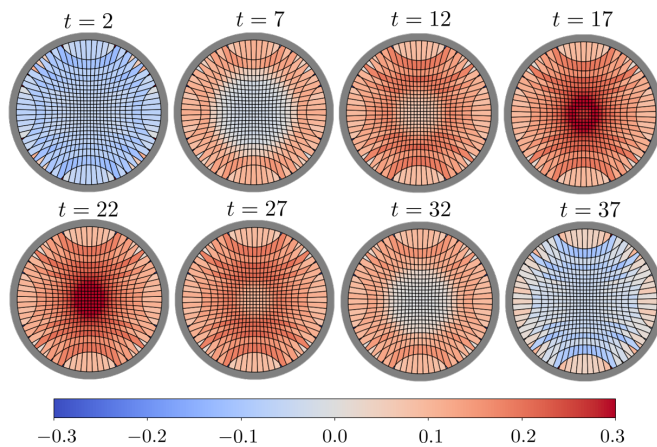
Figure 4.11: Quenched time dynamics of the reconstructed geometry corresponding to a global mass deformation. The boundary data is pre-processed by smoothing to remove small-scale details. The dominant component h_{22} is shown here. The boundary relative error is largely similar to the unsmoothed version except at the initial/final times, where it is significantly reduced by the smoothing.

4.4.3 Random Disorder

Finally, we can take a look at the geometries arising from a random perturbation of the free fermion system. Disorder is introduced by adding a random external field

at each site of the spin chain

$$\hat{H}_{\text{disorder}} = \hat{H}_0 + \sum_i w_i \hat{a}_i^\dagger \hat{a}_i,$$

where each parameter w_i is a random parameter chosen i.i.d. from a uniform random distribution over the interval $[-0.1, 0.1]$.

In Figure 4.12, we show a generic sampling of ground states obtained from such random disorder. As would be naively expected, the resulting ground states have generically large relative errors, with no discernible large scale features (quench dynamics also reveal no discernible large-scale patterns) or symmetries.
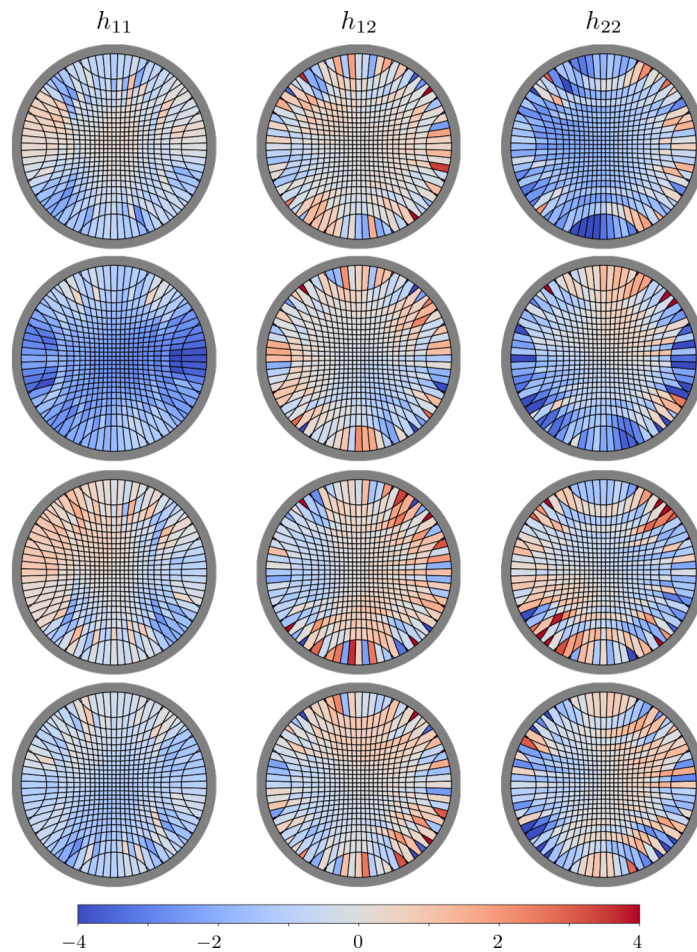
Figure 4.12: A bulk best-fit reconstruction of a state generated with random disorder. Each row is a reconstruction of a particular instance with random disorder. The boundary relative errors vary across a wide range of values, but typically $0.05 \lesssim \mathcal{E}_{\text{bdy}} \lesssim 1$ (see Figure 4.13a).

4.5 Geometry Detection

In this section, we summarize the results of the previous reconstructions and comment on their similarities and differences. We also emphasize that a small relative error of reconstruction is indicative of a classical bulk geometry in which boundary entropies are computed via the RT formula, whereas a large error may indicate some combination of (1) no classical geometry, (2) no classical geometry near the AdS background, or (3) a classical geometry, but where entropies have non-trivial contributions from sources other than the area of RT surfaces, e.g. higher order corrections. For the purpose of this discussion, we refer to all three of these negative cases as non-geometric, but it would clearly be desirable to distinguish them further in future work.

From the results of the mass deformation, we confirm that generic low-energy states of a free fermion system do not appear to have a good geometric reconstruction. Comparing the ground states of the mass deformed 1d free fermion with that of the thermal AdS state, we see that the relative errors for the 1d free fermion are significantly larger than those of the holographic data (see Figures 4.13a, 4.13b, 4.14). Smoothing of the boundary data reduces the relative error of reconstruction by a significant amount (as would be expected due to the reduction of small-scale artifacts). However, even with smoothing, the level of error is clearly distinguishable from the reconstructions of known geometric states such as thermal AdS.

In particular, ground states of the locally deformed 1d free fermion Hamiltonian have a relative error that is of order ~ 1. Smoothing of the boundary data brings this down to $\sim 10^{-1}$, which is still an order of magnitude larger in comparison to the thermal AdS reconstructions, which have a relative error on the order of $\sim 10^{-2}$. The unsmoothed relative error for the global deformation of the 1d free fermion hovers around $\sim 10^{-1}$. Smoothing brings this down significantly to $\sim 10^{-2}$. The most likely explanation for the pronounced effect of smoothing here is due to the global symmetry present in the globally deformed state. We also see that the relative errors for reconstructions corresponding to random disorder in the 1d free fermion tends to interpolate between the results for local and global deformations across different random instances, with the best case relative errors being $\lesssim 10^{-1}$ and the worst case being ~ 1. A summary of these results is plotted in Figure 4.13a, which shows a histogram of the number of instances for each type of reconstruction, as we vary some of the relevant parameters (i.e., temperature, mass, etc). In Figure 4.13b, we also plot the relative error of reconstruction as a function of the mass deformation

parameter m. As would be expected, we see that a larger mass deformation, and hence a larger deviation from criticality, contributes positively to the relative error of reconstruction regardless of smoothing.
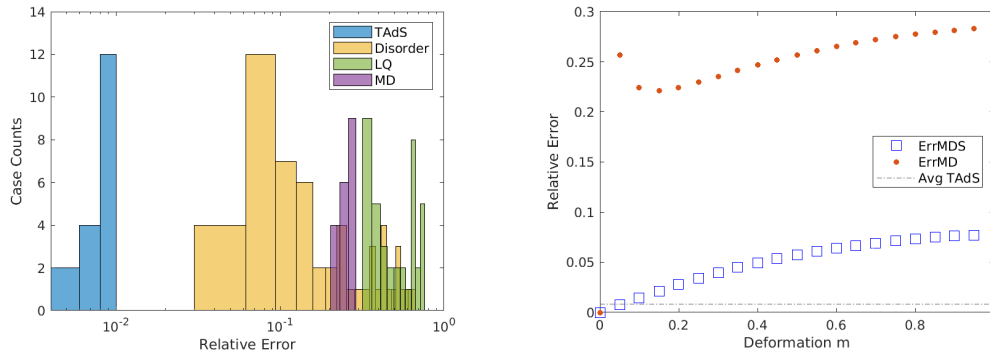
(a) A histogram of relative errors from different boundary data: Thermal AdS (TAdS), global mass deformation (MD), local quench without smoothing (LQ), and system with random disorder (Disorder). Boundary relative error $\mathcal{E}_{\text{bdy}}$ plotted on the x-axis.

(b) Global mass deformation errors, as a function of the deformation parameter m. Red dots denote the boundary relative error of the original fermion data. Blue squares are relative errors after smoothing.

Figure 4.13: Relative errors across various regimes, and global mass deformation errors.

A summary of the time dynamics of the relative errors is shown in Figure 4.14. For the dynamical reconstructions, we see that the relative error generally varies with time. This is most noticeable with the global quench dynamics, where the relative error tends to decrease as entanglement spreads. Although one starts with a mass deformed state with large relative error at $t = 0$, subsequent evolution can effectively wash out non-geometric features the system begins to thermalize. The state at $t \approx 20$ captures basic entanglement properties of a thermal state, thus the reconstruction is thermal AdS-like with small relative errors. Nevertheless, the state cannot actually thermalize due to integrability; the system recurs at later times and the relative error rebounds.

At this point, it is not completely clear why the global quench states have such small relative error at intermediate times. A likely guess that the enhanced symmetries for the globally deformed states contribute to a more geometrical reconstruction, with the effect being especially pronounced at intermediate times when the system is maximally thermal, before it is kicked back by integrability. Such behaviour is

not seen in the local quench disorder, where the magnitude of the relative error is larger, and remains stable across the time evolution. However, symmetry is likely not the full answer either, since otherwise we would expect the disordered models to generically have larger relative error compared to the global mass deformed state, whereas we observer the disordered model to be peaked at smaller errors. While the full characterization for which states appear more geometric than others is currently unknown, it seems likely that there are many contributing factors to a successful reconstruction, among them properties like symmetry and thermality.
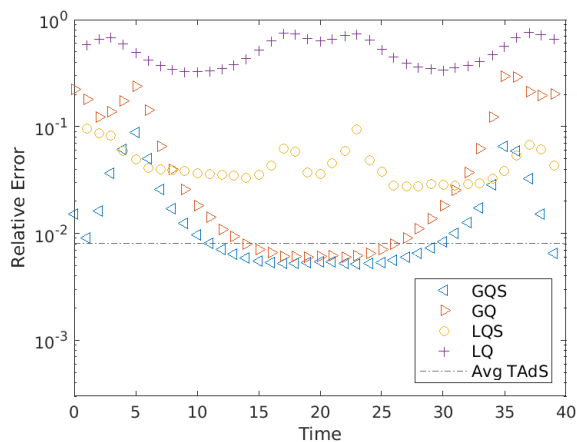
Figure 4.14: Relative boundary error as a function of time for a free fermion system after a quench. GQS, GQ, LQS, LQ denote global quench with smoothing, global quench without smoothing, local quench with smoothing, and local quench without smoothing, respectively. Dashed line marks the average relative error for the thermal AdS entropy data.

Finally, we show the relative errors corresponding to the superposition of thermal states in Figure 4.15. We find two regimes for the geometricality of the superposition, depending on the magnitude of the mixing term $G_N H(p)$. We find the superposition to be non-geometric when the entropy of mixing provides a significant correction to the geodesic lengths. This implies a generally non-geometrical construction when $G_N H(p) \approx 1$, as would be the case when the gravitational coupling G_N is strong, or when the entropy of mixing is made large by superposing a large number of distinct geometries.

Conversely, in the weak coupling limit with few distinct superpositions, the contribution of the mixing term $G_N H(p)$ to the geodesic lengths becomes negligible, causing the entropy operator to be approximately linear in this regime. This leads to a smooth, geometric interpolation between the two distinct geometries as we tune

the probability of mixing. The clear separation between the two cases is clearly illustrated in Figure 4.15, where we plot the relative errors for a state constructed as a mixture of two distinct thermal states.
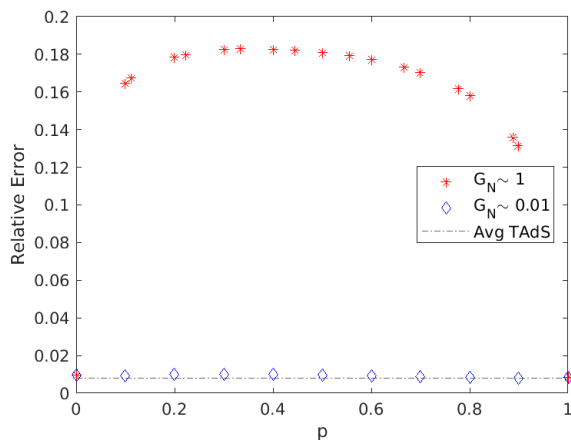
Figure 4.15: Boundary relative errors for superpositions of thermal AdS geometries.

In summary, we find that the boundary relative error provides a useful measure for the extent to which a state is geometric or non-geometric. Through the discretized Radon transform, we confirm some existing expectations for the geometric nature of certain holographic states, as well as states arising from many-body systems that are generally believed to be non-geometric. We find that generic low-lying energy states of a free fermion system are indeed dominated by non-geometric contributions as indicated by our algorithm. However, there also exists states in these systems that have small relative error, such as the configurations following a global quench. It is suspected that the enhanced symmetries for the globally deformed states contributes to a more geometrical reconstruction. Moreover, quench dynamics reveal large-scale patterns of geometric evolution, despite the states themselves being non-geometric as characterized by the relative error. Further advances in the tensor Radon transform and its range characterization may help us understand what kind of properties in the boundary data lead to good reconstructions, and the nature of the large-scale behavior revealed by the numerical transform.

4.6 Discussion

In this work, we took some first steps in addressing the explicit bulk metric reconstruction problem in holography, as well as the question of whether non-geometrical bulk states can be detected from boundary data alone. Motivated by the tensor

Radon transform, we provided – and explicitly implemented – an algorithm that reconstructs the bulk metric tensor without any *a priori* assumptions on the symmetries or form of the metric tensors, other than the fact that they are perturbatively close to AdS. We applied this reconstruction to entanglement data from holographic systems with large N, as well as to that of a $1d$ free fermion system. The reconstruction was also applied, time-slice by time-slice, to time-evolving states. We confirmed that, assuming the linearized Einstein's equations, thermalization following a global quench in holographic systems is consistent with a shell of in-falling matter in the bulk that finally settles into a state that appears to be gravitationally bound deep inside the bulk. This kind of behaviour is absent in free fermion systems where the corresponding shockwave of "in-falling matter" repeatedly oscillates between the boundary and the bulk due to the integrability of the system.

We also provided a partial answer to whether a given state in the conformal field theory has a well-defined geometric dual on the gravity side. We find that boundary reconstruction errors provide a quantitative measure that distinguished "geometrical" states, such as the ones we find in large N theories with semi-classical duals, from "non-geometrical" states like the generic excited states of a free fermion spin chain. This is a precise, albeit coarse, way of understanding what portion of the boundary data lies outside the range of the tensor Radon transform, and therefore cannot be interpreted as a tensor field on a hyperbolic background. In the instances we have examined, this measure was an effective indicator of non-geometricality.

Finally, our initial attempts in developing this algorithm indicate that efficient numerical reconstructions of the bulk metric tensor from entanglement entropy data, at least at the linearized level, are achievable. This is partly because the optimization problem we consider is linear in nature and can be solved in polynomial time with low computational power. As such, it provides an efficiently computable tomographic procedure that translates boundary entropy data, where the underlying spacetime and gravitational dynamics are hidden, to a setting where it is manifest. Although a wealth of previous analytic results are available, the flexibility of numerical studies is also invaluable, as we have found in various areas of physics from quantum many-body systems to numerical general relativity. As we start to move away from the more tractable dynamics of quantum systems that carry simplifying assumptions such as symmetries, the full force of numerical methods will prove to be extremely helpful. Here we provide one such preliminary construction which can hopefully provide a stepping stone for future advances.

Further work is clearly required to improve upon our first efforts. Here we list but a few major directions where progress can be made.

On the front of new results in mathematics, especially related to tensor Radon transform:

1. To put our geometry detector on a firmer mathematical footing, one needs a rigorous characterization of the range of tensor Radon transforms on curved backgrounds, in particular, the hyperbolic background. In the language of holographers, we are in need of a set of necessary[38], [39] and sufficient conditions that characterizes what type of quantum states have semi-classical dual geometries. A complete characterization of the range of Radon transform on hyperbolic space, for example, precisely provides such conditions that checks whether a set of entanglement data is dual to a semi-classical geometry close to AdS.

2. It is also crucial to obtain an explicit reconstruction formula in closed form. Such formulae are known for the Euclidean background, and for the case of scalar or vector Radon transform on curved backgrounds. Development of these results will further enable calculations in the continuum limit, which is relevant for AdS/CFT.

On the front of new results in physics:

1. We hope to extend this method to higher dimensions, which so far has faced the most obstacles in bulk metric reconstruction. Because area variation can already be cast as a tensor Radon transform in arbitrary dimensions [47], the idea of discretization followed by gauge fixing and linear optimization may simply be extended to minimal surfaces. Similarly, X-ray transforms in higher dimensions that makes use of correlation data can also be viable [23]. This may provide a numerically computable alternative to other methods [38], [62], [63].

2. Although we examine dynamical cases, we only reconstruct the spatial metrics for each individual time slice. Reconstruction of the full linearized space-time metric perturbation against the AdS background should also be possible by gluing together these spatial slices. However, one has to be careful about gauge fixing and the role played by the time coordinate. More work is needed to clarify these constructions.

3. It is also worthwhile to go beyond linearized level. A known approach is to apply the inverse Radon transform iteratively, such that the background geometry is updated using the reconstructed metric perturbation h_{ij}. Such method is used in the geophysics community [53], [64]. However, we have to be careful about the change in extremal surface beyond the linear level, especially in the dynamical cases. This is worthwhile though, since working beyond the linearized level should allow a better understanding of non-trivial topologies, for example. As we have seen in this work, they are invisible at the linearized level despite our expectation that collapse of bulk matter should lead to such changes in geometry.

4. Reconstruction of entanglement data from other quantum systems. Thus far, we have only seen limited application of this reconstruction algorithm, largely limited by the availability of entanglement data. Nevertheless, it may be possible to compute these entanglement approximately for smaller but more interesting quantum many-body systems using tensor networks or other classical techniques. It may also be worthwhile to obtain data for geodesic lengths via other more accessible data, such as correlation functions or mutual information.

5. In light of the difficulty in computing von Neumann entropies for quantum systems, we can explore the possibility of using quantum Renyi entropies, which also have a geometric interpretation in holography [65]. A prominent advantage is its computability using numerics and measurability in actual quantum systems. Reconstructions using such data can enable us to directly acquire entropy data from quantum simulations and experimental setups.

6. Similar techniques using Radon transforms to recover the metric tensor of geometry emerged from entanglement are also applicable to near-flat manifolds outside the context of AdS/CFT. In fact, the Radon transform on flat-space is much better understood. Similar reconstructions should be possible for constructions like [18] where, for instance, the relevant quantum states can be low energy states of a gapped local Hamiltonian.

Finally, there is also much progress to be made in our reconstruction algorithm and related numerics.

1. One area of improvement is a data-driven modelling. For our current work, we fixed a particular tiling that is easy to implement. However, it has been shown that a tiling depending on the boundary data can lead to improvement for bulk reconstruction [66]. These include Voronoi partitions or improved modelling using Bayesian inference [67].

2. Our discretization, constraint, and interpolation methods are all extremely simplistic, as appropriate for a proof of concept implementation. Numerous improvements can be made to improve the accuracy of the reconstruction algorithm, for example: proper triangular meshes, finite-element methods, better regularization techniques, etc. We expect that the reconstruction procedure here can be drastically improved in fidelity by adapting proper numerical techniques.

Acknowledgements

4.A   The Tensor Radon Transform

Here, let us formally define the geodesic tensor Radon transform. We begin with an introduction to the tensor Radon transform in general, but will quickly specialize to the special case of the 2-tensor Radon transform on the Poincare disk (with finite cutoff).

In short, the m-tensor geodesic Radon transform R_m is a map which takes a symmetric m-tensor field defined on a sufficiently well-behaved Riemannian manifold M (with boundary) to the space of geodesics on that manifold.

4.A.1 General Definitions

Let (M, g) be an n-dimensional Riemmanian manifold with boundary ∂M and metric g. We say that (M, g) is a *simple* manifold if ∂M is strictly convex[9] and any two points in M are connected by a unique geodesic segment which depends smoothly on the endpoints [68]. Alternatively, a simple manifold is one in which the boundary is strictly convex, and where the exponential map $\exp_p : \exp_p^{-1}(M) \to M$ is a diffeomorphism for every $p \in M$. Fixing some $p \in M$, we may identify a simple manifold with a strictly convex domain Ω of $\mathbb{R}^n$.

The simplicity of a manifold is a sufficient condition for the geodesic Radon transform to be well-defined [69]. There are more general conditions available, but simplicity will generally be sufficient for our purposes. From now on, unless otherwise stated, all of our manifolds will be assumed to be simple.

Let SM denote the unit circle bundle of M. The bundle SM is the collection of all pairs (p, v), where $p \in M$ and where $v \in T_p M$ is a unit tangent vector at p. The boundary of the unit circle bundle, denoted ∂SM, consists of all such pairs where $p \in \partial M$. The boundary of the unit circle bundle naturally splits into two components, $\partial_+ SM$ consisting of all the inward pointing vectors, and $\partial_- SM$ consisting of all the outward pointing vectors. We will define both components to be closed, i.e., vectors tangent to ∂M will be in both $\partial_+ SM$ and $\partial_- SM$.

Given $(p, v) \in \partial_+ SM$, let $\gamma_{p,v} : [0, \tau(p, v)] \to M$ denote the unique unit speed geodesic through (p, v), i.e., the unique geodesic such that

$$\gamma_{p,v}(0) = p, \qquad \text{and} \qquad \dot{\gamma}_{p,v}(0) = v.$$

The parameter $\tau(p, v)$ denotes the *exit time*[10] of $\gamma_{p,v}$, i.e., the first non-zero time such that $\gamma_{p,v}(\tau) \in \partial M$. Note that the exit time is well-defined under the assumption that the underlying manifold is simple.

Now, let $f_{i_1 \cdots i_m}$ be a smooth, symmetric (covariant) m-tensor field on M. Then the Radon transform of f is defined by

$$R_m[f](p, v) = \int_0^{\tau(p,v)} f_{i_1 \cdots i_m}(\gamma_{p,v}(s)) \, \dot{\gamma}_{p,v}^{i_1}(s) \cdots \dot{\gamma}_{p,v}^{i_m}(s) \, ds.$$

Thus, the Radon transform is a map $R_m : S_m(M) \to C^\infty(\partial_+ SM)$ which takes the space $S_m(M)$ of smooth, symmetric (covariant) m-tensors on M to the space

---

[9]Namely, given any two points $p, q \in \partial M$, there exists a geodesic segment connecting p and q that meets ∂M only at p and q.

[10]Or since $\gamma_{p,v}$ is unit speed parameterized, the total arclength of $\gamma_{p,v}$.

$C^\infty(\partial_+ SM)$ of smooth functions on the inward pointing boundary unit circle bundle component $\partial_+ SM$. Since we can uniquely identify each $(p, v) \in \partial_+ SM$ with a corresponding unit speed geodesic $\gamma_{p,v}$, it will be often convenient to consider the Radon transform as a map on the space of boundary anchored geodesics.

4.A.2 s-Injectivity

There are a few natural questions we may ask for the Radon transform, the foremost being the surjectivity and the injectivity of the transform. We will not comment much on the range of the tensor Radon transform, except to note that the tensor Radon transform is generally not surjective. This, of course, corresponds to the well known fact that not all boundary states have a well-defined bulk dual. A useful analytic characterization of the range remains an open problem for the tensor Radon transform on generic manifolds.

Let us now consider the problem of injectivity. The tensor Radon transform has a natural kernel. Let (M, g) be an n-dimensional simple Riemannian manifold. We will let dV denote the canonical volume form, which is locally given by

$$dV = \sqrt{|g|}\, dx^1 \wedge \cdots \wedge dx^n,$$

where $(x^1, \cdots, x^n)$ is some oriented chart, and where $|g|$ is the determinant of the metric g_{ij} in that chart. We will let ∇ denote the Levi-Civita connection on M. Now, let $S_m M$ denote the space of (covariant) symmetric m-tensors on M. We will define the *inner derivative*[11] $\mathbf{d} : S_m M \to S_{m+1} M$ by

$$\mathbf{d}f = \sigma \nabla f,$$

where σ denotes complete symmetrization. In local coordinates, we simply have

$$(\mathbf{d}f)_{i_1 \cdots i_{m+1}} = f_{(i_1 \cdots i_m; i_{m+1})},$$

where parentheses indicate the complete symmetrization of the contained indices as usual. We will likewise define the *divergence*[12] $\delta : S_m M \to S_{m-1} M$ to be

$$\delta f = \text{Tr}_{m,m+1}(\nabla f),$$

---

[11]Despite the confusingly similar notation, $\mathbf{d}$ is *not* the exterior derivative d, which acts on forms and not symmetric tensors. Unfortunately, the notation is somewhat well-established in the integral geometry community, so we will stick to it. We will never need to use the exterior derivative in this paper, but nevertheless we will denote the inner derivative with boldface font $\mathbf{d}$ as a reminder that it is not the exterior derivative.

[12]Again, δ is not the co-exterior derivative. We will use the same boldface convention.

where $\text{Tr}_{m,m+1}$ denotes (Riemannian) contraction between the mth and $(m + 1)$th arguments. In local coordinates, we have

$$(\delta f)_{i_1,\cdots,i_{m-1}} = f_{i_1,\cdots,i_{m-1},j;k}g^{jk}.$$

The operators $\mathbf{d}$ and $-\delta$ are adjoint for compactly supported symmetric tensor fields on M. More generally, for any compact region $D \subseteq M$, we have

$$\int_D \langle \mathbf{d}u, v \rangle + \langle u, \delta v \rangle \, dV = \int_{\partial D} \langle i_\nu u, v \rangle \, dS,$$

where u and v are (sufficiently smooth) symmetric tensors of the appropriate orders, i_ν denotes interior multiplication with respect to an outward pointing normal ν, and where dS is the induced volume form on ∂D. The inner product $\langle \cdot, \cdot \rangle$ on $S_m M$ is given by complete contraction, i.e.,

$$\langle u, v \rangle = g^{i_1 j_1} \cdots g^{i_m j_m} u_{i_1,\cdots,i_m} v_{j_1,\cdots,j_m}.$$

The significance of the operators δ and $\mathbf{d}$ are as follows. Let $H^k(S_m M)$ denote the Sobolev space of m-symmetric tensors, i.e., the space of all sections which are k-times (weakly) differentiable, and such that all derivatives are locally square integrable. Each $H^k(S_m M)$ can be given the structure of a Hilbert space when M is compact. Then we have the generalized Helmholtz decomposition as follows:

Theorem 4.A.1 (Generalized Helmholtz Decomposition [41]). *Let (M, g) be a compact Riemannian manifold with boundary. Let $k \geq 1$ and $m \geq 0$ be integers. Given any section $f \in H^k(S_m M)$, there exists uniquely determined $f^s \in H^k(S_m M)$ and $v \in H^{k+1}(S_{m-1} M)$ such that*

$$f = f^s + \mathbf{d}v, \qquad \text{where} \quad \delta f^s = 0, \quad \text{and} \quad v|_{\partial M} = 0.$$

The fields f^s and $\mathbf{d}v$ are called the solenoidal *and* potential *parts of f.*

Note that in the case $m = 1$, Theorem 4.A.1 is just the usual Helmholtz decomposition

$$\mathbf{F} = -\nabla\varphi + \nabla \times \mathbf{A}$$

after identifying vectors and covectors using the metric.

The decomposition of a symmetric tensor field into solenoidal and potential parts gives us a natural identification of the kernel for the Radon transform. Indeed, we have the following result:

Theorem 4.A.2. *Let (M, g) be a simple manifold and let R_m be the m-tensor Radon transform on M. Let $v \in H^{k+1}(S_{m-1}M)$ be a vector field such that $v|_{\partial M} = 0$. Then*

$$R_m[dv] = 0.$$

Therefore, Theorem 4.A.2 identifies a natural kernel for the Radon transform, namely the space of all potential tensor fields. Since every sufficiently smooth tensor field can be decomposed uniquely into a potential and a solenoidal part, a natural question is whether the solenoidal part is uniquely recoverable from the Radon transform, i.e., whether the space of potential tensor fields exhausts the kernel of the Radon transform. If this is indeed the case, i.e., if $R_m[f] = 0$ implies $f^s = 0$, then we say that the Radon transform is *s-injective*.

The question of the *s*-inectivity of the Radon transform is a fundamental problem in integral geometry. The general case remains open, although the case for 2-dimensional simple manifolds was settled in the affirmative [70]:

Theorem 4.A.3 (*s*-injectivity [70]). *Let (M, g) be a simple 2-dimensional Riemannian manifold. Then the tensor Radon transform R_m on M is s-injective for all $m \geq 0$.*

Note that in the case of the scalar transform for $m = 0$, all scalar functions are automatically solenoidal, so the scalar Radon transform R_0 is injective in the usual sense.

Given the *s*-injectivity of the Radon transform, we can recover the bulk tensor field by imposing the *solenoidal gauge condition*

$$f = f^s.$$

The solenoidal gauge is the most commonly employed gauge condition for the tensor Radon transform, but it comes with some inconvenient features. In particular, it does not respect the decomposition of a tensor into its trace and traceless parts. The 2-tensor Radon transform on a purely trace bulk tensor field is identical to the scalar Radon transform on the corresponding trace function, but the inversion of the 2-tensor Radon transform under the solenoidal gauge introduces extraneous gauge degrees of freedom which causes the recovered field to disagree with the original. This is rather undesirable, since the corresponding scalar transform is purely injective and admits a unique recovery. We will instead employ an alternative gauge condition, which we introduce in Section 4.C, that makes the tensor Radon transform consistent with the scalar Radon transform for pure trace fields.

4.B   The Radon Transform on the Poincare disk

Let us now focus our attention to the case of a single time-slice of AdS$_3$. Such a time-slice is isomorphic to the hyperbolic plane, which we will consider in the Poincare disk model. Concretely, the Poincare disk is the Riemannian manifold of constant scalar curvature $R = -1$, defined within the open unit disk

$$\mathbb{D} = \{z \in \mathbb{C} \mid |z| < 1\}.$$

We will cover the Poincare disk using its natural Cartesian coordinates (x, y), or equivalently, using complex coordinates $z = x + iy$. The metric is then given by

$$g(z) = \frac{4(dx^2 + dy^2)}{(1 - x^2 - y^2)^2} = 4(1 - |z|^2)^{-2} |dz|^2,$$

where $|dz|^2 = dz \cdot d\bar{z} = dx^2 + dy^2$. We will denote the Poincare disk by $\mathbb{H} = (\mathbb{D}, g)$. Geometrically, the geodesics of the Poincare disk are circular segments which are orthogonal to the boundary circle S^1.

Both for physical reasons, and to properly define the Radon transform, we cannot work with the Poincare disk in its entirety.[13] Rather, we must impose a cutoff. We will let $\kappa \in (0, 1)$ be a cutoff radius, and consider the cutoff disk

$$\mathbb{D}_\kappa = \{z \in \mathbb{C} \mid |z| < \kappa\}.$$

Then we consider the cutoff Poincare disk to be the manifold defined by $\mathbb{H}_\kappa = (\mathbb{D}_\kappa, g)$.

We will be mainly interested in the Radon transform for symmetric 2-tensors. The Poincare disk (with cutoff) is a simple manifold. As such, the Radon transform is always well-defined on the Poincare disk. To that end, let f_{ij} be a symmetric 2-tensor field on $\mathbb{H}_\kappa$. The Radon transform is then given by

$$R_2[h](p, v) = \int_0^{\tau(p,v)} f_{ij}(\gamma(s))\dot{\gamma}^i_{p,v}(s)\dot{\gamma}^j_{p,v}(s)\, ds.$$

Note that since the boundary of the (cutoff) Poincare disk is a circle, we may uniquely identify each geodesic $\gamma_{p,v}$ with a connected subregion A (i.e., a circular arc) of the boundary. The corresponding geodesic is then the minimal surface with respect to that boundary subregion. In this way, we may regard the image of the Radon transform as a function on the space of boundary subregions.

---

[13]Technically, the Radon transform does not really care about the fact that the Poincare disk is non-compact, especially since there is still a well-defined notion of (asymptotic) boundary. However, it is much more convenient to work with a compact manifold with an actual boundary.

The metric on the Poincare disk is an isotropic metric. Let us write

$$g(z) = 4(1 - |z|^2)^{-2}|dz|^2 = e^{2\lambda(z)}|dz|^2.$$

We may then write a unit tangent vector as $v = e^{-\lambda}(\cos\theta, \sin\theta)$, where θ is the natural angular coordinate on the unit disk. Then we can decompose the Radon transform as

$$f_{ij}\dot{\gamma}^i\dot{\gamma}^j = e^{-2\lambda}\left(f_{11}\cos^2\theta + 2f_{12}\cos\theta\sin\theta + f_{22}\sin^2\theta\right) \quad (4.8)$$

$$= e^{-2\lambda}\left(\frac{f_{11} + f_{22}}{2} + f_{12}\sin 2\theta + \frac{f_{11} - f_{22}}{2}\cos 2\theta\right)$$

$$= e^{-2\lambda}\left(h_0 + he^{2i\theta} + \overline{h}e^{-2i\theta}\right),$$

where in the last line we have defined the components

$$h_0 = \frac{1}{2}\operatorname{tr} h,$$

$$h = \frac{1}{4}(f_{11} - 2if_{12} - f_{22}),$$

$$\overline{h} = \frac{1}{4}(f_{11} + 2if_{12} - f_{22}).$$

The components h_0, h, and $\overline{h}$ will be called the *trace*, *holomorphic*, and *anti-holomorphic* parts of f, respectively. Note that in this way, we can always identify any symmetric 2-tensor f_{ij} with a function $\tilde{f} \equiv f_{ij}\dot{\gamma}^i\dot{\gamma}^j$ of harmonic content $(-2, 0, 2)$.

We can then equivalently write the tensor Radon transform in the h-components as

$$R_2[g](p, v) = \int_0^{\tau(p,v)} e^{-2\lambda(\gamma(s))}\left(h_0(\gamma(s)) + h(\gamma(s))e^{2i\theta(s)} + \overline{h}(\gamma(s))e^{-2i\theta(s)}\right)\,ds.$$

$$(4.9)$$

Note that the trace component of the transform simply gives the scalar transform (after absorbing the metric into the definition of h_0). Therefore the tensor Radon transform R_2 is essentially equivalent to the scalar Radon transform for tensors which are purely trace.

As previously mentioned, the solenoidal gauge does not respect the decomposition of the tensor into trace and traceless parts. It will therefore be convenient to find a gauge which preserves this decomposition. We do so in Appendix 4.C.

4.C The Holomorphic Gauge

In this section, we present an alternative to the solenoidal gauge, which we call the *holomorphic gauge*, which both uniquely fixes a solution space to the inverse Radon transform and preserves the scalar part of the transform.

To define the holomorphic gauge, we first introduce some background. We employ global isothermal coordinates (x, y, θ) on the unit circle bundle $S\mathbb{H}$ of the Poincare disk, where (x, y) are the usual Poincare coordinates on the base manifold, and where θ is a fiber coordinate indicating the angular direction of a tangent vector.

We can define the geodesic flow X, given in global isothermal coordinates on the unit circle bundle by

$$X = e^{-\lambda} \left(\cos \theta \, \partial_x + \sin \theta \, \partial_y + \left(- \sin \theta \, \partial_x \lambda + \cos \theta \, \partial_y \lambda \right) \partial_\theta \right),$$

where we write the metric as $g = e^{2\lambda}(dx^2 + dy^2)$. Writing the vector field in terms of complex coordinates, we can decompose the geodesic flow as $X = \eta_+ + \eta_-$, where

$$\eta_+ = e^{-\lambda} e^{i\theta} \left(\partial + i\partial\lambda \, \partial_\theta \right),$$
$$\eta_- = e^{-\lambda} e^{i\theta} \left(\overline{\partial} + i\overline{\partial}\lambda \, \partial_\theta \right) = \overline{\eta}_+,$$

where $\partial = \frac{1}{2}(\partial_x - i\partial_y)$ and $\overline{\partial} = \frac{1}{2}(\partial_x + i\partial_y)$ are the Wirtinger derivatives.

Let $\Omega_k = L^2(S\mathbb{H}) \cap \ker(\partial_\theta - ik)$ be the space of square-integrable functions on the unit circle bundle with fixed harmonic content k. Then it can be shown [71] that $\eta_\pm$ are smooth elliptic differential operators such that $\eta_\pm : \Omega_k \to \Omega_{k\pm 1}$ for any $k \in \mathbb{Z}$. In particular, note that $\Delta \equiv \eta_+\eta_- = \eta_-\eta_+$ is also a smooth elliptic partial differential operator.

Note that the geodesic flow X is naturally related to the Radon transform as follows [41]: let f_{ij} be a symmetric 2-tensor field, and let us define the function

$$u(x, y, \theta) = \int_0^{\tau_+(\gamma)} f_{ij} \dot{\gamma}^i \dot{\gamma}^j \, ds,$$

where γ denotes the unique unit speed geodesic through (x, y), with initial angle θ, and where $\tau_+(\gamma)$ denotes the exit time of the geodesic. Note that we have $u|_{\partial_- SM} = 0$ and

$$u|_{\partial_+ SM} = R_2[f],$$

by construction. If $\gamma_{(x_0,y_0,\theta_0)}$ is a geodesic through (x_0, y_0, θ_0), then

$$u\left(\gamma_{(x_0,y_0,\theta_0)}(t), \theta_{(x_0,y_0,\theta_0)}(t)\right) = \int_t^{\tau_+(\gamma)} f_{ij}\dot\gamma^i\dot\gamma^j \, ds,$$

where $\theta_{(x_0,y_0,\theta_0)}(t) \equiv \arg\left[\dot\gamma_{(x_0,y_0,\theta_0)}(t)\right]$ denotes the angle of the tangent vector to $\gamma_{(x_0,y_0,\theta_0)}$ at time t. Differentiating with respect to t, we get

$$\dot\gamma^i \frac{\partial u}{\partial x^i} + \dot\theta \frac{\partial u}{\partial \theta} = -f_{ij}\dot\gamma^i\dot\gamma^j. \tag{4.10}$$

Note that the left-hand side is precisely the expression Xu.[14] Denoting by u^f the unique solution to the *transport equation* (4.C) with boundary condition $u^f|_{\partial_- SM} = 0$, it follows that the Radon transform $R_2[f]$ is given by

$$R_2[f] = u^f|_{\partial_+ SM}.$$

The transport equation can therefore be seen as the differential form of the Radon transform.

The key result leading to the holomorphic gauge is then the following theorem.

Theorem 4.C.1. *For any symmetric 2-tensor $f \in L^2(SM)$, there exists a unique 2-tensor $h \in L^2(SM)$ such that $R_2 f = R_2 h$, and such that h is of the form*

$$h = h_0 + h_2 + h_{-2},$$

where $h_0 \in L^2(M) \cap \Omega_0$, and where $h_{\pm 2} \in \ker \eta_\mp \cap \Omega_{\pm 2}$.

---

[14]Let us explicitly calculate $\dot\theta$ here. We start with the geodesic equation

$$\ddot\gamma^i + \Gamma^i{}_{jk}\dot\gamma^j\dot\gamma^k = 0.$$

Evaluating the x component, for example, we get

$$\frac{d}{ds}(e^{-\lambda}\cos\theta) + e^{-2\lambda}\partial_1\lambda(\cos^2\theta - \sin^2\theta) + 2e^{-2\lambda}\partial_2\lambda\sin\theta\cos\theta = 0.$$

Taking the derivative and expanding, we then have

$$-e^{-2\lambda}(\partial_1\lambda\cos^2\theta + \partial_2\lambda\sin\theta\cos\theta) - e^{-\lambda}\sin\theta \cdot \dot\theta + e^{-2\lambda}\partial_1\lambda(\cos^2\theta - \sin^2\theta) + 2e^{-2\lambda}\partial_2\lambda\sin\theta\cos\theta = 0,$$

which simplifies as

$$-e^{-\lambda}\partial_1\lambda\sin\theta + e^{-\lambda}\partial_2\lambda\cos\theta = \dot\theta,$$

so that equation (4.C) is given by

$$e^{-\lambda}\left[\cos\theta\frac{\partial}{\partial x} + e^{-\lambda}\sin\theta\frac{\partial}{\partial y} + \left(-\sin\theta\frac{\partial\lambda}{\partial x} + \cos\theta\frac{\partial\lambda}{\partial y}\right)\frac{\partial}{\partial\theta}\right]u(x,y,\theta) = Xu(x,y,\theta) = -\tilde{f}(x,y,\theta),$$

where $\tilde{f} \equiv f_{ij}\dot\gamma^i\dot\gamma^j$, as given in Equation (4.B).

Proof. We adapt the proof from Theorem 1 of [48], which covers the case where the underlying metric is the usual flat Euclidean metric. Let $f \in L^2(SM) \cap \Omega_k$ be given. Then consider the differential equation

$$\eta_- f = \Delta v,$$
$$v|_{\partial SM} = 0,$$

where $v \in H^1(SM) \cap \Omega_{k-1}$. Since Δ is a smooth elliptic operator, it follows from the standard theory of elliptic differential equations that the above system admits a unique (weak) solution v. Given such a solution, let us define $g = f - \eta_+ v$. It follows that

$$\eta_- g = \eta_-(f - \eta_+ v) = 0.$$

This shows that each $f \in L^2(SM) \cap \Omega_k$ can be written uniquely as

$$f = \eta_+ v + g,$$

where $v \in H^1(SM) \cap \Omega_{k-1}$, $v|_{\partial SM} = 0$, and where $g \in L^2(SM)$ such that $\eta_- g = 0$.

Next, let f be a given 2-tensor which we may write as

$$f = f_0 + f_{-2} + f_2,$$

where $f_k \in \Omega_k$. The transport equation gives us

$$Xu = -f,$$
$$u|_{\partial_- SM} = 0,$$
$$u|_{\partial_+ SM} = R_2 f.$$

Let us apply the previously derived decomposition to write $f_2 = \eta_+ v_1 + g_2$, where $v_1|_{\partial SM} = 0$ and where $\eta_- g_2 = 0$. This gives us

$$f_2 = \eta_+ v_1 + g_2$$
$$= Xv_1 - \eta_- v_1 + g_2,$$

which allows us to write the transport equation as

$$X(u + v_1) = -(f_0 - \eta_- v_1 + f_{-2} + g_2),$$

where $\eta_- g_2 = 0$. Since $v_1|_{SM} = 0$, it follows that we have $(u + v_1)|_{\partial_+ SM} = u|_{\partial_+ M} = R_2 f$. If we define the 2-tensor h by

$$h = h_0 + h_{-2} + h_2,$$

where $h_0 = f_0 - \eta_- v_1$ and $h_{-2} = f_{-2}$, then h satisfies $R_2 h = R_2 f$, and is such that $\eta_- h_2 = 0$. We may repeat this reasoning with the f_{-2} term (using the complex conjugate of the previous decomposition) to obtain the desired result. $\qquad\square$

Definition 4.C.2. *We will define a 2-tensor whose components satisfy the conditions of Theorem 4.C.1 to be in the* holomorphic gauge.

Importantly, let us note that if a 2-tensor f is purely scalar, i.e., $f = f_0$, then it is trivially already in the holomorphic gauge. Thus the holomorphic gauge is a gauge which respects the scalar part of the transform. This is to be contrasted with the solenoidal gauge, which will introduce spurious off-diagonal components even for scalar tensor fields.

Since $h_{\pm 2} \in \Omega_{\pm 2}$, let us write $h_2 = h e^{2i\theta}$ and $h_{-2} = \overline{h} e^{-2i\theta}$, where $h, \overline{h} \in L^2(M)$. In this notation, the holomorphic gauge condition reads

$$\eta_+ (\overline{h} e^{-2i\theta}) = 0,$$
$$\overline{\eta_+ (\overline{h} e^{-2i\theta})} = 0.$$

We have

$$\eta(\overline{h} e^{-2i\theta}) = e^{-\lambda} e^{i\theta} \left(\partial + i\partial\lambda \, \partial_\theta \right) (\overline{h} e^{-2i\theta}) \tag{4.11}$$
$$= e^{-\lambda} e^{-i\theta} \left(\partial + 2\partial\lambda \right) h, \tag{4.12}$$

so the holomorphic gauge conditions simplify to the Schrodinger type equations

$$(\partial + 2\partial\lambda)\overline{h} = 0,$$
$$(\overline{\partial} + 2\overline{\partial}\lambda)h = 0.$$

We can solve these equations by introducing an integrating factor of $e^{2\lambda}$. Then the equation for h gives us

$$0 = e^{2\lambda}\overline{\partial}h + 2\overline{\partial}\lambda \, e^{2\lambda} h$$
$$= \overline{\partial}(e^{2\lambda} h).$$

This amounts to saying that the solutions are holomorphic functions (up to an exponential factor) on the unit disk $\mathbb{D}$. We can therefore generate any solution as follows: given any continuous function on the unit circle $h|_{\partial\mathbb{D}} : \partial\mathbb{D} \to \mathbb{C}$, we can extend h into the unit disk using Cauchy's integral formula to get

$$h(z) = \frac{c_\lambda e^{-2\lambda(z)}}{2\pi i} \oint_{\partial\mathbb{D}} \frac{h|_{\partial\mathbb{D}}(w)}{w - z} \, dw, \tag{4.13}$$

where we write $c_\lambda = e^{2\lambda(1)}$ to denote the (constant) value of $e^{2\lambda}$ on the unit circle. We will use this convenient reconstruction property of solutions in the holomorphic gauge to benchmark the numerical (inverse) Radon transform in Appendix 4.E.

4.D   The Numerical (Inverse) Radon Transform

In this appendix, we describe the details of the numerical (inverse) Radon transform.

The manifold we work with is the hyperbolic plane $\mathbb{H}_2$, which is modeled as the Poincare disk, i.e., the unit disk equipped with the canonical hyperbolic metric

$$g = \frac{4}{(1 - x^2 - y^2)^2}(dx^2 + dy^2),$$

where (x, y) are global Poincare coordinates. For physical reasons, and to make the Radon transform well-defined, we must impose a cutoff on the Poincare disk. We thus pick a constant $\kappa \in (0, 1)$ and work with the Poincare disk restricted to $r \leq \kappa$. Equivalently, the cutoff Poincare disk can be regarded (after a rescaling of the metric) as a model for a hyperbolic plane with curvature equal to $-\kappa^2$.

To perform the numerical Radon transform, we must first discretize the Poincare disk. A natural first choice would be to use a uniform tiling, a choice which conforms best to the intrinsic symmetries of the Poincare disk. However, the Gauss-Bonnet theorem places limitations on how fine a uniform tiling can be, and the inability to take the tile size to zero is an unwieldy restriction. Instead, we opt for a simple square tessellation which we perform in the Beltrami-Klein model.

The Beltrami-Klein model is related to the Poincare disk model through the change of coordinates

$$(r, \theta) \mapsto (R, \Theta) = \left(\frac{2r}{1 + r^2}, \theta\right),$$

where (r, θ) denotes polar Poincare coordinates, and (R, Θ) denotes polar Beltrami-Klein coordinates. The Beltrami-Klein model has the convenient property that geodesics are straight lines. This makes a square tessellation in the Beltrami-Klein model the closest analogue to a regular Euclidean square tessellation for the hyperbolic plane. The figures present throughout this paper showcase the corresponding square tessellation in the Poincare disk.

Given a tessellation of the Poincare disk, we can discretize any function f by assigning to a given tile $\mathcal{T}$ the value of f at the centroid of $\mathcal{T}$. Ordering the tiles

arbitrarily, we can regard the discretized functions as vectors,

$$\mathbf{f} = \begin{pmatrix} f(\mathcal{T}_1) \\ \vdots \\ f(\mathcal{T}_N) \end{pmatrix}.$$

We must also discretize the (ideal) boundary of the Poincare disk for the Radon transform. We do so by placing M equally spaced boundary sites on the unit circle. We will then consider the collection of all geodesics originating from one boundary site and terminating on another. We can order the boundary sites by their angular position on the unit circle, and the geodesics lexicographically by the angular positions of their endpoints.

For each geodesic of the background geometry, the integral (4.B) can then be discretized by replacing the functions h_0, h and $\overline{h}$ with their piece-wise constant discretizations:

$$I_2[g](\gamma_j) = \int_{\gamma_j} e^{-2\lambda(\gamma_j(s))} \left(h_0(\gamma_j(s)) + h(\gamma_j(s))e^{2i\theta(s)} + \overline{h}(\gamma_j(s))e^{-2i\theta(s)} \right) ds$$

$$\approx \sum_{\mathcal{T}} \left[h_0(\mathcal{T})W_0(\gamma_j, \mathcal{T}) + h(\mathcal{T})W(\gamma_j, \mathcal{T}) + \overline{h}(\mathcal{T})\overline{W}(\gamma_j, \mathcal{T}) \right],$$

where W_0, W, and $\overline{W}$ contain the information on the geodesic and the remaining parts of the integrand. Explicitly, W_0 and W are defined by

$$W_0(\gamma_j, \mathcal{T}) = \int_{\gamma_j} e^{-2\lambda(\gamma_j(s))} \mathbf{1}_{\mathcal{T}}(\gamma_j(s)) \, ds,$$

$$W(\gamma_j, \mathcal{T}) = \int_{\gamma_j} e^{-2\lambda(\gamma_j(s))+2i\theta(s)} \mathbf{1}_{\mathcal{T}}(\gamma_j(s)) \, ds,$$

where $\mathbf{1}_{\mathcal{T}}(\gamma(s))$ is an indicator function such that

$$\mathbf{1}_{\mathcal{T}}(x) = \begin{cases} 1 & x \in \mathcal{T}, \\ 0 & x \notin \mathcal{T}. \end{cases}$$

Note that W_0 is nothing but the arc lengths of the geodesic segments that intersects a tile $\mathcal{T}$. The function W, on the other hand, also contains a complex weight which captures the directionality of the geodesic in the tile.

It is important to note that W_0 and W depend only on the particular choice of discretization, and can be pre-computed.

We can collect all of the quantities into a matrix equation. Let $\mathbf{W}$ be the $K \times 3N$, where $K = \binom{M}{2}$, defined by

$$\mathbf{W} = \begin{pmatrix} W_0(\gamma_1, \mathcal{T}_1) & \cdots & W_0(\gamma_1, \mathcal{T}_N) & W(\gamma_1, \mathcal{T}_1) & \cdots & W(\gamma_1, \mathcal{T}_N) & \overline{W}(\gamma_1, \mathcal{T}_1) & \cdots & \overline{W}(\gamma_1, \mathcal{T}_N) \\ W_0(\gamma_2, \mathcal{T}_1) & \cdots & W_0(\gamma_2, \mathcal{T}_N) & W(\gamma_2, \mathcal{T}_1) & \cdots & W(\gamma_2, \mathcal{T}_N) & \overline{W}(\gamma_2, \mathcal{T}_1) & \cdots & \overline{W}(\gamma_2, \mathcal{T}_N) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ W_0(\gamma_K, \mathcal{T}_1) & \cdots & W_0(\gamma_K, \mathcal{T}_N) & W(\gamma_K, \mathcal{T}_1) & \cdots & W(\gamma_K, \mathcal{T}_N) & \overline{W}(\gamma_K, \mathcal{T}_1) & \cdots & \overline{W}(\gamma_K, \mathcal{T}_N) \end{pmatrix}.$$

Likewise, let $\mathbf{h}$ be the length $3N$ vector defined by

$$\mathbf{h} = \begin{pmatrix} h_0(\mathcal{T}_1) \cdots h_0(\mathcal{T}_N) & h(\mathcal{T}_1) & \cdots & h(\mathcal{T}_N) & \overline{h}(\mathcal{T}_1) & \cdots & \overline{h}(\mathcal{T}_N) \end{pmatrix}^{\mathrm{T}}.$$

Then the discretized Radon transform, which we denote by $\Delta\mathbf{L}$, is given by the matrix equation

$$\Delta\mathbf{L} = \mathbf{W}\mathbf{h}. \tag{4.14}$$

The discrete inverse Radon transform is then just the inverse problem to the system (4.D). However, the inverse problem is complicated by the fact that the Radon transform is neither surjective nor injective (recall that the forward transform has a non-trivial kernel that corresponds to the gauge degrees of freedom).

To make the inverse well-posed, we must supplement it with a set of gauge constraints to pick out a unique inverse (when it exists) in the continuum case. In the discrete analog, we expect the kernel to be visible through the presence of zero eigenvalues in the singular value spectrum of $\mathbf{W}$. Due to discretization and numerical errors, the system (4.D) will either be singular or extremely ill-conditioned. As in the continuum case, we will supplement the system (4.D) with a set of discretized gauge constraints to make the problem well-posed.

For our implementation, we discretize the holomorphic gauge conditions given by equations (4.C) and (4.C). The holomorphic gauge constraints are simple first order partial differential equations. They can be realized discretely by implementing ∂_x and ∂_y are finite-difference operators. We will use a simple three-point stencil for the finite-difference operators, although higher-order variations can be used for increased accuracy. For example, the x partial of a function f at tile i will be approximated by

$$\partial_x f(\mathcal{T}_i) \approx \frac{f(\mathcal{T}_{R(i)}) - f(\mathcal{T}_{L(i)})}{2},$$

where $L(i)$ and $R(i)$ denote the indices of the tiles to the immediate left and right of tile i.[15] The y-partials are analogous. In this way, we can discretize the partial derivatives as matrices $\tilde{\Delta}_x$ and $\tilde{\Delta}_y$.[16] Since the vector $\mathbf{h}$ effectively contains 3 copies stacked on top of each other, we define the operators

$$\Delta_x = \left(\tilde{\Delta}_x \mid \tilde{\Delta}_x \mid \tilde{\Delta}_x\right),$$

$$\Delta_y = \left(\tilde{\Delta}_y \mid \tilde{\Delta}_y \mid \tilde{\Delta}_y\right). \tag{4.15}$$

The discretized holomorphic gauge conditions can then be written as

$$\mathbf{Ch} = (\Delta_x + i\Delta_y + 2\Lambda)\mathbf{h}, \tag{4.16}$$

where Λ is the matrix defined by

$$\Lambda = \left(\tilde{\Lambda} \mid \tilde{\Lambda} \mid \tilde{\Lambda}\right),$$

where $\tilde{\Lambda}$ has entries given by

$$\tilde{\Lambda}_{ij} = \partial\lambda(\mathcal{T}_i)\,\delta_{ij}.$$

4.E Accuracy of the Numerical Reconstruction

In the absence of an analytic reconstruction formula in the continuum case, we need to validate the numerical reconstruction so as to provide confidence that it performs the inverse transform correctly. To do so, we will first benchmark the numerical reconstruction using various test cases obtained by instantiating known rank-2 symmetric tensor fields in the bulk in the holomorhpic gauge. From Equation (4.C), we can see that a bulk solution in the holomorphic gauge can be readily generated by prescribing the boundary values.

As a benchmarking test, we therefore test the reconstruction algorithm as follows:

---

[15]This approximation will work for all non-boundary tiles. The tiles on the boundaries will have to use forwards or backwards finite-differences instead. For example, for a tile on the left boundary, we get

$$\partial_x f(\mathcal{T}_i) \approx \frac{-3f(\mathcal{T}_i) + 4f(\mathcal{T}_{R(i)}) - f(\mathcal{T}_{R(R(i))})}{2}.$$

[16]The entries of $\tilde{\Delta}_x$ are given by

$$(\tilde{\Delta}_x)_{ij} = \frac{\delta_{L(i),j} - \delta_{R(i),j}}{2}$$

if tile i is a non-boundary tile, with the appropriate modifications for the boundaries.

1. We first pick an arbitrary function $h|_{\partial\mathbb{D}} : \partial\mathbb{D} \to \mathbb{C}$ on the unit circle. We also pick another arbitrary scalar function $h_0 : \mathbb{D} \to \mathbb{C}$.

2. Using Equation (4.C), we extend the function $h|_{\partial\mathbb{C}}$ into the bulk to get a holomorphic function $h : \mathbb{D} \to \mathbb{D}$.

3. From the functions h_0 and h, we define the bulk tensor

$$f = h_0 + he^{2i\theta} + \overline{h}e^{-2i\theta},$$

where $\overline{h}$ denotes the complex conjugate of h. The f defined in this way is a real-valued continuum bulk 2-tensor with components fixed in the holomorphic gauge.

4. We run the forward numerical Radon transform to generate boundary data $\mathbf{b}$. From $\mathbf{b}$, we then run the discretized inverse Radon transform with holomorphic gauge constraints to numerically recover a discretized bulk solution $\mathbf{h}$.

5. Finally, we compare the values of h, evaluated at the centroids of the discretized tiling, with the reconstructed value $\mathbf{h}$. Denoting the exact solution at the centroids as $\mathbf{h}_*$, we can evaluate the relative error

$$\mathcal{E}_{\text{bulk}} = \frac{\|\mathbf{h} - \mathbf{h}_*\|}{\|\mathbf{h}_*\|}$$

to get a sense of the reconstruction quality.

In the absence of known analytic results, and before we move on with physically relevant examples, we can run the above test with several choices of boundary functions $h|_{\partial\mathbb{D}}$ as a proof of concept that the numerical inverse Radon transform is performing an adequate reconstruction.

Below, we show some sample test cases for the numerical inverse transform. All of the transforms shown below are performed with 1000 bulk tiles and 100 boundary sites (for a total of 4950 geodesics), following the procedure outlined above.

The reconstructions here show good agreement with the original bulk data. On a qualitative level, the plots of the bulk metric profile are essentially indistinguishable between the original and the reconstruction. The bulk relative errors across various test cases range from 0.01 to 0.1, indicating quantitatively successful reconstructions in general. The boundary relative errors are typically an order of magnitude smaller than the bulk relative errors, ranging from 0.001 to 0.01. The magnitudes of the

boundary relative errors for these known test cases can serve as an estimate for the general magnitude of numerical errors present in the algorithm. See Figures 4.16-4.19.

It should be noted that the relative errors shown here can actually be slightly misleading. Most sources of error in the reconstruction arise due to large fluctuating values of the boundary tiles. Tiles at or near the boundary are generally underconstrained due to the relatively small number of geodesics which pass through any given boundary region. This allows the boundary tiles to take on arbitrary values in order to minimize the relative boundary error as the algorithm is designed to do, although this comes at the cost of bulk accuracy. We can see that if we exclude the values of the boundary tiles from the calculation of relative error, that the relative error of the reconstruction is generally an order of magnitude smaller. This suggests that the bulk reconstruction performs very well deep into the bulk, with larger errors towards the boundary. This is important to keep in mind, since it suggests that the qualitative bulk picture provided by the numerical inverse should be generally trustworthy.
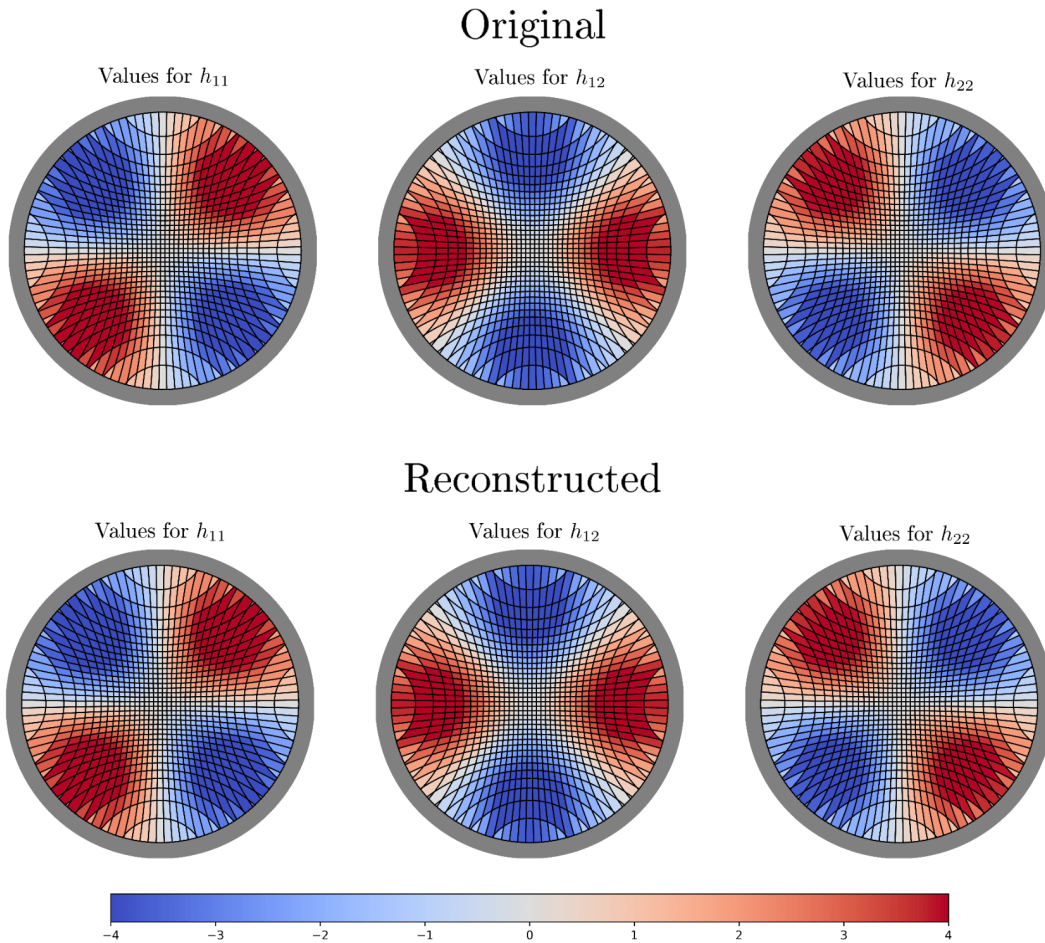
Figure 4.16: Benchmark reconstruction with $h_0(x, y) = xy$ and $h|_{\mathbb{C}}(\theta) = \sin(2\theta)$. **Top:** Original bulk data generated with $h_0(x, y) = xy$ and $h|_{\mathbb{C}}(\theta) = \sin(2\theta)$. **Bottom:** The bulk data reconstructed after running the forward Radon transform to get boundary data. Visually, the two sets of data are identical. The relative error between the two are $\mathcal{E}_{\text{bulk}} \approx 0.0227$ (relative error 0.004643 without boundary tiles).
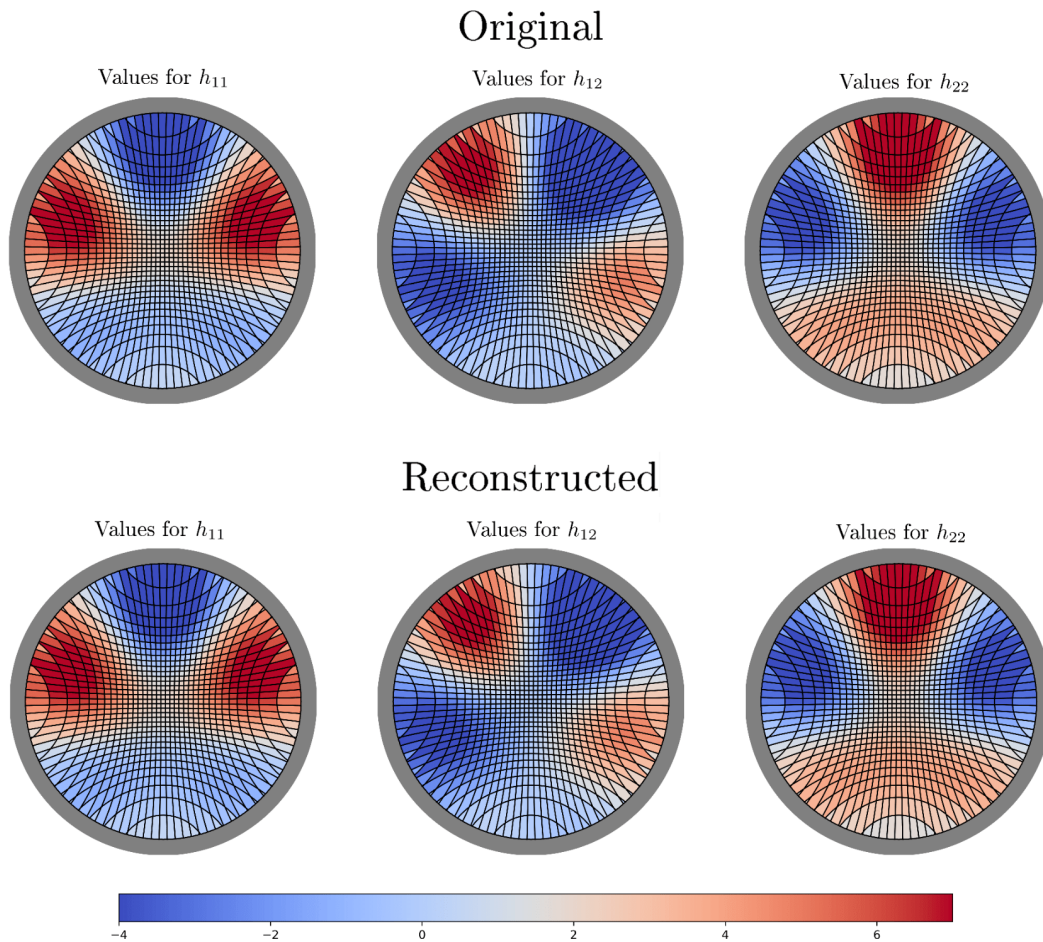
Figure 4.17: Benchmark reconstruction with $h_0(x, y) = 2e^{-(x^2+y^2)}$ and $h|_{\mathbb{C}}(\theta) = \cos(2\theta) + \sin(3\theta)$. **Top:** Original bulk data generated with $h_0(x, y) = 2e^{-(x^2+y^2)}$ and $h|_{\mathbb{C}}(\theta) = \cos(2\theta) + \sin(3\theta)$. **Bottom:** The bulk data reconstructed after running the forward Radon transform to get boundary data. Visually, the two sets of data are identical. The relative error between the two are $\mathcal{E}_{\text{bulk}} \approx 0.0273$ (relative error 0.005516 without boundary tiles).
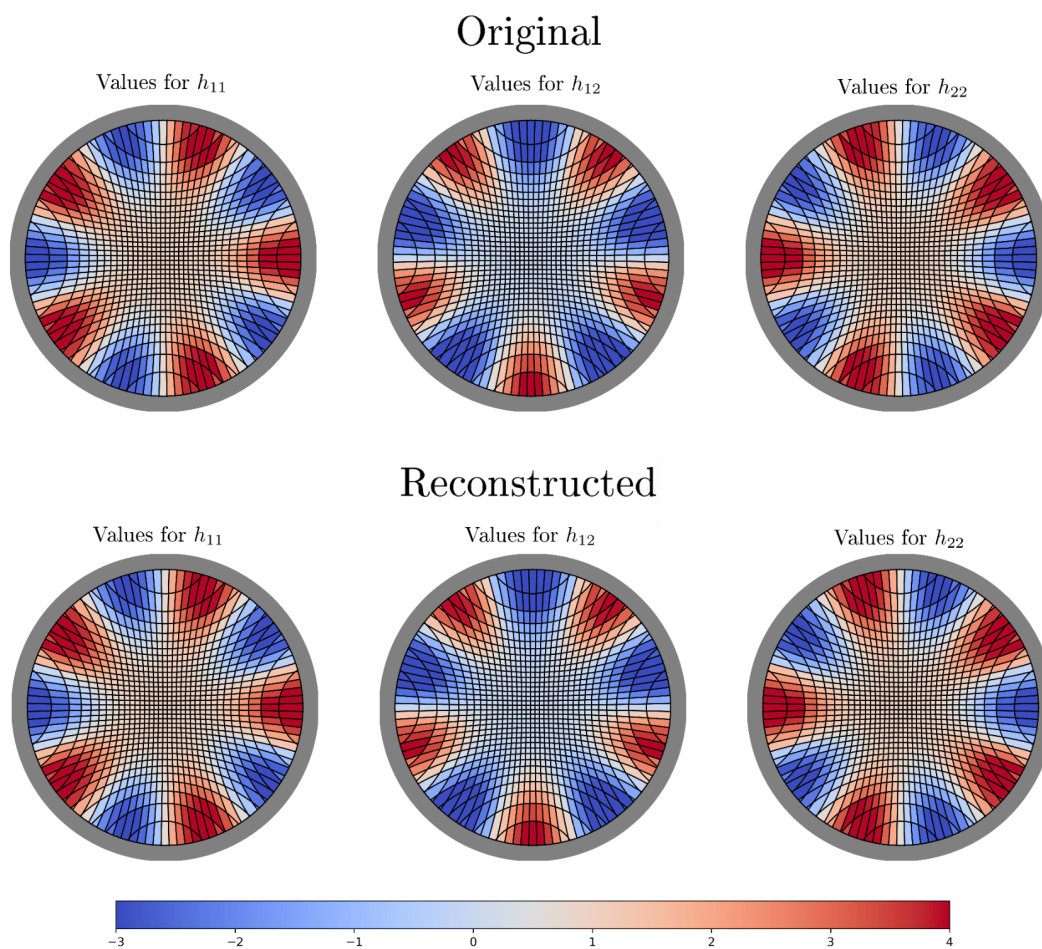
Figure 4.18: Benchmark reconstruction with $h_0(x, y) = 1/(1+x^2+y^2)$ and $h|_{\mathbb{C}}(\theta) = \cos(5\theta)$. **Top:** Original bulk data generated with $h_0(x, y) = 1/(1 + x^2 + y^2)$ and $h|_{\mathbb{C}}(\theta) = \cos(5\theta)$. **Bottom:** The bulk data reconstructed after running the forward Radon transform to get boundary data. Visually, the two sets of data are identical. The relative error between the two are $\mathcal{E}_{\text{bulk}} \approx 0.0885$ (relative error 0.01357 without boundary tiles).

Original

Values for h_{11} Values for h_{12} Values for h_{22}

Reconstructed

Values for h_{11} Values for h_{12} Values for h_{22}
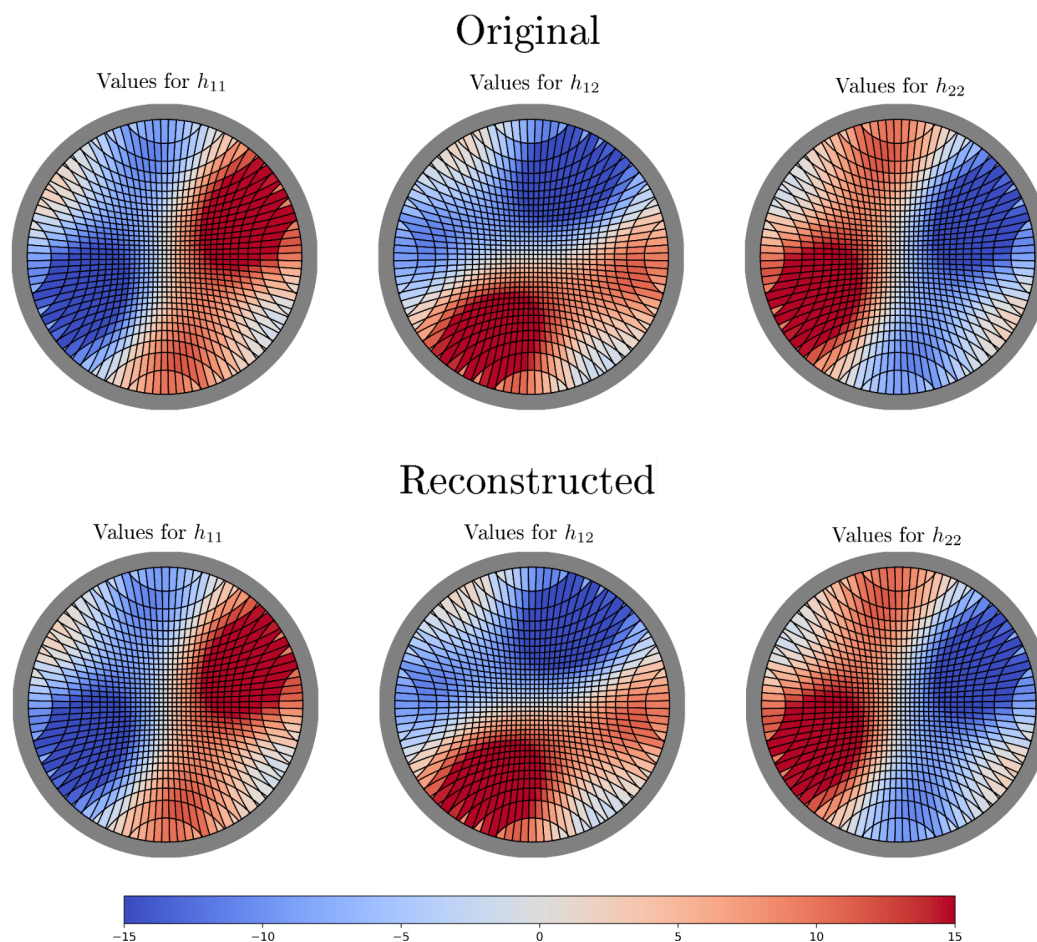
Figure 4.19: Benchmark reconstruction with $h_0(x, y) = x^2 + y^2$ and $h|_{\mathbb{C}}(\theta) = 2\cos(\theta) + 3\sin(3\theta)$. **Top:** Original bulk data generated with $h_0(x, y) = x^2 + y^2$ and $h|_{\mathbb{C}}(\theta) = 2\cos(\theta) + 3\sin(3\theta)$. **Bottom:** The bulk data reconstructed after running the forward Radon transform to get boundary data. Visually, the two sets of data are identical. The relative error between the two are $\mathcal{E}_{\text{bulk}} \approx 0.01886$ (relative error 0.003709 without boundary tiles).

4.E.1 Constrained Optimization

With the discrete Radon transform (4.D) and the discretized holomorphic constraints (4.D), we can solve this linear system for **h** as a constrained optimization problem:

$$\min_{\mathbf{h}} \|\mathbf{W}\mathbf{h} - \mathbf{b}\|, \tag{4.17}$$

$$\text{subject to } \mathbf{C}\mathbf{h} = \mathbf{0}.$$

We look for a best-fit solution $\mathbf{h}_*$ that solves the above system. In general, we do not expect there to exist a solution $\mathbf{h}_*$ such that $\mathbf{W}\mathbf{h}_* - \mathbf{b} = \mathbf{0}$, due to either numerical/discretization errors or the boundary data being non-geometric.

Because the objective function is linear, this problem has a unique global solution that can be obtained using standard constrained least squares. We briefly review the method below.

Theorem 4.E.1. *Consider the constrained least square problem* (4.E.1). *Assuming the stacked matrix*

$$\begin{pmatrix} \mathbf{W} \\ \mathbf{C} \end{pmatrix} \tag{4.18}$$

is left-invertible, and that $\mathbf{C}$ *is right-invertible, a vector* $\mathbf{h}_*$ *uniquely solves the constrained least square problem* (4.E.1) *if and only if there exists some* $\mathbf{g}$ *such that*

$$\begin{pmatrix} \mathbf{W}^\dagger\mathbf{W} & \mathbf{C}^\dagger \\ \mathbf{C} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{h}_* \\ \mathbf{g} \end{pmatrix} = \begin{pmatrix} \mathbf{W}^\dagger\mathbf{b} \\ \mathbf{0} \end{pmatrix}. \tag{4.19}$$

Proof. Suppose that $(\mathbf{h}_*, \mathbf{g})$ satisfies (4.E.1). Clearly $\mathbf{h}_*$ satisfies the constraint $\mathbf{C}\mathbf{h}_* = \mathbf{0}$. Then for any $\mathbf{h}$ that satisfies the constraint $\mathbf{C}\mathbf{h} = \mathbf{0}$, we have

$$\begin{aligned} \|\mathbf{W}\mathbf{h} - \mathbf{b}\|^2 &= \|\mathbf{W}(\mathbf{h} - \mathbf{h}_*) + \mathbf{W}\mathbf{h}_* - \mathbf{b}\|^2 \\ &= \|\mathbf{W}(\mathbf{h} - \mathbf{h}_*)\|^2 + \|\mathbf{W}\mathbf{h}_* - \mathbf{b}\|^2 + 2(\mathbf{h} - \mathbf{h}_*)^\dagger\mathbf{W}^\dagger(\mathbf{W}\mathbf{h}_* - \mathbf{b}) \\ &= \|\mathbf{W}(\mathbf{h} - \mathbf{h}_*)\|^2 + \|\mathbf{W}\mathbf{h}_* - \mathbf{b}\|^2 + 2(\mathbf{h} - \mathbf{h}_*)^\dagger\mathbf{C}^\dagger\mathbf{g} \\ &= \|\mathbf{W}(\mathbf{h} - \mathbf{h}_*)\|^2 + \|\mathbf{W}\mathbf{h}_* - \mathbf{b}\|^2 \\ &\geq \|\mathbf{W}\mathbf{h}_* - \mathbf{b}\|^2, \end{aligned}$$

where in the second line we used the definition of the norm $\|\mathbf{v}\|^2 = \mathbf{v}^\dagger\mathbf{v}$, in the third line we used the fact that

$$\mathbf{W}^\dagger\mathbf{W}\mathbf{h}_* + \mathbf{C}^\dagger\mathbf{g} = \mathbf{W}^\dagger\mathbf{b},$$

and in the fourth line we used the fact that

$$\mathbf{C}(\mathbf{h} - \mathbf{h}_*) = \mathbf{C}\mathbf{h} - \mathbf{C}\mathbf{h}_* = \mathbf{0} - \mathbf{0} = \mathbf{0}.$$

Therefore, $\mathbf{h}_*$ is an optimal solution. Furthermore, the optimal constrained solution is obtained if and only if

$$\begin{pmatrix} \mathbf{W} \\ \mathbf{C} \end{pmatrix} (\mathbf{h} - \mathbf{h}_*) = \mathbf{0}.$$

By assumption, the stacked matrix is left-invertible, therefore the minimum $\mathbf{h} = \mathbf{h}_*$ is also unique.

In the reverse direction, it suffices to show that the matrix in (4.E.1) is invertible. Suppose on the contrary that the matrix is non-invertible. Then there must exist a vector $(\mathbf{h}, \mathbf{g}) \neq \mathbf{0}$ such that

$$\begin{pmatrix} \mathbf{W}^\dagger \mathbf{W} & \mathbf{C}^\dagger \\ \mathbf{C} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{h} \\ \mathbf{g} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}. \tag{4.20}$$

Left multiplying both sides by $(\mathbf{h}, \mathbf{g})^\dagger$, we have

$$\begin{pmatrix} \mathbf{h}^\dagger & \mathbf{0}^\dagger \end{pmatrix} \begin{pmatrix} \mathbf{W}^\mathsf{T} \mathbf{W} & \mathbf{C}^\dagger \\ \mathbf{C} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{h} \\ \mathbf{g} \end{pmatrix} = \mathbf{h}^\dagger \mathbf{W}^\dagger \mathbf{W} \mathbf{h} + \mathbf{h}^\dagger \mathbf{C}^\dagger \mathbf{g} = 0.$$

Noting that $\mathbf{C}\mathbf{h}^\dagger = \mathbf{0}$, we get $\|\mathbf{W}\mathbf{h}\|^2 = 0$, so that $\mathbf{W}\mathbf{h} = \mathbf{0}$. Since the stacked matrix (4.E.1) is injective, it follows that we must have $\mathbf{h} = \mathbf{0}$. The system (4.E.1) then reduces to

$$\mathbf{C}^\dagger \mathbf{g} = \mathbf{0}.$$

We then conclude that $\mathbf{g} = \mathbf{0}$ since $\mathbf{C}$ is right-invertible by assumption. This is in contradiction with the assumption that $(\mathbf{h}, \mathbf{g}) \neq \mathbf{0}$. Hence the matrix in Equation (4.E.1) must be invertible, and a solution for $\mathbf{g}$ must exist. $\qquad\square$

4.E.2   Interpolation and Regularization

In principle, the optimal solution $\mathbf{h}_*$ can be obtained through straightforward matrix inversion of Equation (4.E.1). Suppose $\mathbf{h}$ is a column vector with $3N$ entries, and the constraint matrix is $M \times 3N$. Then any well-known polynomial algorithm for matrix inversion is of $O\left((M + 3N)^3\right)$. The inversion of system (4.E.1) is slightly complicated by the fact that the matrix $\mathbf{W}$ is generally expected to be extremely ill-conditioned. We can get around this by regularizing the system.

To make the least squares problem (4.E.1) better conditioned, and to smooth out small scale fluctuations in the discretized reconstruction, we employ a derivative type Tikhonov regularization. We replace the matrix $\mathbf{W}^\dagger \mathbf{W}$ in (4.E.1) by

$$\mathbf{W}^\dagger \mathbf{W} + \gamma \left(\Delta_x^\dagger \Delta_x + \Delta_y^\dagger \Delta_y \right),$$

where $\gamma > 0$ is a regularization parameter which controls the strength of the regularization, and where Δ_x and Δ_y are the discretized partial differential operators

defined in (4.D). This replacement effectively changes the least squares problem in (4.E.1) to

$$\min_{\mathbf{h}} \left(\|\mathbf{Wh} - \mathbf{b}\|^2 + \gamma\|\Delta_x\mathbf{h}\|^2 + \gamma\|\Delta_y\mathbf{h}\|^2 \right),$$

$$\text{subject to } \mathbf{Ch} = \mathbf{0},$$

which both regularizes the system so that the smallest singular values of $\mathbf{W}$ are of order $\sqrt{\gamma}$ and also takes into account the strength of fluctuations in the resulting solution. Since we expect small scale fluctuations to be mostly due to bulk discretization errors, this choice of regulator serves as a reasonable filter. In this note, all of our reconstructions employ regularization with $\gamma = 10^{-8}$.

In the case of fixed data but variable number of bulk tiles, the reconstruction can also become ill-conditioned when the number of bulk degrees of freedom exceed the number of boundary constraints in the form of geodesic lengths. Roughly speaking, because we have $O(3N)$ number of bulk degrees of freedom, one for each tensor component at a particular tile. Suppose there are K sites on the boundary. Then the number of linear equations from geodesic lengths is of order $O(K^2)$. Hence the reconstruction can become ill-conditioned when $3N > K^2$. While it is possible to decrease the number of bulk tiles, or increase the number of boundary sites, both come at a cost depending on our requirements for reconstruction. As an alternative, we can also interpolate between geodesic data by adding virtual sites and the lengths of geodesics for the extended set of sites.

In the current implementation, we add the virtual sites in between the original lattice sites such that the new lattice scale is half of the original. Let us label all sites sequentially along the counter clockwise direction on the boundary circle from 1 through $2K$ such that $2K + 1 \equiv 1$. To generate the geodesic lengths between a virtual site j and an original site i, we average the lengths of two geodesics that are anchored at $(i, j+1)$ and $(i, j-1)$. For a geodesic that ends on two virtual sites, i, j, we take the 4-point average of the original geodesic lengths for the ones anchored at $(i, j+1), (i, j-1), (i+1, j)$, and $(i-1, j)$.

BIBLIOGRAPHY

[1] M. Van Raamsdonk, "Comments on quantum gravity and entanglement," Jul. 2009. arXiv: 0907.2939.

[2] S. Ryu and T. Takayanagi, "Holographic derivation of entanglement entropy from AdS/CFT," *Physical Review Letters*, vol. 96, p. 181 602, 2006. DOI: 10.1103/PhysRevLett.96.181602.

[3] ——, "Aspects of Holographic Entanglement Entropy," *Journal of High Energy Physics*, no. 08, p. 045, 2006. DOI: 10.1088/1126-6708/2006/08/045.

[4] B. Swingle, "Entanglement Renormalization and Holography," *Physical Review D*, vol. D, p. 065 007, 2012. DOI: 10.1103/PhysRevD.86.065007.

[5] J. Maldacena and L. Susskind, "Cool horizons for entangled black holes," *Fortschritte der Physik*, vol. 61, pp. 781–811, 2013. DOI: 10.1002/prop.201300020.

[6] X.-L. Qi, "Exact holographic mapping and emergent space-time geometry," 2013. arXiv: 1309.6282.

[7] C. Cao, S. M. Carroll, and S. Michalakis, "Space from Hilbert Space: Recovering Geometry from Bulk Entanglement," *Physical Review D*, vol. 95, no. 2, p. 024 031, 2017. DOI: 10.1103/PhysRevD.95.024031.

[8] N. Lashkari, M. B. McDermott, and M. Van Raamsdonk, "Gravitational dynamics from entanglement 'thermodynamics'," *Journal of High Energy Physics*, no. 04, p. 195, 2014. DOI: 10.1007/JHEP04(2014)195.

[9] T. Faulkner, M. Guica, T. Hartman, R. C. Myers, and M. Van Raamsdonk, "Gravitation from Entanglement in Holographic CFTs," *Journal of High Energy Physics*, no. 03, p. 051, 2014. DOI: 10.1007/JHEP03(2014)051.

[10] B. Swingle and M. Van Raamsdonk, "Universality of Gravity from Entanglement," May 2014. arXiv: 1405.2933.

[11] T. Faulkner, F. M. Haehl, E. Hijano, O. Parrikar, C. Rabideau, and M. Van Raamsdonk, "Nonlinear Gravity from Entanglement in Conformal Field Theories," *Journal of High Energy Physics*, no. 08, p. 057, 2017. DOI: 10.1007/JHEP08(2017)057.

[12] N. Lashkari, C. Rabideau, P. Sabella-Garnier, and M. Van Raamsdonk, "Inviolable energy conditions from entanglement inequalities," *Journal of High Energy Physics*, no. 06, p. 067, 2015. DOI: 10.1007/JHEP06(2015)067.

[13] N. Lashkari, J. Lin, H. Ooguri, B. Stoica, and M. Van Raamsdonk, "Gravitational positive energy theorems from information inequalities," *Progress of Theoretical and Experimental Physics*, no. 12, p. 12C109, 2016. DOI: `10.1093/ptep/ptw139`.

[14] A. Almheiri, N. Engelhardt, D. Marolf, and H. Maxfield, "The entropy of bulk quantum fields and the entanglement wedge of an evaporating black hole," *Journal of High Energy Physics*, no. 12, p. 063, 2019. DOI: `10.1007/JHEP12(2019)063`.

[15] G. Penington, "Entanglement Wedge Reconstruction and the Information Paradox," 2019. arXiv: `1905.08255`.

[16] J. M. Maldacena, "The large n limit of superconformal field theories and supergravity," *Advances in Theoretical and Mathematical Physics*, vol. 2, pp. 231–252, 1998. DOI: `10.1023/A:1026654312961`.

[17] O. Aharony, S. S. Gubser, J. M. Maldacena, H. Ooguri, and Y. Oz, "Large N field theories, string theory and gravity," *Physics Reports*, vol. 323, pp. 183–386, 2000. DOI: `10.1016/S0370-1573(99)00083-6`.

[18] C. Cao and S. M. Carroll, "Bulk entanglement gravity without a boundary: Towards finding Einstein's equation in Hilbert space," *Physical Review D*, vol. 97, no. 8, p. 086003, 2018. DOI: `10.1103/PhysRevD.97.086003`.

[19] B. Czech, L. Lamprou, S. McCandlish, and J. Sully, "Tensor Networks from Kinematic Space," *Journal of High Energy Physics*, no. 07, p. 100, 2016. DOI: `10.1007/JHEP07(2016)100`.

[20] N. Bao, C. Cao, S. M. Carroll, and A. Chatwin-Davies, "De Sitter Space as a Tensor Network: Cosmic No-Hair, Complementarity, and Complexity," *Physical Review D*, vol. 96, no. 12, p. 123536, 2017. DOI: `10.1103/PhysRevD.96.123536`.

[21] G. 't Hooft, "Dimensional reduction in quantum gravity," *Conference on Highlights of Particle and Condensed Matter Physics (SALAMFEST)*, vol. C930308, pp. 284–296, 1993. arXiv: `gr-qc/9310026`.

[22] L. Susskind, "The World as a hologram," *Journal of Mathematical Physics*, vol. 36, pp. 6377–6396, 1995. DOI: `10.1063/1.531249`.

[23] M. Porrati and R. Rabadan, "Boundary rigidity and holography," *Journal of High Energy Physics*, vol. 01, p. 034, 2004. DOI: `10.1088/1126-6708/2004/01/034`.

[24] S. R. Roy and D. Sarkar, "Bulk metric reconstruction from boundary entanglement," *Physical Review D*, vol. 98, no. 6, p. 066017, 2018. DOI: `10.1103/PhysRevD.98.066017`.

[25] B. Czech and L. Lamprou, "Holographic definition of points and distances," *Physical Review D*, vol. 90, p. 106005, 2014. DOI: `10.1103/PhysRevD.90.106005`.

[26] F. Pastawski, B. Yoshida, D. Harlow, and J. Preskill, "Holographic quantum error-correcting codes: Toy models for the bulk/boundary correspondence," *Journal of High Energy Physics*, no. 6, p. 149, Jun. 2015. DOI: `10.1007/JHEP06(2015)149`.

[27] P. Hayden, S. Nezami, X.-L. Qi, N. Thomas, M. Walter, and Z. Yang, "Holographic duality from random tensor networks," *Journal of High Energy Physics*, no. 11, p. 009, 2016. DOI: `10.1007/JHEP11(2016)009`.

[28] G. Evenbly, "Hyperinvariant Tensor Networks and Holography," *Physical Review Letters*, vol. 119, no. 14, p. 141 602, 2017. DOI: `10.1103/PhysRevLett.119.141602`.

[29] Y.-Z. You, Z. Yang, and X.-L. Qi, "Machine learning spatial geometry from entanglement features," *Physical Review B*, vol. 97, no. 4, p. 045 153, 2018. DOI: `10.1103/PhysRevB.97.045153`.

[30] D. Kabat and G. Lifschytz, "Emergence of spacetime from the algebra of total modular Hamiltonians," *Journal of High Energy Physics*, no. 05, p. 017, 2019. DOI: `10.1007/JHEP05(2019)017`.

[31] N. Engelhardt and G. T. Horowitz, "Recovering the spacetime metric from a holographic dual," *Advances in Theoretical and Mathematical Physics*, vol. 21, pp. 1635–1653, 2017. DOI: `10.4310/ATMP.2017.v21.n7.a2`.

[32] ——, "Towards a Reconstruction of General Bulk Metrics," *Classical and Quantum Gravity*, vol. 34, no. 1, p. 015 004, 2017. DOI: `10.1088/1361-6382/34/1/015004`.

[33] J. Hammersley, "Extracting the bulk metric from boundary information in asymptotically AdS spacetimes," *Journal of High Energy Physics*, no. 12, p. 047, 2006. DOI: `10.1088/1126-6708/2006/12/047`.

[34] S. Bilson, "Extracting spacetimes using the AdS/CFT conjecture," *Journal of High Energy Physics*, no. 08, p. 073, 2008. DOI: `10.1088/1126-6708/2008/08/073`.

[35] ——, "Extracting Spacetimes using the AdS/CFT Conjecture: Part II," *Journal of High Energy Physics*, no. 02, p. 050, 2011. DOI: `10.1007/JHEP02(2011)050`.

[36] V. Balasubramanian, B. D. Chowdhury, B. Czech, J. de Boer, and M. P. Heller, "Bulk curves from boundary data in holography," *Physical Review D*, vol. 89, no. 8, p. 086 004, 2014. DOI: `10.1103/PhysRevD.89.086004`.

[37] S. Alexakis, T. Balehowsky, and A. Nachman, "Determining a Riemannian Metric from Minimal Areas," Nov. 2017. arXiv: `1711.09379`.

[38] N. Bao, C. Cao, S. Fischetti, and C. Keeler, "Towards Bulk Metric Reconstruction from Extremal Area Variations," *Classical and Quantum Gravity*, vol. 36, no. 18, p. 185 002, 2019. DOI: `10.1088/1361-6382/ab377f`.

[39] N. Bao, S. Nezami, H. Ooguri, B. Stoica, J. Sully, and M. Walter, "The Holographic Entropy Cone," *Journal of High Energy Physics*, no. 09, p. 130, 2015. DOI: `10.1007/JHEP09(2015)130`.

[40] P. Stefanov and G. Uhlmann, "Recent progress on the boundary rigidity problem," *Electron. Res. Announc. Amer. Math. Soc.*, vol. 11, pp. 64–70, 2005. DOI: `10.1090/S1079-6762-05-00148-4`.

[41] V. Sharafutdinov, *Integral Geometry of Tensor Fields*, ser. Inverse and ill-posed problems series. VSP, 1994, ISBN: 9789067641654.

[42] L. Pestov and G. Uhlmann, "Two dimensional compact simple Riemannian manifolds are boundary distance rigid," May 2003. arXiv: `math/0305280`.

[43] P. Calabrese and J. Cardy, "Entanglement entropy and conformal field theory," *Journal of Physics A*, vol. 42, p. 504 005, 2009. DOI: `10.1088/1751-8113/42/50/504005`.

[44] G. Besson, G. Courtois, and S. Gallot, "Entropies et rigidités des espaces localement symétriques de courbure strictment négative," *Geometric and Functional Analysis*, vol. 5, pp. 731–799, 1995.

[45] C. B. Croke, "Rigidity for surfaces of non-positive curvature," *Commentarii Mathematici Helvetici*, vol. 65, pp. 150–169, 1990. DOI: `10.1007/BF02566599`.

[46] V. E. Hubeny, M. Rangamani, and T. Takayanagi, "A Covariant holographic entanglement entropy proposal," *Journal of High Energy Physics*, no. 07, p. 062, 2007. DOI: `10.1088/1126-6708/2007/07/062`.

[47] B. Czech, L. Lamprou, S. McCandlish, B. Mosk, and J. Sully, "Equivalent Equations of Motion for Gravity and Entropy," *Journal of High Energy Physics*, no. 02, p. 004, 2017. DOI: `10.1007/JHEP02(2017)004`.

[48] F. Monard, "Efficient tensor tomography in fan-beam coordinates," Oct. 2015. arXiv: `1510.05132`.

[49] V. P. Krishnan, "On the inversion formulas of Pestov and Uhlmann for the geodesic ray transform," *Journal of Inverse and Ill-posed Problems*, vol. 18, 2010. DOI: `10.1515/jiip.2010.017`.

[50] G. Uhlmann and A. Vasy, "The inverse problem for the local geodesic ray transform," Oct. 2012. arXiv: `1210.2084`.

[51] F. Monard, "On reconstruction formulas for the ray transform acting on symmetric differentials on surfaces," *Inverse Problems*, vol. 30, no. 6, p. 065 001, Jun. 2014. DOI: `10.1088/0266-5611/30/6/065001`.

[52] G. Bal, "Ray transforms in hyperbolic geometry," *Journal de Mathématiques Pures et Appliquées*, vol. 84, no. 10, pp. 1362–1392, 2005, ISSN: 0021-7824. DOI: `10.1016/j.matpur.2005.02.001`.

[53]   F. A. Dahlen and J. Tromp, *Theoretical global seismology*. Princeton University Press, 1999, ISBN: 0691001243.

[54]   I. Heemskerk, J. Penedones, J. Polchinski, and J. Sully, "Holography from Conformal Field Theory," *Journal of High Energy Physics*, no. 10, p. 079, 2009. DOI: 10.1088/1126-6708/2009/10/079.

[55]   M. Banados, C. Teitelboim, and J. Zanelli, "The Black hole in three-dimensional space-time," *Physical Review Letters*, vol. 69, pp. 1849–1851, 1992. DOI: 10.1103/PhysRevLett.69.1849.

[56]   H. Liu and S. J. Suh, "Entanglement growth during thermalization in holographic systems," *Physical Review D*, vol. 89, no. 6, p. 066 012, 2014. DOI: 10.1103/PhysRevD.89.066012.

[57]   J. Couch, S. Eccles, P. Nguyen, B. Swingle, and S. Xu, "The Speed of Quantum Information Spreading in Chaotic Systems," Aug. 2019. arXiv: 1908.06993.

[58]   A. Almheiri, X. Dong, and B. Swingle, "Linearity of Holographic Entanglement Entropy," *Journal of High Energy Physics*, no. 02, p. 074, 2017. DOI: 10.1007/JHEP02(2017)074.

[59]   A. Almheiri, X. Dong, and D. Harlow, "Bulk Locality and Quantum Error Correction in AdS/CFT," *Journal of High Energy Physics*, no. 04, p. 163, 2015. DOI: 10.1007/JHEP04(2015)163.

[60]   P. Di Francesco, P. Mathieu, and D. Sénéchal, *Conformal field theory*, ser. Graduate texts in contemporary physics. New York, NY: Springer, 1997. DOI: 10.1007/978-1-4612-2256-9.

[61]   I. Peschel and V. Eisler, "Reduced density matrices and entanglement entropy in free lattice models," *Journal of Physics A Mathematical General*, vol. 42, no. 50, 504003, p. 504 003, Dec. 2009. DOI: 10.1088/1751-8113/42/50/504003.

[62]   R. C. Myers, J. Rao, and S. Sugishita, "Holographic Holes in Higher Dimensions," *Journal of High Energy Physics*, no. 06, p. 044, 2014. DOI: 10.1007/JHEP06(2014)044.

[63]   V. Balasubramanian and C. Rabideau, "The dual of non-extremal area: Differential entropy in higher dimensions," Dec. 2018. arXiv: 1812.06985.

[64]   V. Lekic, personal communication, 2019.

[65]   X. Dong, "The Gravity Dual of Renyi Entropy," *Nature Commun.*, vol. 7, p. 12 472, 2016. DOI: 10.1038/ncomms12472.

[66]   S. Burdick and V. Lekić, "Velocity variations and uncertainty from trans-dimensional P-wave tomography of North America," *Geophysical Journal International*, vol. 209, no. 2, pp. 1337–1351, Mar. 2017. DOI: 10.1093/gji/ggx091.

[67] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*. USA: Society for Industrial and Applied Mathematics, 2004, ISBN: 0898715725.

[68] L. Pestov and G. Uhlmann, "Two dimensional compact simple riemannian manifolds are boundary distance rigid," *Annals of mathematics*, vol. 161, pp. 1093–1110, May 2003. DOI: `10.4007/annals.2005.161.1093`.

[69] G. Paternain, M. Salo, and G. Uhlmann, "Tensor tomography: Progress and challenges," *Chinese Annals of Mathematics, Series B*, vol. 35, pp. 399–428, May 2014. DOI: `10.1007/s11401-014-0834-z`.

[70] ——, "Tensor tomography on surfaces," *Inventiones mathematicae*, vol. 193, pp. 229–247, Nov. 2013. DOI: `10.1007/s00222-012-0432-1`.

[71] F. Monard, "On reconstruction formulas for the ray transform acting on symmetric differentials on surfaces," *Inverse Problems*, vol. 30, no. 6, p. 065001, May 2014. DOI: `10.1088/0266-5611/30/6/065001`.

Chapter 5

THE GHOST IN THE RADIATION: ROBUST ENCODINGS OF THE BLACK HOLE INTERIOR

We reconsider the black hole firewall puzzle, emphasizing that quantum error-correction, computational complexity, and pseudorandomness are crucial concepts for understanding the black hole interior. We assume that the Hawking radiation emitted by an old black hole is pseudorandom, meaning that it cannot be distinguished from a perfectly thermal state by any efficient quantum computation acting on the radiation alone. We then infer the existence of a subspace of the radiation system which we interpret as an encoding of the black hole interior. This encoded interior is entangled with the late outgoing Hawking quanta emitted by the old black hole, and is inaccessible to computationally bounded observers who are outside the black hole. Specifically, efficient operations acting on the radiation, those with quantum computational complexity polynomial in the entropy of the remaining black hole, commute with a complete set of logical operators acting on the encoded interior, up to corrections which are exponentially small in the entropy. Thus, under our pseudorandomness assumption, the black hole interior is well protected from exterior observers as long as the remaining black hole is macroscopic. On the other hand, if the radiation is not pseudorandom, an exterior observer may be able to create a firewall by applying a polynomial-time quantum computation to the radiation.

5.1 Introduction

The discovery that black holes emit Hawking radiation raised deep puzzles about the quantum physics of black holes [1]. What happens to quantum information that falls into a black hole, if that black hole subsequently evaporates completely and disappears? Is the information lost forever, or does it escape in the radiation emitted by the black hole, albeit in a highly scrambled form that is difficult to decode? And if the information does escape, how? The struggle to definitively answer these

questions has been a major theme of quantum gravity research during the 45 years since Hawking's pivotal discovery.

The AdS/CFT holographic correspondence provides powerful evidence indicating that quantum information really does escape from an evaporating black hole [2]. This correspondence, for which there is now substantial evidence, asserts that the process in which a black hole forms and then completely evaporates in an asymptotically anti-de Sitter bulk spacetime admits a dual description in terms of a conformally-invariant quantum field theory living on the boundary of the spacetime. In this dual description, the system evolves unitarily and therefore the process is microscopically reversible — on the boundary there is no gravity, no black hole, no place for information to hide. Since this observation applies to evaporating black holes that are small compared to the AdS curvature scale, it seems plausible that a similar conclusion should apply to more general spacetimes which are not asymptotically AdS, even though we currently lack a firm grasp of how quantum gravity works in that more general setting.

However, so far the holographic correspondence has not provided a satisfying picture of the *mechanism* that allows the information to escape from behind the black hole's event horizon. It is not even clear how the boundary theory encodes the experience of observers who cross the event horizon and visit the black hole interior.

That describing the inside of a black hole raises subtle issues was emphasized in 2012 by the authors known as AMPS [3]. Following AMPS, consider a black hole H that is maximally entangled with another system E which is outside the black hole, and suppose that B is a thermally occupied Hawking radiation mode which is close to the horizon and moving radially outward. Since the black hole is maximally entangled with E, the highly mixed state of B must be purified by a subsystem of E. But on the other hand, we expect that a freely falling observer who enters the black hole will not encounter any unexpected excitations at the moment of crossing the horizon; since field modes are highly entangled in the vacuum state, this means that B should be purified by a mode $\tilde{B}$ located inside the black hole. Now we have a problem, because it is not possible for the mixed state of B to be purified by both E and $\tilde{B}$. Something has to give! Were we to break the entanglement between B and $\tilde{B}$ for the sake of preserving the entanglement between B and E, the infalling observer would encounter a seething firewall at the horizon. This conclusion is hard to swallow, since for a macroscopic black hole we would expect semiclassical theory to be trustworthy at the event horizon, and the black hole solution to the classical

Einstein equation has a smooth horizon, not a firewall.

To find a way out of this quandary, it is helpful to contemplate the thermofield double (TFD) state of two boundary conformal field theories, which we will refer to as the left and right boundary theories. The TFD is an entangled pure state of the left and right boundaries, with the property that the marginal state of the right boundary (with the left boundary traced out) is a thermal state with temperature T, and likewise the marginal state of the left boundary (with the right boundary traced out) is thermal with the same temperature. The corresponding bulk geometry is a two-sided black hole. Both the left black hole and the right black hole are in equilibrium with a radiation bath at temperature T, and both have smooth event horizons. Furthermore, the two black holes have a shared interior — they are connected in the bulk by a non-traversable wormhole behind the horizon [4]. Here, the right black hole (let us call it H) is purified by another system (the left black hole E), and emits Hawking radiation, yet it has a smooth horizon. How can we reconcile this finding with the AMPS argument?

For the case of the two-sided black hole, there is an instructive answer [5]. The Hawking mode B outside the right black hole can be purified by both $\tilde{B}$ behind the horizon and by a subsystem of E, because E itself lies behind the horizon and $\tilde{B}$ is a subsystem of E! It is very tempting to suggest that a similar resolution of the AMPS puzzle applies to the case of a one-sided black hole H, which is entangled with a system E outside its horizon. That is, we may regard the black hole interior and the exterior system entangled with the black hole as two complementary descriptions of one and the same system. Indeed, we might imagine allowing E to undergo gravitational collapse, thereby obtaining a pair of entangled black holes, which, if we accept a conjecture formulated in [5], would be connected through the bulk by a non-traversable wormhole. The boundary dual of this bulk state, up to a one-sided transformation acting on one of the two boundaries, is a TFD, to which our previous discussion of the entanglement structure of the two-sided black hole ought to apply.

The idea that, for the case of a black hole H purified by the exterior system E, we may regard the black hole interior as related to E by a complicated encoding map, has been advocated, discussed, and criticized in much previous work [5]–[10]. We will revisit this issue in this paper, arguing that a proper resolution of the AMPS puzzle should invoke concepts that have received relatively short shrift in earlier discussions of the firewall problem, namely *quantum error correction, computational complexity, and pseudorandomness*.

The scenario described above, in which the black hole H has become maximally entangled with the exterior system E, might arise because the black hole actually formed long ago, and since then has radiated away more than half of its initial entropy. In that case E would be the Hawking radiation so far emitted during the black hole's lifetime, most of which is by now far away from the black hole. One could object that our proposal, that the black hole interior is related to E by a complicated encoding map, is too wildly non-local to be credible [8], [9], [11]. Why cannot an exterior agent who interacts with the Hawking radiation send instantaneous signals to the black hole interior in flagrant violation of causality? And why cannot such an agent access the encoded system $\tilde{B}$, breaking the entanglement between $\tilde{B}$ and B and hence creating excitations which can be detected by an observer who falls through the horizon?

Our answer is that such non-local operations are in principle possible, but are not accessible to observers whose computational abilities are bounded (a notion we make precise in Section 5.6); the operations required to disturb the interior mode are far too complex to be realizable in practice for any realistic observer. Thus, in spite of the extreme non-locality of the encoding map, violations of the semiclassical causal structure of the black hole spacetime are beyond the reach of any realistic exterior observer. This statement is most conveniently expressed using the language of quantum error correction and computational complexity. We will use $|S|$ to denote the size of a physical system S; by size we mean the number of qubits, so that $2^{|S|}$ is the dimension of the Hilbert space of S. We regard H as the Hilbert space of black hole microstates, and E as the Hilbert space of the previously emitted radiation. For an old black hole H which is nearly maximally entangled with E, we show that a quantum error-correcting code can be constructed, in a subspace of EH, which describes the black hole interior. The logical operators of this code, which preserve the code subspace, are operators acting on the interior. We will argue that a code exists with the following property: any operation on the radiation E that can be performed as a quantum computation whose size is polynomial in $|H|$ will commute with a set of logical operators of the code, up to corrections which are exponentially small in $|H|$. For this encoding, then, an observer outside a black hole can signal the interior only by performing an operation of super-polynomial complexity. Because the encoded interior is for all practical purposes invisible to the agent who roams the radiation system E, we call the code's logical operators *ghost operators*.[1]

---

[1]The word "ghost" is sometimes used to describe unphysical degrees of freedom. That is not what we mean here. The ghost operators act on a system (the interior of a black hole) which is

To reach this conclusion, we make a nontrivial but reasonable assumption — that the radiation system E is *pseudorandom*. Note that if the state of EH is pure, and $|H| \ll |E|$, then the density operator ρ_E of E is not full rank, so that ρ_E is obviously distinguishable from the maximally mixed state σ_E. When we say that ρ_E is pseudorandom, we mean that ρ_E and σ_E are not *computationally* distinguishable. That is, suppose we receive a copy of ρ_E (or even polynomially many copies) and we are asked to determine whether the state is maximally mixed or not using a quantum circuit whose size is polynomial in $|H|$. If ρ_E is pseudorandom, then our probability of answering correctly exceeds $1/2$ by an amount which is exponentially small in $|H|$. Such pseuodrandom quantum states exist, and furthermore it has recently been shown [12] that they can be prepared by efficient quantum circuits, if one accepts a standard (and widely believed) assumption of post-quantum cryptography: That there exist one-way functions which are hard to invert using a quantum computer. Since black holes are notoriously powerful scramblers of quantum information [13], we think the assumption that ρ_E is pseudorandom is plausible, though undeniably speculative. Our main technical result shows that if ρ_E is pseudorandom, then a code with ghost logical operators must exist.

That the existence of quantum-secure one-way functions implies the hardness of *decoding* Hawking radiation had been pointed out earlier in [14] and [15]. But our statement goes further — it indicates that causality is well respected from the viewpoint of computationally bounded observers (as long as $|H|$ is large). The semi-classical causal structure of the evaporating black hole spacetime can be disrupted by an observer with sufficient computational power, but not by an observer whose actions can be faithfully modeled by a quantum circuit with size polynomial in $|H|$. On the other hand, interior observers, who in principle have access to H as well E, could plausibly perform nontrivial operations on the interior which are beyond the reach of the computationally bounded observer who acts on E alone.

Our main result can be regarded as a contribution to the theory of quantum error correction in a nonstandard setting. In the context of fault-tolerant quantum computation, where the goal is to protect a quantum computer from noise due to uncontrolled interactions with the computer's environment, we usually consider noise which is weak and only weakly correlated. For example, we might model the noise using a Hamiltonian describing the interactions of the computer and environment, where each term in the Hamiltonian is small and acts on only a few of the computer's

---

physical but *inaccessible* to observers outside the black hole who have reasonable computational power.

qubits. In our setting the "computer" is the system EH, and the "noise" results from the interactions of the computationally bounded observer with system E, while H is regarded as noiseless. In contrast to conventional quantum error correction, we allow the noise to be strong, highly correlated, and adversarially chosen, yet the logical system $\tilde{B}$ encoded in EH is well protected against this noise. To obtain this result, though, it is essential that the noise acts only on E and not on H, a departure from the usual model of fault tolerance in which all qubits are assumed to be noisy.

We note that the encoding of the black hole interior in EH is state dependent; that is, the way the system $\tilde{B}$ is embedded in EH depends on the initial state that underwent gravitational collapse to form a black hole. This state dependence of the encoding has sparked much discussion and consternation [8], [16]. What seems troubling is that operators which depend on the state to which they are applied are not linear operators acting on Hilbert space, and therefore cannot be regarded as observables as described in the conventional quantum theory of measurement. Our view is that the tension arising from the state dependence of the encoded operator algebra signals that we do not yet have a fully satisfactory way to describe measurements performed inside black holes. We will not rectify this shortcoming in this paper.

Our argument about the robustness of the ghost logical operators makes no direct use of AdS/CFT technology. This may be viewed as either a strength or a weakness. The strength is that our results may be applicable to black holes in spacetimes which are asymptotically flat or de Sitter, and stand independently of any assumptions of holography. The weakness is that we have not presented evidence based on holographic duality which supports our conjecture.

There has been great recent progress toward resolving the discrepancy between Hawking's semiclassical analysis [1] and the Page curve [17]–[19] of an evaporating black hole, including formulas for the entropy of the radiation supported by explicit computations [20]–[22]. These results strengthen the evidence that black hole evaporation is a unitary process, and also point toward a resolution of the firewall problem in which the interior of a partially evaporated black hole is encoded in the Hawking radiation. This beautiful prior work, however, does not directly address how the profoundly nonlocal encoding of the interior in the radiation is compatible with the semiclassical causal structure of the black hole geometry. It is for that purpose that we hope our observations concerning the pseudorandomness of the radiation and the construction of ghost logical operators acting on the interior will prove to be relevant. Our main conclusion is that the encoded interior can be

inaccessible to observers outside the black hole who have reasonable computational power. Establishing closer contact between our work and these recent computations is an important open problem.

The rest of this paper is structured as follows. In Section 5.2, we provide a non-technical summary of the paper. In Sections 5.3 and 5.4, we review the notion of pseudorandomness in both the classical and quantum setting; in Section 5.5, we argue that the Hawking radiation is a pseudorandom quantum state, and we explain in detail our computational model of the black hole.

In the remaining sections, we derive consequences of the pseudorandomness assumption, and explore their potential relevance to the black hole firewall problem. In Section 5.6, we show that it is computationally hard for an observer interacting with the early radiation E to distill the interior mode $\tilde{B}$ and carry it into the black hole. In Section 5.7, we show that the encoded system $\tilde{B}$ is protected against errors inflicted on E by any agent who performs a quantum operation with $\text{poly}(|H|)$ computational complexity and sufficiently small Kraus rank. In Section 5.8, we describe the construction of ghost logical operators acting on the black hole interior; these operators commute with all low-complexity operations applied to E by an agent O, provided that O's quantum memory is not too large. If the observable properties of the black hole interior are described by such ghost operators, we infer that the interior cannot be affected or detected by computationally bounded agents who interact with the Hawking radiation. The theory of ghost operators, which can be constructed for any approximate quantum error-correcting code, may also be of independent interest. In Section 5.9, we show that, if the state of the partially evaporated black hole has been efficiently generated, then an agent with access to both E and H can manipulate the encoded interior efficiently, and efficiently distill the encoded system $\tilde{B}$ to a small quantum memory. Section 5.10 contains our conclusions. Some technicalities are treated in the Appendices, and in Appendix 5.D we discuss via an example how the construction of ghost logical operators may fail if the Hawking radiation is not pseudorandom.

5.2 Probing the Radiation

In this section we will provide a somewhat more explicit explanation of our main result, still skipping over technical details which will be laid out in later sections. The situation we consider is depicted in Figure 5.1. There, the unitary transformation U_{bh} describes the formation and subsequent partial evaporation of a black hole formed

from infalling matter in a pure state $|\phi_{\text{matter}}\rangle$, where E denotes the "early" Hawking radiation which has been emitted so far, H denotes the remaining black hole which has not yet evaporated, and B denotes Hawking quanta of the "late" radiation which has just been emitted from the black hole. We may assume for convenience that B is a single qubit — our conclusions would be the same if we considered B to be any system of constant dimension, independent of the size of E and H. The system P denotes an ancillary system called the "probe", which might represent, for example, ambient dust around the black hole. We will discuss the role of the probe in greater detail shortly, but for simplicity we may ignore its presence right now.
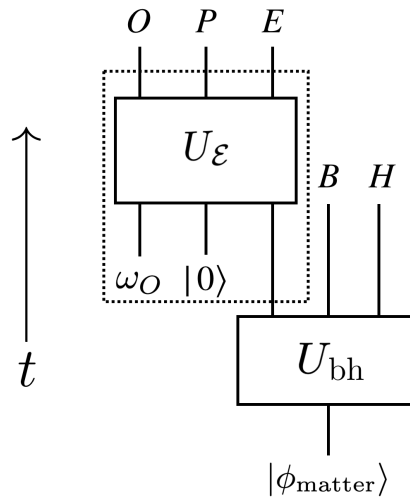
Figure 5.1: A black hole forms due to the gravitational collapse of an infalling state of matter. The black hole then evaporates for a while, emitting the "early" radiation E and the "late" radiation B; the formation and partial evaporation of the black hole are described by the unitary transformation U_{bh}. An observer O interacts with the early radiation and a probe system P, where the unitary transformation $U_{\mathcal{E}}$ (enclosed by the dotted line) has quantum complexity which scales polynomially with the size $|H|$ of the remaining black hole (essentially its entropy S_{bh}). If the radiation is pseudorandom, then O is unable to distinguish E from a perfectly thermal state.

In the case of an "old" black hole H, which has already radiated away over half of its initial entropy and has become nearly maximally entangled with E, we have $|H| < |E|$. Because the lifetime of an evaporating black hole scales like the 3/2 power of its initial system size, we may regard the unitary transformation U_{bh} to be "efficient," meaning that it can be accurately described by a quantum circuit whose size increases only polynomially with $|EHB|$. Our key assumption is that the efficient unitary U_{bh} creates a pseudorandom state of EB (see Section 5.5). The notion of a pseudorandom quantum state will be further discussed in Section 5.4.

In the context of the AMPS puzzle, the recently emitted system B should be purified by a system $\tilde{B}$ behind the horizon. We will explore the idea that this system $\tilde{B}$ is actually encoded in EH, the union of the black hole system H, and the early radiation system E. Let us denote the state prepared by the unitary map U_{bh} as $|\Psi\rangle_{EHB}$, and consider its expansion

$$|\Psi\rangle_{EHB} = \sum_{ijk} \Psi_{ijk} |i\rangle_E \otimes |j\rangle_H \otimes |k\rangle_B.$$

There is a corresponding map $V_\Psi : \tilde{B} \rightarrow EH$ defined by

$$V_\Psi = \sqrt{d_B} \sum_{ijk} \Psi_{ijk} |i\rangle_E \otimes |j\rangle_H \otimes \langle k|_{\tilde{B}},$$

where d_B is the dimension of B. If B is maximally mixed in the state $|\Psi\rangle$, then V_Ψ is an isometric map embedding $\tilde{B}$ in EH. We interpret V_Ψ as the encoding map of a quantum error-correcting code, which maps the interior system $\tilde{B}$ to the subspace of EH with which B is maximally entangled.

If $\tilde{T}_{\tilde{B}}$ is any operator acting on $\tilde{B}$, there is a corresponding "logical" operator T_{EH} acting on EH defined by

$$V_\Psi \tilde{T}_{\tilde{B}} = T_{EH} V_\Psi.$$

This logical operator is not uniquely defined, because Equation (5.2) only specifies its action on the code space, the image of V_Ψ. We may say that T_{EH} is the "mirror operator" of $\tilde{T}_{\tilde{B}}$ determined by $|\Psi\rangle$, whose defining property is that T_{EH} and $\tilde{T}_{\tilde{B}}$ produce the same output when acting on the state $|\Psi\rangle$.

We wish to investigate whether an agent who interacts with only the radiation system E can manipulate the encoded system $\tilde{B}$. For that purpose, we introduce an additional system O to represent an observer outside the black hole who interacts with E. This interaction is modeled by a unitary transformation $U_{\mathcal{E}}$ acting on OE, possibly followed by a simple measurement performed on O; for example one might measure all the qubits of O in a standard basis. The unitary transformation, but not the following measurement, is shown in Figure 5.1. After the interaction, but before O is measured, the joint state of $OEBH$ has evolved to

$$|\Psi'\rangle_{OEBH} = ((U_{\mathcal{E}})_{OE} \otimes I_{BH}) \left(|\omega\rangle_O \otimes |\Psi\rangle_{EBH}\right),$$

where $|\omega\rangle_O$ is the initial state of O before O and E interact.

Our notation $U_{\mathcal{E}}$ for the unitary transformation is motivated by a widely used convention in the theory of quantum channels, in which $\mathcal{E}$ denotes a quantum noisy channel (a trace-preserving completely positive map), with the letter $\mathcal{E}$ indicating an "error" acting on the input to the channel. A quantum channel always admits a dilation (also called a purification), a unitary transformation which acts on the input system and an "environment," after which the environment is discarded. In our context, the noisy channel $\mathcal{E}$ acting on E arises from the action of the observer, and we may regard the observer's system O as the environment in the dilation of $U_{\mathcal{E}}$. In the following discussion, we will often omit the subscript OE on $(U_{\mathcal{E}})_{OE}$, leaving it implicit that $U_{\mathcal{E}}$ acts on the radiation system E and observer O.

Now we can appeal to a standard result in the theory of quantum error correction. In $|\Psi\rangle_{EHB}$, we regard B as a "reference system" which purifies the maximally mixed state of the encoded system $\tilde{B}$. Is there a recovery operator which can be applied to EH to correct the error induced by this noisy channel? In fact a recovery operator that corrects the error *exactly* exists if and only if the marginal state ρ'_{OB} of OB factorizes,

$$\rho'_{OB} = \rho'_O \otimes \rho'_B,$$

in which case we say the reference system B "decouples" from the environment O. Heuristically, the error can be corrected if and only if no information about the state of $\tilde{B}$ leaks to the environment O. There is also an approximate version of this statement [23], [24]. Roughly speaking (we will be more precise in Section 5.6), recovery with fidelity close to one is possible if and only if O and B are nearly uncorrelated after O and E interact.

Now consider the implications of our assumption that the Hawking radiation is pseuodorandom. As stated in Section 5.1, the marginal state ρ_{EB} is pseudorandom if ρ_{EB} cannot be distinguished from a maximally mixed state by any circuit with size polynomial in $|H|$, apart from an error exponentially small in $|H|$. We will show in Section 5.6 that, assuming $|O| \ll |H|$, if ρ_{EB} is pseudorandom and $U_{\mathcal{E}}$ is any polynomial-size unitary transformation, then O and B approximately decouple up to an error exponentially small in $|H|$. Therefore, apart from an exponentially small error, a computationally bounded observer O is unable to inflict an uncorrectable error on the encoded system $\tilde{B}$.

We can make a stronger assertion: it is possible to choose the logical operators acting on the encoded system to be robust *ghost operators*, which (acting on the code

space) nearly commute with any operation applied by the computationally bounded observer O. Returning now for simplicity to the setting of exact correctability, we claim that if the error induced by $U_\mathcal{E}$ is correctable, then for any operator $\tilde{T}_{\tilde{B}}$ acting on system $\tilde{B}$, it is possible to choose the corresponding logical operator T_{EH} satisfying Equation (5.2) such that

$$T_{EH} U_\mathcal{E} \left(I_O \otimes V_\Psi \right) = U_\mathcal{E} T_{EH} \left(I_O \otimes V_\Psi \right).$$

In this sense, the correctable errors have no effect on the ghost logical algebra. This claim is a special case of a more general statement about operator algebra quantum error correction (OAQEC) [25], [26]. Since we do not expect a black hole to provide an exact error-correcting code, we will need to analyze the case of approximate quantum error correction. Unfortunately, it does not seem straightforward to generalize the results of [25], [26] to the approximate setting. Instead, we present a self-contained construction of exact ghost logical operators in Section 5.8.1, without making direct use of known results from the theory of OAQEC, and then generalize the construction to the approximate setting in Section 5.8.2.

We will apply the approximate version of this result to the situation where $U_\mathcal{E}$ induces an approximately correctable error, thus inferring that the logical operators acting on the encoded system $\tilde{B}$ may be chosen so that they nearly commute with the actions of the computationally bounded observer O. We propose that these robust ghost operators are the logical operators acting on the black hole interior, and conclude that the interior is very well protected against the actions of any realistic observer who resides outside the black hole.

The statements about the indistinguishability of ρ_{EB} from a maximally mixed state, the decoupling of O from B, the correctability of $U_\mathcal{E}$, and the commuting action of $\tilde{T}_{EH}$ and $U_\mathcal{E}$ on the code space, are all approximate relations with exponentially small corrections. Therefore, we need to be mindful of these corrections in constructing our arguments. Fortunately, many relevant features of approximate quantum error-correction have been previously studied, and we make use of results from [23], [24] in particular.

For the general argument sketched above we have assumed that the observer system satisfies $|O| \ll |H|$. But it is also instructive to consider a different scenario, in which the observer has access to an auxiliary probe system P. We now imagine that the probe P, which might have a size comparable to or larger than E, is prepared in a simple initial state and then interacts efficiently with E. After this interaction

between E and P, the observer (still satisfying $|O| \ll |H|$), interacts with EP, performing an efficient quantum computation that may be chosen adversarially. In this case, too, we can show under the same pseudorandomness assumption as before that the reference system B decouples from O, and that robust ghost logical operators can be constructed. For example, the probe might cause all of the qubits of E to dephase in a preferred basis, but the entanglement between $\tilde{B}$ and B would still be protected. The modification from the previously considered case is that now $\tilde{B}$ will be encoded in EHP rather than EH, and we conclude that the encoded black hole interior remains inaccessible to any computationally bounded observer O who examines the radiation and the probe, as long as the size of the observer's memory satisfies $|O| \ll |H|$.

Our conclusion that $\tilde{B}$ is difficult to decode or manipulate follows from the pseudorandomness of the Hawking radiation if the observer is computationally bounded and has access only to the radiation system E outside the black hole. But we might imagine that an observer who jumps into the black hole has access to the black hole degrees of freedom H as well as E. We show in Section 5.9 that an observer who has access to EH can efficiently manipulate and decode $\tilde{B}$, assuming only that the state $|\Psi\rangle_{EBH}$ was created by an efficient unitary process. In this sense, an interior observer can interact with the interior degrees of freedom, as one might expect. A similar remark applies to the fully evaporated black hole. If the final state after complete evaporation is a highly scrambled pure state of EB, where $|B| \ll |E|$, then the maximally mixed state of B is purified by a code subspace of E. If B has constant size, then the code state can be efficiently distilled and deposited in a small quantum memory, assuming only that the map from the infalling matter to the outgoing Hawking radiation is an efficient unitary process.

If an efficient measurement of EB *can* detect the correlation between E and B, then we may expect that an observer acting on E is able to interact efficiently with the black hole interior. In Appendix 5.D, we show that, if a product observable $M_E \otimes N_B$ has an expectation value in the state $|\Psi\rangle_{EHB}$ that differs significantly from its expectation value in a maximally mixed state of EB, then there cannot be a complete set of ghost logical operators on EH commuting with M_E. It follows that, if M_E can be realized efficiently, low-complexity operations acting on the Hawking radiation *can* send a signal to the interior.

5.3 Classical Pseudorandomness

Our argument that the black hole interior is inaccessible to computationally bounded exterior observers hinges on the hypothesis that the Hawking radiation emitted by an old black hole is pseudorandom. In this section we will provide background about the concept of pseudorandomness, which some readers might find helpful.

As discussed in Section 5.1, we are interested in a black hole that is still macroscopic, but has already been evaporating for longer than its Page time [27]. The state of the previously emitted radiation system EB is purified by the black hole system H, and by this time EB is much larger than H; therefore the microscopic state ρ_{EB} of EB has far lower rank than a thermal state. It must then be possible, at least in principle, to distinguish ρ_{EB} from a thermal state. But how, operationally, would an observer outside the black hole who interacts with the radiation be able to tell the difference?

To start with, it will be instructive to consider a simple classical model that captures some of the features of this setup — after we understand how the classical model works we will be better prepared to analyze an analogous quantum model. Let us suppose that the emitted Hawking radiation is a classical bit string x of length n, which our observer is permitted to read. But this bit string is not chosen deterministically; rather, when the observer reads the radiation he actually samples from a probability distribution governing n-bit strings. We will say that the state of the black hole is "thermal" if this distribution is the uniformly random distribution $p_I(x)$, where

$$p_I(x) = \frac{1}{2^n}, \qquad \forall x \in \{0, 1\}^n.$$

But suppose that the state of the black hole is described by a distribution that is in principle almost perfectly distinguishable from the uniform distribution. Can this state "fool" the observer, leading him to believe the distribution is uniform even though that is far from the case? See Figure 5.2.

To be more concrete, let us suppose the observer is assured that he is sampling from a distribution which is either the uniform distribution $p_I(x)$, or a different distribution $p_S(x)$ which is uniform on the subset of n-bit strings S:

$$p_S(x) = \begin{cases} 2^{-\alpha n}, & x \in S \\ 0, & x \notin S \end{cases},$$

where $|S| = 2^{\alpha n}$ (with $0 < \alpha < 1$) is the number of strings contained in S. The observer samples once from the distribution, receiving x, and then executes
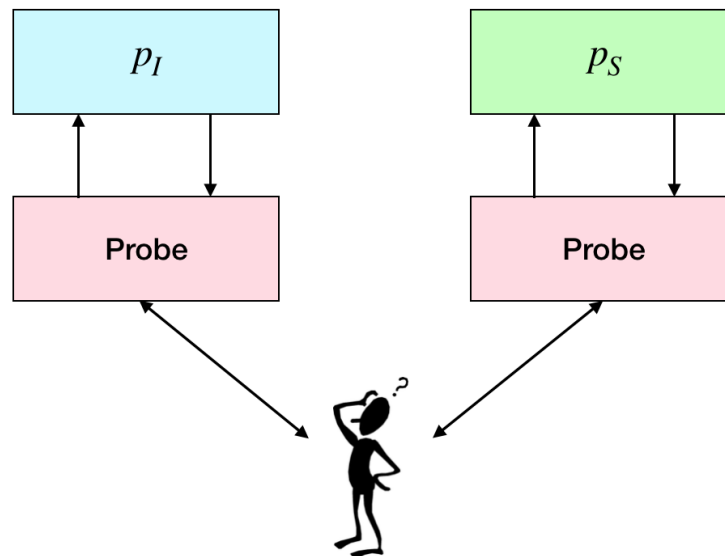
Figure 5.2: An observer samples from a distribution and attempts to decide whether the distribution is uniformly random or not.

a classical circuit C with x as an input, finally producing either the output 1 if he guesses that the distribution is p_S, or the output 0 if he guesses that the distribution is p_I.[2]

For n large, it is clear that C can be chosen so that the observer guesses correctly with a high success probability. Suppose, for example, that he outputs 1 if $x \in S$ and he outputs 0 if $x \notin S$. It the distribution is actually p_S, this guess is correct with probability 1. If the distribution is actually p_I, then the guess is correct unless x happens to lie in S "by accident," which occurs with probability $2^{-(1-\alpha)n}$. Therefore, for fixed α and large n, the probability of an incorrect guess is exponentially small in n.

However, depending on the structure of the set S, the circuit C that distinguishes p_S and p_I might need to be quite complex, making this strategy impractical if the observer has limited computational power. Suppose, for example, that the observer is unable to perform a computation with more than $\Lambda(n)$ gates, where $\Lambda(n)$ grows subexponentially with n — that is, $\Lambda(n) \leq \exp(f(n))$ where $f(n)$ scales sublinearly with n. Then we can show that the set S can be chosen such that this computationally bounded observer has only an exponentially small chance of distinguishing p_S and p_I; that is, his probability of guessing the distribution correctly is no better than

---

[2]The conclusion we reach below would not change much if he were permitted to sample from the distribution a number of times polynomial in n.

$1/2 + 2^{-cn}$, where c is a positive constant. In that case, we say that the distribution p_S is *pseudorandom*.

To show that such a pseudorandom distribution p_S exists, we argue in two steps. In the first step, we consider some fixed circuit C, and denote by L_C the set of n-bit input strings for which C outputs 1 (we say that C "accepts" the strings in L_C). If the input x is chosen by sampling from $p_I(x)$, then C accepts x with probability

$$P_C(I) = \frac{|L_C|}{2^n},$$

while if x is chosen by sampling from $p_S(x)$, then C accepts x with probability

$$P_C(S) = \frac{|S \cap L_C|}{2^{\alpha n}}.$$

Now suppose that S is chosen randomly from among all subsets of n-bit strings with cardinality $|S| = 2^{\alpha n}$. We can envision the possible strings as 2^n balls, of which the balls accepted by C are colored white, and the balls rejected by C are colored black, while S is a random sample containing $|S|$ of these balls. Suppose that the white balls constitute a fraction f of all the balls. Then, for n large, we expect that S also contains a fraction of white balls which is close to f. This intuition can be made precise using Hoeffding's inequality, from which we derive

$$\Pr\left(|P_C(S) - P_C(I)| \geq \epsilon\right) \leq e^{-2|S|\epsilon^2}, \tag{5.1}$$

where the probability is evaluated for the uniform distribution over all subsets with $|S|$ elements. Now we can choose ϵ to have the exponentially small value $\epsilon = |S|^{-1/4}$ (for example) to see that, if S is sampled uniformly with $|S|$ fixed, the probability that C accepts a sample from p_S is exponentially close to the probability that C accepts a sample from p_I. We conclude that, not only is it possible to choose the subset S such that the fixed circuit C can barely distinguish p_S from p_I, but furthermore most choices for S with $|S| = 2^{\alpha n}$ have this property.

We have now completed the first step in our two-step argument. But so far we have only shown that S can be chosen such that p_S and p_I are hard to distinguish for one fixed circuit C. We wish to make a much stronger claim, that there is a choice for S such that p_S and p_I are nearly indistinguishable by *any* circuit with a number of gates subexponential in n.

To prove this stronger claim, we proceed with the second step in the argument. For a collection of circuits $\{C_1, \ldots, C_N\}$, what is the probability that $|P_{C_i}(S) - P_{C_i}(I)| \geq$

ϵ, for at least one i? An upper bound on this probability follows from the union bound, which asserts that

$$P(A_1 \cup \cdots \cup A_N) \leq \sum_{i=1}^{N} P(A_i),$$

where $\{A_1, A_2, \ldots, A_N\}$ is any set of events. Using Equation (5.3), we conclude that the probability that at least one of the N circuits distinguishes p_S from p_I with probability at least ϵ is no larger than $Ne^{-2|S|\epsilon^2}$.

How many possible circuits are there which act on the n-bit input x and contain m computation steps? In each step of the computation, we either input one of the bits of x or we execute a gate which is chosen from a set of G possible gates, where G is a constant. Our claim will hold if each gate in G has a constant number of input and output bits, so for simplicity let us assume that each gate has at most two input bits and generates a single output bit (like a NAND gate for example). Each two-bit gate acts on a pair of bits which are outputs from previous gates; this pair can be chosen in fewer than m^2 ways. Therefore, the total number $N(m)$ of size-m circuits can be bounded as

$$N(m) \leq \left((n+G)m^2 \right)^m,$$

which implies

$$\log N(m) \leq m \left(2 \log m + \log(n+G) \right).$$

Even if we choose an exponentially large circuit size $m = 2^{\gamma n}$ and an exponentially small error $\epsilon = 2^{-\delta n}$, we find that $N(m)e^{-2|S|\epsilon^2}$ is doubly exponentially small in n for $|S| = 2^{\alpha n}$ and $\alpha > \gamma + 2\delta$. Hence, if S is randomly chosen, it's extremely likely that the distributions p_S and p_I are indistinguishable by circuits of size $2^{\gamma n}$, up to an exponentially small error.

To summarize, we have shown that the set S can be chosen so that the probability distribution p_S has these properties:

1. Its entropy per bit α is a positive constant less than 1.

2. It is statistically distinguishable from p_I with an exponentially small failure probability.

3. If $\gamma < \alpha$, any circuit using at most $2^{\gamma n}$ gates almost always fails to distinguish p_S and p_I.

If n is macroscopic, the task of distinguishing p_S from p_I can be absurdly difficult, even if the entropy density α is quite small. Suppose, for example, that $n = 10^{23}$ is comparable to Avogadro's number, and $\alpha = 10^{-12}$. Choosing $\gamma = \delta = 10^{-13}$, we conclude that a circuit with $m = 2^{10^{10}}$ gates can distinguish p_S from p_I with a success probability no larger than $\epsilon = 2^{-10^{10}}$. Even if we could perform one gate per unit of Planck time and Planck volume, an unimaginably large spacetime region would be required to execute so large a circuit.

The existence of pseudorandom distributions was first suggested by Yao [28], and the construction we have described was discussed by Goldreich and Krawcyzk [29]. In our analysis, we have assumed that the observer executes a deterministic circuit, but it turns out that giving the observer access to a random number generator does not make his task any easier [29].

Up until now, we assumed that the observer performs a computation whose output is a single bit. But what if he obtains a k-bit output instead? Can we choose the set S so that for all circuits of bounded size, the probability distribution governing the k output bits is very similar for input strings drawn from p_S and p_I? Our previous reasoning does not have to be modified much to handle this case. Now for the fixed circuit C, we denote by $L_C[y]$ the set of n-bit input strings for which C outputs the k-bit string y, and we denote by $P_C(I)[y]$, $P_C(S)[y]$ the probability that C outputs y when receiving as input a sample from p_I and p_S, respectively. Now we envision the n-bit input strings as balls which can be colored in 2^k possible ways, corresponding to the 2^k possible values of the output y. Applying the previous argument to each color, we find that when S is chosen at random from among all subsets with cardinality $|S|$,

$$\text{Prob}\left[|P_C(S)[y] - P_C(I)[y]| \geq \epsilon\right] \leq e^{-2|S|\epsilon^2},$$

for each output y. From the union bound, the probability that $|P_C(S)[y] - P_C(I)[y]|$ exceeds ϵ for at least one value of y is bounded above by $2^k e^{-2|S|\epsilon^2}$, which also provides an upper bound on the probability that the total variation distance between $P_C(S)$ and $P_C(I)$ exceeds $\epsilon' = 2^k \epsilon$. Therefore the total variation distance will be no larger than ϵ' with high probability as long as $2^{\alpha n} 2^{-2k} \epsilon'^2$ is large, which means ϵ' can be exponentially small in n provided $2k < \alpha n$.

We conclude that the pseudorandom input distribution and the uniformly random input distribution will yield exponentially close output distributions as long as the observer's output register is small compared to the entropy of S.

The preceding argument shows that, indeed, there are probability distributions which are computationally indistinguishable from the uniformly random distribution. But can we make such a distribution *efficiently*? It turns out that our argument for the computational hardness of distinguishing a pseudorandom distribution from a uniformly random distribution can be used to show that such distributions are typically hard to produce with polynomial-sized circuits.[3] However, there are distributions that can be created using polynomial-sized circuits which, under reasonable complexity-theoretic assumptions, are difficult to differentiate from the uniformly random distribution, for any polynomial-sized circuit. Such distributions can be generated by pseudorandom generators [30]. We will give a more complete description of such constructions for the quantum case in Section 5.4.

Efficient sampling from a (classical) pseudorandom distribution is analogous to the formation and partial evaporation of a black hole. Pseudorandom number generators consult a random "key" which is hidden from the adversary, and then compute a function which depends on the key. This function is chosen so that an output drawn from the resulting family of outputs indexed by the key is computationally indistinguishable from the output of a truly random function. In the case of the partially evaporated black hole, the key becomes a black hole microstate, and the key-dependent function evaluation becomes the chaotic unitary evolution of the evaporating black hole. An adversary samples the Hawking radiation, and attempts to determine whether the sample is drawn from a thermal distribution or not.

To properly discuss the evaporating black hole, we will need to consider the quantum version of pseudorandomness, to which we turn in the next two sections. But our simplified classical model of "Hawking radiation" is instructive. It teaches us that the (classical) adversary can interact with the (classical) radiation for a subexponential time (or even for the exponential time $2^{\gamma n}$ if γ is sufficiently small), without ever suspecting that the radiation is far from uniformly random. On the other hand, that conclusion may no longer apply if the adversary collects k bits of information where k is sufficiently large ($k > \alpha n/2$). Both of these features will pertain to the quantum version of our story.

5.4   Quantum Pseudorandomness

Now consider the quantum version of the task described in the Section 5.3. Our observer receives a quantum state ρ, and is challenged to guess whether ρ is maxi-

---

[3]We thank Adam Bouland for emphasizing this point.

mally mixed or not. For that purpose, he performs a quantum computation with ρ as input, and he outputs a single bit: 0 if he guesses ρ is maximally mixed, and 1 otherwise.

Following our analysis of the classical case, let us first suppose that the observer executes a particular fixed quantum circuit. That means the observer measures a particular Hermitian observable A with unit operator norm. Suppose we try to fool the observer by providing as input a pure state $\rho = |\psi\rangle\langle\psi|$. How well can the observable A distinguish this pure state from the maximally mixed state?

Suppose that $|\psi\rangle$ is chosen uniformly at random from among all n-qubit pure states. Then Levy's lemma [31] says that

$$\Pr\left(\left|\langle\psi|A|\psi\rangle - \frac{\text{Tr}(A)}{2^n}\right| \geq \epsilon\right) \leq e^{-c\,2^n\epsilon^2} \tag{5.2}$$

for some constant c, where the probability is evaluated with respect to the invariant Haar measure on the n-qubit Hilbert space. This means that, for n large, the pure state $|\psi\rangle$ can be chosen so that $|\psi\rangle$ and the maximally mixed state are exponentially difficult to distinguish using the observable A. Furthermore, most pure states have this property. Even a pure quantum state can pretend to be maximally mixed, and the observer will not know the difference![4]

As in the classical case we can strengthen this claim: the state $|\psi\rangle$ can be chosen so that $|\psi\rangle$ is hard to distinguish from the maximally mixed state not just for one fixed quantum circuit, but for *any* quantum circuit of reasonable size. To carry out this step of the argument, we will need an upper bound on the number of quantum circuits of specified size; here we confront the subtlety that quantum circuits, unlike classical ones, form a continuum, but this wrinkle poses no serious obstacle to completing the argument. If we settle for specifying the unitary transformation realized by a circuit with m gates to constant accuracy, it suffices to specify each gate to $O(\log m)$ bits of precision. Therefore, as in the classical case, the complete circuit can be specified by $O(m\log m)$ bits. It follows that, if m is subexponential in n, then the number $N(m)$ of circuits with size m is the exponential of a function which is subexponential in n. In contrast, the right-hand side of Equation (5.4) is the exponential of an exponential function of n. Using the union bound, we conclude that if the pure state $|\psi\rangle$ is chosen uniformly at random, it will, with high probability,

---

[4]If two identical copies of $|\psi\rangle$ are available, then it is easy to distinguish the pure state $|\psi\rangle$ from the maximally mixed state by conducting a swap test. Here we assumed that only a single copy is available.

be hard to distinguish $|\psi\rangle$ from the maximally mixed state using *any* circuit of size subexponential in n.

On the other hand, if we were not concerned about the complexity of the observer's task, then it would be easy to distinguish $|\psi\rangle$ from the maximally mixed state. The observer could perform a projective measurement with the two outcomes $\{E_0 = I - |\psi\rangle\langle\psi|,\ E_1 = |\psi\rangle\langle\psi|\}$, guessing that the input state is $|\psi\rangle$ if he obtains the outcome E_1, and guessing that the input state is maximally mixed if he obtains the outcome E_0. This strategy always succeeds if the input is $|\psi\rangle$, and fails with the exponentially small probability 2^{-n} if the input is maximally mixed. The trouble is that, for a typical pure state $|\psi\rangle$, this measurement is far too complex to carry out in practice.

A typical pure quantum state is somewhat analogous to the distribution p_S we described in Section 5.3. In both cases, it is hard for an observer who is limited to performing polynomial-size computations to tell that the state is not uniformly random, even though an observer with unlimited computational power can tell the difference. Furthermore, both examples are subject to the same criticism — it is computationally hard to sample uniformly from Haar measure (that is, to prepare a "typical" pure state), just as it is computationally hard in the classical setting to sample from the distribution p_S. In the quantum setting, as for the classical setting, we may ask a more nuanced question: can quantum states be prepared *efficiently* which are hard to distinguish from maximally mixed states? This more nuanced question is the relevant one as we contemplate the properties of the radiation emitted by a partially evaporated black hole, because the formation and subsequent complete evaporation of a black hole can occur in a time that scales like $S_{\text{bh}}^{3/2}$, where S_{bh} is the initial black hole entropy. Hence, the preparation of the Hawking radiation can be simulated accurately by an efficient quantum circuit.

The answer is yes (under a reasonable assumption), as was shown recent by Ji, Liu, and Song [12]; pseudorandom quantum states *can* be prepared efficiently. The assumption we need is the existence of a family of *quantum-secure pseudorandom functions* $\{\text{PRF}_k\}_{k \in K}$. This means that each PRF_k can be efficiently computed, but it is difficult to distinguish a randomly sampled member of $\{\text{PRF}_k\}$ from a truly random function with any efficient quantum algorithm. The set K is called the *key space* of the function family. The existence of such pseudorandom functions follows from the existence of quantum-secure one-way functions, an assumption which is standard in cryptography.

The key idea is that we can construct a pseudorandom quantum state as a superposition of computational basis states, where all basis states appear with equal weight except for a phase, and the phases appear to be random to a computationally bounded observer. Specifically, we consider a family of states $\{|\phi_k\rangle\}_{k \in K}$

$$|\phi_k\rangle = \frac{1}{\sqrt{N}} \sum_{x \in X} \omega_N^{\mathrm{PRF}_k(x)} |x\rangle,$$

where $N = 2^n$, $\omega_N = e^{2\pi i/N}$, $X = \{0, 1, 2, \ldots, N-1\}$, and $\{\mathrm{PRF}_k : X \to X\}_{k \in K}$ is a family of quantum-secure pseudorandom functions. We can show that a uniform mixture of the states $\{|\phi_k\rangle\}$ is computationally indistinguishable from the maximally mixed state.[5]

We may argue as follows. First ,we consider the family of *all* functions $f_{k'} : X \to X\}_{k' \in K'}$ indexed by key space K', and the corresponding family of pure states

$$|f_{k'}\rangle = \frac{1}{\sqrt{N}} \sum_{x \in X} \omega_N^{f_{k'}(x)} |x\rangle. \tag{5.3}$$

The first thing to note is that $\{|f_{k'}\rangle\}$ is information-theoretically indistinguishable from Haar-random; we state this fact for the reader's convenience in Lemma 5.4.1.

Lemma 5.4.1 ([12], Lemma 1). *Let $\{|f_{k'}\rangle\}$ be the family of states defined in Equation (5.4). Then, for m polynomial in n, the state ensemble $\{|f_{k'}\rangle^{\otimes m}\}$ is statistically indistinguishable from the ensemble $\{|\psi\rangle^{\otimes m}\}$ where $|\psi\rangle$ is Haar-random, up to a negligible error.*

Furthermore, the ensemble $\{|\phi_k\rangle\}$ cannot be efficiently distinguished from the ensemble $\{|f_{k'}\rangle\}$. If it could be, then we could leverage this fact to efficiently distinguish $\{\mathrm{PRF}_k\}$ from a family of random functions [12], contradicting our assumption that $\{\mathrm{PRF}_k\}$ is a quantum-secure pseudorandom function family. It now follows that the ensemble $\{|\phi_k\rangle\}$ cannot be efficiently distinguished from a maximally mixed state.

So far, we have shown that a uniform mixture of the states $\{|\phi_k\rangle\}$ is pseudorandom; it remains to show that this mixture can be prepared efficiently. We start with a product of qubits, each in the state $|0\rangle$, and apply a Hadamard gate to each qubit to obtain the state

$$\frac{1}{\sqrt{N|K|}} \sum_{x \in X} \sum_{k \in K} |x\rangle |k\rangle.$$

---

[5]In fact, we can simplify the construction. It was shown in [32] that the same family of states is still pseudorandom if we replace the root of unity ω_N by -1.

Next, we apply the quantum Fourier transform (which has complexity polynomial in n) to another n-qubit register that is initialized in the state $|00\ldots01\rangle$, obtaining

$$\frac{1}{N\sqrt{|K|}} \sum_{x,y\in X} \sum_{k\in K} |x\rangle|k\rangle\omega_N^y|y\rangle.$$

Now, we compute $\mathrm{PRF}_k(x)$ and subtract modulo N from the y register. This computation can be done efficiently because by assumption the PRF_k is an efficiently computable function; the resulting state is

$$\frac{1}{N\sqrt{|K|}} \sum_{x,y\in X} \sum_{k\in K} |x\rangle|k\rangle\omega_N^y|y - \mathrm{PRF}_k(x)\rangle.$$

After shifting the summation index y, we have, up to a global phase,

$$\frac{1}{N\sqrt{|K|}} \sum_{x,y\in X} \sum_{k\in K} \omega_N^{\mathrm{PRF}_k(x)}|x\rangle|k\rangle\omega_N^y|y\rangle = \frac{1}{\sqrt{|K|}} \sum_{k} |\phi_k\rangle|k\rangle|\mathrm{QFT}\rangle,$$

where $|\mathrm{QFT}\rangle = N^{-1/2} \sum_{y\in X} \omega_N^y|y\rangle$. After the key $|k\rangle$ is discarded, the marginal state over the first register is the uniform mixture of $\{|\phi_k\rangle\}$. Thus, we have prepared this mixture efficiently.

The definition of a pseudorandom quantum state in reference [12] is really overkill for our purposes. Those authors are concerned with cryptographic applications, and therefore consider a definition (as stated in Lemma 5.4.1) where, for each value k of the key, m identical copies of $|\phi_k\rangle$ are available where m is polynomial in n. We will not encounter such scenarios in this paper. Therefore, we may instead adopt a simplified definition of pseudorandomness which is more suitable for the application to black hole physics. In Definition 5.6.1 below, the size $|H|$ of the remaining black hole parametrizes how difficult it is to distinguish radiation emitted from the partially evaporated black hole from the maximally mixed state. In this sense, the remaining black hole H serves as the key space of the pseudorandom radiation state. Even if $|H|$ is less than half of the initial black hole entropy, this task remains difficult so long as the remaining black hole H is macroscopic. Our hypothesis that the Hawking radiation is pseudorandom provides a way to formalize the idea that the Hawking radiation is effectively thermal even when the state of E has relatively low rank because $|H| \ll |E|$.

5.5 Is Hawking Radiation Pseudorandom?

We have now seen, in both the classical and quantum settings, that pseudorandom states exist. Though in principle these states are almost perfectly distinguishable

from maximally mixed states, in practice no observer with reasonable computational power can tell the difference. Moreover, under standard cryptographic assumptions, there exist constructions of such states which can be efficiently prepared. But up to this point, we have not addressed whether pseudorandom quantum states can be efficiently prepared in plausible physical processes like the evaporation of a black hole.

In the case of a black hole which forms from gravitational collapse, and then *completely* evaporates, the resulting state of the emitted Hawking radiation, though highly scrambled, would *not* be pseudorandom. We take it for granted that the time evolution of the quantum state can be accurately approximated by a quantum circuit, which has size polynomial in the initial black hole entropy S_{bh} because the evaporation process takes a time $O\left(S_{bh}^{3/2}\right)$. We may consider a toy model of this process, in which the initial state $|\phi_{matter}\rangle$ of the collapsing matter is a product state of n qubits $|\phi_{matter}\rangle = |0\rangle^{\otimes n}$, and the final state after complete evaporation is $|\Psi_{fin}\rangle = U|\phi_{matter}\rangle$, where U is a unitary transformation constructed as a polynomial-size circuit. In this case, an observer could just execute this circuit in reverse, hence applying $U^\dagger$ to $|\Psi_{fin}\rangle$, and then measure the qubits in the standard basis, thus easily distinguishing $|\Psi\rangle$ from the maximally mixed state. We see that, if ρ is a state that can be prepared by a polynomial-size quantum circuit, yet is hard to distinguish from maximally mixed by polynomial-size circuits, then ρ cannot be pure.

Instead, we consider a *partially* evaporated black hole as in Figure 5.3. We imagine that the n-qubit state ρ_{EB} is prepared by applying a polynomial-size unitary circuit U_{bh} to the initial state $|0\rangle^{\otimes n}|0\rangle^{\otimes k}$ of EBH, where H is a k-qubit system, and then discarding H. In our toy model, EB is the Hawking radiation that has been emitted so far, and H is the remaining black hole. (Recall that B is a small portion of the emitted Hawking radiation whose properties we will investigate later; for the purpose of the present discussion, we are only interested in the state of EB, the full radiation system.) If our observer had access to H as well as EB, he could easily tell that the state is not maximally mixed, but what if H is inaccessible?

We are particularly interested in the case where $1 \ll k = |H| < n = |EB|$, so that ρ_{EB} fails to have full rank, and must therefore be information-theoretically distinguishable from the maximally mixed state; this situation resembles the classical model discussed in Section 5.3, where the entropy of the distribution p_S is substantial but not maximal. Could the state ρ_{EB} of the Hawking radiation, which is prepared by unitary evolution of EBH for a time which is polynomial in $|EB|$, be pseudorandom?

$$E \quad H \quad B \qquad\qquad E \quad H \quad B$$

$$\Psi \quad\equiv\quad U_{\mathrm{bh}}$$
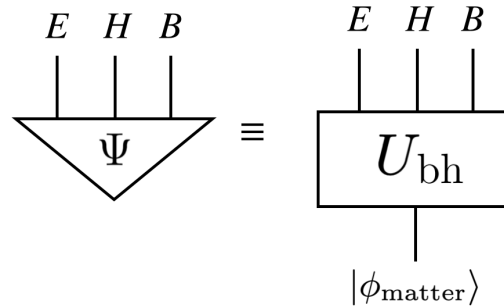
$$|\phi_{\mathrm{matter}}\rangle$$

Figure 5.3: Our toy model of a partially evaporated black hole, where EB is the Hawking radiation emitted so far, and H is the remaining black hole. The initial state $|\phi_{\mathrm{matter}}\rangle$ of the gravitationally collapsing matter is modeled as a product state. We conjecture that the unitary black hole dynamics prepares a pseudorandom state of EB.

This is a question about quantum gravity, and we do not know the answer for sure, but we can make a reasonable guess. We have already seen in Section 5.4 that quantum circuits exist that efficiently prepare pseudorandom quantum states. Since black holes are believed to be particularly potent scramblers of quantum information, it is natural to conjecture that the internal dynamics of a black hole can produce pseudorandom states as well. Indeed, we may expect similar behavior for the radiation emitted by other strongly chaotic quantum systems aside from black holes. Only for the case of black holes, though, where we face the daunting firewall puzzle, will our constructions of robust logical operators acting on EH seem to have a natural interpretation.

To better understand why the state of EB might be hard to distinguish from a maximally mixed state, we may suppose, for example, that ρ_H is maximally mixed so that the pure state of EBH has the form

$$|\Psi\rangle_{EBH} = \frac{1}{2^{|H|/2}} \sum_i |\psi_i\rangle_{EB} \otimes |i\rangle_H,$$

where the states $\{|\psi_i\rangle_{EB}\}$ are orthonormal. The marginal state of EB is then

$$\rho_{EB} = \frac{1}{2^{|H|}} \sum_i |\psi_i\rangle\langle\psi_i|.$$

Suppose the observer receives a state which is either ρ_{EB} or the maximally mixed state $\sigma_{EB} = I_{EB}/2^{|EB|}$. A natural test is as follows: The observer augments EB with the maximally mixed state of H (which is easy to prepare), and then measures the projection onto $|\Psi\rangle$. This can be done efficiently by applying U_{bh}^{-1} and then

measuring in the standard basis. If the input state is σ_{EB}, the projection onto $|\Psi\rangle$ succeeds with probability $2^{-(|EB|+|H|)}$, while if the input state is ρ_{EB}, the success probability is

$$\langle\Psi|\frac{\sum_i |\psi_i\rangle\langle\psi_i|}{2^{|H|}} \otimes \frac{I_H}{2^{|H|}}|\Psi\rangle = \frac{1}{2^{2|H|}}.$$

Thus this test distinguishes ρ_{EB} and σ_{EB}, but only with a probability that is exponentially small in H.

To conduct a better test, we would somehow need to exploit the structure of the ensemble $\{|\psi_i\rangle_{EB}\}$. But if, as we expect, black holes are especially effective information scramblers, it is reasonable to suppose that the ensemble lacks any special properties that can be exploited by an observer who is limited to performing a polynomial-time quantum computation. If so, the Hawking radiation is pseudorandom, and the test we have described may be nearly optimal.

In the example above, we have assumed that the radiation has infinite temperature. The actual behavior of a black hole evaporating in asymptotically flat spacetime is more complicated — the temperature is actually finite, and in fact becomes hotter and hotter as the evaporation proceeds. Conceptually, though, the situation is similar to the idealized case of an black hole evaporating at infinite temperature. At early times, when $|H| \gg |EB|$, we expect the radiation emitted at a specified time to be information-theoretically indistinguishable from precisely thermal radiation at the same temperature. At late times, when $|H| \ll |EB|$, the global state of the radiation is distinguishable in principle from a thermal state (with temperature varying according to the time of emission), but we assume that telling the difference is computationally hard because the radiation is highly scrambled.

We also note that the constructions in [12] reinforce earlier observations concerning the computational hardness of *decoding* the Hawking radiation [14], [15]. These authors considered the quantum state $|\Psi\rangle_{EBH}$ of an old black hole, and analyzed the task of extracting from the early radiation E the subsystem which is entangled with the recently emitted Hawking mode B. This task would be easy for an observer who has access to both E and H, but one can argue that there are efficiently preparable states of EBH for which this decoding task cannot be achieved by an observer who performs a polynomial-size quantum computation on E alone. Here, too, the hardness of decoding cannot be proven from first principles, but it follows from plausible complexity assumptions which are standard in "post-quantum" cryptography [14], [15]. Again, the existence of states that are hard to decode does not guarantee that a black hole creates such states, but we take it on

faith that if efficient preparation of such states is possible, then a black hole will be up to the job.

To summarize, on the basis of these (admittedly speculative) considerations, we propose that for the quantum state $|\Psi\rangle_{EBH}$ of an old black hole, the state ρ_{EB} of the Hawking radiation is pseudorandom. If $|H| < |EB|$, then the rank of ρ_{EB} is not maximal, so that ρ_{EB} is distinguishable from a thermal state. In fact, an observer with access to H as well as EB could efficiently check that ρ_{EB} is not thermal. Furthermore, an observer without access to H could check that ρ_{EB} is not thermal by performing a quantum computation of exponential size on EB alone. But an observer outside the black hole, who performs a polynomial-size quantum computation on EB without access to H, will be able to distinguish ρ_{EB} from a thermal state with a success probability that is at best exponentially small in $|H|$. Our analysis of the robustness of the encoded black hole interior in the remainder of this paper will rest on this assumption.

This discussion highlights the importance of distinguishing the von Neumann entropy of the Hawking radiation from its thermodynamic entropy. After the Page time, the Von Neumann entropy of EB becomes far smaller than the von Neumann entropy of a perfectly thermal state, so one could in principle verify that the Hawking radiation is not perfectly thermal by measuring its von Neumann entropy. The existence of pseudorandom quantum states then implies that measuring the von Neumann entropy with a small error requires an operation of superpolynomial complexity [33]. One could imagine trying to measure the entropy of the radiation by, for example, withdrawing its thermal energy to operate a heat engine. If the radiation is pseudorandom, though, the radiation would be indistinguishable from thermal radiation in any efficient process, despite its low von Neumann entropy.

Recalling the construction of the pseudorandom state recounted in Section 5.4, we note [34] that the quantum Fourier transform can be executed with circuit depth $O(\log n)$, and that under plausible cryptographic assumptions the function PRF_k can be computed in depth polylog n [12]. Thus a pseudorandom state can be prepared in polylog n time. Plausibly, the state preparation can be achieved in a time comparable to the $O(\log n)$ scrambling time of a black hole, as one might naively expect.

5.6 Pseudorandomness and Decoupling

In this section, we formalize our hypothesis that Hawking radiation is pseudorandom, and explore its implications regarding the firewall paradox [3]. Our analysis can be viewed as a refinement of the Harlow-Hayden argument [14], [15].

As formulated in [3] and summarized in Section 5.1, the firewall paradox highlights a conflict between the unitarity of black hole evaporation and the monogamy of entanglement. A possible resolution is that the interior mode $\tilde{B}$ that purifies a recently emitted Hawking mode B may actually be encoded in the radiation. On the face of it, this resolution flagrantly violates locality, and one wonders whether this violation of locality can be detected by an agent who first interacts with the radiation and then falls through the event horizon to visit the interior. We will argue that, provided the Hawking radiation is pseudorandom and the size of the observer is small compared to the black hole, the nonlocality is undetectable in practice because it would take an exponentially long time for the observer to distill the encoded interior mode before falling into the black hole.

We will first present a sketch of the argument in a simplified setting where the radiation interacts with a single observer who is significantly smaller than the remaining black hole. Later on we will extend the argument to the case where the observer has access to a large probe outside the horizon, whose size may be comparable to or even larger than the remaining black hole.

Recall our conventions: let O denote the observer, H the remaining black hole, B the late outgoing mode, E the early radiation, and P the external probe. We will also refer to the joint system EB as the exterior radiation. All subsystems can be decomposed in terms of qubits, and our statements about computational complexity concern the number of steps in a computation executed by a universal quantum computer. Below and throughout the remainder of the paper, given an operator A, we will use $\|A\|_1 = \mathrm{Tr}(\sqrt{A^\dagger A})$ to denote the trace norm, $\|A\|_F = \sqrt{\mathrm{Tr}(A^\dagger A)}$ to denote the Frobenius norm, and $\|A\|$ to denote the operator norm.

First, we define what it means for the external radiation of the black hole to be pseudorandom.

Definition 5.6.1. *Let $|\Psi\rangle_{EBH}$ be the state of the black hole and the radiation. Let $\sigma_{EB} = I_{EB}/d_{EB}$ be the maximally mixed state of EB, and let $\rho_{EB} = \mathrm{Tr}_H (|\Psi\rangle\langle\Psi|)$. We say that the state $|\Psi\rangle_{EBH}$ is pseudorandom on the radiation EB, if there exists*

some $\alpha > 0$ such that

$$|\Pr\left(\mathcal{M}(\rho_{EB}) = 1\right) - \Pr\left(\mathcal{M}(\sigma_{EB}) = 1\right)| \leq 2^{-\alpha|H|}, \tag{5.4}$$

for any two-outcome measurement $\mathcal{M}$ with quantum complexity polynomial in $|H|$, the size of the remaining black hole.

This definition captures the notion that no feasible measurement can tell the difference between ρ_{EB} and the maximally mixed state. A few remarks will help to clarify the definition. (1) When we say a measurement of ρ_{EB} has polynomial quantum complexity, we mean it can be performed by executing a quantum circuit of polynomial size acting on EB, followed by a qubit measurement in the standard computational basis. Use of ancilla systems is also permitted in the measurement process, provided the ancilla is initialized in a product state. (2) Of particular interest is the value of the constant α that makes the bound in Equation (5.6.1) tight for asymptotically large black holes. But because black holes are such effective information scramblers, we would expect a comparable value of α to apply also for black holes of moderate size. There is no obvious small parameter in the problem that would lead us to expect α to be small compared to 1. (3) This definition is appropriate for the case where the Hawking radiation has infinite temperature. As we remarked in Section 5.5, we expect the realistic case of finite-temperature radiation to be conceptually similar, and for similar conclusions to apply in that case. But we will stick with the infinite-temperature case for the rest of the paper to simplify our analysis.[6]

Let us now deduce a consequence of Definition 5.6.1. We introduce an observer subsystem O initialized in a state ω_O, and an ancilla subsystem P initialized in the product state $|0\rangle_P$. The main result of this section is the following: suppose that $|\Psi\rangle_{EBH}$ is pseudorandom, and let ρ_{OPE} be any state of OPE obtained by applying a unitary of polynomial complexity to $\omega_O \otimes |0\rangle_P \otimes |\Psi\rangle_{EBH}$. Then the correlation between the observer and the early radiation is exponentially small in $|H|$ for any such state; *i.e.*,

$$\|\rho_{OB} - \rho_O \otimes \rho_B\|_1 \leq 6 \cdot 2^{-(\alpha|H|-|O|)}, \tag{5.5}$$

---

[6]In the finite temperature case, the entanglement between B and the rest of the system is no longer maximal. This causes an extra complication when we use the quantum error-correction technology in Section 5.7, because the encoding map V from B to EH defined by the state Ψ_{EBH} need not be exponentially close to an isometry. Instead we may replace V by the approximate isometry $V\rho_B^{-1/2}$, which slightly modifies the error bounds derived in Section 5.7 and 5.8. Similar techniques have been used in [7], [8], [10].

where we have now assumed that B is a single qubit. We will call (5.6) the *decoupling bound*, because it states that the observer O nearly decouples from the exterior radiation mode B, and therefore gains negligible information about the interior mode $\tilde{B}$ which is entangled with B. In Section 5.8 we leverage (5.6) to show that the interior mode $\tilde{B}$ can be regarded as an encoded subsystem of EH which is protected against all "low-complexity" errors, where "low-complexity" is shorthand for polynomial complexity.

Prior work [14], [15] has suggested that the decoupling bound holds when the size of the remaining black hole is an $O(1)$ fraction of the initial black hole entropy S_{bh}. However, our conclusion goes further. Even if the majority of the initial black hole has evaporated, so that $|H| \ll |EB|$, the observer O and the late radiation B remain decoupled as long as the remaining black hole H is macroscopic and the observer's system O obeys $|O| \ll |H|$.

To derive the decoupling bound, we apply the pseudorandomness assumption to the setup described in Figure 5.1. The unitary $U_{\mathcal{E}}$ is applied to the radiation, probe, and the observer. Because the evaporation time of the black hole is polynomial in its size, and $U_{\mathcal{E}}$ is applied before the evaporation is complete, we may assume that $U_{\mathcal{E}}$ is applied in a polynomial time and therefore has polynomial complexity. We also assume that the initial state of the observer ω_O is of low complexity, although this assumption is not crucial; we may take the state ω_O to be arbitrary, at the cost of a slightly weaker decoupling bound.

In order to bound the correlation between B and O, we consider a complete set of operators acting on OB. A convenient choice is the set of Pauli operators P_i acting on OB. By a Pauli operator acting on n qubits we mean a tensor product of n 2×2 Pauli matrices; there are 4^n such operators $\{P_i, i = 0, 1, 2, \ldots, 4^n - 1\}$ (where $P_0 = I$) whose phases can be chosen so that each P_i for $i \neq 0$ has eigenvalues ± 1, and the $\{P_i\}$ are orthogonal in the Froebenius norm: $\text{Tr}\left(P_i P_j\right) = 2^n \delta_{ij}$. Here n is the number of qubits in OB. Because measurement of P_i is a low-complexity two-outcome measurement, it follows from the assumption that Ψ_{EBH} is pseudorandom that

$$|\text{Tr}((\rho_{OB} - \sigma_{OB})P_i)| \leq 2^{-\alpha|H|}$$

for any Pauli operator P_i, where σ_{OB} is the state that results when the state ρ_{EB} measured by the observer is replaced by the maximally mixed state; see Figure 5.4.

To understand why Equation (5.6) follows from pseudorandomness, note that we

are modeling a measurement of EB by the observer O as a low-complexity unitary interaction between EB and O, followed by a simple measurement of the O register. Strictly speaking, then, we should allow the Pauli operator P_i to act only on O, not on OB. In effect, we are assuming that the observer's quantum memory contains $|OB|$ qubits rather than $|O|$ qubits, so that measuring a Pauli operator acting on OB is permitted. In our formulation of the pseudorandomness assumption, there is no restriction on the size of the observer's memory, only on the complexity of his operation. Therefore, assuming that the state of EB is pseudorandom, the observer's measurement will not distinguish ρ_{OB} from σ_{OB} even if the observer is permitted to measure B as well as O.

Using the completeness and orthogonality of the Pauli operators, we can bound the Frobenius distance between the two states as

$$\|\rho_{OB} - \sigma_{OB}\|_F^2 = \mathrm{Tr}\left((\rho_{OB} - \sigma_{OB})^2\right)$$
$$= 2^{-(|OB|)} \sum_i |\mathrm{Tr}((\rho_{OB} - \sigma_{OB})P_i)|^2 .$$

Because there are $4^{|OB|}$ Pauli operators, the right hand side is bounded by $2^{-2\alpha|H|}2^{|OB|}$. The trace distance is bounded by the Frobenius norm as

$$\|\rho\|_1 \leq \sqrt{\mathrm{rank}(\rho)} \, \|\rho\|_F,$$

for any operator ρ. Therefore we have

$$\|\rho_{OB} - \sigma_{OB}\|_1 \leq 2^{|OB|}2^{-\alpha|H|}, \tag{5.6}$$

because the rank of ρ_{OB} can be no larger than $2^{|OB|}$. From (5.6), one finds that

$$\begin{aligned}
\|\rho_{OB} - \rho_O \otimes \rho_B\|_1 &\leq \|\rho_{OB} - \sigma_{OB}\|_1 + \|\sigma_{OB} - \rho_O \otimes \rho_B\|_1 \\
&\leq 2^{|OB|-\alpha|H|} + \|\sigma_{OB} - \rho_O \otimes \rho_B\|_1 \\
&= 2^{|OB|-\alpha|H|} + \|\sigma_O \otimes \sigma_B - \rho_O \otimes \rho_B\|_1 \\
&\leq 2^{|OB|-\alpha|H|} + \|\sigma_O - \rho_O\|_1 + \|\sigma_B - \rho_B\|_1 \\
&\leq 3 \times 2^{(|OB|-\alpha|H|)}.
\end{aligned} \tag{5.7}$$

The first line is the triangle inequality. From the first line to the second line, we used (5.6). From the second line to the third line, we used the fact that σ_{OB} is a product state over O and B. From the third line to the fourth line, we used the fact that

$$\begin{aligned}
\|\sigma_O \otimes \sigma_B - \rho_O \otimes \rho_B\|_1 &\leq \|(\sigma_O - \rho_O) \otimes \sigma_B\|_1 + \|\rho_O \otimes (\sigma_B - \rho_B)\|_1 \\
&\leq \|\sigma_O - \rho_O\|_1 + \|\sigma_B - \rho_B\|_1,
\end{aligned} \tag{5.8}$$

where the first line of (5.6) follows from the triangle inequality, and the second from the property that tracing out a subsystem cannot increase the trace distance. To reach the last line of (5.6), we again used the property that tracing out a subsystem cannot increase the trace distance. Finally, in the case where B is a single qubit, so that $|OB| = |O| + 1$, (5.6) becomes the decoupling bound (5.6). More generally, decoupling is satisfied whenever $|OB| \ll \alpha|H|$.
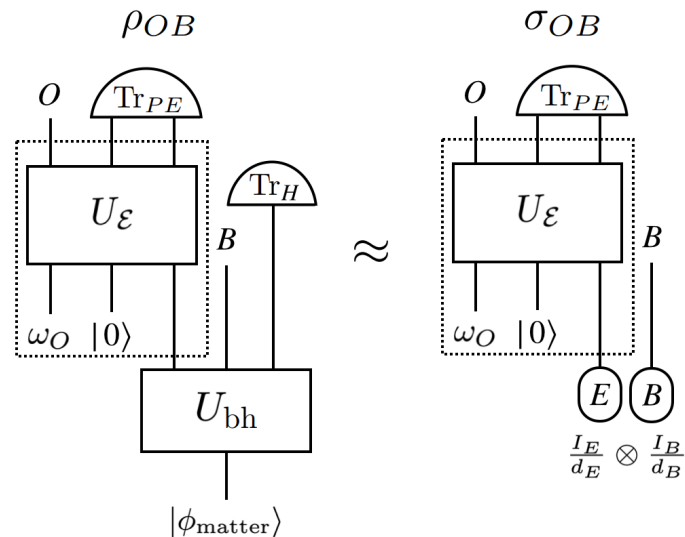
Figure 5.4: Graphical depiction of the decoupling bound, which follows from the pseudorandomness of the Hawking radiation emitted by an old black hole. On the left, a black hole forms from collapse and partially evaporates; the emitted radiation is EB and the remaining black hole is H. Then an observer O and probe P interact with the radiation subsystem E for a time that scales polynomially with the initial black hole entropy S_{bh}. On the right, the unitary transformation describing the interaction of OPE is the same as on the left, but the state of the Hawking radiation is replaced by a maximally mixed state of EB. The decoupling bound asserts that the final state of OB is the same in both cases, up to an error that is exponentially small in $|H|$, the size of the remaining black hole, provided that $|O| \ll |H|$.

Because two states close in trace distance cannot be distinguished well by any measurement, the decoupling bound implies that the state ρ_{OB} cannot be distinguished from the state σ_{OB} assuming that $|O| \ll |H|$. We thus conclude that any subsystem small compared to the remaining black hole H, even after interacting with the early radiation E, cannot be correlated with B; see Figure 5.4. In particular, an observer outside the black hole who interacts with E for a polynomially bounded time remains decoupled from B, assuming that the Hawking radiation is pseudorandom.

This conclusion about the hardness of decoding follows from the pseudorandomness

assumption for any computationally bounded observer who can access only system E. However, the decoding becomes easy if the observer has access to both E and H, as long as the state $|\Psi\rangle_{EBH}$ has polynomial complexity. For this case, we will describe an explicit decoding protocol in Section 5.9.

5.7 Black Holes as Quantum Error-Correcting Codes

In this section, we recast the findings in Section 5.6 in the language of quantum error correction. The quantum error correction point of view will prove to be useful in understanding more subtle thought experiments studied in Section 5.8. We will see that an old black hole, together with its previously emitted Hawking radiation, is a quantum error-correcting code with exotic properties that have not been noted in previous discussions of holographic quantum error-correcting codes [35], [36]. These properties hold if the Hawking radiation is pseudorandom. That a black hole can be viewed as a quantum error-correcting code is not new [17], [36]–[38]. What is new is that a black hole can protect quantum information against seemingly pernicious errors; we refer to these as "low-complexity errors," meaning errors inflicted by a malicious agent who performs a quantum computation on the Hawking radiation with complexity scaling polynomially in the size of the remaining black hole.

To explain this claim, it is useful to view the state of the black hole and the radiation as an encoding map from the interior mode $\tilde{B}$ into EH. That is, $|\Psi\rangle_{EHB}$ defines an isometric embedding of $\tilde{B}$ into EH. Recall that E denotes the early radiation, H denotes the remaining black hole, and B denotes a late outgoing mode. For simplicity, we assume that B is a single qubit, but the following results remain essentially unchanged for B of any constant size (small compared to H). The encoded system $\tilde{B}$ describes the mode in the black hole interior that is entangled with B.

We can define an (approximate) isometric embedding $V_\Psi : \mathcal{H}_{\tilde{B}} \to \mathcal{H}_{EH}$ of a single qubit $\tilde{B}$ into the subspace EH by

$$V_\Psi |i\rangle_{\tilde{B}} = 2(I_{EH} \otimes \langle \omega|_{B\tilde{B}})(|\Psi\rangle_{EHB} \otimes |i\rangle_{\tilde{B}}), \tag{5.9}$$

where $|\omega\rangle_{B\tilde{B}} = 2^{-1/2}(|00\rangle_{B\tilde{B}} + |11\rangle_{B\tilde{B}})$ denotes an EPR pair on $B\tilde{B}$; see Figure 5.5. While V_Ψ itself is not precisely an isometric embedding, it is exponentially close to one under the assumption that $|\Psi\rangle_{EBH}$ is pseudorandom on the exterior system EB, as specified in Definition 5.6.1. In Appendix 5.A, we show that there exists an

isometric embedding V such that

$$\|V - V_\Psi\| \leq 2 \cdot 2^{-\alpha|H|}, \tag{5.10}$$

where $\|\cdot\|$ denotes the operator norm. The isometry V then defines a code subspace that encodes $\tilde{B}$. For macroscopic observers (*i.e.*, $|O| \gg 1$), the error in (5.7) is negligible compared to the error in the decoupling bound (5.6). Although the norm in Equation (5.7) is the operator norm rather than the trace norm, that distinction need not concern us if $|B|$ is sufficiently small compared to $|H|$. Therefore we can ignore any differences between V and V_Ψ and use them interchangeably.
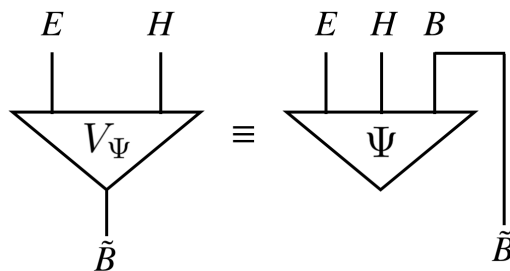
Figure 5.5: The definition of the encoding of $\tilde{B}$ into EH, with Ψ defined as in Figure 5.3.

We will now show that the isometry $V_\Psi : \mathcal{H}_{\tilde{B}} \to \mathcal{H}_{EH}$ defined above embeds $\tilde{B}$ into EH as a code subspace for which any low-complexity noise model acting on E is (approximately) correctable. By low-complexity error, we mean that the unitary process $U_{\mathcal{E}}$ in Figure 5.6 has complexity at most polynomial in $|H|$. Here the external observer O plays the role of the "environment" for the noise process acting on E and the probe P.

The error model depicted in Figure 5.6 is rather exotic compared to error models that are typically considered in discussions of quantum gravity and fault-tolerant quantum computing. For example, one widely studied error model is the "erasure model," wherein each qubit may be removed with some probability, and we know which qubits are removed. The performance of quantum codes against erasure errors arises, in particular, in studies of the holographic AdS/CFT dictionary [35], [36]; if a logical bulk operator can be "reconstucted" on a portion of the boundary, that means that erasure of the complementary portion of the boundary is correctable for that logical operator. By the no-cloning theorem, no code can tolerate erasure of more than 50% of the qubits in the code block. In contrast, in our setup, erasure
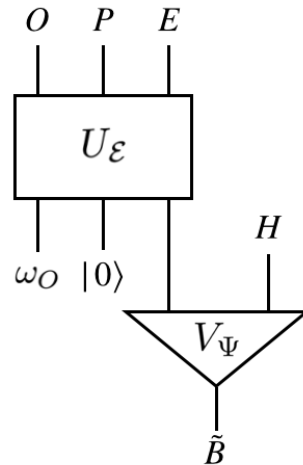
Figure 5.6: A black hole can be viewed as a quantum error-correcting code. By tracing out the observer O, we obtain a "noise model" $\mathcal{E}$ on the early radiation and the probe.

is correctable even if most of the qubits are removed. The catch is that the erased qubits must lie in E; removal of qubits in H is not allowed.

In studies of fault-tolerant quantum computing, the noise afflicting the physical qubits is usually assumed to be weak and weakly correlated. In a Hamiltonian formulation of the noise model, this means that each qubit in the computer is weakly coupled to a shared environment [39]. In contrast, for the noise model described by $U_\mathcal{E}$, the noise may act strongly on all the qubits in E; the only restriction is that the noise has quantum complexity scaling polynomially with $|H|$. Furthermore, how the noise acts depends on the initial state ω_O of the observer O, which may be chosen adversarially. Again, what makes successful error correction possible is that the subsystem H is assumed to be noiseless, an assumption that would be unrealistic for typical quantum computing hardware.

Codes that can protect against this malicious typle of noise are central to our proposed resolution of the firewall paradox. An old black hole provides such a code if its previously emitted radiation is pseudorandom. The code corrects errors successfully if the noise acting on E has low complexity and the remaining black hole H is noiseless, provided that the observer O is small compared to H.

5.7.1 Correcting Low-Complexity Errors

For simplicity, we will first consider a scenario without the probe P shown in Figure 5.6. We will see that the error applied to the radiation system E is (approxi-

mately) correctable. In Section 5.7.2, we will explain how our conclusion changes when the probe P is included.

A central result in the theory of quantum error correction is the *information-disturbance relation*, which states that a code can protect quantum information from noise if and only if the "environment" of the noise channel $\mathcal{E}$ learns nothing about the logical information. More precisely, there is a physical process $\mathcal{R}$, the recovery process, which reverses the error:

$$\mathcal{R} \circ \mathcal{E} \approx \mathcal{I}, \tag{5.11}$$

where $\mathcal{I}$ is the identity operation, if and only if the "reference system" that purifies the quantum error-correcting code decouples from the environment . In Figure 5.6 (neglecting the probe P), the environment of the noise channel $\mathcal{E}$ acting on E is the observer O, and B is the reference system that purifies the encoded interior mode $\tilde{B}$. Therefore the necessary and sufficient condition for (approximate) correctability is the (approximate) decoupling of B and O,

$$\rho_{OB} \approx \rho_O \otimes \rho_B \tag{5.12}$$

where ρ_{OB} is the reduced density operator for OB. Here the approximation errors of Equation (5.7.1) and Equation (5.7.1) are related to each other by a constant factor. Therefore, using the decoupling bound (5.6), we can conclude that there exists a recovery process $\mathcal{R}$ that reverses $\mathcal{E}$ up to an error exponentially small in $|H|$, as long as $|O| \ll |H|$ and assuming that the Hawking radiation is pseudorandom.

Various formal statements that imply the existence of $\mathcal{R}$ in Equation (5.7.1) are known; for the reader's convenience, we reproduce some of these results below. First, let us properly define what it means for a channel to be approximately correctable with respect to some code subspace — a more comprehensive discussion can be found in [23]. Let $S(\mathcal{H})$ denote the set of states on a Hilbert space $\mathcal{H}$. Suppose that we are given channels $\mathcal{E}, \mathcal{N} : S(\mathcal{H}) \to S(\mathcal{H})$. Fixing a state $\rho \in S(\mathcal{H})$, we define the *entanglement fidelity* between $\mathcal{E}$ and $\mathcal{N}$ with respect to ρ to be

$$F_\rho(\mathcal{E}, \mathcal{N}) = f\left[(\mathcal{E} \otimes \mathcal{I})(|\psi\rangle\langle\psi|), (\mathcal{N} \otimes \mathcal{I})(|\psi\rangle\langle\psi|)\right],$$

where $|\psi\rangle$ is a purification of ρ, and where

$$f(\rho, \tau) = \text{Tr}\left(\sqrt{\sqrt{\tau}\rho\sqrt{\tau}}\right)$$

is the usual fidelity between states ρ and τ. To quantify the closeness of two channels, we use the worst-case entanglement fidelity to define the *Bures distance*, given by

$$\mathfrak{B}(\mathcal{E}, \mathcal{N}) = \max_{\rho} \sqrt{1 - F_{\rho}(\mathcal{E}, \mathcal{N})}; \tag{5.13}$$

we sometimes define a more restricted notion of the Bures distance, where we maximize over states in some specified subspace. In discussions of error correction, we say that a noise channel $\mathcal{E}$ is ϵ-correctable with respect to a code subspace $C \subseteq \mathcal{H}$ if there exists a recovery channel $\mathcal{R}$ such that

$$\mathfrak{B}(\mathcal{R} \circ \mathcal{E}, \mathcal{I}) \leq \epsilon, \tag{5.14}$$

where the maximization in the Bures metric is over all code states ρ with support on C.

The Bures metric is bounded above and below by the trace norm as

$$2\mathfrak{B}^2(\mathcal{E}, \mathcal{N}) \leq \max_{\rho} \|(\mathcal{E} \otimes \mathcal{I})(|\psi\rangle\langle\psi|) - (\mathcal{N} \otimes \mathcal{I})(|\psi\rangle\langle\psi|)\|_1$$

$$\leq 2\sqrt{2}\mathcal{B}(\mathcal{E}, \mathcal{N}). \tag{5.15}$$

The norm in the middle is essentially the diamond-norm distance between the channels $\mathcal{E}$ and $\mathcal{N}$ [40], except that for the purpose of characterizing error correction the maximization is over code states only. Applying this inequality and tracing out the purifying system, the ϵ-correctability of a channel $\mathcal{E}$ implies that we have

$$\max_{\rho} \|(\mathcal{R} \circ \mathcal{E})(\rho) - \rho\|_1 \leq 2\sqrt{2}\epsilon, \tag{5.16}$$

where again the maximization is over code states.

As mentioned previously, an important result characterizing approximate correctability is the information-disturbance trade-off, which we now state quantitatively. Let $\mathcal{E} : S(\mathcal{H}_A) \rightarrow S(\mathcal{H}_A)$ be a noise channel acting on a system A, and let $V : \mathcal{H}_A \rightarrow \mathcal{H}_F \otimes \mathcal{H}_A$ be an isometry which purifies $\mathcal{E}$; i.e.,

$$\mathcal{E}(\rho) = \mathrm{Tr}_F(V\rho V^{\dagger}).$$

Hence F is the environment of the channel; we have resisted the temptation to denote the environment by E to avoid confusion with our convention that E denotes a subsystem of the Hawking radiation. Then the *complementary channel* $\widehat{\mathcal{E}} : S(\mathcal{H}_A) \rightarrow S(\mathcal{H}_F)$ is defined by

$$\widehat{\mathcal{E}}(\rho) = \mathrm{Tr}_A(V\rho V^{\dagger}).$$

A special case of interest is the identity channel $\mathcal{I}$. Taking the environment to be 1-dimensional, the complementary channel to the identity channel is simply the (partial) trace

$$\widehat{\mathcal{I}}(\rho) = \mathrm{Tr}_A(\rho).$$

Then the information-disturbance trade-off states the following:

Theorem 5.7.1 ([23], Theorem 1). *Let $C \subseteq \mathcal{H}_A$ be a code subspace. Let $\mathcal{E} : S(\mathcal{H}_A) \to S(\mathcal{H}_A)$ be an error channel. Then*

$$\inf_{\mathcal{R}} \mathfrak{B}\left(\mathcal{R} \circ \mathcal{E}, \mathcal{I}\right) = \inf_{\mathcal{R}'} \mathfrak{B}\left(\widehat{\mathcal{E}}, \mathcal{R}' \circ \mathrm{Tr}\right), \tag{5.17}$$

where the infimums are taken over all channels $\mathcal{R} : S(\mathcal{H}_A) \to S(\mathcal{H}_A)$, and $\mathcal{R}' : \mathbb{R} \to S(\mathcal{H}_F)$.

Note that a channel $\mathcal{R}' : \mathbb{R} \to S(\mathcal{H}_F)$ is just state preparation on the channel environment $\mathcal{H}_F$, i.e., every such channel $\mathcal{R}'$ is uniquely identified with a state $\sigma_F \in S(\mathcal{H}_F)$ such that

$$(\mathcal{R}' \circ \mathrm{Tr})(\rho) = \mathrm{Tr}(\rho)\,\sigma_F,$$

so we can equivalently write

$$\inf_{\mathcal{R}'} \mathfrak{B}\left(\widehat{\mathcal{E}}, \mathcal{R}' \circ \mathrm{Tr}\right) = \inf_{\sigma_F} \mathfrak{B}\left(\widehat{\mathcal{E}}, \sigma_F \otimes \mathrm{Tr}\right).$$

Now let us see what Equation (5.7.1) tells us in the context of the black hole error-correcting code defined by the (approximate) isometry V_Ψ. Let $\tilde{\rho}_{\tilde{B}}$ be a logical state and let $\tilde{\rho}_{\tilde{B}B}$ be a purification. The isometry V_Ψ then embeds $\tilde{\rho}_{\tilde{B}B}$ as a (purified) code state ρ_{EHB}:

$$\rho_{EHB} = V_\Psi \tilde{\rho}_{\tilde{B}B} V_\Psi^\dagger.$$

Let $\mathcal{E} : S(\mathcal{H}_E) \to S(\mathcal{H}_E)$ be an arbitrary channel acting on E such that some purification $U_\mathcal{E}$ of $\mathcal{E}$ has low-complexity (see the set-up described in Figure 5.1). Let

$$\sigma_{OEHB} = U_\mathcal{E}(\omega_O \otimes \rho_{EHB})U_\mathcal{E}^\dagger \tag{5.18}$$

denote the overall post-evolution state. To apply Theorem 5.7.1, let us consider the error channel $\mathcal{E} \otimes \mathcal{I}_H$. Then the environment of the channel $\mathcal{E} \otimes \mathcal{I}_H$ is the

observer subsystem O, and the complementary channel $\widehat{\mathcal{E} \otimes \mathcal{I}_H}$ maps $S(\mathcal{H}_{EH})$ to $S(\mathcal{H}_O)$. From (5.7.1), the state obtained from ρ_{EHB} after the application of $\widehat{\mathcal{E} \otimes \mathcal{I}_H}$ is precisely given by

$$\left(\widehat{\mathcal{E} \otimes \mathcal{I}_H} \otimes \mathcal{I}_B\right)(\rho_{EHB}) = \mathrm{Tr}_{EH}(\sigma_{OEHB}) = \sigma_{OB}.$$

Since σ_{OEHB} was a state obtained through acting on the black hole code state ρ_{EHB} with a low-complexity unitary, it follows by the pseudorandom hypothesis that the decoupling bound (5.6) holds. Therefore we have

$$\|\sigma_{OB} - \sigma_O \otimes \sigma_B\|_1 \leq 6 \cdot 2^{-(\alpha|H|-|O|)}.$$

Finally, since $U_{\mathcal{E}}$ is supported away from B, we have $\sigma_B = \rho_B$, and so

$$\left\|\left(\widehat{\mathcal{E} \otimes \mathcal{I}_H} \otimes \mathcal{I}_B\right)(\rho_{EHB}) - \sigma_O \otimes \rho_B\right\|_1 \leq 6 \cdot 2^{-(\alpha|H|-|O|)}. \tag{5.19}$$

This holds for all code states, so (5.7.1), together with the first inequality in (5.7.1), implies that we have

$$\inf_{\sigma_O} \mathfrak{B}\left(\widehat{\mathcal{E} \otimes \mathcal{I}_H}, \sigma_O \circ \mathrm{Tr}_{EH}\right) \leq \sqrt{3} \cdot 2^{-(\alpha|H|-|O|)/2}.$$

Therefore, the channel $\mathcal{E}$ is approximately correctable by Theorem 5.7.1. We state this as a Lemma.

Lemma 5.7.2. *Let V_Ψ be the approximate isometric embedding defined by the state Ψ_{EHB}. Let $\mathcal{E}$ be an error channel on E with purification $U_{\mathcal{E}}$. Suppose that the decoupling bound (5.6) holds. Then $\mathcal{E}$ is ϵ-correctable for V_Ψ, where*

$$\epsilon = \sqrt{3} \cdot 2^{-(\alpha|H|-|O|)/2},$$

if B is a single qubit. For general $|B|$, we have

$$\epsilon = \sqrt{\frac{3}{2}} \cdot 2^{-(\alpha|H|-|OB|)/2}.$$

Note that the recovery operator $\mathcal{R}$ acts on EH rather than E. The same will be true for the ghost logical operators we construct in Section 5.8.

5.7.2   Including the Probe

We would now like to consider a modified scenario in which both the observer O and a probe P interact with the Hawking radiation system E, as indicated in Figure 5.1.

We cannot simply absorb P into O, because we will continue to insist that O is small compared to H, while we wish to allow P to be comparable to H in size, or even larger. In this modified scenario, the unitary purification $U_{\mathcal{E}}$ of the noise model acts on OPE rather than OE. This change does not alter the conclusion that O and B decouple if $U_{\mathcal{E}}$ has low complexity. Therefore, just as before, there is a recovery map that reverses the effect of the noise on the encoded state. What changes is that now the recovery map acts on PEH rather than EH.

We emphasize that if the probe P is sufficiently large, then P need not decouple from B, even if $U_{\mathcal{E}}$ has low complexity. To understand why not, suppose P has the same size as the system E and that the channel $\mathcal{E}$ swaps P and E. Before this swap, B is entangled with the code space embedded in EH; therefore after the swap (a low-complexity operation), B is entangled with PH. More realistically, we might imagine that P is a cloud of dust surrounding the black hole, and that $|P| \gg |E|$. After the dust interacts with the Hawking radiation, the encoding of $\tilde{B}$ will be modified, so that B is entangled with a code subspace of PEH rather than a subspace of EH [9].

However, any subsystem of OP which is small compared to H will decouple from B, as long as $U_{\mathcal{E}}$ has low complexity, and assuming that the Hawking radiation is pseudorandom. The only way to distill the encoded state into a small subsystem is to perform a high complexity operation. Hence, if only low-complexity operations are allowed, we need not worry about a scenario in which the encoded version of $\tilde{B}$ outside the horizon is decoded into a small system, and then falls into the black hole to meet its twin in the interior. This is essentially the observation of Harlow and Hayden [14], later extended by Aaronson [15]. Our analysis goes further by clarifying that the encoded state is hard to distill even when the remaining black hole H is much smaller than E, as long as H is macroscopic and assuming that the Hawking radiation is pseudorandom.

One might wonder whether the encoded mode can be easily extracted if the probe system P is prepared in a carefully chosen state [41]. Our conclusion is that any such initial state of P would need to have exponential complexity, an unlikely property for the dust surrounding an evaporating black hole. One might also ask what happens if all the qubits in the early radiation system E are measured in the standard basis by the observer. Surely this *would* disrupt the encoded interior of the black hole. But in our model, the number of radiation qubits that can be measured is limited by the size $|O|$ of the observer's memory, and the interior will stay well protected as long

as $|O|$ is much smaller than $|H|$.

It is also instructive to view the system O in a different way. Up until now, we have regarded O as a potentially malicious agent who attempts to damage the encoded interior of the black hole by acting on its exterior. More prosaically, we can think of O as an abstract purifying space which is introduced for convenience so that we can describe the noise channel $\mathcal{E}$ using its purification, the unitary transformation $U_{\mathcal{E}}$. From that point of view, limiting the size $|O|$ of the "observer" O is just a convenient way of restricting the form of the quantum channel $\mathcal{E}$. Specifically, the rank of the marginal density operator ρ_O, after $U_{\mathcal{E}}$ is applied, is called the Kraus rank (or simply the rank) of the channel $\mathcal{E}$. This rank can be no larger than the dimension of system O, namely $2^{|O|}$, which we have assumed to be small compared to the dimension $2^{|H|}$ of the Hilbert space of black hole microstates. Thus our conclusion can be restated: if the Hawking radiation is pseudorandom and H is macroscopic, then the quantum error-correcting code protects the encoded version of $\tilde{B}$ against any noise channel acting on PE that has both low complexity *and* low rank.

An advantage of this viewpoint is that one might otherwise be misled into interpreting $|O|$ as the physical size of an actual observer. More accurately, it can be regarded as the effective size of the quantum memory of a physical object. This distinction is significant. For an object of specified mass, the largest possible quantum memory is achieved by a black hole of that mass, but the memory size of a quantum computer typically falls far short of that optimal value, because most of its mass is locked into the rest mass of atomic nuclei and unavailable for information processing purposes. Furthermore, the mass per unit volume of a typical quantum computer is far smaller than a black hole's. Therefore it is reasonable to expect that the effective Hilbert space dimension of system O (and hence the Kraus rank of the channel $\mathcal{E}$) is far smaller than the Hilbert space dimension of a black hole with the same circumference as system O.

Up until now, we have mostly focused on the hardness of decoding the black hole interior mode by acting on the Hawking radiation outside the black hole, concluding that distilling the encoded system to a small quantum memory is computationally hard if the remaining black hole is macroscopic. In Section 5.8 we will turn to a more subtle question: can a low-complexity operation acting on the Hawking radiation system E create an excitation near the black hole horizon that could be detected by an infalling observer who falls into the black hole? Here too, we will argue that the answer is no. This is a nontrivial extension beyond what we have

found so far — on the face of it, perturbing a quantum state is a far easier task than depositing the state in a compact quantum memory.

Bousso emphasized that if the interior mode $\tilde{B}$ is encoded in EH, and if effective quantum field theory on curved spacetime is a good approximation in regions of low curvature, then the vacuum near the black hole horizon would need to be "frozen" [9]. That is, neither a small agent O acting on E nor a large probe P interacting with E could disrupt the entanglement of $\tilde{B}$ with B and hence create an excitation localized near the horizon. We agree with this conclusion, provided that $|H| \gg 1$ and that the interactions of OP with E have quantum complexity scaling polynomially with $|H|$. Interactions with the large probe may alter how the black hole interior is encoded in the radiation and probe, but they do not disrupt the frozen vacuum.

Once $|H|$ is O(1), large corrections to effective field theory may be expected. Furthermore, the semiclassical structure of spacetime may no longer be applicable in the regime where operations of superpolynomial complexity are allowed; these high-complexity operations could tear spacetime apart. In particular, our expectation that an agent acting on E should be unable to influence the black hole interior might be flagrantly violated if the agent can perform high-complexity operations. We should grow accustomed to the notion that for effective field theory to be an accurate approximation, we require not only geometry with low curvature and states with low energy, but also operations with low complexity and low Kraus rank.

To investigate whether the semiclassical causal structure is robust with respect to low-complexity operations, we will need to develop some additional formalism, specifically the theory of *ghost logical operators*; in the context of an old black hole, these may be viewed as operators which act on the black hole interior. We would like to understand, given that the interior is encoded in the Hawking radiation outside the black hole, why low-complexity operations acting on the Hawking radiation produce no detectable excitations inside the black hole. We turn to that task next.

5.8 Theory of Ghost Logical Operators

So far, we have argued that the late radiation system B remains decoupled from any sufficiently small subsystem of the early radiation E and the probe P, when the observer O performs a low-complexity operation on EP. Therefore an infalling observer with reasonable computational power is prevented from extracting the encoded interior mode before jumping into the black hole. But what if the observer settles for the seemingly easier task of disrupting the interior rather than decoding

it? In this section, we will show that an algebra of *ghost logical operators* can be constructed acting on the interior mode, with the property that low-complexity operations performed outside the black hole nearly commute with the ghost algebra. Hence, if these ghost operators are regarded as operations that can be performed by an observer inside the black hole, we may conclude that the interior is well protected against the actions of malicious agents outside the black hole.

Following arguments from [11], consider an operator T which acts on the interior mode. Because the corresponding encoded operator acting on the Hawking radiation is highly scrambled, the commutator of this encoded operator with a generic simple operator acting on the radiation has no reason to be small. It seems, then, that an external observer should be able to perturb the interior mode easily [9], [11]. Can this conclusion be evaded by constructing the encoded operators suitably? For two-side black holes in AdS/CFT, Papadodimas and Raju argued that "mirror operators" with the desired properties can be constructed [7], [10], but no satisfactory construction is known for evaporating black holes.

Within our simple toy model of evaporating black hole, we can construct analogues of the mirror operators. Assume that the decoupling bound (5.6) holds. Then, as we will see, for every operator $\tilde{T}_B$ acting on some outgoing mode B, there exists a "mirror operator" T_{EH} acting on EH which satisfies the following conditions:

$$
\begin{aligned}
\tilde{T}_B|\Psi\rangle &\approx T_{EH}|\Psi\rangle, \\
[T_{EH}, E_a]|\Psi\rangle &\approx 0,
\end{aligned}
\tag{5.20}
$$

where $\{E_a\}$ is a set of operators that a computationally bounded external observer can apply on the radiation, and $|\Psi\rangle$ is the state of the radiation and the black hole. The equations (5.8) hold up to an error exponentially small in $|H|$. The first line implies that one can (in principle if not in practice) certify entanglement between an outgoing radiation mode and an abstract subsystem specified by the operators $\{T_{EH}\}$.[7] Therefore, these operators satisfy the right measurement statistics expected for sensibly defined interior operators. The second line implies that these operators approximately commute with all the operators that the external observer can apply. The fact that T_{EH} commutes with $\{E_a\}$ holds as an operator equation on all the states in the code subspace. Therefore, the subsystem specified by the mirror operators $\{T_{EH}\}$ is fully entangled with the late outgoing radiation modes while also being effectively "space-like separated" from the external observer. That is, the external

---

[7]For example, one could perform Bell tests using the Pauli operators acting on B and its mirror.

observer can disrupt the semiclassical causal structure of the black hole only by applying operations of superpolynomial complexity to the radiation.

In our construction, it is important to properly characterize the set $\{E_a\}$ of operators that the exterior observer can apply to the radiation. If we view the observer, the black hole, and the exterior radiation as a closed system, we ought to model the entire evolution as a unitary process. In order to enforce the unitarity of this process, the operator applied by the observer to the radiation should depend on the initial state of the observer, as in Figure 5.7.
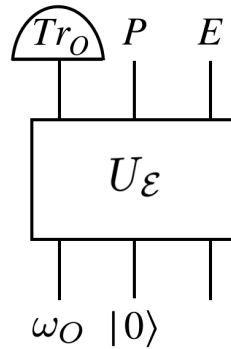
Figure 5.7: The operator applied by an exterior observer to the Hawking radiation depends on the observer's initial state ω_O, the probe's initial state $|0\rangle$, and the joint unitary transformation $U_{\mathcal{E}}$.

In this scenario, the set of operations that the observer can apply to the radiation is not completely arbitrary. Specifically, any such operation must be of the following form:

$$\rho_{PE} \mapsto \text{Tr}_O\big(U_{\mathcal{E}}(\omega_O \otimes \rho_{PE})U_{\mathcal{E}}^\dagger\big), \tag{5.21}$$

wherein the only freedom available to the observer is the choice of the initial state ω_O. Because the observer is part of a system that is governed by the laws of physics, the observer's actions are determined entirely by that initial state, not by the global unitary process. One may view Equation (5.8) as a quantum channel that acts on PE with a Kraus representation and corresponding dilation given by

$$\rho_{PE} \mapsto \sum_a E_a \rho_{PE} E_a^\dagger$$

$$= \text{Tr}_O\left(\sum_{a,b}(|a\rangle_O \otimes E_a)\,\rho_{PE}\,(\langle b|_O \otimes E_b^\dagger)\right),$$

where $\sum_a E_a^\dagger E_a = I$, and $\{|a\rangle\}$ is an orthonormal basis for O. Therefore, $E_a \rho_{PE} E_a^\dagger$ can be thought as a (subnormalized) post-selected state in which the state of the

observer after interacting with the radiation is $|a\rangle_O$. Up to normalization, the operator that the observer applied on the radiation would be E_a in that case. While we do not know the exact details about $\{E_a\}$, within our model we have the following non-trivial constraints:

1. The cardinality of the set $\{E_a\}$ is bounded above by d_O, where $d_O = 2^{|O|}$ is the dimension of the observer's Hilbert space.

2. The global unitary evolution $U_\mathcal{E}$ has a complexity polynomial in the black hole entropy $\sim |H|$.

The construction of the mirror operators rests on the observation that V_Ψ defines the embedding map of a quantum error-correcting code that can protect quantum information against "environmental noise" caused by the observer O; see Figure 5.6. The error model induced by the observer is different from conventional error models that are typically considered in discussions of fault-tolerant quantum computing. For one, the error is applied only on the radiation E and probe P, not the remaining black hole H. Secondly, $U_\mathcal{E}$ can apply any operation to the radiation with complexity polynomial in $|H|$. In contrast, more conventional noise models such as the depolarizing channel or the amplitude damping channel typically result from a brief interaction between the environment and the system of interest.

We have already seen in Section 5.7 that the encoding map V_Ψ protects quantum information against this exotic error model; this conclusion follows from the decoupling condition, which in turn is a consequence of the pseudorandomness of Hawking radiation as discussed in Section 5.6. Our next task is to relate this robustness against low-complexity noise to the claim in Equation (5.8). The formalization and proof of this statement is the main technical contribution of this section.

Before diving into details in the following subsections, let us summarize the conclusion. Consider an error model in which one applies either a channel $\mathcal{E}(\cdot) = \sum_a E_a(\cdot)E_a^\dagger$ or the identity channel, each occurring with nonzero probability. If a quantum error-correcting code V_Ψ can correct such errors, then there is a complete set of logical operators that commutes with all the errors $\{E_a\}$ when acting on the code space; see Figure 5.8. That is, for any operator $\tilde{T}$ acting on the abstract logical space, there exists a corresponding logical operator T acting identically on the code subspace such that T satisfies the following intertwining condition for all E_a:

$$TE_aV_\Psi \approx E_aTV_\Psi \approx E_aV_\Psi\tilde{T}. \tag{5.22}$$

These logical operators are special because the commutation relation holds as an operator equation acting on all the states in the code subspace. By mapping the isometry V_Ψ back to the state $|\Psi\rangle$, we arrive at Equation (5.8) and Figure 5.8. Note that this is a stronger statement than saying that the commutator of T and E_a has a vanishing expectation value in the code subspace, *i.e.*,

$$V_\Psi^\dagger T E_a V_\Psi \approx V_\Psi^\dagger E_a T V_\Psi.$$

In Section 5.8.1, we will prove (5.8) in the exactly correctable setting. We will then generalize the construction to the approximate case in Section 5.8.2.

Equation (5.8) also arises in the theory of Operator Algebra Quantum Error-Correction (OAQEC) [25], [26]. However, in that context, one normally considers a logical operator T which annihilates the orthogonal complement of the code space. A novelty of our discussion is that we will allow T to have support extending beyond the code space. In that case, it is delicate to ensure that the action of T on states outside the code space is consistent with (5.8). More importantly, OAQEC was formulated in [25], [26] for the case of exact quantum error-correction. Our discussion in Section 5.8.1 is self-contained, and generalizes readily to the approximate setting, as we show in Section 5.8.2.
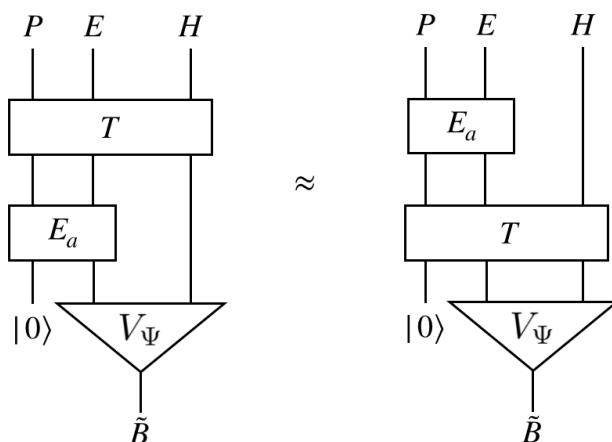
Figure 5.8: Acting on any code state, the ghost logical operator T (approximately) commutes with any "error" in the set $\{E_a\}$.

5.8.1 Exact Ghost Operators

Let $\tilde{\mathcal{H}}$ be an abstract logical Hilbert space, and consider an encoding $V : \tilde{\mathcal{H}} \to C \subseteq \mathcal{H}$, where C denotes the code subspace embedded within the larger physical Hilbert space $\mathcal{H}$. Given a Hilbert space $\mathcal{H}$, we will let $S(\mathcal{H})$ denote the state space of $\mathcal{H}$,

i.e., the set of all density operators supported on $\mathcal{H}$. Let $\mathcal{E}$ be a correctable error channel for C, which we can write in a Kraus representation as

$$\mathcal{E}(\rho) = \sum_{a=1}^{|K|} E_a \rho E_a^\dagger,$$

where we denote the set of Kraus operators as $K = \{E_a\}$. A given channel will of course have many different Kraus representations; the choice of representation will not matter in the exact case, since the set of exactly correctable errors is closed under linear combinations, but we will have to be careful in the analysis of the approximate case in section 5.8.2. In this section, we will fix an arbitrary Kraus representation K for $\mathcal{E}$.

As a convention, we will denote quantities in $\tilde{\mathcal{H}}$ with tildes, and quantities in $\mathcal{H}$ without. Let

$$\tilde{T} = \sum_{k=1}^{r} \lambda_k \tilde{P}_k$$

be a normal operator on $\tilde{\mathcal{H}}$, with (distinct) eigenvalues $\{\lambda_k\}$, where each $\mathcal{P}_k$ is the spectral projector onto the corresponding eigenspace. For ease of notation, given any projector P, we will denote the corresponding range subspace as $[P]$, i.e., $[P] = \mathrm{Im}(P)$.

Consider the encoded subspace $F_k = \mathrm{Im}(V\tilde{P}_k)$ of each eigenspace, and define

$$[P_k] = \mathrm{span}\left\{ E_a|\phi\rangle \,\Big|\, E_a \in K,\ |\phi\rangle \in F_k \right\}.$$

Note that $[P_k]$ is the subspace generated by the set of all correctable errors, i.e., the span of K, acting on the encoded eigenspace F_k. These subspaces are well-defined since linear combinations of correctable errors remain correctable, and the Knill-Laflamme conditions [42] imply that subspaces corresponding to distinct eigenvalues will be orthogonal. We can then define a normal operator $T : \mathcal{H} \to \mathcal{H}$ by

$$T = \sum_{k=1} \lambda_k P_k,$$

where each P_k is the corresponding projector onto $[P_k]$.

Definition 5.8.1. *Given any normal operator $\tilde{T} : \tilde{\mathcal{H}} \to \tilde{\mathcal{H}}$, we will call the operator $T : \mathcal{H} \to \mathcal{H}$ obtained through the above construction the* pseudo-ghost *operator corresponding to $\tilde{T}$.*

For any $|\chi_j\rangle \in F_j$ and any error operator $E_a \in K$, the action of the pseudo-ghost operator T is such that

$$TE_a|\chi_j\rangle = \sum_{k=1}^{r} \lambda_k P_k E_a|\chi_j\rangle = \lambda_j E_a|\chi_j\rangle = E_a \hat{T}|\chi_j\rangle.$$

Here $\hat{T}$ can be any logical operator for $\tilde{T}$, which therefore satisfies $\hat{T}|\chi_j\rangle = \lambda_j|\chi_j\rangle$.

These pseudo-ghost operators satisfy $TE_a = E_a\hat{T}$ acting on the code space, and so do the ghost operators that we wish to construct. However, note that a pseudo-ghost operator T will not necessarily act as a logical operator for $\tilde{T}$ since we might not have $F_k \subseteq [P_k]$ if the identity is not among the Kraus operators. The operator T will not act correctly on the code subspace unless each of the encoded eigenspaces for $\tilde{T}$ are contained within the corresponding eigenspace for T. Our definition of a ghost logical operator should stipulate that T is logical, as well as requiring $[T, E_a] = 0$ acting on the code space.

Definition 5.8.2. *Let $T : \mathcal{H} \to \mathcal{H}$ be a logical operator for $\tilde{T}$. We say that T is a ghost logical operator for $\tilde{T}$ if*

$$TE_a|\psi\rangle = E_a T|\psi\rangle \tag{5.23}$$

for all $E_a \in K$ and $|\psi\rangle \in C$. Given a pseudo-ghost operator T, we say that T is extensible if it admits an extension onto $\mathcal{H}$ such that it becomes a logical operator for $\tilde{T}$.

Clearly the extension of any extensible pseudo-ghost operator will define a corresponding ghost logical operator. With the above definitions, it is simple to give a concise criterion for when pseudo-ghost operators extend to ghost logical operators in the exact setting.

Lemma 5.8.3. *Let T be a pseudo-ghost operator. Then T is extensible if and only if*

$$\langle \chi_j|E|\chi_i\rangle = 0, \qquad (i \neq j) \tag{5.24}$$

for all $E \in K$, $|\chi_i\rangle \in F_i$, $|\chi_j\rangle \in F_j$.

Proof. To see necessity, suppose that T is extensible, and let T' denote its logical extension. Because T' is a logical operator for $\tilde{T}$, it must satisfy

$$T'|\chi_i\rangle = \lambda_i|\chi_i\rangle.$$

Let $E \in K$ be arbitrary. Left multiplying by $\langle \chi_j | E^\dagger$, we get

$$\lambda_i \langle \chi_j | E^\dagger | \chi_i \rangle = \langle \chi_j | E^\dagger T' | \chi_i \rangle = \sum_{k=1}^{r} \lambda_k \langle \chi_j | E^\dagger P_k | \chi_i \rangle = \lambda_j \langle \chi_j | E^\dagger | \chi_i \rangle.$$

Here we have used $E|\chi_j\rangle \in [P_k]$, and noted that T and T' have the same action on $[P_k]$; we also used $P_k E|\chi_j\rangle = \delta_{kj} \lambda_j E|\chi_j\rangle$. If $\lambda_i \neq \lambda_j$, then we must have $\langle \chi_j | E^\dagger | \chi_i \rangle = 0$. Taking the complex conjugate, we obtain Equation (5.8.3).

Conversely, suppose that for all $i \neq j$ and all correctable errors we have $\langle \chi_i | E | \chi_j \rangle = 0$. We must extend the action of T to each encoded eigenvector $|\chi_i\rangle \in C$. The relations $\langle \chi_i | E | \chi_j \rangle = 0$ imply that $|\chi_i\rangle$ is orthogonal to the subspaces $[P_j]$ for $j \neq i$. There are two possible cases, either $|\chi_i\rangle \in [P_i]$, for which $T|\chi_i\rangle = \lambda_i |\chi_i\rangle$ is already well-defined and we are done, or else there exists a component of $|\chi_i\rangle$ lying in the subspace orthogonal to $\bigoplus_{k=1}^{r} [P_k]$.

Let $|\chi_i^\perp\rangle$ denote the normalized component of $|\chi_i\rangle$ orthogonal to $[P_i]$. Then we extend the subspace $[P_i]$ to $[P_i']$ by defining the projector

$$P_i' = P_i + |\chi_i^\perp\rangle\langle\chi_i^\perp|.$$

Note that the new subspace $[P_i']$ contains within it $[P_i]$ and remains orthogonal to $[P_j]$ for $j \neq i$. Moreover, we have $|\chi_i\rangle \in [P_i']$. We can now define an extension of T with the projector P_i' in place of P_i. Then the extension T' satisfies

$$T'|\chi_i\rangle = \lambda_i |\chi_i\rangle.$$

We may repeat this procedure with an orthogonal basis $\{|\chi_k\rangle\}$ for C until we are left with an extension which acts as a logical operator for $\tilde{T}$. $\qquad \square$

We will be primarily interested in the case where there exists a full set of ghost logical operators. We say that there exists a *complete set* of ghost logical operators if for every normal operator $\tilde{T} : \tilde{\mathcal{H}} \to \tilde{\mathcal{H}}$, there exists a corresponding ghost logical operator T. In what follows, given a channel $\mathcal{E}$, we will let $\mathcal{E}_\mathcal{I}$ denote the channel

$$\mathcal{E}_\mathcal{I} = \mathcal{I}/2 + \mathcal{E}/2,$$

where $\mathcal{I}$ is the identity channel. That is, in the channel $\mathcal{E}_\mathcal{I}$, with probability $1/2$ $\mathcal{E}$ is applied, and with probability $1/2$ nothing happens.

Theorem 5.8.4. *Let $\mathcal{E}$ be a correctable channel with Kraus operators K. Then a complete set of ghost logical operators for $\mathcal{E}$ exists if and only if $K \cup \{I\}$ is a correctable set, i.e., if and only if $\mathcal{E}_\mathcal{I}$ is a correctable channel.*

Proof. Suppose that $K \cup \{I\}$ is a correctable set. Then the Knill-Laflamme conditions for $K \cup \{I\}$ imply that the hypotheses of Lemma 5.8.3 are satisfied so that every pseudo-ghost operator is extensible to a ghost logical operator. It follows that there exists a complete set of ghost logical operators.

Conversely, suppose that there exists a complete set of ghost logical operators. Let $|\psi\rangle, |\phi\rangle \in C$ be two mutually orthogonal code states, and let $|\tilde{\psi}\rangle = V^\dagger |\psi\rangle$ and $|\tilde{\phi}\rangle = V^\dagger |\phi\rangle$ be the corresponding pre-images in $\tilde{\mathcal{H}}$. Define the operators

$$\tilde{T}_1 = |\tilde{\phi}\rangle\langle\tilde{\phi}| - |\tilde{\psi}\rangle\langle\tilde{\psi}|,$$

and

$$\tilde{T}_2 = |\tilde{\phi} + \tilde{\psi}\rangle\langle\tilde{\phi} + \tilde{\psi}| - |\tilde{\phi} - \tilde{\psi}\rangle\langle\tilde{\phi} - \tilde{\psi}|,$$

where $|\tilde{\phi} \pm \tilde{\phi}\rangle = 2^{-1/2}(|\tilde{\phi}\rangle \pm |\tilde{\psi}\rangle)$. By assumption, there exist ghost logical operators T_1 and T_2 corresponding to $\tilde{T}_1$ and $\tilde{T}_2$. Now let $E_a, E_b \in K \cup \{I\}$. Then we have

$$\langle\psi|E_a^\dagger E_b|\phi\rangle = \langle\psi|E_a^\dagger E_b T_1|\phi\rangle$$
$$= \langle\psi|T_1 E_a^\dagger E_b|\phi\rangle$$
$$= -\langle\psi|E_a^\dagger E_b|\phi\rangle,$$

where the first line follows due to the fact that $|\phi\rangle$ is an eigenvector for T_1 with eigenvalue 1, the second line follows from the defining equations (5.8.2) for the ghost operators, together with the fact that T_1 is self-adjoint, and the last line follows from the fact that $|\psi\rangle$ is an eigenvector for T_1 with eigenvalue -1. This implies that $\langle\psi|E_a^\dagger E_b|\phi\rangle = 0$.

Repeating the same argument for T_2, we have

$$\langle\phi - \psi|E_a^\dagger E_b|\phi + \psi\rangle = \langle\phi - \psi|E_a^\dagger E_b T_2|\phi + \psi\rangle$$
$$= \langle\phi - \psi|T_2 E_a^\dagger E_b|\phi + \psi\rangle$$
$$= -\langle\phi - \psi|E_a^\dagger E_b|\phi + \psi\rangle,$$

which implies that

$$0 = \langle\phi|E_a^\dagger E_b|\phi\rangle - \langle\psi|E_a^\dagger E_b|\psi\rangle.$$

Since ϕ and ψ were arbitrary, this holds for any pair of orthogonal states.

Let $\{|j\rangle\}$ be an orthonormal basis for C and define $\lambda_{ab} = \langle\psi|E_a^\dagger E_b|\psi\rangle$ for an arbitrary state $|\psi\rangle \in C$. Then it follows that we have

$$\langle i|E_a^\dagger E_b|j\rangle = \lambda_{ab}\delta_{ij},$$

so that the Knill-Laflamme conditions for $K \cup \{I\}$ are satisfied. Therefore, $K \cup \{I\}$ is a correctable set of errors. $\qquad\square$

5.8.2 Approximate Ghost Operators

In this section, we discuss how the ghost logical operators can be constructed for *approximate* quantum error-correcting codes. We need to consider this case because we inferred in Section 5.7 that the errors due to low-complexity operations on the radiation system E are correctable approximately (with a residual error exponentially small in $|H|$) rather than exactly. Although the uncorrected error is exponentially small, the Hilbert space is exponentially large, so we need to do a careful analysis to check that the ghost logical operators commute with the errors apart from exponentially small effects.

It turns out the strategy that we pursued in the exact setting also works in the approximate setting. To get started, we will construct *approximate ghost projectors* $\{\mathcal{P}_i\}$ that play the same role as the $\{P_i\}$ in the previous section.

Definition 5.8.5. *Let $\{|\tilde{i}\rangle\}$ be an orthonormal basis for $\tilde{\mathcal{H}}$, and suppose V is an encoding isometry. We define (approximate)* ghost projectors *with respect to this basis, denoted $\mathcal{P}_i$, to be the orthogonal projectors onto the* positive *eigenspace of*

$$\mathcal{E}(|i\rangle\langle i| - \rho_{i,\perp}),$$

where $|i\rangle = V|\tilde{i}\rangle$ for $|\tilde{i}\rangle \in \tilde{\mathcal{H}}$, and where

$$\rho_{i,\perp} = \frac{1}{\dim \tilde{\mathcal{H}} - 1} \sum_{j \neq i} |j\rangle\langle j|.$$

The motivation behind this definition follows from the fact that $\mathcal{P}_i$ is an operator that can optimally distinguish $\mathcal{E}(|i\rangle\langle i|)$ from $\mathcal{E}(\rho_{i,\perp})$, according to the Holevo-Helstrom theorem [40]. Because the effect of the channel $\mathcal{E}$ can be reversed up to a small error, it nearly preserves the orthogonality of $|i\rangle\langle i|$ and $\rho_{i,\perp}$; therefore, $\mathcal{P}_i$ can distinguish the two states almost perfectly. This suggests that $\mathcal{P}_i$, up to a small error, projects $\mathcal{E}(|i\rangle\langle i|)$ to a state close to $\mathcal{E}(|i\rangle\langle i|)$ and nearly annihilates $\mathcal{E}(\rho_{i,\perp})$. In the following two lemmas, we prove these claims rigorously. In Lemma 5.8.6, we show that $\mathcal{P}_i E_a |i\rangle \approx E_a |i\rangle$, and in Lemma 5.8.7, we show that $\mathcal{P}_i E_a |j\rangle \approx 0$, for $i \neq j$, where E_a is any Kraus operator of the channel $\mathcal{E}$.

If $\mathcal{E}$ is an ϵ-correctable channel, then we have

$$\max_{\rho} \|(\mathcal{R} \circ \mathcal{E})(\rho) - \rho\|_1 \leq 2\sqrt{2}\epsilon := \tilde{\epsilon},$$

as given by Equation (5.7.1). Let us define $\tilde{\epsilon} = 2\sqrt{2}\epsilon$ to minimize factors of $2\sqrt{2}$. Then we can obtain the following bound:

Lemma 5.8.6. *Let $\mathcal{E}$ be an ϵ-correctable channel and let $\mathcal{P}_i$ be the corresponding ghost projector with respect to some basis. Then we have*

$$\||E_a|i\rangle - \mathcal{P}_i E_a |i\rangle\|_2^2 \leq 2\sqrt{2}\epsilon := \tilde{\epsilon}, \tag{5.25}$$

where $\||\phi\rangle\|_2 := \sqrt{\langle\phi|\phi\rangle}$.

Proof. Note that, by the monotonicity of the trace norm, we have

$$\|\mathcal{E}(|i\rangle\langle i| - \rho_{i,\perp})\|_1 \geq \|(\mathcal{R} \circ \mathcal{E})(|i\rangle\langle i| - \rho_{i,\perp})\|_1.$$

We can use the fact that the recovery map $\mathcal{R}$ nearly succeeds in recovering the original state. By the triangle inequality,

$$
\begin{aligned}
2 = \||i\rangle\langle i| - \rho_{i,\perp}\|_1 &\leq \|(\mathcal{R} \circ \mathcal{E})(|i\rangle\langle i| - \rho_{i,\perp})\|_1 \\
&\quad + \||i\rangle\langle i| - (\mathcal{R} \circ \mathcal{E})(|i\rangle\langle i|)\|_1 \\
&\quad + \|\rho_{i,\perp} - (\mathcal{R} \circ \mathcal{E})(\rho_{i,\perp})\|_1.
\end{aligned}
$$

Therefore,

$$\|\mathcal{E}(|i\rangle\langle i| - \rho_{i,\perp})\|_1 \geq 2 - 2\tilde{\epsilon}. \tag{5.26}$$

Moreover, we have

$$
\begin{aligned}
\|\mathcal{E}(|i\rangle\langle i| - \rho_{i,\perp})\|_1 &= \text{Tr}(2\mathcal{P}_i \mathcal{E}(|i\rangle\langle i| - \rho_{i,\perp})) \\
&\leq 2\text{Tr}(\mathcal{P}_i \mathcal{E}(|i\rangle\langle i|)).
\end{aligned}
\tag{5.27}
$$

The first line above follows by decomposing $\mathcal{E}(|i\rangle\langle i| - \rho_{i,\perp})$ into its positive and negative parts. Because the operator is traceless, the trace of the positive part is equal to the trace of the negative part, up to a minus sign. Since the trace distance is equal to the sum of the absolute value of the positive and negative trace, and because these values are the same, we arrive at the first identity. The second line then follows from the fact that $\text{Tr}(\mathcal{P}_i \mathcal{E}(\rho_{i,\perp})) \geq 0$.

Therefore, we get the following bound:

$$
\begin{aligned}
1 - \tilde{\epsilon} &\leq \text{Tr}(\mathcal{P}_i \mathcal{E}(|i\rangle\langle i|)) \\
&= \sum_a \langle i|E_a^\dagger \mathcal{P}_i E_a |i\rangle \\
&= \sum_a q_{ia} \langle\psi_{ia}|\mathcal{P}_i|\psi_{ia}\rangle \\
&= 1 - \sum_a q_{ia}(1 - \langle\psi_{ia}|\mathcal{P}_i|\psi_{ia}\rangle),
\end{aligned}
\tag{5.28}
$$

where we define

$$|\psi_{ia}\rangle = \frac{E_a|i\rangle}{\sqrt{\langle i|E_a^\dagger E_a|i\rangle}},$$

and $q_{ia} = \langle i|E_a^\dagger E_a|i\rangle$. Note that $\sum_a q_{ia} = 1$ since $\mathcal{E}$ is trace-preserving. Therefore, we get

$$1 - \langle\psi_{ia}|\mathcal{P}_i|\psi_{ia}\rangle \le \frac{\tilde{\epsilon}}{q_{ia}}$$

by noting that the last line of Equation (5.8.2) contains a sum of non-negative terms. Since the sum is $\le \tilde{\epsilon}$, each individual term must be $\le \tilde{\epsilon}$ as well. Substituting in the expressions for q_{ia} and $|\psi_{ia}\rangle$, this inequality becomes

$$\langle i|E_a^\dagger E_a|i\rangle - \langle i|E_a^\dagger \mathcal{P}_i E_a|i\rangle \le \tilde{\epsilon},$$

which is equivalent to Equation (5.8.6). $\qquad\square$

Lemma 5.8.7. *Under the same hypothesis as Lemma 5.8.6, if $i \ne j$, then*

$$\|\mathcal{P}_i E_a|j\rangle\|_2^2 \le (\dim C)\,\tilde{\epsilon}. \qquad (5.29)$$

Proof. Note that

$$
\begin{aligned}
2 - 2\tilde{\epsilon} &\le \|\mathcal{E}(|i\rangle\langle i| - \rho_{i,\perp})\|_1 \\
&= \mathrm{Tr}(2\mathcal{P}_i\mathcal{E}(|i\rangle\langle i| - \rho_{i,\perp})) \\
&\le 2 - 2\mathrm{Tr}(\mathcal{P}_i\mathcal{E}(\rho_{i,\perp})),
\end{aligned}
$$

where we have used Equation (5.8.2) in the first line, and Equation (5.8.2) in the second. The last line follows from the fact that $\mathcal{P}_i \le I$. It follows that

$$\frac{1}{\dim C - 1}\sum_{j \ne i}\mathrm{Tr}(\mathcal{P}_i\mathcal{E}(|j\rangle\langle j|)) = \mathrm{Tr}(\mathcal{P}_i\mathcal{E}(\rho_{i,\perp})) \le \tilde{\epsilon},$$

and therefore, we have

$$\mathrm{Tr}(\mathcal{P}_i\mathcal{E}(|j\rangle\langle j|)) \le (\dim C)\,\tilde{\epsilon},$$

for all $j \ne i$. Expanding in terms of the Kraus operators of the channel $\mathcal{E}$, this becomes

$$\sum_a \mathrm{Tr}(\mathcal{P}_i E_a|j\rangle\langle j|E_a^\dagger \mathcal{P}_i) = \sum_a \|\mathcal{P}_i E_a|j\rangle\|_2^2 \le (\dim C)\tilde{\epsilon},$$

where the first equality holds because $\mathcal{P}_i$ is a projector. Equation (5.8.7) then follows. $\qquad\square$

At this point, we can follow the construction we used for the case of exact ghost operators. Let C be a code subspace and let $\mathcal{E}$ be an error channel such that $\mathcal{E}_I$ is ϵ-correctable. Then by Lemmas 5.8.6 and 5.8.7, we have

$$\||E_a|i\rangle - \mathcal{P}_i E_a |i\rangle\|_2^2 \leq 2\tilde{\epsilon}, \quad \text{and} \quad \|\mathcal{P}_i E_a|j\rangle\|_2^2 \leq 2(\dim C)\,\tilde{\epsilon},$$

where each E_a is a Kraus operators for $\mathcal{E}$, or the identity. Note that the extra factor of 2 comes from the fact that the Kraus operators for $\mathcal{E}_I$ are given by $\{E_a/\sqrt{2}\} \cup \{I/\sqrt{2}\}$, where each E_a is a Kraus operator for $\mathcal{E}$.

Given a normal operator $\tilde{T} : \tilde{\mathcal{H}} \to \tilde{\mathcal{H}}$ defined by

$$\tilde{T} = \sum_k \lambda_k |\tilde{k}\rangle\langle\tilde{k}|,$$

we define the operator

$$T = \sum_k \lambda_k \mathcal{P}_k, \tag{5.30}$$

where each $\mathcal{P}_k$ is a ghost projector with respect to the given eigenbasis for $\tilde{T}$. Then the operator T satisfies

$$
\begin{aligned}
\|T E_a |j\rangle - \lambda_j E_a |j\rangle\|_2 &= \left\| \sum_k \lambda_k \mathcal{P}_k E_a |j\rangle - \lambda_j E_a |j\rangle \right\|_2 \\
&\leq \left\| \sum_{k \neq j} \lambda_k \mathcal{P}_k E_\ell |j\rangle \right\|_2 + \left\| \lambda_j \mathcal{P}_j E_\ell |j\rangle - \lambda_j E_a |j\rangle \right\|_2 \\
&\leq \sum_{k \neq j} |\lambda_k| \, \|_2 \mathcal{P}_k E_a |j\rangle\|_2 + |\lambda_j| \, \|\mathcal{P}_j E_a |j\rangle - E_a |j\rangle\|_2 \\
&\leq \sqrt{2(\dim C)\tilde{\epsilon}} \sum_{k \neq j} |\lambda_k| + |\lambda_j| \sqrt{2\tilde{\epsilon}} \\
&\leq \sqrt{2(\dim C)\tilde{\epsilon}} \|\tilde{T}\|_1,
\end{aligned}
$$

where $\|\tilde{T}\|_1$ is the trace norm of $\tilde{T}$. Now, let $\hat{T} = V\tilde{T}V^\dagger$, where V is the code embedding. Then for a general code state $|\psi\rangle = \sum_j c_j |j\rangle$, we have

$$
\begin{aligned}
\left\| T E_a |\psi\rangle - E_a \hat{T} |\psi\rangle \right\|_2 &\leq \sum_j |c_j| \cdot \left\| T E_a |j\rangle - \lambda_j E_a |j\rangle \right\|_2 \\
&\leq \sqrt{2(\dim C)\tilde{\epsilon}} \|\tilde{T}\|_1 \sum_j |c_j| \\
&\leq (\dim C)\|\tilde{T}\|_1 \sqrt{2\tilde{\epsilon}}.
\end{aligned}
$$

A slightly weaker, but more convenient bound in terms of the operator norm of $\tilde{T}$ can be given as

$$\left\|TE_a|\psi\rangle - E_a\hat{T}|\psi\rangle\right\|_2 \leq (\dim C)^2\|\tilde{T}\|\sqrt{\left(4\sqrt{2}\right)\epsilon},$$

which we can also express as a bound on the difference of two operators in the operator norm:

$$\left\|TE_aV - E_aV\tilde{T}\right\| \leq 2^{5/4}(\dim C)^2\|\tilde{T}\|\sqrt{\epsilon}, \tag{5.31}$$

where we have used $\tilde{\epsilon} = 2\sqrt{2}\epsilon$. Note that the bound (5.8.2) holds for any Kraus representation $\{E_a\}$ of $\mathcal{E}$.

The bound (5.8.2) motivates the following definition:

Definition 5.8.8. *Let $(\mathcal{E}, K)$ be a noise channel $\mathcal{E}$ equipped with a given Kraus representation $K = \{E_a\}$. Let $\tilde{T} : \tilde{\mathcal{H}} \to \tilde{\mathcal{H}}$ be a normal operator. We say that T is a δ-approximate ghost operator for $\tilde{T}$ with respect to $(\mathcal{E}, K)$ if we have*

$$\|TE_aV - E_aV\tilde{T}\| \leq \|\tilde{T}\|\delta, \tag{5.32}$$

where $E_a \in K \cup \{I\}$ is either a Kraus operator for $\mathcal{E}$, or the identity. We say that a ghost operator T is universal *if Equation (5.8.8) holds for every Kraus representation of $\mathcal{E}$.*

Now we are ready to prove the analog of Theorem 5.8.4 in the approximate setting. As before, we say that there exists a complete set of ϵ-approximate ghost operators if there exists an ϵ-approximate ghost logical operator for every normal operator on $\tilde{\mathcal{H}}$.

Theorem 5.8.9. *Let C be a code subspace and suppose that $\mathcal{E}_I$ is ϵ-correctable for C. Then there exists a complete set of δ-approximate universal ghost operators, where*

$$\delta = 2^{5/4}(\dim C)^2\sqrt{\epsilon}. \tag{5.33}$$

For the sake of completeness, we also prove a converse of this result (Theorem 5.B.3) in Appendix 5.B. These results collectively can be seen as a generalization of the standard theorems of operator algebra quantum error-correction [25], [26] to the approximate setting.

Proof. Suppose that $\mathcal{E}_{\mathcal{I}}$ is ϵ-correctable for C. Then Equation (5.8.2) shows that T defined by Equation (5.8.2) is a δ-approximate ghost operator for any normal operator $\tilde{T}$, where $\delta = 2^{5/4} (\dim C)^2 \sqrt{\epsilon}$. The construction of the ghost projector $\mathcal{P}_k$, and therefore also the construction of T, depends only on the channel $\mathcal{E}$ and not on any particular Kraus representation; it follows that T is universal. $\qquad\square$

5.8.3 Firewalls Revisited

We have now seen that, by assuming that the state of the Hawking radiation system EB is pseudorandom, we may infer that low-complexity operations on E are approximately correctable; the code space $\tilde{B}$ that purifies the late radiation system B is protected against low-complexity operations on E. Correctability in turn implies that a complete set of ghost logical operators acting on EH, which nearly commute with all low-complexity operations on E, can be constructed.

Let us now reconsider the potential implications of the existence of ghost logical operators in the context of the black hole firewall problem. First, we assemble the results we have derived thus far to determine the value of δ for which the ghost logical operators are δ-approximate. Under the pseudorandomness assumption Equation (5.6.1), we saw in Lemma 5.7.2 that low-complexity operations are ϵ-correctable for $\epsilon = \sqrt{3/2} \cdot 2^{-(\alpha|H|-|OB|)/2}$. Since the code space dimension is $\dim C = 2^{|B|}$, Equation (5.8.9) says that the ghost operators are δ-approximate for

$$\delta = 2^{5/4} 2^{2|B|} \sqrt{\epsilon} = 2 \cdot 3^{1/4} \, 2^{2|B|} \, 2^{-(\alpha|H|-|OB|)/4} = 2 \cdot 3^{1/4} \, 2^{-(\alpha|H|-|O|-9|B|)/4}. \quad (5.34)$$

Thus, δ becomes exponentially small for asymptotically large $|H|$, $|O|$, and $|B|$, provided $|O|, |B| \ll |H|$. We could, for example, consider an encoded interior and an observer with size scaling linearly with $|H|$, and still have a complete set of ghost logical operators commuting with all low-complexity operations on E, up to exponentially small errors.

This conclusion followed only from the assumption that the state of EB is pseudorandom — we needed no other special properties of black holes to derive it. We might, in fact, expect the same pseudorandomness assumption to hold not just for black holes but also for other strongly chaotic quantum systems. But a black hole *is* special, because it has an event horizon, and it is because of the event horizon that we expect the late radiation system B to be entangled with modes behind the horizon as well as with a subspace of EH; thus arises the black hole firewall problem. To ease the firewall problem, we propose using the ghost logical operators to describe

(a portion of) the black hole interior. We would not make such a proposal for describing the "interior" of a burning lump of coal.

Pleasingly, under this proposal, it is hard for an agent who acts on the radiation to create a firewall, or to otherwise influence the black hole interior apart from exponentially small effects. To create an excitation behind the horizon, the agent outside the black hole must perform an operation of superpolynomial complexity.

We might want to allow the observer to perform a quantum computation on E which is chosen from a long list of possible unitary transformations. The observer's freedom to choose can be encoded in the observer's initial state ω_O, as depicted in Figure 5.7. If there are multiple observers $\{O_1, O_2, \ldots O_m\}$, all interacting with E, we can group them all together into a collective observer $O = O_1 O_2 \ldots O_m$. We may construct a complete set of ghost logical operators acting on the encoded black hole interior, consistently shared by all the observers, provided that $|O|, |B| \ll |H|$.

To be more concrete, suppose we want the black hole interior to be protected against any unitary transformation acting on E chosen from amongst a collection of N possible unitaries $\mathcal{U} = \{U_a\}_{a=1}^N$. We can model this situation by considering a conditional unitary transformation, controlled by an ancilla register in the observer's possession. To ensure that we can apply Theorem 5.8.9, we will add the identity transformation $U_0 = I_E$ to the list of possibilities, and envision that the observer applies

$$U_\mathcal{U} = \sum_{a=0}^N |a\rangle\langle a|_O \otimes (U_a)_E, \tag{5.35}$$

where each $|a\rangle_O$ is a computational basis state and $2^{|O|} = N+1$. Thus U_a is applied by fixing the initial state of the O register to be $|a\rangle_O$; see Figure 5.9.
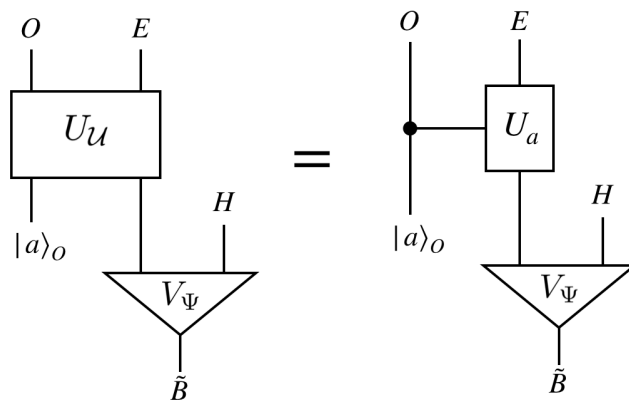

Figure 5.9: The action of the observer as a controlled unitary tranformation.

Our construction of a complete set of ghost logical operators applies — assuming the Hawking radiation is pseudorandom — if $U_{\mathcal{U}}$ has complexity polynomial in $|H|$. This will be assured if the cardinality N of the list of unitaries is polynomial in $|H|$. The unitary

$$\Lambda_a(U_a) = |a\rangle\langle a|_O \otimes (U_a)_E$$

for which a non-trivial unitary acting on E is triggered only by the basis state $|a\rangle_O$, has polynomial quantum complexity if U_a does — we show in Lemma 5.C.1 that, if we fix the complexity of U_a, then $\Lambda_a(U_a)$ can be implemented to precision ϵ with a circuit of $O(N^2 \log^4(1/\epsilon))$ two-qubit gates. Furthermore, the overall operator

$$U_{\mathcal{U}} = \prod_{a=0}^{N} \Lambda_a(U_a)$$

is a product of $N + 1$ such unitaries, and thus has complexity at worst a factor of $N + 1$ larger. Therefore, if $N = \mathrm{poly}(|H|)$, then $U_{\mathcal{U}}$ can be executed to exponential precision with a circuit of size $\mathrm{poly}(|H|)$.

The unitary transformation $U_{\mathcal{U}}$ is a dilation of the quantum channel

$$\mathcal{E}_{\mathcal{U}}(\rho) = \frac{1}{N+1} \sum_{a=0}^{N} U_a \rho U_a^\dagger \tag{5.36}$$

acting on E, with Kraus operators $\{U_a\}_{a=0}^{N}$. Because $U_{\mathcal{U}}$ has polynomial complexity, under the pesudorandomness assumption, a complete set of δ-approximate ghost logical operators can be constructed, with δ given by Equation (5.8.3). In other words, for each unitary U_a that the observer might apply, U_a commutes with all ghost logical operators up to an exponentially small error. Hence no matter which low-complexity unitary the observer applies, the encoded black hole interior is hardly affected at all.

This conclusion is summarized by the following theorem:

Theorem 5.8.10. *Suppose that the decoupling bound (5.6) holds. Let $V : \mathcal{H}_{\tilde{B}} \to \mathcal{H}_{EH}$ denote the black hole code embedding. Let $\mathcal{U} = \{U_a\}_{a=1}^{N}$ denote an arbitrary set of $N = \mathrm{poly}(|H|)$ unitaries acting on the early radiation E, where each unitary has complexity $\mathrm{poly}(|H|)$. Then there exists a complete set of logical operators $\mathcal{L} \subseteq \mathcal{B}(\mathcal{H}_{EH})$ for the black hole code such that for all $T \in \mathcal{L}$, and all $U_a \in \mathcal{U}$, we have*

$$\|[U_a, T]V\| \le 2\delta' \|\tilde{T}\|, \tag{5.37}$$

and

$$\|TV - V\tilde{T}\| \le \delta'\|\tilde{T}\|, \tag{5.38}$$

where $\tilde{T}$ is the operator on $\tilde{B}$ corresponding to T, and

$$\delta' = 8 \cdot 6^{1/4} \, 2^{-\alpha|H|/4}(N+1)^{3/4} \tag{5.39}$$

if $\tilde{B}$ is a single qubit ($|B| = 1$).

Proof. Let us model the observer O on the Hilbert space $\mathcal{H}_O = \mathbb{C}^{N+1}$, so that $2^{|O|} = N + 1$. In Lemma 5.C.1, we show that the conditional unitary $U_{\mathcal{U}}$ defined in Equation (5.8.3) can be approximated to exponential accuracy with a circuit of size $\mathrm{poly}(|H|)$ if each U_a has complexity $\mathrm{poly}(|H|)$ and $N = \mathrm{poly}(|H|)$; therefore, under the pseudorandomness assumption, $U_{\mathcal{U}}$ is ϵ-correctable with

$$\epsilon = \sqrt{\frac{3}{2}} \cdot 2^{-(\alpha|H|-|OB|)/2},$$

and hence there exists a complete set of δ-approximate ghost logical operators for $U_{\mathcal{U}}$ with

$$\delta = 2 \cdot 3^{1/4} \, 2^{-(\alpha|H|-|O|-9|B|)/4},$$

or

$$\delta = 8 \cdot 6^{1/4} \, 2^{-\alpha|H|/4}2^{|O|/4} = 8 \cdot 6^{1/4} \, 2^{-\alpha|H|/4}(N+1)^{1/4}$$

if $|B| = 1$. The Kraus operators for the channel $\mathcal{E}_{\mathcal{U}}$ in Equation (5.8.3) are $\{U_a/\sqrt{N+1}\}$; hence

$$\|TU_aV - U_aV\tilde{T}\| \le \|\tilde{T}\|\delta\sqrt{N+1} = \|\tilde{T}\|\delta'.$$

This, together with Lemma 5.B.2, gives the desired results in Equation (5.8.10) and Equation (5.8.10). $\square$

Note that, although δ' in Equation (5.8.10) could be exponentially small even for superpolynomial N, we required $N = \mathrm{poly}(|H|)$ because only in that case have we shown that the conditional unitary $U_{\mathcal{U}}$ has complexity $\mathrm{poly}(|H|)$; we needed this property for the pseudorandomness assumption to imply that the observer is unable to distinguish the state of EB from a maximally mixed state.

We have inferred the existence of ghost logical operators which act on EH. It should also be possible to realize a non-trivial logical operator as a physical operator acting

on E alone, but only if that operator is computationally complex to construct. For instance, suppose that $W : \mathcal{H}_E \to \mathcal{H}_E$ is a unitary logical operator that can be accurately approximated by a quantum circuit of polynomial size. Then there exists a ghost logical operator T that fails to commute with W acting on the code space. Since W has polynomial complexity, this contradicts Theorem 5.8.10, and we conclude that no such W can exist. This conclusion resonates with the observations of Bouland, Fefferman, and Vazirani, who argued that in the context of AdS/CFT duality, the dictionary relating the black hole exterior and interior should be computationally complex [43], [44].

On the other hand, if a quantum circuit is allowed to act on H as well as E, and if B has constant size, then any logical operator on the code space *can* be realized efficiently. We show this in Section 5.9.

5.8.4 State Dependence

The (approximate) encoding isometry $V_\Psi : \mathcal{H}_{\tilde{B}} \to \mathcal{H}_{EH}$ is determined by the pure quantum state Ψ_{EHB} of the black hole H and its emitted Hawking radiation EB. This state, and hence the encoding map, depends on the initial microstate of the infalling matter that collapsed to form the black hole. Therefore, the encoded interior of the black hole is said to be "state dependent" [7], [10].

If black hole evaporation is unitary, and the event horizon is smooth because the black hole interior is encoded in the radiation, then state dependence of the encoding seems to be unavoidable; if the quantum information encoded in the initial state is preserved in the final state of the fully evaporated black hole, then how the late radiation emitted after the Page time is entangled with the early radiation emitted before the Page time must depend on that initial state. This state dependence of the encoding is nonetheless troubling [8], [9], [16]. If the experiences of observers who fall through the event horizon are described by the logical operators of the code, and these logical operators are state dependent, then the observers inside the black hole seem to be capable of measuring nonlinear operators acting on Ψ_{EHB}, rather than linear operators as in the standard theory of quantum measurement. This ability to measure nonlinear properties of the state could lead to inconsistencies. We regard this as an unresolved issue, reflecting our incomplete understanding of how to describe measurements conducted behind black hole horizons.

But the state-dependent encoding of the black hole interior is not sufficient by itself

to solve the black hole firewall problem.[8] If the Hawking radiation is thoroughly scrambled, then we expect that the interior mode that purifies B can be decoded by acting on E alone after the Page time [45], and therefore that the logical operators of the code may also be chosen to act on E alone. If T and S are two noncommuting logical operators, where S acts on E, then an observer (Bob) outside the black hole who applies S could in principle alter the outcome of a measurement of T performed by an observer (Alice) inside the black hole. Thus Bob can send an instantaneous message to Alice, in apparent violation of relativistic causality.

While we agree that such acausal signaling is possible in principle, we insist that the computational complexity of the task should be considered. Under the assumption that the Hawking radiation is pseudorandom, we have found that, in order to signal Alice, Bob must apply an operation to E with complexity superpolynomial in $|H|$, if Alice's observables are the ghost logical operators we have constructed. Though possible, such an operation is infeasible in practice if the black hole H is macroscopic; therefore the semiclassical causal structure of the spacetime is respected.

5.9 Inside the Black Hole

Under our pseudorandomness assumption, an observer who acts on the early radiation system E can affect the encoded interior of a black hole only by applying an operation with superpolynomial complexity. However, an agent who has access to the black hole system H as well as E can manipulate the interior efficiently. Here we construct an efficient unitary circuit $\bar{U}_{EH}$, acting on EH, that perturbs the encoded interior. Our construction makes use of an efficient quantum circuit that realizes the unitary U_{bh} that describes the formation and partial evaporation of a black hole. This unitary creates a state in which B is maximally entangled with a subspace of EH; if the circuit that implements U_{bh} is efficient, then $\bar{U}_{EH}$ can be implemented efficiently as well. We will also see that an agent with access to EH can efficiently decode the interior, distilling the code subspace of EH to a small quantum memory.

Suppose we are given a unitary operator U_{BEH} which realizes the map

$$U_{BEH}|0\rangle_B|0\ldots0\rangle_{EH} = \frac{1}{\sqrt{2}}(|0\rangle_B|\psi_0\rangle_{EH} + |1\rangle_B|\psi_1\rangle_{EH}),$$

where B is a single qubit, and EH is n qubits. By applying the circuits that implement U_{BEH} and $U_{BEH}^{\dagger}$ on an ancillary register, together with some additional gates acting

---

[8]We thank Raphael Bousso for raising this issue.

on the ancilla and EH, we will apply a unitary operator $\bar{U}_{EH}$ acting on EH with the property that

$$\bar{U}_{EH}|\psi_0\rangle_{EH} = v_{00}|\psi_0\rangle_{EH} + v_{10}|\psi_1\rangle_{EH},$$

$$\bar{U}_{EH}|\psi_1\rangle_{EH} = v_{01}|\psi_0\rangle_{EH} + v_{11}|\psi_1\rangle_{EH},$$

where

$$v = \begin{pmatrix} v_{00} & v_{01} \\ v_{10} & v_{11} \end{pmatrix}$$

is some chosen 2×2 unitary matrix. That is, $\bar{U}_{EH}$ applies an arbitrary "logical" unitary transformation on the two-dimensional "code space" spanned by $\{|\psi_0\rangle_{EH}, |\psi_1\rangle_{EH}\}$.

The protocol is explained in two steps. First, we describe a probabilistic protocol which applies $\bar{U}_{EH}$ with success probability $\frac{1}{4}$. Next, using the probabilistic protocol, we build a deterministic protocol which applies $\bar{U}_{EH}$ with probability 1. The first protocol applies a unitary $U_{a_1a_2}$ and $U_{a_1a_2}^{\dagger}$ once each. Here, $U_{a_1a_2}$ is a unitary acting on an ancillary register $a = a_1a_2$ and can be realized by applying the circuit that implements U_{BEH} on register a_1 and a_2. The register B is replaced with a_1 and the register EH is replaced with a_2. The second protocol applies $U_{a_1a_2}$ and $U_{a_1a_2}^{\dagger}$ three times each. We also use some additional gates, which are also efficient.

For the probabilistic protocol, consider the following sequence of operations:

1. Initialize a in the $|0\ldots0\rangle$ state.

2. Apply $U_{a_1a_2}$.

3. Apply a swap between a_2 and EH.

4. Apply the single-qubit operation v^T to a_1.

5. Apply $U_{a_1a_2}^{\dagger}$.

6. Measure the a register in the computational basis, and postselect on measuring the all-0 bit string.

Applying this protocol for $U_{BEH} = U_{bh}$, and taking the initial state to be $|\phi_{matter}\rangle = |00\ldots0\rangle$, we obtain the circuit diagram in Figure 5.10.

Let us analyze what happens when this protocol is executed. Suppose the state of EH is an arbitrary pure quantum state $|\psi\rangle_{EH}$. After the second step, we have

$$\frac{1}{\sqrt{2}}\left(|0\rangle_{a_1}|\psi_0\rangle_{a_2} + |1\rangle_{a_1}|\psi_1\rangle_{a_2}\right)|\psi\rangle_{EH}.$$
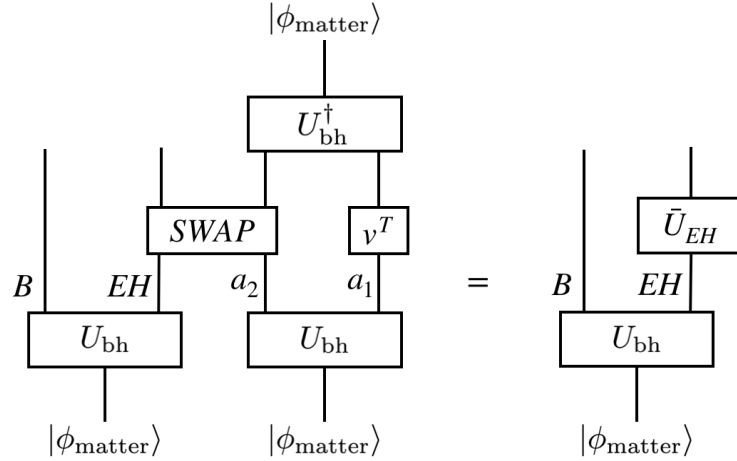
Figure 5.10: A probabilistic protocol which (with success probability $1/4$) applies an arbitrary unitary operator v to $\tilde{B}$, the encoded interior partner of B. Here $\bar{U}_{EH}$ denotes v acting on the code subspace of EH.

Now expand $|\psi\rangle_{EH}$ in an orthonormal basis that includes both $|\psi_0\rangle$ and $|\psi_1\rangle$; the remaining $2^n - 2$ elements of the basis set are labeled $|\psi_i\rangle$ from $i = 2$ to $i = 2^n - 1$, so that

$$|\psi\rangle = \sum_i \lambda_i |\psi_i\rangle.$$

After the third step, we obtain

$$\frac{1}{\sqrt{2}}(|0\rangle_{a_1}|\psi_0\rangle_{EH} + |1\rangle_{a_1}|\psi_1\rangle_{EH})|\psi\rangle_{a_2},$$

which after the fourth step becomes

$$\frac{1}{\sqrt{2}}((v_{00}|0\rangle_{a_1} + v_{01}|1\rangle_{a_1})|\psi_0\rangle_{EH} + (v_{10}|0\rangle_{a_1} + v_{11}|1\rangle_{a_1})|\psi_1\rangle_{EH})|\psi\rangle_{a_2}$$

$$= \frac{1}{\sqrt{2}}(|0\rangle_{a_1}(v_{00}|\psi_0\rangle_{EH} + v_{10}|\psi_1\rangle_{EH}) + |1\rangle_{a_1}(v_{01}|\psi_0\rangle_{EH} + v_{11}|\psi_1\rangle_{EH}))|\psi\rangle_{a_2}.$$

Now we want to study what happens after we carry out the fifth and the sixth step. Instead of explicitly applying $U^{\dagger}_{a_1 a_2}$, it is more convenient to think about an orthogonal measurement in a basis that includes $U_{a_1 a_2}|0\ldots0\rangle_a = \frac{1}{\sqrt{2}}(|0\rangle_{a_1}|\psi_0\rangle_{a_2} + |1\rangle_{a_1}|\psi_1\rangle_{a_2})$. After projecting onto this state, we obtain the (subnormalized) state

$$\frac{1}{2}(\lambda_0(v_{00}|\psi_0\rangle_{EH} + v_{10}|\psi_1\rangle_{EH}) + \lambda_1(v_{01}|\psi_0\rangle_{EH} + v_{11}|\psi_1\rangle_{EH}))$$

$$= \frac{1}{2}((v_{00}\lambda_0 + v_{01}\lambda_1)|\psi_0\rangle_{EH} + (v_{10}\lambda_0 + v_{11}\lambda_1)|\psi_1\rangle_{EH}),$$

which, aside from the normalization factor of $1/2$, is equivalent to applying v to the code vector $\lambda_0|\psi_0\rangle_{EH} + \lambda_1|\psi_1\rangle_{EH}$. Hence, $\bar{U}_{EH}$ is applied with success probability $1/4$.

Now we explain how to upgrade this probabilistic operation to a unitary quantum circuit that applies $\bar{U}_{EH}$ deterministically. For this purpose, we use the *oblivious amplitude amplification* technique introduced by Berry *et al.*; see Lemma 3.6 of [46]. For the reader's convenience, we restate this result.

Lemma 5.9.1. *(Oblivious amplitude amplification) Let V' and V be unitary matrices on $\mu + n$ qubits and n qubits respectively, and let $\theta \in (0, \pi/2)$. Suppose that for any n-qubit state $|\psi\rangle$,*

$$V'|0^\mu\rangle|\psi\rangle = \sin(\theta)|0^\mu\rangle V|\psi\rangle + \cos(\theta)|\Phi^\perp\rangle,$$

where $(|0^\mu\rangle\langle 0^\mu| \otimes I)|\Phi^\perp\rangle = 0$. Let $R = 2|0^\mu\rangle\langle 0^\mu| \otimes I - I$ and $S = -V'RV'^\dagger R^\dagger$. Then,

$$S^\ell V'|0^\mu\rangle|\psi\rangle = \sin((2\ell+1)\theta)|0^\mu\rangle V|\psi\rangle + \cos((2\ell+1)\theta)|\Phi^\perp\rangle.$$

In our case, V' is the unitary process described in the first five steps, V is $\bar{U}_{EH}$, $|0^\mu\rangle$ is $|0\dots 0\rangle_a$, and $\sin(\theta) = \frac{1}{2}$. Therefore, $\theta = \frac{\pi}{6}$, and we can choose $\ell = 1$ to apply V deterministically. For this choice of ℓ, it suffices to apply V' twice and its inverse once to achieve V. For each V', we apply $U_{a_1a_2}$ and its inverse $U_{a_1a_2}^\dagger$ once each (as well as other simple unitary operations). In total, then, we can deterministically apply $\bar{U}_{EH}$ by using $U_{a_1a_2}$ three times and $U_{a_1a_2}^\dagger$ three times. In particular, the entire circuit is efficient if $U_{a_1a_2}$ is. Applying this protocol for $U_{BEH} = U_{\text{bh}}$, we obtain the circuit diagram in Figure 5.11.

More generally, suppose that the register B contains $|B| > 1$ qubits, so that the code subspace of EH has dimension $2^{|B|}$. A probabilistic protocol for applying an arbitrary unitary transformation to the code space can be constructed that closely follows the construction for a single qubit, but now with success probability $2^{-2|B|}$. In particular, using the probabilistic protocol and oblivious amplitude amplification we can approximate any two-qubit gate ($|B| = 2$) acting on the code space accurately and efficiently.

From a universal set of such two-qubit gates, we can build a logical unitary circuit. Hence any low-complexity operation on the code space can be realized as a low-complexity quantum circuit acting on EH.
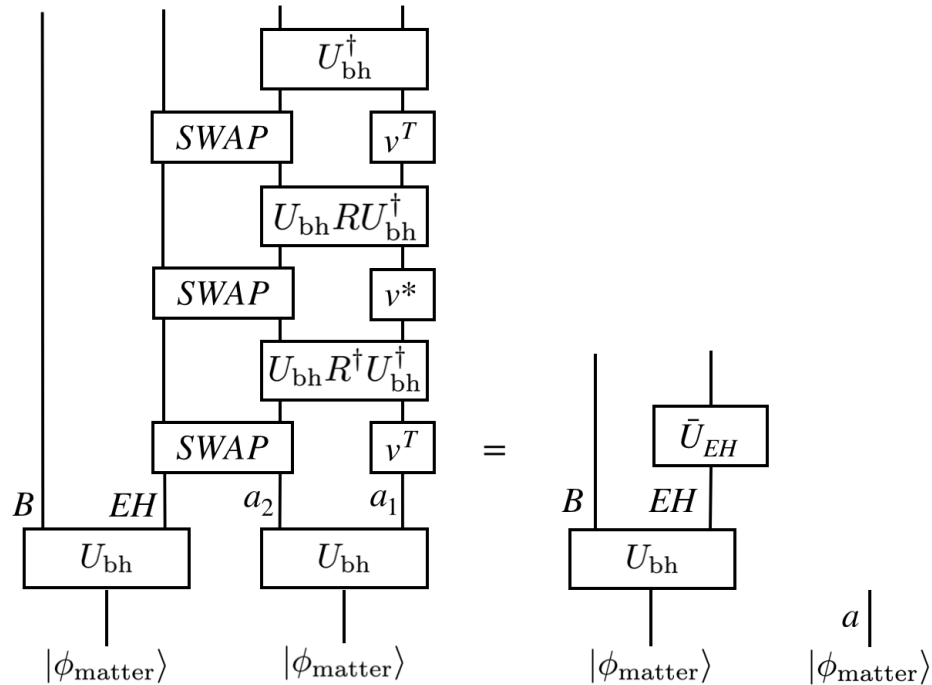
Figure 5.11: A deterministic circuit which applies an arbitrary unitary operator v to $\tilde{B}$, the encoded interior partner of B. Here $\bar{U}_{EH}$ denotes v acting on the code subspace of EH. Note that the final $U_{\text{bh}}^\dagger$ and v^T acting on the ancilla can be removed without changing how the circuit acts on the code space.

If we can perform logical gates on the code space, then we can also decode the logical state, distilling it to a small quantum memory in our possession. To be concrete, suppose the code space is two-dimensional. To decode, it suffices to prepare an ancilla qubit b in an arbitrary state, and then perform a SWAP operation on b and the encoded qubit. For this purpose, we can use the quantum circuit identity shown in Figure 5.12, where SWAP is constructed from controlled-X, controlled-Z, and Hadamard gates. The Hadamard gates act on b, and the C-X and C-Z gates act with b as the control qubit and the code space as the target qubit.
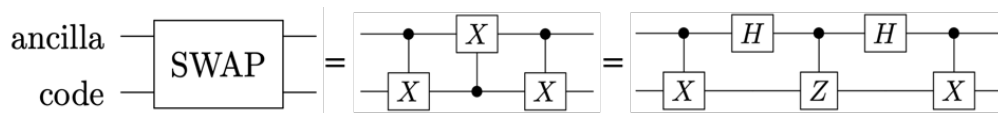

Figure 5.12: A two-qubit SWAP gate can be expressed in terms of Hadamard gates, controlled-X gates, and a controlled-Z gate. If there are efficient circuits for the X and Z gates acting on the code space, we may replace the gates in these circuits by gates controlled by an ancilla qubit, and use this identity to build a circuit that swaps the logical qubit in the code space with the ancilla qubit.

Suppose we have a circuit acting on EH that applies X to the code. We can replace each gate in that circuit by a controlled gate, with b as the control qubit. The resulting circuit applies C-X with b as the control qubit, and if the circuit for X is efficient, so is the circuit for C-X. Likewise, we can turn an efficient circuit acting on EH that applies Z to the code into an efficient circuit for C-Z. Using the circuit identity, we obtain an efficient circuit acting on b and EH that swaps the encoded information into b. Using this realization of the SWAP gate, the entangled state of B with the encoded interior mode $\tilde{B}$ becomes an entangled state of B and b.

Note that this construction of logical gates, and of the decoding circuit, can also be applied to the fully evaporated black hole. After the evaporation is complete, H is gone, but any Hawking radiation qubit B is entangled with a highly scrambled subspace of E, a large system composed of all the other radiation quanta. Because the evolution of the initial infalling matter to the final outgoing Hawking radiation is described by an efficient unitary transformation U, we have seen how $U_{a_1 a_2}$ and $U_{a_1 a_2}^{\dagger}$ can be used three times each to construct either X or Z acting on the encoded qubit. By replacing the gates in U by controlled gates, we can construct the SWAP operator, and hence distill the encoded qubit which is entangled with B into a small quantum memory efficiently.

5.10   Conclusion

From a purely quantum information perspective, the results in this paper apply to a tripartite pure state Ψ_{EHB}, where $|E| \gg |H| \gg |B|$. Our central assumption, from which all else follows, is that the marginal state ρ_{EB} is *pseudorandom* — *i.e.*, cannot be distinguished from a maximally mixed state with a bias better than $2^{-\alpha|H|}$ by any quantum computation with complexity polynomial in $|H|$. From this assumption, it follows that if a unitary transformation with complexity poly($|H|$) acts on E and an observer O, then B and O *decouple* in the resulting state Ψ'_{OEHB}, *i.e.*, $\rho'_{OB} \approx \rho'_O \otimes \rho'_B$ up to an error $O\left(2^{-\alpha|H|+|O|+|B|}\right)$. Here $\alpha = O(1)$ is a positive constant.

The state Ψ_{EHB} also defines an encoding map $V_\Psi : \mathcal{H}_{\tilde{B}} \to \mathcal{H}_{EH}$, whose image is a subspace of EH that is nearly maximally entangled with B. From the decoupling condition, we can infer that the encoded system $\tilde{B}$ is hard to decode if $\alpha|H|-|O|-|B| \gg 1$; the observer can distill $\tilde{B}$ to a small subsystem only by performing an operation with complexity superpolynomial in $|H|$. Furthermore, if the observer O performs any quantum computation on E with complexity poly($|H|$), there is a recovery operator $\mathcal{R}$ acting on EH that corrects this "error" with fidelity

$F = 1 - \epsilon$ where $\epsilon = O\left(2^{-\alpha|H|/2+|O|/2+|B|/2}\right)$. Here the size $|O|$ of the observer O may be interpreted as the number of qubits in O's quantum memory, or equivalently as the Kraus rank of the quantum channel applied to E by O.

The existence of such a recovery operator $\mathcal{R}$ has a further implication. We can construct a complete set of *ghost logical operators* for $\tilde{B}$ acting on EH; if O applies a quantum channel to E with complexity poly($|H|$), then these ghost operators commute with all the Kraus operators of the channel, up to an error $O\left(2^{-\alpha|H|/4+|O|/4+9|B|/4}\right)$. Thus the ghost operators fail to detect the action of any observer who performs an operation on E with complexity poly($|H|$).

For quantum informationists, these results may be viewed as a contribution to the theory of operator algebra quantum error-correcction in the approximate setting. What can be said about their potential physical consequences?

The existence of pseudorandom quantum states that can be prepared by quantum circuits with depth $O(\text{polylog}|H|)$ follows from standard assumptions used in post-quantum cryptography [12]. Because black holes are efficient scramblers of quantum information, it is plausible that a pseudorandom state can be efficiently prepared by an evaporating black hole, where the black hole microstates of H provide the concealed "key" of the state. A similar remark may apply to other strongly chaotic systems as well. In the setting of black holes, our conclusion about the hardness of decoding the Hawking radiation of an old black hole builds on the work of Harlow and Hayden [14] by highlighting the role of pseudorandomness, and by clarifying that the condition $|H| \gg 1$ already ensures that decoding is hard — even if $|H|$ is much smaller than $|E|$.

We require in addition that $|H|$ is sufficiently large compared to the size $|O|$ of the observer's quantum memory, though we may allow the observer to wield a large probe system P which interacts with E, where $|P| \gg |H|, |O|$. In that case, the system $\tilde{B}$ becomes encoded in PEH rather than EH. However, the conclusion that $\tilde{B}$ cannot be efficiently distilled to a subsystem of size $|O|$ still applies for $|O| \ll \alpha|H|$, if B has constant size. Therefore, no agent with reasonable computational power can decode $\tilde{B}$ and carry it into the black hole without incurring a substantial backreaction on the black hole geometry.

To evade the black hole firewall problem, it has been proposed that (part of) the interior of an old black hole past its Page time is actually encoded in the radiation system E emitted long ago. This encoding is profoundly nonlocal and therefore

potentially problematic — why cannot an agent far outside the black hole who acts on E send instantaneous messages to observers who are inside, or even create a firewall at the event horizon? Our view is that computational complexity should be invoked to reconcile the nonlocal encoding of the interior with the semiclassical causal structure of the black hole geometry.

The finding that ghost logical operators can be constructed when the Hawking radiation is pseudorandom fits neatly with this viewpoint. We propose that the observables accessible to observers inside the black hole are described by these ghost logical operators, though admittedly we have no compelling general basis for this claim other than to address the firewall problem. If we accept the claim, it follows that an agent outside the black hole can create detectable excitations behind the horizon only by performing operations of superpolynomial complexity. This conclusion, though based on different arguments, meshes with the proposal by Bouland *et al.* [43], [44], that the dictionary relating the black hole interior to its exterior in the context of AdS/CFT duality must be computationally complex.

In our discussion, the encoding map relating the interior system $\tilde{B}$ to the early radiation E and remaining black hole H depends on the microstate of the initial collapsing body from which the black hole formed. It can also depend on how the observer interacts with the radiation [47]. Specifically, an observer who controls a large probe system P that comes into contact with E is empowered to alter the encoding substantially. But modifying the code does not help the observer to decode the radiation or to send a message to the interior — achieving either task by acting on E requires an operation with complexity superpolynomial in $|H|$.

Once an observer falls through the event horizon, the interior of the black hole should become accessible. From our point of view, this interior observer can interact not just with E but also with H, which makes the task of manipulating the interior far easier. Indeed, for a code space of constant dimension, arbitrary unitary transformations on the code space can be realized by quantum circuits acting on EH with complexity $\text{poly}(|EH|)$.

It is a familiar notion that, even in a theory of quantum gravity, local effective field theory on a curved background can provide an excellent approximation when the spacetime curvature is sufficiently small and the energy is sufficiently low. The story of ghost logical operators indicates that further constraints may need to be satisfied for physics to be approximately local: operations must have sufficiently low complexity and Kraus rank. Operations with high complexity and/or high rank

can tear spacetime apart.

Our description of the robust encoded interior of an old black hole highlights the effectiveness of quantum error-correction against a nonstandard noise model. In the setting of fault-tolerant quantum computing, we normally seek an encoding that can protect against weakly correlated errors with a relatively low error rate. Here, though, the "noise" inflicted by our observer O on the early radiation system E is strong and chosen adversarially. As long as this noise process has computational complexity poly$(|H|)$ and sufficiently small Kraus rank, the encoded system $\tilde{B}$ can be restored with high fidelity, and the ghost logical operators are barely affected at all. What makes this protection possible is that, although E is treated very harshly, the "key space" H is assumed to be noiseless. Perhaps related ideas can be exploited to protect quantum information in other physically relevant settings.

Acknowledgements

5.A   Approximate Embedding

Lemma 5.A.1. *Let $|\Psi\rangle_{EBH}$ be a pseudo-random state (see Definition 5.6.1), where B is a single qubit. Then the operator V_Ψ defined by Equation (5.7) is an approximate embedding, i.e., there exists an embedding V such that*

$$\|V - V_\Psi\| \leq 2 \cdot 2^{-\alpha|H|}.$$

Proof. Let $\rho_{EBH} = |\Psi\rangle\langle\Psi|_{EBH}$. Applying the decoupling inequality (5.6) without the presence of an observer (i.e., taking $|O| = 0$), we see that ρ_B is nearly maximally

mixed, i.e.,

$$\|\rho_B - \frac{1}{2}I_B\|_1 \leq 2^{-\alpha|H|}.$$

Equivalently, this implies that

$$\|V_\Psi^\dagger V_\Psi - I_{\tilde{B}}\|_1 \leq 2 \cdot 2^{-\alpha|H|} := \epsilon. \tag{5.40}$$

Now, let $U\Sigma_\Psi W^\dagger = V_\Psi$ be the singular value decomposition for V_Ψ. Let us denote the singular values of V_Ψ as $\{\sigma_k\}$. Then (5.A) implies that we have $|\sigma_k^2 - 1| \leq \epsilon$. Since $|\sigma_k + 1| \geq 1$ (the singular values are nonnegative real numbers), we then also have

$$|\sigma_k - 1| \leq \epsilon \cdot |\sigma_k + 1|^{-1} \leq \epsilon.$$

Now, let Σ denote the matrix with the same shape as Σ_Ψ whose diagonal values are all equal to 1. Define $V = U\Sigma W^\dagger$, and note that V is an isometric embedding since $V^\dagger V = I_{\tilde{B}}$. Finally, we have

$$\begin{aligned}
\|V - V_\Psi\| &= \|U(\Sigma - \Sigma_\Psi)W^\dagger\| \\
&\leq \|U\| \cdot \|W^\dagger\| \cdot \|\Sigma - \Sigma_\Psi\| \\
&\leq \epsilon,
\end{aligned}$$

where the last inequality follows since all singular values of $\Sigma - \Sigma_\Psi$ are bounded above by ϵ by construction. $\qquad\square$

5.B Complete Set of Ghost Operators Implies Correctability

In this Appendix, we prove a converse to Theorem 5.8.9, showing that if a quantum error-correcting code C has a complete set of δ-approximate ghost logical operators for a channel

$$\mathcal{E}(\rho) = \sum_{a=1}^{r} E_a \rho E_a^\dagger$$

with a set of r Kraus operators $K = \{E_a\}$, then the channel $\mathcal{E}_I$ with Kraus operators $K \cup \{I\}$ is ϵ-correctable for C, where $\epsilon = O(|K|\sqrt{(\dim C)\,\delta})$.

For this purpose, we will use the approximate version of the Knill-Laflamme error-correction conditions studied by Bény and Oreshkov [23]; these may be expressed in the form

$$PE_a^\dagger E_b P = \lambda_{ab} P + B_{ab},$$

where P is the projector on the code space C, λ_{ab} is a density matrix (a non-negative Hermitian operator with trace 1), and for each a and b, B_{ab} is an operator mapping C to C. For $B_{ab} = 0$, these are the usual Knill-Laflamme conditions for exact correctability [42]. If B_{ab} is small, the Knill-Laflamme conditions are approximately satisfied, and a recovery operator $\mathcal{R}$ exists that corrects the channel $\mathcal{E}$ acting on the code space, up to a small error ϵ as in Equation (5.7.1).

A relation between B_{ab} and ϵ was derived in [23]. We define maps Λ and $\mathcal{B}$ by

$$\Lambda(\rho) = \sum_{a,b=1}^{r} \lambda_{ab}\mathrm{Tr}(\rho)|a\rangle\langle b|, \quad \text{and} \quad \mathcal{B}(\rho) = \sum_{a,b=1}^{r} \mathrm{Tr}(\rho B_{ab})|a\rangle\langle b|, \quad (5.41)$$

respectively. Consider the Bures distance $\mathfrak{B}(\Lambda + \mathcal{B}, \Lambda)$ defined as in Equation (5.7.1), with the maximum taken over all code states ρ. Then the noise channel $\mathcal{E}$ is ϵ-correctable for the code C if and only if $\mathfrak{B}(\Lambda + \mathcal{B}, \Lambda) \le \epsilon$ [23].

We may estimate this Bures distance as in Equation (5.7.1), finding

$$2\mathfrak{B}^2(\Lambda + \mathcal{B}, \Lambda) \le \max_{\rho} \|(\mathcal{B} \otimes \mathcal{I})(|\psi\rangle\langle\psi|)\|_1, \quad (5.42)$$

where $|\psi\rangle$ is a purification of the logical density operator ρ. Using Equation (5.B), we obtain

$$\begin{aligned}
\|(\mathcal{B} \otimes \mathcal{I})(|\psi\rangle\langle\psi|)\|_1 &= \left\| \sum_{a,b=1}^{r} \langle\psi|B_{ab}|\psi\rangle|a\rangle\langle b| \right\|_1 \\
&\le r^2 \max_{a,b} |\langle\psi|B_{ab}|\psi\rangle| \\
&\le r^2 \max_{a,b} \|B_{ab}\| \\
&\le r^2(\dim C) \max_{a,b} \|B_{ab}\|_{\max}.
\end{aligned} \quad (5.43)$$

Here the entry-wise max norm of $\|A\|_{\max}$ of a matrix A is defined as the largest (in absolute value) entry of the matrix in the computational basis; *i.e.*,

$$\|A\|_{\max} = \max_{i,j} |\langle i|A|j\rangle|,$$

and we used an inequality relating the operator and max norms,

$$\|B_{ab}\| \le (\dim C)\|B_{ab}\|_{\max}.$$

We can now prove:

Lemma 5.B.1. *The channel*

$$\mathcal{E}(\rho) = \sum_{a=1}^{r} E_a \rho E_a^{\dagger}$$

is ϵ-correctable with respect to the code C, with

$$\epsilon = r\sqrt{\frac{1}{2}(\dim C)\,\delta},$$

if there is a density operator λ_{ab} and an orthonormal basis $\{|i\rangle\}$ for the code space such that for all i and j,

$$\left|\langle i|E_a^{\dagger}E_b|j\rangle - \delta_{ij}\lambda_{ab}\right| \le \delta. \tag{5.44}$$

Proof. According to the Bény-Oreshkov criterion [23], the channel is ϵ-correctable if $\mathfrak{B}^2(\Lambda + \mathcal{B}, \Lambda) \le \epsilon^2$, and from Equations (5.B) and (5.B), we have

$$\mathfrak{B}^2(\Lambda + \mathcal{B}, \Lambda) \le \frac{1}{2}r^2(\dim C)\max_{a,b}\|B_{ab}\|_{\max} \le \frac{1}{2}r^2(\dim C)\delta,$$

where we derived the last inequality from the definition of the $\|\cdot\|_{\max}$ norm and Equation (5.B.1). This proves the Lemma. $\square$

We will use the following Lemma in the proof of Theorem 5.8.10, as well as in the proof of Theorem 5.B.3 below.

Lemma 5.B.2. *Let C be a code subspace with code projector P. Let T be an δ-approximate ghost operator for the channel $\mathcal{E}$ and the set of Kraus operators K. Then*

$$\|[T, E]P\| \le 2\delta\|\tilde{T}\| \tag{5.45}$$

for all $E \in K$.

Proof. Let V be the code embedding. By definition of the ghost operator, we have

$$\|TEV - EV\tilde{T}\| \le \delta\|\tilde{T}\|,$$

for all $E \in K \cup \{I\}$. Taking $E = I$ gives

$$\|TV - V\tilde{T}\| \le \delta\|\tilde{T}\|. \tag{5.46}$$

Then we have

$$
\begin{aligned}
\|[T, E]V\| = \|TEV - ETV\| = \|TEV - EV\tilde{T} + EV\tilde{T} - ETV\| \\
\leq \|TEV - EV\tilde{T}\| + \|EV\tilde{T} - ETV\| \\
\leq 2\delta\|\tilde{T}\| + \|E\| \cdot \|V\tilde{T} - TV\| \\
\leq 2\delta\|\tilde{T}\|,
\end{aligned}
\tag{5.47}
$$

where in the last line we used Equation (5.B) and the fact that $\|E\| \leq 1$ since $E^\dagger E \leq I$ implies $\|E^\dagger E\| = \|E\|^2 \leq 1$. We can now obtain Equation (5.B.2) if we can replace V in Equation (5.B) by P. This is justified because, for any operator A, we have

$$
\|AP\| = \|AVV^\dagger\| \leq \|AV\| \cdot \|V^\dagger\| \leq \|AV\|,
$$

where we have used $\|V^\dagger\| \leq 1$ in the last line since V is an isometric embedding. $\qquad\square$

With these Lemmas in hand, we can proceed to prove:

Theorem 5.B.3. *Suppose that there exists a complete set of δ-approximate ghost logical operators for the channel $\mathcal{E}$ and its set of Kraus operators $K = \{E_a\}$. Then $\mathcal{E}_{\mathcal{I}}$ is ϵ-correctable for the code C, where*

$$
\epsilon = (|K| + 1)\sqrt{2(\dim C)\,\delta}.
$$

Proof. Suppose that there exists a complete set of δ-approximate ghost logical operators for $\mathcal{E}$ with respect to some Kraus decomposition $K = \{E_a\}_{a=1}^r$. We will also define $E_0 = I$.

Given any two orthogonal code states $|\psi\rangle, |\phi\rangle \in C$, let us define the operators $\tilde{T}_1$ and $\tilde{T}_2$ as in the proof of Theorem 5.8.4. Note that $\|\tilde{T}_1\| = \|\tilde{T}_2\| = 1$. Let T_1 and T_2 be their respective δ-approximate ghost operators. Then, for $0 \leq a, b \leq r$, we get

$$
\begin{aligned}
\left|2\langle\psi|E_a^\dagger E_b|\phi\rangle\right| &= \left|\langle\psi|E_a^\dagger E_b T_1|\phi\rangle - \langle\psi|T_1 E_a^\dagger E_b|\phi\rangle\right| \\
&= \left|\langle\psi|E_a^\dagger E_b T_1|\phi\rangle - \langle\psi|E_a^\dagger T_1 E_b|\phi\rangle + \langle\psi|E_a^\dagger T_1 E_b|\phi\rangle - \langle\psi|T_1 E_a^\dagger E_b|\phi\rangle\right| \\
&\leq \left|\langle\psi|E_a^\dagger E_b T_1|\phi\rangle - \langle\psi|E_a^\dagger T_1 E_b|\phi\rangle\right| + \left|\langle\psi|E_a^\dagger T_1 E_b|\phi\rangle - \langle\psi|T_1 E_a^\dagger E_b|\phi\rangle\right| \\
&\leq \|\left(E_b T_1 - T_1 E_b\right)|\phi\rangle\|\|E_a|\psi\rangle\| + \|E_b|\phi\rangle\|\|\left(T_1 E_a - E_a T_1\right)|\psi\rangle\| \\
&\leq 4\delta,
\end{aligned}
$$

where in the second-to-last line we used the Schwarz inequality, and in the last line we used Lemma 5.B.2 and the fact that $\|E_a\| \leq 1$. Therefore we have

$$
\left|\langle\psi|E_a^\dagger E_b|\phi\rangle\right| \leq 2\delta.
\tag{5.48}
$$

Repeating the same argument for $\tilde{T}_2$, we likewise get

$$\left|\langle\phi-\psi|E_a^\dagger E_b|\phi+\psi\rangle\right| \le 2\delta.$$

Then we have

$$\left|\langle\phi|E_a^\dagger E_b|\phi\rangle - \langle\psi|E_a^\dagger E_b|\psi\rangle\right|$$
$$= \left|\langle\phi|E_a^\dagger E_b|\phi\rangle - \langle\psi|E_a^\dagger E_b|\psi\rangle + \langle\phi|E_a^\dagger E_b|\psi\rangle - \langle\psi|E_a^\dagger E_b|\phi\rangle - \langle\phi|E_a^\dagger E_b|\psi\rangle + \langle\psi|E_a^\dagger E_b|\phi\rangle\right|$$
$$\le \left|\langle\phi|E_a^\dagger E_b|\phi\rangle - \langle\psi|E_a^\dagger E_b|\psi\rangle + \langle\phi|E_a^\dagger E_b|\psi\rangle - \langle\psi|E_a^\dagger E_b|\phi\rangle\right| + 2\left|\langle\phi|E_a^\dagger E_b|\psi\rangle\right|$$
$$\le 2\left|\langle\phi-\psi|E_a^\dagger E_b|\phi+\psi\rangle\right| + 4\delta$$
$$\le 8\delta. \tag{5.49}$$

Now consider an orthonormal basis $\{|i\rangle, i = 0, 1, 2, \ldots, \dim C - 1\}$, for the code space and define $\lambda_{ab} = \langle 0|E_a^\dagger E_b|0\rangle$. Noting that in the equations (5.B) and (5.B), $|\phi\rangle$ and $|\psi\rangle$ can be any two elements of the orthonormal basis, we see that

$$|\langle i|E_a^\dagger E_b|j\rangle| \le 2\delta$$

for $i \ne j$, while

$$|\langle i|E_a^\dagger E_b|i\rangle - \lambda_{ab}| \le 8\delta.$$

Thus we find that the approximate Knill-Laflamme conditions for $\mathcal{E}_I$ are satisfied:

$$\frac{1}{2}\left|\langle i|E_a^\dagger E_b|j\rangle - \lambda_{ab}\delta_{ij}\right| \le 4\delta.$$

Note that the factor of $1/2$ comes from the normalization of the Kraus operators for $\mathcal{E}_I$. From Lemma 5.B.1, this implies that $\mathcal{E}_I$ is ϵ-correctable for C, where

$$\epsilon = (|K| + 1)\sqrt{2(\dim C)\,\delta}.$$

$\square$

5.C   Complexity of Controlled Unitary

Lemma 5.C.1. *Let U be a unitary of circuit complexity k with respect to some universal 2-qubit gate set $\mathcal{G}$. Given an ancillary system of n qubits, let $\Lambda_m(U)$ be the operator controlled on the state $|m\rangle$, where $0 \le m < 2^n$, i.e.,*

$$\Lambda_m(U)(|\ell\rangle \otimes |x\rangle) = |\ell\rangle \otimes U^{\delta_{\ell m}}|x\rangle.$$

Then given any $\epsilon > 0$, the operator $\Lambda_m(U)$ can be implemented with ϵ-precision with circuit complexity $O\left(4^n k \log^4(k/\epsilon)\right)$.

Proof. Let $U = U_k \cdots U_1$ be a decomposition of U into elements of $\mathcal{G}$. To implement $\Lambda_m(U)$ to ϵ-precision, it suffices to implement $\Lambda_m(U_i)$ to ϵ/k-precision for each $1 \leq i \leq k$. Since each U_i is a 2-qubit gate, it follows that $\Lambda_m(U_i)$ is supported on at most $n + 2$ qubits. By the Solovay-Kitaev theorem [48], each $\Lambda_m(U_i)$ can be implemented to ϵ/k-precision with $O(4^n \log^4(k/m))$ gates from $\mathcal{G}$. It follows that U itself can be implemented to ϵ-precision with $O(4^n k \log^4(k/\epsilon))$ gates. $\qquad\square$

The scaling with n can be considerably improved using circuit constructions from [49], but Lemma 5.C.1 will suffice for our purposes.

5.D What if the Radiation is Not Pseudorandom?

The central assumption of this paper is that the state of the Hawking radiation EB emitted by a partially evaporated black hole is pseudorandom. Here we ask what happens if this assumption is broken in a particular way.

Suppose B is a single qubit and the pure state of EBH is

$$|\Psi\rangle_{EBH} = \frac{1}{\sqrt{2}}(|0\rangle_B|\psi_0\rangle_{EH} + |1\rangle_B|\psi_1\rangle_{EH}).$$

Consider a Hermitian operator M_E acting on E such that $M_E \otimes Z_B$ can be efficiently measured, where Z_B is the Pauli-Z operator acting on B. Suppose that

$$\langle M_E \otimes Z_B\rangle_\Psi - \langle M_E\rangle_\Psi \langle Z_B\rangle_\Psi = c,$$

where the subscript Ψ indicates that the expectation value is evaluated in the global state $|\Psi\rangle_{EBH}$, or equivalently in the marginal state ρ_{EB}. Note that $c = 0$ if ρ_{EB} is maximally mixed. Therefore, by definition, if ρ_{EB} is pseudorandom, then c must be exponentially small in $|H|$. It follows that if c is a nonzero constant, independent of $|H|$, then ρ_{EB} is not pseudorandom (though the converse is not necessarily true).

We will now show that, if $c \neq 0$, there cannot be a complete set of logical operators that commute with M_E acting on the code space spanned by $\{|\psi_0\rangle_{EH}, |\psi_1\rangle_{EH}\}$. Note that because the marginal state ρ_B is maximally mixed, we have $\langle Z_B\rangle_\Psi = 0$, and therefore

$$2c = 2\langle M_E \otimes Z_B\rangle_\Psi = \langle\psi_0|M_E|\psi_0\rangle - \langle\psi_1|M_E|\psi_1\rangle.$$

Consider a Hermitian operator X_L on EH that acts on the code basis states $\{|\psi_0\rangle_{EH}, |\psi_1\rangle_{EH}\}$ like the Pauli-X operator:

$$X_L|\psi_0\rangle = |\psi_1\rangle, \quad X_L|\psi_1\rangle = |\psi_0\rangle,$$

and notice that

$$\langle\psi_1|[X_L, M_E]|\psi_0\rangle = \langle\psi_0|M_E|\psi_0\rangle - \langle\psi_1|M_E|\psi_1\rangle = 2c \neq 0.$$

This shows that the commutator $[X_L, M_E]$ is $O(1)$ acting on the code space. Thus no logical Pauli-X operator commutes with M_E acting on the code space, and in particular there can be no complete set of ghost logical operators commuting with M_E.

For this argument, we chose the operator acting on B to be Z_B, but a similar argument works for any Hermitian operator acting on B. Suppose N_B is a Hermitian operator acting on B such that

$$\langle M_E \otimes N_B\rangle_\Psi - \langle M_E\rangle_\Psi\langle N_B\rangle_\Psi = c \neq 0. \tag{5.50}$$

Since N_B is Hermitian, we can diagonalize it in a certain basis, and we may assume, without loss of generality, that N_B is traceless. (If N_B is not traceless, we may replace N_B by $N_B' = N_B - \text{Tr}(N_B)(I/2)$ without modifying Equation (5.D).) In the basis in which it is diagonal, then N_B is equal to Z_B up to a nonzero multiplicative constant.

BIBLIOGRAPHY

[1]   S. W. Hawking, "Particle creation by black holes," *Communications in Mathematical Physics*, vol. 43, no. 3, pp. 199–220, 1975. DOI: `10.1007/BF02345020`.

[2]   J. M. Maldacena, "The large n limit of superconformal field theories and supergravity," *Advances in Theoretical and Mathematical Physics*, vol. 2, pp. 231–252, 1998. DOI: `10.1023/A:1026654312961`.

[3]   A. Almheiri, D. Marolf, J. Polchinski, and J. Sully, "Black holes: Complementarity or firewalls?" *Journal of High Energy Physics*, no. 2, p. 62, Feb. 2013. DOI: `10.1007/JHEP02(2013)062`.

[4]   J. Maldacena, "Eternal black holes in anti-de sitter," *Journal of High Energy Physics*, no. 04, pp. 021–021, Apr. 2003. DOI: `10.1088/1126-6708/2003/04/021`.

[5]   J. Maldacena and L. Susskind, "Cool horizons for entangled black holes," *Fortschritte der Physik*, vol. 61, no. 9, pp. 781–811, 2013. DOI: `10.1002/prop.201300020`.

[6]   L. Susskind, "Computational complexity and black hole horizons," Feb. 23, 2014. arXiv: `1402.5674`.

[7]   K. Papadodimas and S. Raju, "An infalling observer in AdS/CFT," *Journal of High Energy Physics*, no. 10, p. 212, Oct. 2013. DOI: `10.1007/JHEP10(2013)212`.

[8]   D. Harlow, "Aspects of the papadodimas-raju proposal for the black hole interior," *Journal of High Energy Physics*, no. 11, p. 55, Nov. 2014. DOI: `10.1007/JHEP11(2014)055`.

[9]   R. Bousso, "Violations of the equivalence principle by a nonlocally reconstructed vacuum at the black hole horizon," *Physical Review Letters*, vol. 112, p. 041 102, 4 Jan. 2014. DOI: `10.1103/PhysRevLett.112.041102`.

[10] K. Papadodimas and S. Raju, "Remarks on the necessity and implications of state-dependence in the black hole interior," *Physical Review D*, vol. 93, p. 084 049, 8 Apr. 2016. DOI: `10.1103/PhysRevD.93.084049`.

[11] A. Almheiri, D. Marolf, J. Polchinski, D. Stanford, and J. Sully, "An apologia for firewalls," *Journal of High Energy Physics*, no. 9, p. 18, Sep. 2013. DOI: `10.1007/JHEP09(2013)018`.

[12] Z. Ji, Y.-K. Liu, and F. Song, "Pseudorandom quantum states," in *Advances in Cryptology – CRYPTO 2018*, Cham: Springer International Publishing, 2018, pp. 126–152, ISBN: 978-3-319-96878-0.

[13] Y. Sekino and L. Susskind, "Fast scramblers," *Journal of High Energy Physics*, no. 10, pp. 065–065, Oct. 2008. DOI: 10.1088/1126-6708/2008/10/065.

[14] D. Harlow and P. Hayden, "Quantum computation vs. firewalls," *Journal of High Energy Physics*, no. 6, p. 85, Jun. 2013. DOI: 10.1007/JHEP06(2013)085.

[15] S. Aaronson, "The complexity of quantum states and transformations: From quantum money to black holes," 2016. arXiv: 1607.05256.

[16] D. Marolf and J. Polchinski, "Violations of the born rule in cool state-dependent horizons," *Journal of High Energy Physics*, no. 1, p. 8, Jan. 2016. DOI: 10.1007/JHEP01(2016)008.

[17] P. Hayden and G. Penington, "Learning the Alpha-bits of black holes," *Journal of High Energy Physics*, no. 12, p. 7, Dec. 2019, ISSN: 1029-8479. DOI: 10.1007/JHEP12(2019)007.

[18] G. Penington, "Entanglement wedge reconstruction and the information paradox," 2019. arXiv: 1905.08255.

[19] A. Almheiri, N. Engelhardt, D. Marolf, and H. Maxfield, "The entropy of bulk quantum fields and the entanglement wedge of an evaporating black hole," *Journal of High Energy Physics*, no. 12, p. 063, 2019. DOI: 10.1007/JHEP12(2019)063.

[20] A. Almheiri, R. Mahajan, J. Maldacena, and Y. Zhao, "The page curve of hawking radiation from semiclassical geometry," 2019. arXiv: 1908.10996.

[21] G. Penington, S. H. Shenker, D. Stanford, and Z. Yang, "Replica wormholes and the black hole interior," 2019. arXiv: 1911.11977.

[22] A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian, and A. Tajdini, "Replica wormholes and the entropy of hawking radiation," 2019. arXiv: 1911.12333.

[23] C. Bény and O. Oreshkov, "General conditions for approximate quantum error correction and near-optimal recovery channels," *Physical Review Letters*, vol. 104, p. 120 501, 12 Mar. 2010. DOI: 10.1103/PhysRevLett.104.120501.

[24] S. T. Flammia, J. Haah, M. J. Kastoryano, and I. H. Kim, "Limits on the storage of quantum information in a volume of space," *Quantum*, vol. 1, p. 4, Apr. 2017. DOI: 10.22331/q-2017-04-25-4.

[25] C. Bény, A. Kempf, and D. W. Kribs, "Generalization of quantum error correction via the heisenberg picture," *Physical Review Letters*, vol. 98, p. 100 502, 10 Mar. 2007. DOI: 10.1103/PhysRevLett.98.100502.

[26] ——, "Quantum error correction of observables," *Physical Review A*, vol. 76, p. 042 303, 4 Oct. 2007. DOI: 10.1103/PhysRevA.76.042303.

[27] D. N. Page, "Average entropy of a subsystem," *Physical Review Letters*, vol. 71, pp. 1291–1294, 9 Aug. 1993. DOI: `10.1103/PhysRevLett.71.1291`.

[28] A. C. Yao, "Theory and application of trapdoor functions," in *23rd Annual Symposium on Foundations of Computer Science (sfcs 1982)*, Nov. 1982, pp. 80–91. DOI: `10.1109/SFCS.1982.45`.

[29] O. Goldreich and H. Krawczyk, "Sparse pseudorandom distributions," in *Proceedings of the 9th Annual International Cryptology Conference on Advances in Cryptology*, ser. CRYPTO '89, Berlin, Heidelberg: Springer-Verlag, 1990, pp. 113–127, ISBN: 3-540-97317-6.

[30] S. Arora and B. Barak, *Computational Complexity: A Modern Approach*, 1st. New York, NY, USA: Cambridge University Press, 2009, ISBN: 0521424267, 9780521424264.

[31] P. Hayden, D. W. Leung, and A. Winter, "Aspects of generic entanglement," *Communications in Mathematical Physics*, vol. 265, pp. 95–117, 2006. DOI: `10.1007/s00220-006-1535-6`.

[32] Z. Brakerski and O. Shmueli, "(pseudo) random quantum states with binary phase," in *TCC*, 2019.

[33] A. Gheorghiu and M. J. Hoban, "Estimating the entropy of shallow circuit outputs is hard," 2020. arXiv: `2002.12814`.

[34] R. Cleve and J. Watrous, "Fast parallel circuits for the quantum fourier transform," in *Proceedings 41st Annual Symposium on Foundations of Computer Science*, IEEE, 2000, pp. 526–536. DOI: `10.1109/SFCS.2000.892140`.

[35] A. Almheiri, X. Dong, and D. Harlow, "Bulk locality and quantum error correction in AdS/CFT," *Journal of High Energy Physics*, no. 4, p. 163, Apr. 2015. DOI: `10.1007/JHEP04(2015)163`.

[36] F. Pastawski, B. Yoshida, D. Harlow, and J. Preskill, "Holographic quantum error-correcting codes: Toy models for the bulk/boundary correspondence," *Journal of High Energy Physics*, no. 6, p. 149, Jun. 2015. DOI: `10.1007/JHEP06(2015)149`.

[37] E. Verlinde and H. Verlinde, "Black hole entanglement and quantum error correction," *Journal of High Energy Physics*, no. 10, p. 107, Oct. 2013. DOI: `10.1007/JHEP10(2013)107`.

[38] B. Yoshida, "Firewalls vs. scrambling," *Journal of High Energy Physics*, no. 10, p. 132, Oct. 2019. DOI: `10.1007/JHEP10(2019)132`.

[39] J. Preskill, "Sufficient condition on noise correlations for scalable quantum computing," *Quantum Information and Computation*, vol. 13, no. 3-4, pp. 0181–0194, 2013, ISSN: 1533-7146.

[40] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*, 10th. New York, NY, USA: Cambridge University Press, 2011, ISBN: 1107002176, 9781107002173.

[41] J. Oppenheim and B. Unruh, "Firewalls and flat mirrors: An alternative to the amps experiment which evades the harlow-hayden obstacle," *Journal of High Energy Physics*, no. 3, p. 120, Mar. 2014. DOI: 10.1007/JHEP03(2014)120.

[42] E. Knill and R. Laflamme, "Theory of quantum error-correcting codes," *Physical Review A*, vol. 55, no. 2, p. 900, 1997. DOI: 10.1103/PhysRevLett.84.2525.

[43] A. Bouland, B. Fefferman, and U. Vazirani, "Computational pseudorandomness, the wormhole growth paradox, and constraints on the AdS/CFT duality," 2019. arXiv: 1910.14646.

[44] L. Susskind, "Horizons protect church-turing," 2020. arXiv: 2003.01807.

[45] P. Hayden and J. Preskill, "Black holes as mirrors: Quantum information in random subsystems," *Journal of High Energy Physics*, no. 09, p. 120, 2007. DOI: 10.1088/1126-6708/2007/09/120.

[46] D. W. Berry, A. M. Childs, R. Cleve, R. Kothari, and R. D. Somma, "Exponential improvement in precision for simulating sparse hamiltonians," in *Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing*, ser. STOC '14, New York, New York: Association for Computing Machinery, 2014, pp. 283–292. DOI: 10.1145/2591796.2591854.

[47] B. Yoshida, "Observer-dependent black hole interior from operator collision," 2019. arXiv: 1910.11346.

[48] A. Y. Kitaev, A. Shen, M. N. Vyalyi, and M. N. Vyalyi, *Classical and quantum computation*, 47. American Mathematical Society, 2002, ISBN: 0821832298.

[49] A. Barenco, C. H. Bennett, R. Cleve, D. P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. A. Smolin, and H. Weinfurter, "Elementary gates for quantum computation," *Physical Review A*, vol. 52, pp. 3457–3467, 5 Nov. 1995. DOI: 10.1103/PhysRevA.52.3457.