Learning the Quantum, Scrambling the Universe

by

Shuwei Liu

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Physics

Waterloo, Ontario, Canada, 2025

© Shuwei Liu 2025

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	Bartlomiej Czech Member, Institute for Advanced Study, Tsinghua University
Supervisor(s):	Beni Yoshida Senior Faculty, Perimeter Institute for Theoretical Physics Robert Myers BMO Financial Group Isaac Newton Chair in Theoretical Physics, Perimeter Institute for Theoretical Physics
Internal Member:	Niayesh Afshordi Professor, Dept. of Physics and Astronomy, University of Waterloo
Internal-External Member:	Eduardo Martin-Martinez Professor, Dept. of Applied Mathematics, University of Waterloo
Other Member(s):	Alex May Junior Faculty, Perimeter Institute for Theoretical Physics

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

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

Statement of Contributions

This thesis comprises contributions from three research manuscripts, detailed as follows:

Chapter 1 is based on material published in Ref. [93]. This research was conducted under the supervision of Beni Yoshida.

Chapter 2 contains material from ongoing research conducted in collaboration with Charlie Cummings, Bruno de S. L. Torres, and Beni Yoshida, which is yet to be published.

Chapter 3 incorporates material from Ref. [94], which is presently available as an arXiv preprint. This study is a collaborative effort with Shiyu Zhou, Zi-Wen Liu, and Jinmin Yi. Shuwei Liu shares co-first authorship on this manuscript.

The contributions presented in each chapter demonstrate extensive collaborative work and valuable mentorship, which have significantly enriched the depth and breadth of the research.

Abstract

This thesis explores how quantum information behaves in extreme physical settings, from black hole interiors to noisy quantum devices. First, we derive a thermodynamic relation linking gravitational shockwaves to microscopic deformations of the black hole horizon, illuminating the connection between quantum chaos and horizon area deformation. Next, we explore the black hole information problem through the lens of holography, demonstrating how scrambling and recoverability emerge from gravitational backreaction in shockwave geometries. Finally, we shift to quantum technologies, introducing noise-strength-adapted (NSA) quantum error-correcting codes discovered via hybrid machine learning. These non-stabilizer codes outperform conventional designs under amplitude damping and generalize to larger systems. Together, these works reveal how quantum information unifies seemingly disparate domains, offering both conceptual insights into spacetime and practical tools for building resilient quantum systems.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor, Beni Yoshida, for his invaluable guidance throughout my PhD. His mentorship has shaped not only my academic growth but also allowed me the freedom to explore my own research interests.

I am also sincerely thankful to my committee members—Niayesh Afshordi, Bartek Czech, Eduardo Martin-Martinez, Alex May, and Rob Myers—for their insightful feedback, stimulating discussions, and thoughtful suggestions that have helped refine my work.

I gratefully acknowledge my collaborators—Charlie Cummings, Zi-Wen Liu, Bruno Torres, Jinmin Yi, Beni Yoshida, and Shiyu Zhou—for the enriching collaborations and the many ideas we developed together.

I would like to thank 1QBit and Mitacs for supporting my internship in the quantum industry, and Mohamed Hibat Allah for his kind assistance in helping me secure that opportunity. I would also like to thank my co-supervisor, Rob Myers, for his generosity in serving as my Mitacs academic supervisor and for his guidance and support throughout these years.

I am deeply grateful to my friends—Yushao Chen, Jinmin Yi, Yijian Zou, Zhiren Wang, Weicheng Ye and Jackie Caminiti, Yangrui Hu, Hassan Khalvati, Chen Li, Ruizhi Liu, Zi-Wen Liu, Shan-Ming Ruan, Shengqi Sang, Hao Xu, Shiyu Zhou, Zheng Zhou—for their friendship, encouragement, and all the moments we shared during this journey. Special thanks also go to Shelly Pavlic for her therapy sessions, which provided meaningful support during particularly challenging times.

Finally, and most importantly, I would like to thank my parents for their unwavering support, patience, and understanding. Their belief in me has been the foundation of everything I have accomplished.

Table of Contents

Examining Committee Membership ii

Author's Declaration iii

Statement of Contributions iv

Abstract v

Acknowledgements vi

List of Figures xi

List of Abbreviations xiii

1 Soft thermodynamics of gravitational shock wave 1

1.1 Introduction 1

1.1.1 Main result 3

1.1.2 Relation to previous works 5

1.1.3 Organization 6

1.2 Review of the Noether charge method 6

1.2.1 Variation of Lagrangian 7

1.2.2 Wald's entropy formula 8

1.2.3 Example: general relativity . . . 10
1.3 Two refinements to Wald's entropy formula . . . 11
1.3.1 Matter current . . . 12
1.3.2 Higher-order matter variation . . . 13
1.3.3 Comparison . . . 15
1.4 Gravitational shock wave in General Relativity . . . 17
1.5 Horizon translation vector . . . 19
1.5.1 Timelike vector . . . 20
1.5.2 Horizon translation . . . 21
1.6 Microscopic thermodynamics on horizon . . . 22
1.6.1 Noether charge from shifted vector . . . 22
1.6.2 Thermodynamic relation . . . 23
1.6.3 Pre-symplectic form . . . 25
1.7 Shock wave from area minimization . . . 27
1.7.1 Area variation . . . 27
1.7.2 Intrinsic metric . . . 28
1.8 Gauss-Bonnet gravity . . . 31
1.8.1 Noether charge in Gauss-Bonnet gravity . . . 31
1.8.2 Isotropic space . . . 33
1.9 Measuring shock waves . . . 33
1.10 Discussions . . . 36

2 Information recovery and shock waves **37**
2.1 Introduction . . . 37
2.1.1 Overview of results . . . 39
2.2 Hayden-Preskill thought experiment . . . 41
2.2.1 Setup . . . 42
2.2.2 Minimal surface calculus . . . 43

2.2.3 Projective measurement . . . 45
2.3 Scrambling and Information Recoverability . . . 46
2.3.1 Out-of-time order correlation . . . 46
2.3.2 Weakly scrambling regime . . . 47
2.3.3 Scrambling and shockwaves . . . 49
2.4 Shockwave Geometry for Infalling/Outgoing Matter . . . 50
2.4.1 Quantum extremal surface and maxmin prescription . . . 51
2.4.2 Infalling quantum state . . . 52
2.4.3 A detour . . . 54
2.4.4 Failed approaches for outgoing mode . . . 54
2.4.5 Outgoing mode . . . 55
2.5 Information Recovery from Shockwave Geometry . . . 57
2.5.1 Scrambling case . . . 58
2.5.2 Non-scrambling case . . . 60
2.5.3 Weakly scrambling regime . . . 61
2.5.4 On Σ_3 . . . 61
2.5.5 Measuring a Hawking radiation . . . 63
2.6 Interior Partner from Shockwave Geometry . . . 65
2.6.1 Setup . . . 66
2.6.2 Boundary picture . . . 67
2.6.3 Bulk picture . . . 68
2.7 Outlook . . . 68

3 Noise-strength-adapted approximate quantum codes inspired by machine learning 70

3.1 Introduction . . . 70
3.1.1 Organization . . . 72
3.2 Preliminaries . . . 72

3.2.1 Quantum error correction . . . 72
3.2.2 Approximate quantum error correction . . . 74
3.2.3 Worst-case fidelity and loss function for AQEC codes . . . 74
3.2.4 Amplitude damping noise . . . 75
3.3 Machine learning experiments . . . 76
3.3.1 Details of variational quantum learning . . . 77
3.4 Basic analytical results . . . 81
3.5 Recovery protocol and fidelity calculation . . . 82
3.5.1 Recovery and worst-case fidelity of $((4,1))$ self-complementary NSA code . . . 82
3.5.2 Recovery and worst-case fidelity of $((4,1))$ pair-complementary NSA code . . . 86
3.5.3 Entanglement fidelity of $((4,1))$ pair-complementary NSA code . . . 89
3.6 Generalizations . . . 90
3.6.1 Fidelity of self-complementary qubit NSA code generalization . . . 92
3.6.2 Self-complementary qudit NSA code generalization . . . 94
3.6.3 Fidelity of pair-complementary qubit NSA code generalization . . . 102
3.6.4 Recovery and worst-case fidelity of binomial NSA codes . . . 107
3.7 Discussions . . . 108

References **110**

List of Figures

1.1 Gravitational shock wave at $u = 0$, moving in the v-direction. The shaded region experiences metric variations. 18

1.2 (a) The timelike Killing vector $\xi^{(t)}$. (b) The shifted timelike vector χ. . . 21

1.3 The Noether charge integral. A redline represents the gravitational shock wave running at $u = 0$. Blue lines represent the null-shifted timelike vector. The shaded region ($u > 0$) experiences the metric variations $\delta g_{\mu\nu}$ and the energy momentum variations $\delta T_{\mu\nu}$ which result from the shift $v \to v + \theta(u)\alpha(x)$. In addition, at $u = 0$, the variation $\delta T^{(P)}_{uu}$ from the shock wave is introduced. 23

1.4 The area deformation in the original unperturbed geometry. 29

2.1 Schematics of the HP thought experiment and labelling of Hilbert spaces. 42

2.2 Two candidate surfaces, homologous to $\overline{B}D$, with "Areas" being $n_B + n_D$ and $n_C + n_A$ respectively. 45

2.3 Operator growth and scrambling. $V(0)$ and $W(t)$ will eventually overlap, leading to the decay of OTOC $\langle V(0)W(t)V(0)W(t)\rangle$. 46

2.4 The Penrose diagram of the black hole geometry after including the perturbation from asymptotic infinity. The dashed lines are the horizon, and the infalling matter shockwave is in red. The horizon is disjoint because of the coordinate shift. 50

2.5 Shockwave geometry for an infalling quantum state. 53

2.6 Failed approaches for geometries corresponding to catching the Hawking radiation. (a) Sending a negative energy shockwave. (b) Shockwave exiting the asymptotic boundary. 55

2.7 (a) Shockwave geometry for catching a Hawking radiation. (b) Shockwave geometry for measuring a Hawking radiation. 56

2.8 Shockwave geometry for the HP recovery via catching the Hawking radiation. 58

2.9 (a) Approximate locations of candidate surfaces and two kinds of Cauchy slices. (b) Generalized entropies associated with the Cauchy slice Σ_1. Here, $M_D \geq M_A$ guarantees $S^{(III)}_{\overline{B}D} \leq S^{(I)}_{\overline{B}D}$. The entropy barrier is $S^{(II)}_{\overline{B}D} - S^{(I)}_{\overline{B}D} \approx 2S_A$. 60

2.10 Weakly scrambling regime where the traversable wormhole algorithms work efficiently. The Python's lunch disappears by tuning the time separation Δt so that $S^{(III)}_{\overline{B}D} \leq S^{(II)}_{\overline{B}D} \leq S^{(I)}_{\overline{B}D}$. Dotted paths correspond to $\Delta t \gtrapprox t_{\rm scr}$ and $\Delta t \ll t_{\rm scr}$. 62

2.11 The effect of coordinate shift from A's shockwave. The Cauchy slice Σ_3 is not well defined. 63

2.12 Shockwave geometry for the HP recovery via measuring the Hawking radiation. 65

3.1 Cartoon of quantum error correction. 73

3.2 (a) Loss $\mathcal{L}$ as a function of γ is shown. The solid green line shows the smooth γ^2 dependence of LNCY code. The solid blue and orange lines show the losses of two codes optimized for $\gamma = 10^{-1.5}$. The blue and orange dashed lines, instead, show the losses for NSA codes with adaptive code states. (b) shows the fidelity plotted as a function of γ for $((4,1))$ NSA SC, $((4,1))$ NSA PC, and LNCY codes. 78

3.3 Encoding circuit $\mathcal{E}$ with trainable parameters for searching $((4,1))$ code. . . 79

3.4 Codeword obtained through variational quantum learning, corresponding to the self-complementary NSA code. 80

3.5 Codeword obtained through variational quantum learning, corresponding to the pair-complementary NSA code. 81

3.6 **(A)** Circuit for error syndrome detection. The measurement outcomes (M_2, M_4) determine which of the three recovery operations W_k to apply. **(B)**, **(C)**, and **(D)** show the corresponding recovery circuits for measurement results 00, 10, and 01, respectively. The controlled-rotation gates are CR_y gates with rotation angle $2\cos^{-1}\left(\sqrt{\frac{2}{(1-\gamma)^2+(1-\gamma)^{-2}}}\right)$. 84

List of Abbreviations

ADM Arnowitt–Deser–Misner 4

AdS anti-de Sitter 2

AdS/CFT anti-de Sitter/conformal field theory correspondence 1

AMPS Almheiri-Marolf-Polchinski-Sully 65

AQEC approximate quantum error correction 70

BMS Bondi–Metzner–Sachs 2

dS de Sitter 2

HP Hayden-Preskill 38

HRT Hubeny-Rangamani-Takayanagi 51

JT Jackiw-Teitelboim 5

ML machine learning 71

OTOC out-of-time order correlator 33

PC pair-complementary 71

QEC quantum error correction 70

QES quantum extermal surface 38

RT Ryu-Tayanagi 51

SC self-complementary 71

Chapter 1

Soft thermodynamics of gravitational shock wave

1.1 Introduction

Recent developments at the interface between quantum gravity and quantum information theory in the AdS/CFT correspondence have provided useful tools to address conceptual puzzles concerning quantum aspects of black holes on less ambiguous settings [107, 64, 135, 5, 109, 67, 32, 111, 6]. At the heart of these progresses is the improved understanding on how the structure of quantum entanglement changes dynamically under the gravitational backreaction from the infalling and outgoing matter [132, 131, 136, 133, 75, 118, 55, 50, 121, 81, 114, 122, 82, 151, 113, 149, 148, 150]. The gravitational shock wave geometries [3, 38, 37, 68, 120, 18, 19, 4, 9] provide a particularly useful family of analytically tractable models of the gravitational backreaction that can be induced by infalling massless particles, or equivalently, perturbations on the boundary by light operators.

Despite extensive studies in the past decades and recent revivals in the AdS/CFT correspondence, the quantum origin of the gravitational shock waves still remains elusive. The gravitational shock waves in the black hole background are often interpreted as low-energy excitations which constitute the microscopic degrees of freedom of the quantum black hole. It is, however, unclear how the gravitational shock waves are ever quantized and account for the finite value of Bekenstein-Hawking entropy in a concrete manner. Also, if one hopes to study geometries beyond the AdS/CFT correspondence (*e.g.* asymptotically flat spaces or systems without the spatial infinity as in the de Sitter space), additional

technical and conceptual subtleties often hinder naive application of holographic intuitions [1] [2]. Hence, as a first step, it is desirable to develop a universal framework to characterize the gravitational shock waves through some microscopic lens in a way applicable to geometries beyond the AdS/CFT correspondence as well.

Recent rapid progress on the soft hair of black holes in an asymptotically flat space provide useful hints toward such a goal with series of interesting derivations of the Bekenstein-Hawking entropy [24, 23, 57, 2, 8, 27] [3]. It has been also pointed out that the linearized gravitational shockwaves can be realized as BMS supertranslations via certain choices of parameters [128, 62, 63]. Furthermore, some general expressions of thermodynamic relations for soft charges have been derived for the Schwarzschild black hole and several other geometries [35].

While these developments may provide a further insight into quantum aspects of the gravitational shock waves, it remains unclear how the soft charges may be measured in a concrete setting. In addition, the derived thermodynamic relations, associated with soft charges, are in rather abstract forms without concrete physical interpretations. As such, the implications of the soft hair physics on conceptual puzzles of quantum black holes still remain vague, in comparison with concrete developments within the framework of the AdS/CFT correspondence. What is currently missing is an effort to tie these two developments. Hence, deriving a microscopic thermodynamic characterization of the gravitational shock waves on a unified footing may be the necessary first step.

In this chapter, we derive the microscopic thermodynamic relations which are localized on the near horizon region and are sensitive to the gravitational shock wave in a concrete form. Our treatment is applicable to arbitrary gravitational shock waves at a bifurcating horizon in generic static spacetime (including AdS, dS, and asymptotically flat spaces), and is valid for arbitrary covariant theory of gravity which may be beyond general relativity (*e.g.* higher-derivative gravity theories). Furthermore, our thermodynamic relation provides a concrete physical interpretation of the gravitational shock waves, or soft charges. In particular, we derive a simple formula which relates the horizon area deformation induced by two intersecting gravitational shock waves to the incoming and outgoing energy sources. We will then find that the soft charge associated with the gravitational shock waves is nothing but the horizon area in the presence of two intersecting gravitational shock waves.

Our main technical machinery is the covariant phase space formalism, namely Wald's

[1] For instance, the absence of the spatial infinity may suggest that the structure of quantum entanglement in the de Sitter space differs crucially from that of the AdS space or the asymptotically flat space.

[2] See [48, 49] for other studies of entanglement structure in the de Sitter space.

[3] See [14, 117, 21, 22, 72, 33, 36, 34, 126, 29] for samples of earlier works and relevant developments.

Noether charge method. It turns out that the original treatment by Wald, which focuses on contributions at the linear order of field variations, is not sufficient to probe the gravitational shock wave. For this reason, we will extend the Noether charge method to include variations of the matter fields at *arbitrary order*. This refinement enables us to obtain the microscopic thermodynamic relations (as in Eq. (1.7)) which directly associate the area deformation to the energy momentum tensor variation. One benefit of the covariant phase space formalism is that it is applicable to arbitrary covariant theory of gravity. As an application, we demonstrate that, in the Gauss-Bonnet gravity [97, 127, 17, 1], the gravitational shock wave equations indeed follow from the microscopic thermodynamic relation with much simplified calculations.

Our result also suggests a concrete framework to physically characterize (or, even measure) soft charges via out-of-time order correlation (OTOC) functions. The leading-order (in $1/G_N$) behaviours of OTOCs are dominated by the gravitational scattering unitary matrix which typically takes the following form; $\mathcal{U} = \exp\left[i \iint d\Omega_1 d\Omega_2 P_{\text{out}}(\Omega_1) f(\Omega_1, \Omega_2) P_{\text{in}}(\Omega_2)\right]$. Here $f(\Omega_1, \Omega_2)$ is the green function for the shock wave equation, and P_{in} and P_{out} account for the incoming and outgoing energy fluxes. Our thermodynamic characterization enables us to rewrite the above scattering matrix into a simple form:

$$\mathcal{U} = \exp\left(i\text{Area}\right) , \tag{1.1}$$

up to some multiplicative factor in the phase. Here "Area" represents the horizon area in the presence of two intersecting gravitational shockwaves due to P_{in} and P_{out}. Thus, the phase factor of the scattering matrix is proportional to the horizon area (or equivalently, the soft charge of an infalling matter measured by an outgoing matter).

1.1.1 Main result

Here we present a summary of our main result. We study a bifurcating Killing horizon in generic static spacetime of the following form:

$$\begin{aligned} ds^2 &= 2F(u,v)dudv + G(u,v)h_{ij}dx^i dx^j, \\ T &= 2T_{uv}dudv + T_{uu}dudu + T_{vv}dvdv + T_{ij}dx^i dx^j , \end{aligned} \tag{1.2}$$

where $u = 0$ and $v = 0$ corresponds to the bifurcating horizons. We assume that the timelike vector $\partial_t = \kappa(-u\partial_u + v\partial_v)$ is a Killing vector of the spacetime. The gravitational shock wave can be generated by shifting this spacetime as

$$\tilde{v} = v + \theta(u)\alpha(x) , \tag{1.3}$$

where $\theta(u)$ is a step function and x collectively denotes angular profiles. At the leading order, the metric variation can be computed by considering the horizon translation vector $\hat{\alpha}$:

$$\delta g_{\mu\nu} = \theta(u)\mathcal{L}_{\hat{\alpha}} g_{\mu\nu} \ , \qquad \hat{\alpha} \equiv \alpha(x)\partial_v \ , \tag{1.4}$$

where $\mathcal{L}_{\hat{\alpha}}$ is the Lie derivative with respect to $\hat{\alpha}$. To satisfy the Einstein equation, an additional energy momentum tensor needs to be introduced:

$$T^{(p)\,vv} = T^{(p)}(x)\delta(u) \ . \tag{1.5}$$

For general relativity (*i.e.* $L^{(G)} = \frac{1}{16\pi}(R - 2\Lambda)$), Sfetsos derived the shock wave equation of motion which relates the shift profile $\alpha(x)$ and the energy momentum profile $T^{(p)}(x)$ via a certain differential equation [120]. In this chapter, we will work on arbitrary covariant theory of gravity (*i.e.* arbitrary covariant Lagrangian $L^{(G)}$) which may be beyond general relativity.

We will investigate the shock wave geometry by using the covariant phase space formalism [138, 31, 90, 139, 74, 140] [4] and derive a microscopic thermodynamic formula which captures low-energy excitations on the Killing horizon. The standard method utilizes the timelike Killing vector $\xi^{(t)} = \partial_t$ to construct the Noether charge $\mathbf{Q}_\xi$ and relate the horizon area (at $u = v = 0$) to the ADM thermodynamic parameters at asymptotic infinity. It turns out, however, that the area variation due to the gravitational shock wave vanishes at $u \to 0$, and hence non-trivial thermodynamic relations cannot be derived.

In this chapter, we study the gravitational shock wave by using the Noether charge $\mathbf{Q}_{\hat{\beta}}$ constructed from another horizon translation vector $\hat{\beta}$ in the u direction:

$$\hat{\beta} = \beta(x)\partial_u \ . \tag{1.6}$$

In particular, we probe the gravitational shock wave, which was generated by the shift $v \to v + \alpha(x)$ in the v direction, by another shift $u \to u + \beta(x)$ in the u direction. For general relativity, this enables us to derive the following microscopic thermodynamic relation:

$$\text{Area}(\beta(x), \alpha(x)) - \text{Area}(0, 0) = \int \sqrt{-g}\beta(x)T^{(p)}(x)dx \ , \tag{1.7}$$

up to a multiplicative factor. Here $\text{Area}(\beta(x), \alpha(x))$ corresponds to the total area of the sphere at $u = \beta(x)$ and $v = \alpha(x)$. For a generic covariant Lagrangian $\mathbf{L}^{(G)}$, the area term

[4] See [78] for a review.

would correspond to the Wald's geometric entropy. This thermodynamic relation can be rewritten as

$$E_{\text{shock}}\beta(x) = 0 \ , \tag{1.8}$$

such that $E_{\text{shock}} = 0$ corresponds to the shock wave equation of motion for $\alpha(x)$ and $T^{(p)}(x)$.

We will also study the gravitational scattering unitary matrix:

$$\mathcal{U} = \exp\left[i \iint d\Omega_1 d\Omega_2 P_{\text{out}}(\Omega_1) f(\Omega_1, \Omega_2) P_{\text{in}}(\Omega_2) \right] \ . \tag{1.9}$$

Let us assume that P_{out} and P_{in} induce the shifts $\alpha(x)$ and $\beta(x)$. Then, for general relativity, the above thermodynamic characterization enables us to rewrite $\mathcal{U}$ as follows:

$$\mathcal{U} = \exp\Big(i(\text{Area}(\alpha(x), \beta(x)) - \text{Area}(0,0)) \Big) \ , \tag{1.10}$$

up to a multiplicative factor for the area. Hence, the scattering matrix is the exponential of the horizon area (or equivalently, the soft charge).

1.1.2 Relation to previous works

The microscopic relation between the area deformation and the gravitational shock wave for the Schwarzschild black hole, for the case with $T^{\mu\nu} = 0$ (*i.e.* a "sourceless" shock wave), was initially communicated to us from Yoni BenTov [5]. Our contribution is to include the energy momentum tensor and extend this observation to a bifurcate Killing horizon in generic static spacetime for arbitrary covariant theory of gravity. The covariant phase space formalism has been recently applied to the Jackiw-Teitelboim (JT) gravity which is a tractable model of the gravitational backreaction [59].

In an asymptotically flat space, Hawking, Perry and Strominger have identified a concrete way of generating the gravitational shock wave from BMS supertranslation and superrotation on the boundary of the asymptotic flat spacetime [63]. Namely they provided a generic formula to prove the shock wave can be created by a supertranslation, by using the Noether charge constructed from a supertoration. This result was further investigated and extended, see [35, 137, 110, 80] for instance. It is worth clarifying our contributions and novelties in comparison with these previous works. First, previous works study the linearized shock wave near the horizon whereas our treatment applies to the shock wave

[5] BenTov told us that he learned it from Alexei Kitaev.

which runs exactly on the horizon as well. Second, our thermodynamic relation is cast in a concrete form. The shock waves described by these relations can be considered low-energy excitations, which span the black hole microstates. So these thermodynamic relations give useful hints to the Hilbert space of a black hole.

Third, as already noted above, our treatment applies to any bifurcate Killing horizon in generic static spacetime for arbitrary covariant theory of gravity.

It should be noted that, while our primary motivation stems from quantum aspects of black holes, the treatments in the present chapter are entirely classical. Also, our treatment is localized near the horizon and does not consider asymptotic infinity or boundaries. As such, we do not discuss BMS transformations in a direct manner.

1.1.3 Organization

The chapter is organized as follows. In section 1.2, we present a review of Wald's Noether charge method. In section 1.3, we will make certain refinements to the Wald's entropy formula so that it is valid up to higher order of matter field variations. In section 1.4, we present a review of gravitational shock wave in general relativity on a bifurcate Killing horizon in generic static spacetime. In section 1.5, we investigate the variations of the metric due to horizon translations. In section 1.6, we derive a microscopic thermodynamic relation on the horizon. In section 1.7, we apply our formalism to general relativity and demonstrate that the shockwave equations of motion indeed follows from the area deformation. In section 1.8, we apply our formalism to the Gauss-Bonnet gravity. In section 1.9, we briefly discuss the gravitational scattering matrix. In section 1.10, we conclude with discussions.

1.2 Review of the Noether charge method

In this section, we present a self-contained review of the Noether charge method in the presence of matter fields. Readers who are familiar with the method may skip to the next section.

1.2.1 Variation of Lagrangian

Bold letters are used to represent differential forms. Consider the following diffeomorphism covariant Lagrangian d-form on a d-dimensional manifold [6]:

$$\mathbf{L} = \mathbf{L}^{(G)} + \mathbf{L}^{(M)} \,, \tag{1.11}$$

where $\mathbf{L}^{(G)}$ is the gravitational Lagrangian d-form and $\mathbf{L}^{(M)}$ is the matter Lagrangian d-form. Here we assumed that the Lagrangian can be separated into two parts $\mathbf{L}^{(G)}$ and $\mathbf{L}^{(M)}$ [7]. At the linear order in $\delta g_{\mu\nu}, \delta\phi$, the variations can be expressed as

$$\begin{aligned} \delta\mathbf{L}^{(G)} &= \mathbf{E}^{(G)\mu\nu}\delta g_{\mu\nu} + d\mathbf{\Theta}^{(G)}(g_{\mu\nu}, \delta g_{\mu\nu}) \,; \\ \delta\mathbf{L}^{(M)} &= \frac{1}{2}\boldsymbol{\epsilon} T^{\mu\nu}\delta g_{\mu\nu} + \mathbf{E}^{(\phi)}\delta\phi + d\mathbf{\Theta}^{(M)}(g_{\mu\nu}, \phi, \delta\phi) \,, \end{aligned} \tag{1.12}$$

where ϕ collectively denotes all the matter fields. Here $\boldsymbol{\epsilon}$ denotes the volume form $\boldsymbol{\epsilon} = \mathbf{d^d x}\sqrt{-g}$ and $\mathbf{\Theta}^{(G)}, \mathbf{\Theta}^{(M)}$ are called symplectic potentials [8]. These variations lead to equations of motion:

$$\mathbf{E}^{(G)\mu\nu} + \frac{1}{2}\boldsymbol{\epsilon} T^{\mu\nu} = 0, \qquad \mathbf{E}^{(\phi)} = 0 \,, \tag{1.13}$$

where the energy momentum tensor is defined by [9]

$$T^{\mu\nu} \equiv \frac{2}{\sqrt{-g}}\frac{\delta(\sqrt{-g}L^{(M)})}{\delta g_{\mu\nu}}, \qquad \mathbf{L}^{(M)} = \boldsymbol{\epsilon} L^{(M)} \,. \tag{1.14}$$

[6] The diffeomorphism covariance condition is $\mathcal{L}_\xi \mathbf{L}(\psi) = \frac{\partial \mathbf{L}}{\partial \psi}\mathcal{L}_\xi \psi$.

[7] More precisely, we assume that $\mathbf{L}^{(G)}$ consists of terms which do not involve matter fields ϕ and their covariant derivatives $\nabla_{\mu_1}\dots\nabla_{\mu_m}\phi$ whereas $\mathbf{L}^{(M)}$ consists of terms which do not involve Riemann tensors R_{abcd} and their covariant derivatives $\nabla_{\mu_1}\dots\nabla_{\mu_m}R_{abcd}$. As such, Riemann tensors R_{abcd} do not directly couple to matter fields (e.g. terms like $R\phi^2$ do not appear) in the Lagrangian. Such matter fields are often said to be minimally coupled to gravity. Gravitational shock waves with non-minimal couplings have been studied in [84].

[8] Since the matter Lagrangian does not involve R_{abcd} and its covariant derivatives, the matter symplectic potential $\mathbf{\Theta}^{(M)}$ can be constructed so that it does not depend on $\delta g_{\mu\nu}$. See Section 3 of Iyer and Wald for details [74].

[9] The Euler-Lagrange derivative is defined by

$$\frac{\delta L}{\delta\phi} = \frac{\partial L}{\partial\phi} - \partial_\mu\frac{\partial L}{\partial\partial_\mu\phi} + \partial_\mu\partial_\nu\frac{\partial L}{\partial\partial_\mu\partial_\nu\phi} + \cdots \,.$$

Field configurations $\{g_{\mu\nu}, \phi\}$ which satisfy equations of motion Eq. (1.13) are called on shell.

Now consider variations under an infinitesimal diffeomorphism by an arbitrary vector field ξ. According to Eq. (1.12), the total on shell variation can be expressed by

$$\hat{\delta}_\xi \mathbf{L} = \hat{\delta}_\xi \mathbf{L}^{(G)} + \hat{\delta}_\xi \mathbf{L}^{(M)} = d\mathbf{\Theta}^{(G)}_\xi + d\mathbf{\Theta}^{(M)}_\xi \ , \qquad \{g_{\mu\nu}, \phi\} \text{ on shell} \ , \tag{1.15}$$

where $\mathbf{\Theta}^{(G)}_\xi \equiv \mathbf{\Theta}^{(G)}(g_{\mu\nu}, \hat{\delta}_\xi g_{\mu\nu})$ and $\mathbf{\Theta}^{(M)}_\xi \equiv \mathbf{\Theta}^{(M)}(g_{\mu\nu}, \phi, \hat{\delta}_\xi \phi)$ with $\hat{\delta}_\xi g_{\mu\nu} \equiv \mathcal{L}_\xi g_{\mu\nu}$ and $\hat{\delta}_\xi \phi \equiv \mathcal{L}_\xi \phi$. Here $\mathcal{L}_\xi$ denotes the Lie derivative with respect to ξ. It is worth emphasizing that the variation $\hat{\delta}_\xi \mathbf{L}$ is induced by a diffeomorphism $\hat{\delta}_\xi$ acting on each field $g_{\mu\nu}, \phi$.

The variation $\hat{\delta}_\xi \mathbf{L}$ can be expressed in another form due to the diffeomorphism covariance. The covariance of the gravity and matter Lagrangians implies

$$\hat{\delta}_\xi \mathbf{L}^{(G)} = \mathcal{L}_\xi \mathbf{L}^{(G)}, \qquad \hat{\delta}_\xi \mathbf{L}^{(M)} = \mathcal{L}_\xi \mathbf{L}^{(M)} \ , \tag{1.16}$$

where $\mathcal{L}_\xi \mathbf{L}^{(G)}, \mathcal{L}_\xi \mathbf{L}^{(M)}$ are the Lie derivatives of $\mathbf{L}^{(G)}, \mathbf{L}^{(M)}$. We have

$$\hat{\delta}_\xi \mathbf{L}^{(G)} = \mathcal{L}_\xi \mathbf{L}^{(G)} = d(\xi \cdot \mathbf{L}^{(G)}), \qquad \hat{\delta}_\xi \mathbf{L}^{(M)} = \mathcal{L}_\xi \mathbf{L}^{(M)} = d(\xi \cdot \mathbf{L}^{(M)}) \ , \tag{1.17}$$

where we made use of the Cartan's magic formula[10]

$$\mathcal{L}_\xi \mathbf{\Lambda} = \xi \cdot d\mathbf{\Lambda} + d(\xi \cdot \mathbf{\Lambda}) \ , \qquad \mathbf{\Lambda} : \text{ arbitrary} \ . \tag{1.18}$$

Note that $\hat{\delta}_\xi \mathbf{L}^{(G)}, \hat{\delta}_\xi \mathbf{L}^{(M)}$ must be total derivatives according to Eq. (1.17).

1.2.2 Wald's entropy formula

Wald introduced the following Noether current $(d-1)$-form

$$\mathbf{J}_{\xi,\text{Wald}} \equiv \mathbf{\Theta}^{(G)}_\xi - \xi \cdot \mathbf{L}^{(G)} + \mathbf{\Theta}^{(M)}_\xi - \xi \cdot \mathbf{L}^{(M)} \ . \tag{1.19}$$

One can verify

$$d\mathbf{J}_{\xi,\text{Wald}} = 0 \ , \qquad \{g, \phi\} \text{ on shell} \ . \tag{1.20}$$

[10]Here $\xi \cdot \mathbf{\Lambda}$ indicates insertion of ξ into the first argument of $\mathbf{\Lambda}$. It is also called interior product and denoted as $\iota_\xi \mathbf{\Lambda}$.

Here we used g to denote the metric tensor $g_{\mu\nu}$ for brevity of notation. One can actually show that $\mathbf{J}_{\xi,\text{Wald}}$ is exact due to the fact that $\mathbf{J}_{\xi,\text{Wald}}$ is closed for all ξ [90]. As such, the Noether charge $(d-2)$-form $\mathbf{Q}_{\xi,\text{Wald}}$ can be constructed [11]

$$\exists \mathbf{Q}_{\xi,\text{Wald}} \text{ such that } \mathbf{J}_{\xi,\text{Wald}} = d\mathbf{Q}_{\xi,\text{Wald}} , \qquad \{g,\phi\} \text{ on shell} . \tag{1.21}$$

Let us now fix ξ and consider the variation of the Noether current Eq. (1.19) at the linear orders of $\delta g_{\mu\nu}, \delta\phi$:

$$\begin{aligned} \delta\mathbf{J}_{\xi,\text{Wald}} &= \delta[\mathbf{\Theta}^{(G)}(\mathcal{L}_\xi g)] - \xi\cdot\delta\mathbf{L}^{(G)} + \delta[\mathbf{\Theta}^{(M)}(\mathcal{L}_\xi\phi)] - \xi\cdot\delta\mathbf{L}^{(M)} \\ &= \delta[\mathbf{\Theta}^{(G)}(\mathcal{L}_\xi g)] - \xi\cdot d\mathbf{\Theta}^{(G)}(\delta g) + \delta[\mathbf{\Theta}^{(M)}(\mathcal{L}_\xi\phi)] - \xi\cdot d\mathbf{\Theta}^{(M)}(\delta\phi) , \end{aligned} \tag{1.22}$$

where we used the on shell conditions Eq. (1.13) in the second line. Let us define the symplectic currents $\mathbf{\Omega}^{(G)}(g,\delta g,\mathcal{L}_\xi g)$, $\mathbf{\Omega}^{(M)}(g,\delta g,\mathcal{L}_\xi g,\phi,\delta\phi,\mathcal{L}_\xi\phi)$ by

$$\begin{aligned} \mathbf{\Omega}^{(G)}(g,\delta g,\mathcal{L}_\xi g) &\equiv \delta[\mathbf{\Theta}^{(G)}(g,\mathcal{L}_\xi g)] - \mathcal{L}_\xi[\mathbf{\Theta}^{(G)}(g,\delta g)] ; \\ \mathbf{\Omega}^{(M)}(g,\delta g,\mathcal{L}_\xi g,\phi,\delta\phi,\mathcal{L}_\xi\phi) &\equiv \delta[\mathbf{\Theta}^{(M)}(g,\phi,\mathcal{L}_\xi\phi)] - \mathcal{L}_\xi[\mathbf{\Theta}^{(M)}(g,\phi,\delta\phi)] . \end{aligned} \tag{1.23}$$

By using the Cartan's formula Eq. (1.18), we obtain

$$\begin{aligned} \delta\mathbf{J}_{\xi,\text{Wald}} = {} & \mathbf{\Omega}^{(G)}(\delta g,\mathcal{L}_\xi g) + d(\xi\cdot\mathbf{\Theta}^{(G)}(\delta g)) \\ & + \mathbf{\Omega}^{(M)}(\delta g,\mathcal{L}_\xi g,\delta\phi,\mathcal{L}_\xi\phi) + d(\xi\cdot\mathbf{\Theta}^{(M)}(\delta\phi)) , \quad \{g,\phi\} \text{ on shell} . \end{aligned} \tag{1.24}$$

Finally we shall focus on variations $\delta g, \delta\phi$ which satisfy the linearized equations of motion. This allows us to replace $\mathbf{J}_{\xi,\text{Wald}}$ and its variation $\delta\mathbf{J}_{\xi,\text{Wald}}$ with the Noether charge $d\mathbf{Q}_{\xi,\text{Wald}}$ and its variation $\delta d\mathbf{Q}_{\xi,\text{Wald}}$. Hence, by focusing on on shell variations, we obtain

$$\begin{aligned} \delta d\mathbf{Q}_{\xi,\text{Wald}} = {} & \mathbf{\Omega}^{(G)}(\delta g,\mathcal{L}_\xi g) + d\big(\xi\cdot\mathbf{\Theta}^{(G)}(\delta g)\big) \\ & + \mathbf{\Omega}^{(M)}(\delta g,\mathcal{L}_\xi g,\delta\phi,\mathcal{L}_\xi\phi) + d\big(\xi\cdot\mathbf{\Theta}^{(M)}(\delta\phi)\big) , \quad \{\delta g,\delta\phi\} \text{ on shell} . \end{aligned} \tag{1.25}$$

The above expression can be further simplified when ξ is a Killing vector:

$$\mathcal{L}_\xi g_{\mu\nu} = 0 . \tag{1.26}$$

Let us further assume a similar relation for the matter field:

$$\mathcal{L}_\xi\phi = 0 . \tag{1.27}$$

[11] A systematic algorithm to construct the Noether charge $\mathbf{Q}_{\xi,\text{Wald}}$ can be found in Iyer and Wald [74].

Then at the linear order in δg and $\delta\phi$, we have

$$\mathbf{\Omega}^{(G)}(\delta g, \mathcal{L}_\xi g) = 0, \qquad \mathbf{\Omega}^{(M)}(\delta g, \mathcal{L}_\xi g, \delta\phi, \mathcal{L}_\xi\phi) = 0 \ . \tag{1.28}$$

Hence we arrive at the well-celebrated result of Wald:

$$\delta d\mathbf{Q}_{\xi,\text{Wald}} = d\big(\xi \cdot \mathbf{\Theta}^{(G)}(\delta g)\big) + d\big(\xi \cdot \mathbf{\Theta}^{(M)}(\delta\phi)\big) \ , \quad \xi : \text{Killing vector, } \mathcal{L}_\xi\phi = 0, \{\delta g, \delta\phi\} \text{ on shell} \ . \tag{1.29}$$

1.2.3 Example: general relativity

We illustrate the covariant phase space formalism and the Noether charge by looking at general relativity with scalar fields. The Lagrangian is given by

$$\mathbf{L}^{(G)} = \frac{1}{16\pi}\boldsymbol{\epsilon}(R - 2\Lambda), \qquad \mathbf{L}^{(M)} = -\frac{1}{2}\boldsymbol{\epsilon}(\nabla^\mu\phi)(\nabla_\mu\phi) \ . \tag{1.30}$$

The equations of motion are

$$\mathbf{E}^{(G)\mu\nu} = -\frac{1}{16\pi}\boldsymbol{\epsilon}G^{\mu\nu}, \qquad \mathbf{E}^{(\phi)} = \boldsymbol{\epsilon}\nabla_\mu\nabla^\mu\phi \ , \tag{1.31}$$

where $G^{\mu\nu} \equiv R^{\mu\nu} - \frac{1}{2}Rg^{\mu\nu} + \Lambda g^{\mu\nu}$. The symplectic potentials are given by

$$\mathbf{\Theta}^{(G)}_{abc} = \boldsymbol{\epsilon}_{dabc}\frac{1}{16\pi}g^{de}g^{fh}(\nabla_f\delta g_{eh} - \nabla_e\delta g_{fh}) \ , \qquad \mathbf{\Theta}^{(M)}_{abc} = -\boldsymbol{\epsilon}_{dabc}(\nabla^d\phi)\delta\phi \ . \tag{1.32}$$

and the energy momentum tensor is

$$T_{\mu\nu} = \nabla_\mu\phi\nabla_\nu\phi - \frac{1}{2}g_{\mu\nu}\nabla_\alpha\phi\nabla^\alpha\phi \ . \tag{1.33}$$

For variations by diffeomorphism, we have

$$\mathbf{\Theta}^{(G)}_{\xi\,abc} = \boldsymbol{\epsilon}_{dabc}\frac{1}{16\pi}(\nabla_e\nabla^d\xi^e + \nabla_e\nabla^e\xi^d - 2\nabla^d\nabla_e\xi^e) \ . \tag{1.34}$$

To compute the Noether current, it is useful to make use of the following relation:

$$\nabla_e\nabla^d\xi^e - \nabla^d\nabla_e\xi^e = -R^{\beta\ \ d}_{\ \alpha\ \ \beta}\xi^\alpha = R_\alpha{}^d\xi^\alpha \ . \tag{1.35}$$

We can then rewrite $\boldsymbol{\Theta}^{(G)}_{\xi}$ as

$$\boldsymbol{\Theta}^{(G)}_{\xi}{}_{abc} = \boldsymbol{\epsilon}_{dabc}\frac{1}{16\pi}(\nabla_e\nabla^e\xi^d - \nabla_e\nabla^d\xi^e + 2R_\alpha{}^d\xi^\alpha) \ , \tag{1.36}$$

and obtain

$$\mathbf{J}_{\xi,\mathrm{Wald}}{}_{abc} = \boldsymbol{\epsilon}_{dabc}\frac{1}{16\pi}(\nabla_e\nabla^e\xi^d - \nabla_e\nabla^d\xi^e + 2R_\alpha{}^d\xi^\alpha) - \frac{1}{16\pi}\boldsymbol{\epsilon}_{dabc}\xi^d(R-2\Lambda) - \boldsymbol{\epsilon}_{dabc}T^{de}\xi_e \ . \tag{1.37}$$

By using the Einstein equation:

$$T_{\mu\nu} = \frac{1}{8\pi}\Big(R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}(R-2\Lambda)\Big) \ , \tag{1.38}$$

we arrive at

$$\mathbf{J}_{\xi,\mathrm{Wald}}{}_{abc} = \boldsymbol{\epsilon}_{dabc}\frac{1}{16\pi}(\nabla_e\nabla^e\xi^d - \nabla_e\nabla^d\xi^e) \ . \tag{1.39}$$

Hence the expression of $\mathbf{J}_{\xi,\mathrm{Wald}}$ is written entirely with the metric and does not contain Λ, ϕ or $T_{\mu\nu}$ explicitly. Finally we obtain the Noether charge

$$\mathbf{Q}_{\xi,\mathrm{Wald}}{}_{ab} = -\frac{1}{16\pi}\boldsymbol{\epsilon}_{abcd}\nabla^c\xi^d \ . \tag{1.40}$$

1.3 Two refinements to Wald's entropy formula

The original Noether charge method is not particularly suitable for studying the gravitational shock wave geometries. In this section, we will make two refinements to the Wald's entropy formula.

First, we will derive an alternative expression of the Noether current $\mathbf{J}_{\xi,\mathrm{field}}$, which differs from the Wald's construction $\mathbf{J}_{\xi,\mathrm{Wald}}$, by evaluating the Lie derivative of the matter Lagrangian $\mathcal{L}_\xi\mathbf{L}^{(M)}$ explicitly and relating it to the energy momentum tensor $T_{\mu\nu}$. Our construction of the Noether charge $\mathbf{Q}_{\xi,\mathrm{field}}$ is particularly useful for studying the gravitational backreaction as it relates the energy momentum tensor variation directly to the metric variation.

Second, we will extend the linear order analysis by Wald to the higher order matter field variations $\delta\phi$ while keeping the metric variation δg at the linear order. This is due to the fact that contributions from the linear order variation $\delta\phi$ vanish for the gravitational shock wave geometries, and hence the linear order variation of $T^{\mu\nu}$ is induced by the higher order variation of ϕ. Our construction of the Noether current $\mathbf{Q}_{\xi,\mathrm{field}}$ enables us to evaluate its variation $\delta\mathbf{Q}_{\xi,\mathrm{field}}$ under the matter field variation $\delta\phi$ at *any order* in a systematic manner.

1.3.1 Matter current

Our derivation deviates from Wald's where we evaluate the Lie derivative $\mathcal{L}_\xi \mathbf{L}^{(M)}$ of the matter Lagrangian in Eq. (1.17). Diffeomorphism variation of the matter Lagrangian can be explicitly evaluated as follows:

$$\begin{aligned}
\hat{\delta}_\xi \mathbf{L}^{(M)} &= \mathcal{L}_\xi \mathbf{L}^{(M)} = \mathcal{L}_\xi g_{\mu\nu} \frac{\delta \mathbf{L}^{(M)}}{\delta g_{\mu\nu}} + \mathcal{L}_\xi \phi \frac{\delta \mathbf{L}^{(M)}}{\delta \phi} + d\mathbf{\Theta}_\xi^{(M)} \\
&= 2\nabla_{(\mu}\xi_{\nu)} \frac{\delta \mathbf{L}^{(M)}}{\delta g_{\mu\nu}} + \mathcal{L}_\xi \phi \frac{\delta \mathbf{L}^{(M)}}{\delta \phi} + d\mathbf{\Theta}_\xi^{(M)} \\
&= \mathbf{d^d x} \sqrt{-g} T^{\mu\nu} \nabla_{(\mu}\xi_{\nu)} + \mathcal{L}_\xi \phi \frac{\delta \mathbf{L}^{(M)}}{\delta \phi} + d\mathbf{\Theta}_\xi^{(M)} \\
&= \mathbf{d^d x} \sqrt{-g} \nabla_\mu (T^{\mu\nu}\xi_\nu) - \mathbf{d^d x} \sqrt{-g} (\nabla_\mu T^{\mu\nu}) \xi_\nu + \mathcal{L}_\xi \phi \ \mathbf{E}^{(\phi)} + d\mathbf{\Theta}_\xi^{(M)} \\
&= -\boldsymbol{\epsilon} (\nabla_\mu T^{\mu\nu}) \xi_\nu + \mathbf{d^d x} \partial_\mu (\sqrt{-g} T^{\mu\nu} \xi_\nu) + d\mathbf{\Theta}_\xi^{(M)} \ ,
\end{aligned} \tag{1.41}$$

where we made use of the metric compatibility

$$\nabla_\xi g_{\mu\nu} = 0, \qquad \mathcal{L}_\xi g_{\mu\nu} = \nabla_\mu \xi_\nu + \nabla_\nu \xi_\mu \ , \tag{1.42}$$

the evaluation of the integral by parts, and on shell condition of matter fields. Let us define the matter current $(d-1)$-form by

$$\mathbf{J}^{(M)}_{\xi,\text{field}} \equiv J^{(M)\ \mu}_{\xi,\text{field}} \sqrt{-g} (\mathbf{d^{d-1} x})_\mu, \qquad J^{(M)\ \mu}_{\xi,\text{field}} \equiv -T^{\mu\nu}\xi_\nu \ . \tag{1.43}$$

Here $(\mathbf{d^{d-1} x})_\mu$ is a $(d-1)$-form that does *not* contain $\mathbf{dx}^\mu$ [12]. We then obtain

$$\mathcal{L}_\xi \mathbf{L}^{(M)} = -\boldsymbol{\epsilon} (\nabla_\mu T^{\mu\nu}) \xi_\nu - d\mathbf{J}^{(M)}_{\xi,\text{field}} + d\mathbf{\Theta}^{(M)}_\xi \ , \qquad \phi \text{ on shell} \ . \tag{1.45}$$

Recall that $\mathcal{L}_\xi \mathbf{L}^{(M)}$ can be expressed as a total derivative, according to Eq. (1.17). Thus, from Eq. (1.45), we can deduce the conservation of the energy momentum tensor:

$$\nabla_\mu T^{\mu\nu} = 0 \ , \qquad \phi \text{ on shell} \ . \tag{1.46}$$

[12] Explicitly, we have

$$(\mathbf{d^{d-p}} x)_{\mu_1 \dots \mu_p} = \frac{1}{p!(n-p)!} \epsilon_{\mu_1 \dots \mu_p \nu_{p+1} \dots \nu_d} \ \mathbf{dx}^{\nu_{p+1}} \wedge \cdots \wedge \mathbf{dx}^{\nu_d} \ , \tag{1.44}$$

where ϵ is the Levi-Civita symbol. Here $\mathbf{J}^{(M)}_\xi$ should be understood as the Hodge dual of the 1-form current.

Then, the Lie derivative of the matter Lagrangian $\mathcal{L}_\xi \mathbf{L}^{(M)}$ can be expressed as

$$\mathcal{L}_\xi \mathbf{L}^{(M)} = -d\mathbf{J}^{(M)}_{\xi,\text{field}} + d\mathbf{\Theta}^{(M)}_\xi \ , \qquad \phi \text{ on shell} \ . \tag{1.47}$$

Also, the Lie derivative of the gravity Lagrangian is given by

$$\mathcal{L}_\xi \mathbf{L}^{(G)} = d\mathbf{J}^{(M)}_{\xi,\text{field}} + d\mathbf{\Theta}^{(G)}_\xi \ , \qquad \{g, \phi\} \text{ on shell} \ . \tag{1.48}$$

Here it is worth emphasizing that $\mathcal{L}_\xi \mathbf{L}^{(G)} \neq d\mathbf{\Theta}^{(G)}_\xi$ and $\mathcal{L}_\xi \mathbf{L}^{(M)} \neq d\mathbf{\Theta}^{(M)}_\xi$ due to the matter current.

We define the Noether current $(d-1)$-form by

$$\mathbf{J}_{\xi,\text{field}} \equiv \mathbf{\Theta}^{(G)}_\xi - \xi \cdot \mathbf{L}^{(G)} + \mathbf{J}^{(M)}_{\xi,\text{field}} \ . \tag{1.49}$$

One can verify that $\mathbf{J}_{\xi,\text{field}}$ is closed when $\{g, \phi\}$ are on shell:

$$\begin{aligned} d\mathbf{J}_{\xi,\text{field}} &= d\mathbf{\Theta}^{(G)}_\xi - d(\xi \cdot \mathbf{L}^{(G)}) + d\mathbf{J}^{(M)}_{\xi,\text{field}} \\ &= d\mathbf{\Theta}^{(G)}_\xi - \mathcal{L}_\xi \mathbf{L}^{(G)} + d\mathbf{J}^{(M)}_{\xi,\text{field}} \\ &= 0 \ . \end{aligned} \tag{1.50}$$

Then one can construct the Noether charge $\mathbf{Q}_{\xi,\text{field}}$:

$$\exists \mathbf{Q}_{\xi,\text{field}} \ \text{ such that } \ \mathbf{J}_{\xi,\text{field}} = d\mathbf{Q}_{\xi,\text{field}} \ , \qquad \{g, \phi\} \text{ on shell} \ . \tag{1.51}$$

Note that our definition $\mathbf{J}_{\xi,\text{field}}$ differs from Wald's $\mathbf{J}_{\xi,\text{Wald}}$ since, instead of $\mathbf{\Theta}^{(M)}_\xi - \xi \cdot \mathbf{L}^{(M)}$, we have used $\mathbf{J}^{(M)}_{\xi,\text{field}}$. We will compare two constructions in section 1.3.3

1.3.2 Higher-order matter variation

Until this point, our derivation of the Noether current Eq. (1.49) is an exact calculation which does not rely on perturbative analysis [13]. The remaining task is to evaluate the variation of the Noether current Eq. (1.49) under $\delta g_{\mu\nu}, \delta\phi$.

[13] Recall that it suffices to use the linearized variation formula Eq. (1.12) to evaluate the Lie derivative.

We will focus on the linear order metric variation $\delta g_{\mu\nu}$ while we do not impose such restriction on the matter field variation $\delta\phi$. We then obtain

$$\begin{aligned}\delta\mathbf{J}_{\xi,\text{field}} &= \delta[\mathbf{\Theta}^{(G)}(\mathcal{L}_\xi g)] - \xi\cdot\delta\mathbf{L}^{(G)} + \delta\mathbf{J}^{(M)}_{\xi,\text{field}} \\ &= \delta[\mathbf{\Theta}^{(G)}(\mathcal{L}_\xi g)] - \xi\cdot d\mathbf{\Theta}^{(G)}(\delta g) + \delta\mathbf{J}^{(M)}_{\xi,\text{field}} - \xi\cdot\mathbf{E}^{(G)\mu\nu}\delta g_{\mu\nu} \\ &= \delta[\mathbf{\Theta}^{(G)}(\mathcal{L}_\xi g)] - \mathcal{L}_\xi[\mathbf{\Theta}^{(G)}(\delta g)] + d(\xi\cdot\mathbf{\Theta}^{(G)}(\delta g)) + \delta\mathbf{J}^{(M)}_{\xi,\text{field}} - \xi\cdot\mathbf{E}^{(G)\mu\nu}\delta g_{\mu\nu} \\ &= \delta[\mathbf{\Theta}^{(G)}(\mathcal{L}_\xi g)] - \mathcal{L}_\xi[\mathbf{\Theta}^{(G)}(\delta g)] + d(\xi\cdot\mathbf{\Theta}^{(G)}(\delta g)) + \delta\mathbf{J}^{(M)}_{\xi,\text{field}} + \frac{1}{2}(\xi\cdot\boldsymbol{\epsilon})T^{\mu\nu}\delta g_{\mu\nu}\ ,\end{aligned} \tag{1.52}$$

where we used $\mathbf{E}^{(G)\mu\nu} = -\frac{1}{2}\boldsymbol{\epsilon}T^{\mu\nu}$. By using the gravity part of the symplectic current $\mathbf{\Omega}^{(G)}(\delta g, \mathcal{L}_\xi g) \equiv \delta[\mathbf{\Theta}^{(G)}(\mathcal{L}_\xi g)] - \mathcal{L}_\xi[\mathbf{\Theta}^{(G)}(\delta g)]$, we obtain [14]

$$\delta\mathbf{J}_{\xi,\text{field}} = \mathbf{\Omega}^{(G)}(\delta g, \mathcal{L}_\xi g) + d(\xi\cdot\mathbf{\Theta}^{(G)}(\delta g)) + \delta\mathbf{J}^{(M)}_{\xi,\text{field}} + \frac{1}{2}(\xi\cdot\boldsymbol{\epsilon})T^{\mu\nu}\delta g_{\mu\nu}\ , \qquad \{g,\phi\}\text{ on shell}\ . \tag{1.53}$$

It is worth emphasizing again that Eq. (1.53) is valid up to the linear order in $\delta g_{\mu\nu}$ and up to *any order* in $\delta\phi$. Note that the matter field variation $\delta\phi$ enters through the matter current variation $\delta\mathbf{J}^{(M)}_{\xi,\text{field}}$. Finally we shall focus on variations $\delta g, \delta\phi$ which satisfy the equations of motion up to the linear order in δg and up to any desired order in $\delta\phi$.

Finally, we arrive at

$$\delta d\mathbf{Q}_{\xi,\text{field}} = \mathbf{\Omega}^{(G)}(\delta g, \mathcal{L}_\xi g) + d\big(\xi\cdot\mathbf{\Theta}^{(G)}(\delta g)\big) + \delta\mathbf{J}^{(M)}_{\xi,\text{field}} + \frac{1}{2}(\xi\cdot\boldsymbol{\epsilon})T^{\mu\nu}\delta g_{\mu\nu}\ , \quad \{\delta g,\delta\phi\}\text{ on shell}\ . \tag{1.54}$$

We will use this formula to derive the microscopic thermodynamic relation.

The above expression can be further simplified when ξ is a Killing vector of the spacetime:

$$\mathcal{L}_\xi g_{\mu\nu} = 0\ . \tag{1.55}$$

Then we have $\mathbf{\Omega}^{(G)}(\delta g, \mathcal{L}_\xi g) = 0$ in the linear order in δg.

[14] In fact, we do not need the on shell condition for the matter fields ϕ for this equation to hold. We would need it to make sure $d\mathbf{J}^{(M)}_{\xi,\text{field}} = 0$ so that the matter charge $\mathbf{Q}^{(M)}_{\xi,\text{field}}$ can be constructed.

When ξ is a Killing vector, the matter current is conserved:

$$\nabla_\mu J^{(M)\ \mu}_{\xi,\text{field}} = -\nabla_\mu(T^{\mu\nu}\xi_\nu) = -(\nabla_\mu T^{\mu\nu})\xi_\nu - T^{\mu\nu}\nabla_\mu\xi_\nu = -\frac{1}{2}T^{\mu\nu}(\nabla_\mu\xi_\nu + \nabla_\nu\xi_\mu) = 0\ , \tag{1.56}$$

or equivalently

$$d\mathbf{J}^{(M)}_{\xi,\text{field}} = 0\ . \tag{1.57}$$

Thus, we can construct the matter charge $(d-2)$-form $\mathbf{Q}^{(M)}_{\xi,\text{field}}$ such that

$$\mathbf{J}^{(M)}_{\xi,\text{field}} = d\mathbf{Q}^{(M)}_{\xi,\text{field}}\ . \tag{1.58}$$

Hence we arrive at the following expression:

$$\delta d\mathbf{Q}_{\xi,\text{field}} = d\big(\xi\cdot\mathbf{\Theta}^{(G)}(\delta g)\big) + \delta d\mathbf{Q}^{(M)}_{\xi,\text{field}} + \frac{1}{2}(\xi\cdot\boldsymbol{\epsilon})T^{\mu\nu}\delta g_{\mu\nu}\ , \xi:\ \text{Killing vector},\ \{\delta g, \delta\phi\}\ \text{on shell}\ , \tag{1.59}$$

which is valid up to linear order in δg and any order in $\delta\phi$.

1.3.3 Comparison

Let us explicitly compare two possible constructions of the matter current

$$\begin{aligned} \mathbf{J}^{(M)}_{\xi,\text{Wald}} &\equiv \mathbf{\Theta}^{(M)}_\xi(\phi, \mathcal{L}_\xi\phi) - \xi\cdot\mathbf{L}^{(M)}\ ; \\ \mathbf{J}^{(M)}_{\xi,\text{field}} &\equiv J^{(M)\ \mu}_{\xi,\text{field}}\sqrt{-g}(\mathbf{d^{d-1}x})_\mu, \qquad \text{where} \quad J^{(M)\ \mu}_{\xi,\text{field}} \equiv -T^{\mu\nu}\xi_\nu\ . \end{aligned} \tag{1.60}$$

To begin, we compare their total derivatives:

$$\begin{aligned} d\mathbf{J}^{(M)}_{\xi,\text{Wald}} &= d\mathbf{\Theta}^{(M)}_\xi - d(\xi\cdot\mathbf{L}^{(M)}) \\ &= -\mathbf{E}^{(\phi)}\mathcal{L}_\xi\phi - \frac{1}{2}\boldsymbol{\epsilon}T^{\mu\nu}\mathcal{L}_\xi g_{\mu\nu}\ , \end{aligned} \tag{1.61}$$

and

$$\begin{aligned} d\mathbf{J}^{(M)}_{\xi,\text{field}} &= -\mathbf{d^d x}\sqrt{-g}\nabla_\mu(T^{\mu\nu}\xi_\nu) \\ &= -\boldsymbol{\epsilon}\left((\nabla_\mu T^{\mu\nu})\xi_\nu + T^{\mu\nu}\nabla_{(\mu}\xi_{\nu)}\right) \\ &= -\boldsymbol{\epsilon}\left((\nabla_\mu T^{\mu\nu})\xi_\nu + \frac{1}{2}T^{\mu\nu}\mathcal{L}_\xi g_{\mu\nu}\right)\ . \end{aligned} \tag{1.62}$$

Their difference indeed vanishes for matter fields on shell:

$$d\mathbf{J}^{(M)}_{\xi,\text{field}} - d\mathbf{J}^{(M)}_{\xi,\text{Wald}} = \mathbf{E}^{(\phi)}\mathcal{L}_\xi\phi - \boldsymbol{\epsilon}(\nabla_\mu T^{\mu\nu})\xi_\nu = 0 \ , \qquad \phi \text{ on shell} \ . \tag{1.63}$$

Hence both $\mathbf{J}^{(M)}_{\xi,\text{Wald}}$ and $\mathbf{J}^{(M)}_{\xi,\text{field}}$ lead to valid constructions of the Noether charge $\mathbf{Q}_{\xi,\text{Wald}}$ and $\mathbf{Q}_{\xi,\text{field}}$.

Next, let us look at the scalar field Lagrangian

$$\mathbf{L}_{\text{scalar}} = -\frac{1}{2}\boldsymbol{\epsilon} g^{\mu\nu}\nabla_\mu\phi\nabla_\nu\phi \ . \tag{1.64}$$

We have

$$\begin{aligned}
\mathbf{J}^{(M)}_{\xi,\text{field}} &= -T^{\mu\nu}\xi_\nu\sqrt{-g}(\mathbf{d^{d-1}x})_\mu \\
&= -(\nabla^\mu\phi\nabla^\nu\phi\ \xi_\nu - \frac{1}{2}\xi^\mu\nabla_\alpha\phi\nabla^\alpha\phi)\sqrt{-g}(\mathbf{d^{d-1}x})_\mu \ ; \\
\mathbf{J}^{(M)}_{\xi,\text{Wald}} &= -\big((\nabla^\mu\phi)\mathcal{L}_\xi\phi - \frac{1}{2}\xi^\mu\nabla_\alpha\phi\nabla^\alpha\phi\big)\sqrt{-g}(\mathbf{d^{d-1}x})_\mu \\
&= -\big(\xi_\nu\nabla^\nu\phi\nabla^\mu\phi - \frac{1}{2}\xi^\mu\nabla_\alpha\phi\nabla^\alpha\phi\big)\sqrt{-g}(\mathbf{d^{d-1}x})_\mu \ .
\end{aligned} \tag{1.65}$$

Hence, we have

$$\mathbf{J}^{(M)}_{\xi,\text{field}} = \mathbf{J}^{(M)}_{\xi,\text{Wald}} \ . \tag{1.66}$$

For the electromagnetic Lagrangian

$$\mathbf{L}_{\text{EM}} = -\frac{1}{4}\boldsymbol{\epsilon} F_{\mu\nu}F^{\mu\nu} \ , \tag{1.67}$$

we have

$$T_{\mu\nu} = F_{\mu\alpha}F_\nu{}^\alpha - \frac{1}{4}g_{\mu\nu}F_{\alpha\beta}F^{\alpha\beta} \ , \tag{1.68}$$

$$\boldsymbol{\Theta}^{(M)}_{abc}(A,\delta A) = -\boldsymbol{\epsilon}_{\mu abc}F^{\mu\alpha}\delta A_\alpha \ , \tag{1.69}$$

$$\mathbf{E}^{(A),\alpha} = \frac{\delta\mathbf{L}_{\text{EM}}}{\delta A_\alpha} = \boldsymbol{\epsilon}\nabla_\mu F^{\mu\alpha} \ , \tag{1.70}$$

and

$$\begin{aligned}
\mathbf{J}^{(M)}_{\xi,\text{field}} &= -T^{\mu\nu}\xi_\nu\sqrt{-g}(\mathbf{d^{d-1}x})_\mu \\
&= -(F^\mu{}_\alpha F^{\nu\alpha} - \frac{1}{4}g^{\mu\nu}F_{\alpha\beta}F^{\alpha\beta})\xi_\nu\sqrt{-g}(\mathbf{d^{d-1}x})_\mu \ ; \\
\mathbf{J}^{(M)}_{\xi,\text{Wald}} &= -(F^{\mu\alpha}\mathcal{L}_\xi A_\alpha - \frac{1}{4}\xi^\mu F_{\alpha\beta}F^{\alpha\beta})\sqrt{-g}(\mathbf{d^{d-1}x})_\mu \\
&= -(F^{\mu\alpha}\xi^\nu F_{\nu\alpha} + F^{\mu\alpha}\nabla_\alpha(\xi^\nu A_\nu) - \frac{1}{4}\xi^\mu F_{\alpha\beta}F^{\alpha\beta})\sqrt{-g}(\mathbf{d^{d-1}x})_\mu \ ,
\end{aligned} \tag{1.71}$$

where we used the Cartan's magic formula for the Lie derivative. Hence we obtain

$$\mathbf{J}^{(M)}_{\xi,\text{field}} - \mathbf{J}^{(M)}_{\xi,\text{Wald}} = F^{\mu\alpha}\nabla_\alpha(\xi^\nu A_\nu)\sqrt{-g}(\mathbf{d^{d-1}x})_\mu \ , \tag{1.72}$$

which, by imposing on shell conditions $\nabla_\alpha F^{\mu\alpha} = 0$, becomes

$$\mathbf{J}^{(M)}_{\xi,\text{field}} - \mathbf{J}^{(M)}_{\xi,\text{Wald}} = \nabla_\alpha(F^{\mu\alpha}\xi^\nu A_\nu)\sqrt{-g}(\mathbf{d^{d-1}x})_\mu \ , \qquad (A_\nu \text{ on shell}) \ . \tag{1.73}$$

Note that two definitions of the matter current for the electromagnetic field differ by a total derivative. Thus we can write

$$\mathbf{J}^{(M)}_{\xi,\text{field}} - \mathbf{J}^{(M)}_{\xi,\text{Wald}} = d\mathbf{Q}^{(\Delta)}_\xi \ , \tag{1.74}$$

which leads to the following difference of two constructions:

$$\mathbf{Q}_{\xi,\text{field}} - \mathbf{Q}_{\xi,\text{Wald}} = \mathbf{Q}^{(\Delta)}_\xi \ . \tag{1.75}$$

Here it is worth emphasizing that the charge difference $\mathbf{Q}^{(\Delta)}_\xi$ is linear in the vector field ξ, and thus vanishes as $\xi \to 0$.

In fact, by following Iyer and Wald (Lemma 3.1 in [74]), one can show that, for an arbitrary matter Lagrangian,

(a) $\mathbf{J}^{(M)}_{\xi,\text{field}}$ and $\mathbf{J}^{(M)}_{\xi,\text{Wald}}$ differ only by a total derivative.

(b) The charge difference $\mathbf{Q}^{(\Delta)}_\xi$ vanishes as $\xi \to 0$.

For our application to the gravitational shock wave geometries, terms which vanish as $\xi \to 0$ do not contribute to the surface integral of the Noether charge. As such, our construction of the Noether charge $\mathbf{Q}_{\xi,\text{field}}$ bears the same geometrical meaning as the Wald's construction $\mathbf{Q}_{\xi,\text{Wald}}$.

1.4 Gravitational shock wave in General Relativity

Having reviewed and refined the Noether charge formula, let us shift the gear a bit. In this section, we present a brief review of gravitational shock wave geometries in general gravity by following the works by Dray, 't Hooft and Sfetsos [38, 120] who derived the gravitational shock wave equations of motion from the Einstein equation.

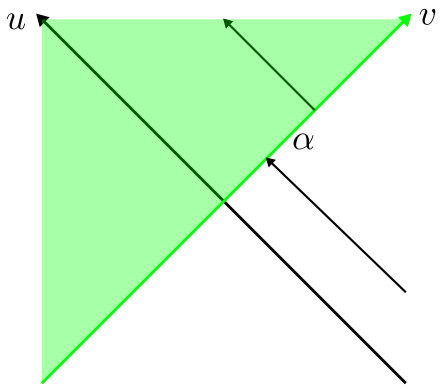


Figure 1.1: Gravitational shock wave at $u = 0$, moving in the v-direction. The shaded region experiences metric variations.

Consider the following family of d-dimensional static spacetime, expressed in the Kruskal-type coordinate and its corresponding energy momentum tensor:

$$\begin{aligned} ds^2 &= 2F(u,v)dudv + G(u,v)h_{ij}dx^i dx^j \ ; \\ T &= 2T_{uv}dudv + T_{uu}dudu + T_{vv}dvdv + T_{ij}dx^i dx^j \ . \end{aligned} \tag{1.76}$$

Following Sfetsos [120], we assume that the metric and matter fields satisfy the following conditions:

$$G_{,v} = F_{,v} = T_{vv} = 0 \ , \qquad (u = 0) \ , \tag{1.77}$$

$$G_{,u} = F_{,u} = T_{uu} = 0 \ , \qquad (v = 0) \ . \tag{1.78}$$

The above conditions follow from the assumption that the timelike vector ∂_t is a Killing vector of the spacetime as we will see in the next subsection. Non-vanishing Christoffel symbols are

$$\begin{aligned} &\Gamma^u_{uu} = \frac{F_{,u}}{F} \qquad \Gamma^u_{ij} = -\frac{G_{,v}}{2F}\ h_{ij} \qquad \Gamma^v_{vv} = \frac{F_{,v}}{F} \qquad \Gamma^v_{ij} = -\frac{G_{,u}}{2F}\ h_{ij} \\ &\Gamma^i_{uj} = \frac{G_{,u}}{2G}\delta^i_j \qquad \Gamma^i_{vj} = \frac{G_{,v}}{2G}\delta^i_j \qquad \Gamma^i_{jk} = \frac{1}{2}h^{il}(h_{lk,j} + h_{lj,k} - h_{jk,l}). \end{aligned} \tag{1.79}$$

Let us shift this spacetime by

$$v \quad \to \quad \tilde{v} = v + \theta(u)\alpha(x) \ , \tag{1.80}$$

where x represents x^j collectively and $\theta(u)$ is a step function. The resulting metric and energy momentum tensor are

$$d\tilde{s}^2 = 2F(u, v+\theta\alpha)du(dv + \theta\alpha_{,i}dx^i) + G(u, v+\theta\alpha)h_{ij}dx^i dx^j \ , \tag{1.81}$$

and

$$\begin{aligned}\tilde{T} = &2T_{uv}(u, v+\theta\alpha, x)du(dv + \theta\alpha_{,i}dx^i) + T_{uu}(u, v+\theta\alpha, x)du^2 \\ &+ T_{vv}(u, v+\theta\alpha, x)(dv + \theta\alpha_{,i}dx^i)^2 + T_{ij}(u, v+\theta\alpha, x)dx^i dx^j \ .\end{aligned} \tag{1.82}$$

Note that the shifted metric and the shifted energy momentum tensor still satisfy the Einstein equation except at $u = 0$ where an additional source of the energy momentum tensor is needed. Consider the extra contribution to the energy momentum tensor from a massless particle, located at $u = 0$ and moving with the speed of light in the v-direction:

$$T^{(P)} = T^{(P)}_{uu}du^2 = 4p^{(v)}F^2\delta(x)\delta(u)du^2 \ , \tag{1.83}$$

where $p^{(v)}$ is the momentum of the particle; $T^{(P),vv} = 4p^{(v)}\delta(u)\delta(x)$ [15]. The Einstein equation is satisfied if Eq. (1.77) holds *and* the following equation holds:

$$\Delta_{h_{ij}}\alpha(x) - \frac{d-2}{2}\frac{G_{,uv}}{F}\alpha(x) = 32\pi p^{(v)}GF\delta(x) \ . \tag{1.84}$$

Here the Laplacian is defined as $\Delta_{h_{ij}} \equiv \frac{1}{\sqrt{h}}\partial_i\sqrt{h}h^{ij}\partial_j$. We will derive this equation by using the Noether charge method later [16].

1.5 Horizon translation vector

In this section, we investigate the symmetry properties of the spacetime. The gravitational shock wave solutions require certain consistency conditions on the fields $g_{\mu\nu}, T_{\mu\nu}$ as we reviewed in the previous section. These conditions actually follow from the existence of the timelike Killing vector $\xi^{(t)} = \partial_t$ as we shall show in this section. We will also study the effect of a vector field $\hat{\beta}$ which introduce the horizon translation. In later sections, we will use the horizon translation vector $\hat{\beta}$ to construct the Noether charge which is sensitive to the gravitational shock wave. Note that the argument in this and next section holds not only for general relativity, but also for generic covariant theories of gravity.

[15] Our choice of $T^{(P),vv} = 4p^{(v)}\delta(u)\delta(x)$ differs from Sfetsos' by a factor of (-1) due to the sign difference in the Einstein equation.

[16] We note that Eq. (1.78) is not necessary to derive Eq. (1.84). We will however assume both Eq. (1.77) and Eq. (1.78) in this chapter since we are interested in static spacetime.

1.5.1 Timelike vector

In this subsection, we verify that Eq. (1.77) and Eq. (1.78) follow from the fact that the timelike $\xi^{(t)} = \partial_t$ vector is a Killing vector.

A vector field ξ is a Killing vector if $\mathcal{L}_\xi g_{\mu\nu} = 0$. Let us begin by computing $\mathcal{L}_\xi g_{\mu\nu}$ for the metric given in Eq. (1.76). Looking at $(\mu, \nu) = (u, v)$, we have

$$\begin{aligned} \mathcal{L}_\xi g_{uv} &= \xi^a(\partial_a g_{uv}) + (\partial_u \xi^a) g_{av} + (\partial_v \xi^a) g_{ua} \\ &= \xi^u F_{,u} + \xi^v F_{,v} + (\partial_u \xi^u + \partial_v \xi^v) F \ . \end{aligned} \tag{1.85}$$

Looking at $(\mu, \nu) = (i, j)$, we have

$$\begin{aligned} \mathcal{L}_\xi g_{ij} &= \xi^a(\partial_a g_{ij}) + (\partial_i \xi^a) g_{aj} + (\partial_j \xi^a) g_{ia} \\ &= \xi^u G_{,u} h_{ij} + \xi^v G_{,v} h_{ij} + \xi^k(\partial_k g_{ij}) + (\partial_i \xi^k) g_{kj} + (\partial_j \xi^k) g_{ik} \ . \end{aligned} \tag{1.86}$$

Looking at $(\mu, \nu) = (u, i)$, we have

$$\begin{aligned} \mathcal{L}_\xi g_{ui} &= \xi^a(\partial_a g_{ui}) + (\partial_u \xi^a) g_{ai} + (\partial_i \xi^a) g_{ua} \\ &= (\partial_u \xi^j) g_{ji} + (\partial_i \xi^v) g_{uv} \ . \end{aligned} \tag{1.87}$$

Looking at $(\mu, \nu) = (v, i)$, we have

$$\begin{aligned} \mathcal{L}_\xi g_{vi} &= \xi^a(\partial_a g_{vi}) + (\partial_v \xi^a) g_{ai} + (\partial_i \xi^a) g_{va} \\ &= (\partial_v \xi^j) g_{ji} + (\partial_i \xi^u) g_{uv} \ . \end{aligned} \tag{1.88}$$

Consider the following timelike vector field:

$$\xi^{(t)} \equiv \kappa(-u\partial_u + v\partial_v) \ , \tag{1.89}$$

where κ represents the surface gravity. One can verify

$$\mathcal{L}_{\xi^{(t)}} g_{uv}, \ \mathcal{L}_{\xi^{(t)}} g_{ij} \to 0 \ , \qquad u \to 0 \text{ or } v \to 0 \ , \tag{1.90}$$

and $\mathcal{L}_{\xi^{(t)}} g_{ui}, \mathcal{L}_{\xi^{(t)}} g_{vi} = 0$, suggesting that $\xi^{(t)}$ asymptotically becomes a Killing vector on the horizon at $u = 0$ and $v = 0$ due to Eq. (1.77) (1.78). Furthermore, we see that the norm of $\xi^{(t)}$ vanishes on the horizon, and $\xi^{(t)} = 0$ at $(u, v) = (0, 0)$ [17]. Hence, $u = 0$ and $v = 0$ hypersurfaces are Killing horizons of $\xi^{(t)}$. In static black holes such as the Schwarzschild black hole, $\xi^{(t)}$ corresponds exactly to the timelike Killing vector ∂_t [18].

[17] The norm is given by $2\xi^u g_{uv} \xi^v = -2\kappa^2 uvF \to 0$ for $u, v \to 0$ where we assumed $F(u, v)$ is not divergent at $u = 0$ or $v = 0$. For non-static black holes, such as the Kerr black hole, g_{uv} in the Kruskal-like coordinate contains a term proportional to $(uv)^{-1}$. As such, $u = 0$ or $v = 0$ hypersurfaces are not horizons with respect to the timelike Killing vector ∂_t.

[18] As for $T_{\mu\nu}$, we were not able to verify $\mathcal{L}_\xi T_{\mu\nu} = 0$ from the conditions $T_{vv} = 0$ $(u = 0)$ and $T_{uu} = 0$

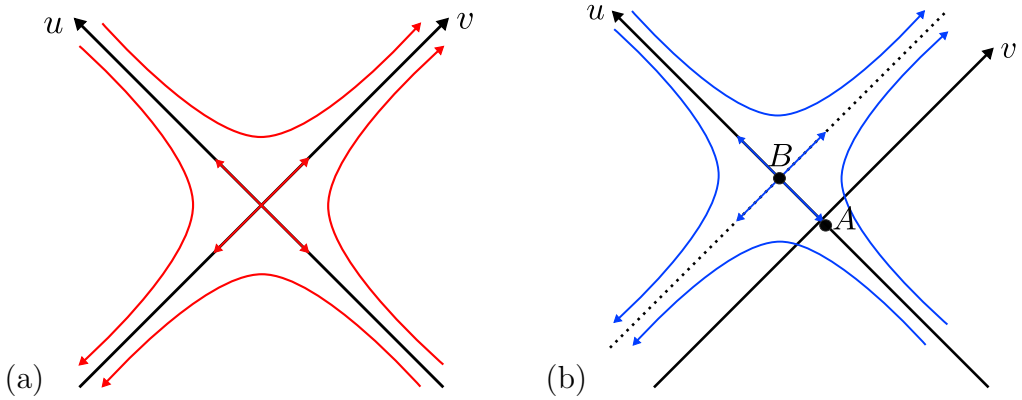


Figure 1.2: (a) The timelike Killing vector $\xi^{(t)}$. (b) The shifted timelike vector χ.

1.5.2 Horizon translation

Next, consider the following horizon translation vector field:

$$\hat{\beta} \equiv \beta(x)\partial_u \ , \tag{1.91}$$

where $\beta(x)$ is an arbitrary function of x^i. With direct calculations, one finds

$$\begin{aligned} \mathcal{L}_{\hat{\beta}} g_{uv} &= \beta(x)\big(\partial_u F\big), \qquad \mathcal{L}_{\hat{\beta}} g_{ij} = \beta(x) h_{ij}\big(\partial_u G\big) \ , \\ \mathcal{L}_{\hat{\beta}} g_{vi} &= \big(\partial_i \beta(x)\big) F, \qquad \mathcal{L}_{\hat{\beta}} g_{ui} = 0 \ . \end{aligned} \tag{1.92}$$

Namely, at the limit of $v \to 0$, we have

$$\mathcal{L}_{\hat{\beta}} g_{uv}, \mathcal{L}_{\hat{\beta}} g_{ij}, \mathcal{L}_{\hat{\beta}} g_{ui} \to 0 \ , \qquad \mathcal{L}_{\hat{\beta}} g_{vi} = \big(\partial_i \beta(x)\big) F \not\to 0 \ . \tag{1.93}$$

In the next section, we will use this horizon shift vector $\hat{\beta}$ to construct the Noether charge and derive the microscopic thermodynamic relations as in Eq. (1.7). It is worth recalling that, in the original treatment by Wald, the vector field ξ in the Noether charge $\mathbf{Q}_\xi$ was chosen to be a Killing vector. This was crucial in the derivation of Wald's Noether charge relation since the pre-symplectic form $\boldsymbol{\Omega}^{(G)}(g, \delta g, \mathcal{L}_\xi g)$ vanishes when ξ is a Killing vector. In the above calculation, we found that the horizon shift vector $\hat{\beta}$ is not a Killing vector at $v = 0$ (unless $\beta(x)$ does not depend on x). Nevertheless, the pre-symplectic form $\boldsymbol{\Omega}^{(G)}(g, \delta g, \mathcal{L}_{\hat{\beta}} g)$ vanishes at $v = 0$ as we shall see in section 1.6.3.

($v = 0$) in Eqs. (1.77)(1.78). Note, however, that we did not need to impose $\mathcal{L}_\xi \phi = 0$ or $\mathcal{L}_\xi T_{\mu\nu} = 0$ in deriving the Noether charge relation in section 1.3.

1.6 Microscopic thermodynamics on horizon

In this section, we will derive microscopic thermodynamic relations which are localized on the horizon.

1.6.1 Noether charge from shifted vector

Let us briefly recall the setup. The original static spacetime was given by Eq. (1.76) which is reprinted below

$$\begin{aligned} ds^2 &= 2F(u,v)dudv + G(u,v)h_{ij}dx^i dx^j \ ; \\ T &= 2T_{uv}dudv + T_{uu}dudu + T_{vv}dvdv + T_{ij}dx^i dx^j \ . \end{aligned} \tag{1.94}$$

The shifted metric and the energy momentum tensor are given by Eq. (1.81) and Eq. (1.82) which are obtained by the substitution $v \to \tilde{v} = v + \theta(u)\alpha(x)$. It is worth noting that, at the leading order in $\alpha(x)$, the metric variation is given by

$$\delta g_{\mu\nu} = \theta(u)\mathcal{L}_{\hat{\alpha}} g_{\mu\nu} \ , \qquad \hat{\alpha} \equiv \alpha(x)\partial_v \ , \tag{1.95}$$

which can be verified from results in section 1.5. In addition, there is an extra source of the energy momentum tensor $T^{(P)}_{uu}$ at $u = 0$ as in Eq. (1.83). See Fig. 1.3 for summary.

Let us specify how the integral of the Noether charge is taken. We will consider an integral of the Noether charge relation Eq. (1.54) over the $(d-1)$-dimensional surface Λ at $v = 0$ that connects two $(d-2)$-dimensional boundaries Σ_A and Σ_B (see Fig. 1.2(b) or Fig. 1.3):

$$\Sigma_A = \{(u,v) = (0_-, 0) \ , \qquad \Sigma_B = \{(u,v) = (\beta(x), 0)\} \ . \tag{1.96}$$

Here 0_- means that u approaches zero from below. For our purpose of deriving the shock wave equations of motion, it actually suffices to choose Σ_A as any boundary with $u < 0$.

Finally, let us construct the Noether charge. As a vector field, we shall take the shifted timelike vector:

$$\chi \equiv \xi^{(t)} + \kappa\hat{\beta} = \kappa\Big(\big(\beta(x) - u\big)\partial_u + v\partial_v\Big) \ , \tag{1.97}$$

in order to construct the Noether charge. It is essential to observe

$$\chi = 0 \qquad \text{at } \Sigma_B \ , \tag{1.98}$$

which will be useful in simplifying the thermodynamic relation.

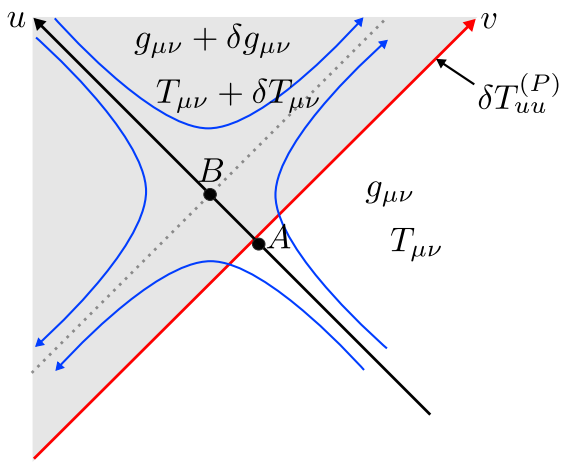


Figure 1.3: The Noether charge integral. A redline represents the gravitational shock wave running at $u = 0$. Blue lines represent the null-shifted timelike vector. The shaded region ($u > 0$) experiences the metric variations $\delta g_{\mu\nu}$ and the energy momentum variations $\delta T_{\mu\nu}$ which result from the shift $v \to v + \theta(u)\alpha(x)$. In addition, at $u = 0$, the variation $\delta T^{(P)}_{uu}$ from the shock wave is introduced.

1.6.2 Thermodynamic relation

Our central formula is Eq. (1.54) which is reprinted below:

$$\delta d\mathbf{Q}_{\chi,\text{field}} = \mathbf{\Omega}^{(G)}(\delta g, \mathcal{L}_\chi g) + d\big(\chi \cdot \mathbf{\Theta}^{(G)}(\delta g)\big) + \delta \mathbf{J}^{(M)}_{\chi,\text{field}} + \frac{1}{2}(\chi \cdot \boldsymbol{\epsilon}) T^{\mu\nu}\delta g_{\mu\nu} \ , \quad \{\delta g, \delta\phi\} \text{ on shell} \ . \tag{1.99}$$

Integrating it on the $v = 0$ surface generates

$$\delta \int_\Lambda d\mathbf{Q}_{\chi,\text{field}} = \int_\Lambda \mathbf{\Omega}^{(G)}(\delta g, \mathcal{L}_\chi g) + \int_\Lambda d\big(\chi \cdot \mathbf{\Theta}^{(G)}(\delta g)\big) + \delta \int_\Lambda d\mathbf{Q}^{(M)}_{\chi,\text{field}} \ . \tag{1.100}$$

Two comments follow. First, the last term $\frac{1}{2}(\chi \cdot \boldsymbol{\epsilon}) T^{\mu\nu}\delta g_{\mu\nu}$ vanishes since the integral is along the vector field χ and thus $\int_\Lambda (\chi \cdot \boldsymbol{\epsilon}) = 0$. Second, we have introduced the matter charge $\mathbf{Q}^{(M)}_{\chi,\text{field}}$. Here it is useful to note that, while $\hat{\beta}$ is not a Killing vector of the spacetime, one can still construct the matter charge on the $v = 0$ surface since

$$\nabla_\mu({J_{\chi,\text{field}}}^\mu) = -\frac{1}{2} T^{\mu\nu}(\mathcal{L}_\chi g_{\mu\nu}) = -T^{vi}(\mathcal{L}_\chi g_{vi}) = 0 \ , \qquad (v = 0). \tag{1.101}$$

Then, by using the Stokes' theorem, we obtain

$$\begin{aligned}\int_{\Sigma_A}\left(\delta\mathbf{Q}_{\chi,\text{field}}-\delta\mathbf{Q}^{(M)}_{\chi,\text{field}}-\chi\cdot\mathbf{\Theta}^{(G)}(\delta g)\right)+\int_{\Lambda}\mathbf{\Omega}^{(G)}(\delta g,\mathcal{L}_\chi g)\\ =\int_{\Sigma_B}\left(\delta\mathbf{Q}_{\chi,\text{field}}-\delta\mathbf{Q}^{(M)}_{\chi,\text{field}}-\chi\cdot\mathbf{\Theta}^{(G)}(\delta g)\right)\ .\end{aligned}\tag{1.102}$$

Below, we will further simplify this expression by using the properties of the shifted vector field χ.

Let us evaluate each term in Eq. (1.102) at Σ_A and Σ_B. First, we observe

$$\chi\cdot\mathbf{\Theta}^{(G)}(\delta g)=0\qquad\text{at }\Sigma_A,\Sigma_B\ ,\tag{1.103}$$

by noting that $\chi=0$ at Σ_B and $\delta g=0$ at Σ_A. Second, we observe

$$\delta\mathbf{Q}_{\chi,\text{field}}=0\qquad\text{at }\Sigma_A\ ,\tag{1.104}$$

since $\delta g=0$ at Σ_A. Third, we can construct the matter charge $\mathbf{Q}^{(M)}_{\chi,\text{field}}$ as an integral over $v=0$ surface in the u direction :

$$\mathbf{Q}^{(M)}_{\chi,\text{field}}\equiv\int_{-\infty}^{u}\mathbf{J}^{(M)}_{\chi,\text{field}}=-\int_{-\infty}^{u}T^{\mu\nu}\chi_\nu\sqrt{-g}(\mathbf{d^{d-1}x})_\mu=-\int_{-\infty}^{u}T^{vv}\chi_v\sqrt{-g}(\mathbf{d^{d-1}x})_v\ .\tag{1.105}$$

Due to the constraint $T^{vv}=0$ at $v=0$ in Eq. (1.78), we have

$$\delta\mathbf{Q}^{(M)}_{\chi,\text{field}}=-\int_{-\infty}^{u}(\delta T^{vv})\chi_v\sqrt{-g}(\mathbf{d^{d-1}x})_v\ .\tag{1.106}$$

There are two potential contributions to the variation δT^{vv}. The first contribution is due to the metric shift which is explicitly given by

$$\delta T^{vv}=T^{vv}(u,v+\alpha)-T^{vv}(u,v)\ ,\qquad u>0\ .\tag{1.107}$$

It turns out that this contribution is subleading in α. Recall that the timelike vector $\xi^{(t)}$ is the symmetry of the spacetime. We then have

$$\mathcal{L}_{\xi^{(t)}}T^{vv}=-u\partial_uT^{vv}+v\partial_vT^{vv}-2T^{vv}=0\ .\tag{1.108}$$

Let us Taylor expand T^{vv} for small u and v. Then we find that only the terms of the form u^av^{a+2} are allowed. This suggests that T^{vv} is $\sim v^2$ at most, and hence δT^{vv} is $\sim v\alpha+\alpha^2$.

Integrating it along $v = 0$ only generates a contribution of $O(\alpha^2)$. The second contribution comes from the shock wave. Considering $\delta T^{vv} = 4p^{(v)}\delta(u)\delta(x)$ at $v = 0$, we find

$$\delta \mathbf{Q}^{(M)}_{\chi,\text{field}} = 0 \qquad \text{at } \Sigma_A \ , \tag{1.109}$$

and

$$\begin{aligned} \delta \mathbf{Q}^{(M)}_{\chi,\text{field}} &= -\int_{-\infty}^{u=\beta(x)} 4p^{(v)} F \delta(u)\delta(x)\kappa(\beta(x)-u)\sqrt{-g}(\mathbf{d^{d-1}x})_v \\ &= (\mathbf{d^{d-2}x})_{uv}\sqrt{\mathcal{G}} 4p^{(v)} F \delta(x)\kappa\beta(x) \qquad \text{at } \Sigma_B \ . \end{aligned} \tag{1.110}$$

Note that $\sqrt{\mathcal{G}}$ is the volume element intrinsic on the $(d-2)$-dimensional surface $(u, v) = (0, 0)$. Hence, we obtain

$$\delta \int_{\Sigma_B} \left(\mathbf{Q}_{\chi,\text{field}} - \mathbf{Q}^{(M)}_{\chi,\text{field}} \right) - \int_{u=0_-}^{u=\beta(x)} \mathbf{\Omega}^{(G)}(\delta g, \mathcal{L}_\chi g) = 0 \ . \tag{1.111}$$

1.6.3 Pre-symplectic form

The remaining task is to evaluate the pre-symplectic form $\mathbf{\Omega}^{(G)}(\delta g, \mathcal{L}_\chi g)$. Here we will show that

$$\int_{u=0_-}^{u=\beta(x)} \mathbf{\Omega}^{(G)}(\delta g, \mathcal{L}_\chi g) = 0 \ . \tag{1.112}$$

Since the timelike vector $\xi^{(t)}$ is the symmetry of the original spacetime, we have $\mathbf{\Omega}^{(G)}(\delta g, \mathcal{L}_{\xi^{(t)}} g) \simeq 0$, and

$$\mathbf{\Omega}^{(G)}(\delta g, \mathcal{L}_\chi g) \simeq \kappa \mathbf{\Omega}^{(G)}(\delta g, \mathcal{L}_{\hat{\beta}} g) \ , \tag{1.113}$$

where $\simeq$ denotes that we evaluate the differential form on the $v = 0$ surface. Recalling $\delta g = \theta(u)\mathcal{L}_{\hat{\alpha}} g$ at the linear order in $\alpha(x)$, we arrive at

$$\int_{u=0_-}^{u=\beta(x)} \mathbf{\Omega}^{(G)}(\delta g, \mathcal{L}_\chi g) = \kappa \int_{u=0}^{u=\beta(x)} \mathbf{\Omega}^{(G)}(\mathcal{L}_{\hat{\alpha}} g, \mathcal{L}_{\hat{\beta}} g) \ . \tag{1.114}$$

Let us recall that the variation of the gravity Lagrangian is given by $\delta \mathbf{L}^{(G)} = \mathbf{E}^{(G)\mu\nu}\delta g_{\mu\nu} + d\mathbf{\Theta}(\delta g)$. Using the equations of motion and taking $\delta = \mathcal{L}_{\hat{\beta}}$, we obtain

$$\mathcal{L}_{\hat{\beta}} \mathbf{L}^{(G)} = -\frac{1}{2}\boldsymbol{\epsilon} T^{\mu\nu} \mathcal{L}_{\hat{\beta}} g_{\mu\nu} + d\mathbf{\Theta}_{\hat{\beta}} \ . \tag{1.115}$$

Taking another variation $\mathcal{L}_{\hat{\alpha}}$, integrating it and using the Cartan's formula, we have

$$\begin{aligned}\int \mathcal{L}_{\hat{\alpha}}\mathcal{L}_{\hat{\beta}}\mathbf{L}^{(G)} &= -\frac{1}{2}\int \mathcal{L}_{\hat{\alpha}}(\boldsymbol{\epsilon}T^{\mu\nu}\mathcal{L}_{\hat{\beta}}g_{\mu\nu}) + \int d\mathcal{L}_{\hat{\alpha}}\boldsymbol{\Theta}_{\hat{\beta}} \\ &= -\frac{1}{2}\int d(\hat{\alpha}\cdot\boldsymbol{\epsilon}T^{\mu\nu}\mathcal{L}_{\hat{\beta}}g_{\mu\nu}) + \int d\mathcal{L}_{\hat{\alpha}}\boldsymbol{\Theta}_{\hat{\beta}} \\ &= -\frac{1}{2}(\hat{\alpha}\cdot\boldsymbol{\epsilon}T^{\mu\nu}\mathcal{L}_{\hat{\beta}}g_{\mu\nu}) + \mathcal{L}_{\hat{\alpha}}\boldsymbol{\Theta}_{\hat{\beta}} \\ &= -\frac{1}{2}\alpha(x)\sqrt{-g}T^{\mu\nu}\mathcal{L}_{\hat{\beta}}g_{\mu\nu}(\mathbf{d^{d-1}x})_v + \mathcal{L}_{\hat{\alpha}}\boldsymbol{\Theta}_{\hat{\beta}} \\ &\simeq -\alpha(x)\sqrt{-g}T^{vi}\mathcal{L}_{\hat{\beta}}g_{vi}(\mathbf{d^{d-1}x})_v + \mathcal{L}_{\hat{\alpha}}\boldsymbol{\Theta}_{\hat{\beta}} = \mathcal{L}_{\hat{\alpha}}\boldsymbol{\Theta}_{\hat{\beta}}\ .\end{aligned} \tag{1.116}$$

Similarly, we have

$$\int \mathcal{L}_{\hat{\beta}}\mathcal{L}_{\hat{\alpha}}\mathbf{L}^{(G)} = -\frac{1}{2}\beta(x)\sqrt{-g}T^{\mu\nu}\mathcal{L}_{\hat{\alpha}}g_{\mu\nu}(\mathbf{d^{d-1}x})_u + \mathcal{L}_{\hat{\beta}}\boldsymbol{\Theta}_{\hat{\alpha}}\ . \tag{1.117}$$

Since $\mathcal{L}_{\hat{\alpha}}$ and $\mathcal{L}_{\hat{\beta}}$ commute, we have $\int \mathcal{L}_{\hat{\beta}}\mathcal{L}_{\hat{\alpha}}\mathbf{L}^{(G)} = \int \mathcal{L}_{\hat{\alpha}}\mathcal{L}_{\hat{\beta}}\mathbf{L}^{(G)}$. Hence we obtain [19]

$$\boldsymbol{\Omega}^{(G)}(g, \mathcal{L}_{\hat{\alpha}}g, \mathcal{L}_{\hat{\beta}}g) \equiv \mathcal{L}_{\hat{\alpha}}\boldsymbol{\Theta}_{\hat{\beta}} - \mathcal{L}_{\hat{\beta}}\boldsymbol{\Theta}_{\hat{\alpha}} \tag{1.118}$$

$$\simeq -\frac{1}{2}\beta(x)\sqrt{-g}T^{\mu\nu}\mathcal{L}_{\hat{\alpha}}g_{\mu\nu}(\mathbf{d^{d-1}x})_u\ . \tag{1.119}$$

By integrating $\boldsymbol{\Omega}^{(G)}(g, \mathcal{L}_{\hat{\alpha}}g, \mathcal{L}_{\hat{\beta}}g)$ on the $v = 0$ surface, we obtain

$$\int_{u=0}^{u=\beta(x)} \boldsymbol{\Omega}^{(G)}(\mathcal{L}_{\hat{\alpha}}g, \mathcal{L}_{\hat{\beta}}g) = 0\ , \tag{1.120}$$

since the term with $(\mathbf{d^{d-1}x})_u$ does not contribute to the integral.

This enables us to obtain the following simple constraint that governs microscopic thermodynamics of gravitational shock waves on the horizon:

$$\delta\int_{\Sigma_B}\mathbf{Q}_{\chi,\text{field}} - \delta\int_{\Sigma_B}\mathbf{Q}^{(M)}_{\chi,\text{field}} = 0\ . \tag{1.121}$$

This is the central result of this chapter.

As we will see in the next section, the first term corresponds to the area deformation under two shifts $u \to u + \beta(x)$ and $v \to v + \alpha(x)$. The second term corresponds to

[19] For general relativity, we have checked that $\mathcal{L}_{\hat{\alpha}}\boldsymbol{\Theta}_{\hat{\beta}} \simeq \mathcal{L}_{\hat{\beta}}\boldsymbol{\Theta}_{\hat{\alpha}} \simeq 0$ via brute force calculation.

the matter charge variation due to the gravitational shock wave, probed by the horizon translation vector field $\hat{\beta}$.

While the formula itself is remarkably simple (and might look unsurprising), the underlying reason behind the simplicity of the formula is rather non-trivial. Here it is useful to recall the generic relation concerning the Noether charge from Eq. (1.54) with on shell $\{\delta g, \delta\phi\}$, as reprinted below:

$$\delta d\mathbf{Q}_{\chi,\text{field}} - \delta d\mathbf{Q}^{(M)}_{\chi,\text{field}} = \mathbf{\Omega}^{(G)}(\delta g, \mathcal{L}_\chi g) + d\big(\chi \cdot \mathbf{\Theta}^{(G)}(\delta g)\big) - \chi \cdot \mathbf{E}^{(G)\mu\nu}\delta g_{\mu\nu} \ . \tag{1.122}$$

In the original analysis due to Wald, the vector field χ was chosen to be the timelike Killing vector. As a result, we have $\mathcal{L}_\chi g = 0$ everywhere in the spacetime and $\xi = 0$ at the bifurcation surface which would eliminate the contributions from three terms on the right hand side. In our treatment, however, ξ is *not* a Killing vector, and $\xi \neq 0$ at the bifurcation surface. Still, the contributions from three terms vanish due to careful choice of our vector field ξ and the interval for the integral.

Finally, it is worth emphasizing that, if the original unshifted timelike vector $\xi^{(t)}$ were used, we would not have any interesting thermodynamic relation since $\delta\mathbf{Q}^{(M)}_{\xi,\text{field}} = \delta\mathbf{Q}_{\xi,\text{field}} = 0$. This is related to the fact that the area variation by $v \to v+\alpha(x)$ vanishes at the bifurcate surface as we shall explicitly see in the next section.

1.7 Shock wave from area minimization

In this section, we derive the gravitational shock wave equations of motion in general relativity by using the microscopic thermodynamic relation on the horizon.

1.7.1 Area variation

We have evaluated the matter charge $\delta\mathbf{Q}^{(M)}_{\chi,\text{field}}$ in the previous section. Here we evaluate $\delta\mathbf{Q}_{\chi,\text{field}}$ at Σ_B. For general relativity with scalar fields, the Noether charge is given by [20]

$$\mathbf{Q}_{\chi,\text{field}_a} = -\frac{1}{16\pi}\boldsymbol{\epsilon}_{abc}\nabla^b\chi^c \ , \tag{1.123}$$

[20] For scalar fields, we have $\mathbf{Q}_{\chi,\text{field}} = \mathbf{Q}_{\chi,\text{Wald}}$. The difference between $\mathbf{Q}_{\chi,\text{field}}$ and $\mathbf{Q}_{\chi,\text{Wald}}$ generically vanishes as $\chi \to 0$. See section 1.3 for discussions.

where $a = a_1 \cdots a_{d-2}$ collectively denotes $(d-2)$ indices. Recall that χ is held fixed in variations. Also $\chi = 0$ at Σ_B. Hence we have

$$\delta \mathbf{Q}_{\chi,\text{field}_a} = -\frac{1}{16\pi}(\delta \boldsymbol{\epsilon}_{abc})\nabla^b \chi^c \qquad \text{at } \Sigma_B \ . \tag{1.124}$$

Non-trivial contributions come from $b, c = u, v$. We thus have

$$\delta \mathbf{Q}_{\chi,\text{field}} = -\frac{\kappa}{8\pi}\frac{1}{F}(\mathbf{d^{d-2}x})_{uv}\delta\sqrt{\mathcal{G}} \qquad \text{at } \Sigma_B \ , \tag{1.125}$$

where $(\mathbf{d^{d-2}x})_{uv}\sqrt{\mathcal{G}}$ is the volume element intrinsic on the $(d-2)$-dimensional surface Σ_B.

It is worth relating the variation of the Noether charge explicitly to the area deformation (Fig. 1.4). Let $\text{Area}(\beta, \alpha)$ denote the total area of the hypersphere at $(u, v) = (\beta, \alpha)$ in the original unperturbed metric ds^2. The Noether charge integral gives

$$\begin{aligned} \int_{\Sigma_B} \delta \mathbf{Q}_{\chi,\text{field}} &= -\frac{\kappa}{8\pi}\frac{1}{F}\Big(\text{Area}(\beta,\alpha) - \text{Area}(\beta,0)\Big) \\ &= -\frac{\kappa}{8\pi}\frac{1}{F}\Big(\text{Area}(\beta,\alpha) - \text{Area}(0,0)\Big) \ , \end{aligned} \tag{1.126}$$

where we have used $\text{Area}(\beta, 0) = \text{Area}(0, 0)$, since $(u, v) = (\beta, 0)$ is still on the horizon. Hence the Noether charge variation is proportional to the increase of the area of the bifurcating horizon by two null-shifts $u \to u + \beta$ and $v \to v + \alpha$. Note that α comes from the shift by the gravitational shock wave while β is set by the shift in the vector field χ.

One comment follows. If we used the timelike Killing vector ξ instead of the shifted one χ, we would have the contribution proportional to $\text{Area}(0, \alpha) - \text{Area}(0, 0)$ which vanishes since $(u, v) = (0, \alpha)$ is still on the horizon. Hence the Noether charge relation would be trivial.

1.7.2 Intrinsic metric

The remaining task is to evaluate the intrinsic metric on Σ_B. A spacetime point on Σ_B can be expressed as

$$X^\mu(\vec{y}) \equiv (\beta(\vec{y}), \alpha(\vec{y}), \vec{y}) \ , \tag{1.127}$$

where $\alpha(\vec{y})$ results from the metric variation. The intrinsic metric on Σ_B is given by

$$d\sigma^2 \equiv ds^2|_{(u,v,x)=X(\vec{y})} \equiv \mathcal{G}_{ij}(\vec{y})dy^i dy^j \ , \tag{1.128}$$

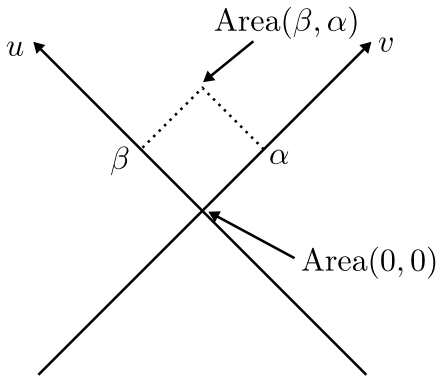


Figure 1.4: The area deformation in the original unperturbed geometry.

where

$$\mathcal{G}_{ij}(\vec{y}) = g_{\mu\nu}(X(\vec{y}))\frac{\partial X^\mu}{\partial y^i}\frac{\partial X^\nu}{\partial y^j} \ . \tag{1.129}$$

We shall Taylor expand $\mathcal{G}_{ij}(\vec{y})$ for small α, β. Looking at $(\mu,\nu) = (u,v)$, we have the following contribution to $\mathcal{G}_{ij}(\vec{y})$:

$$g_{uv}(X(\vec{y}))\frac{\partial\beta}{\partial y^i}\frac{\partial\alpha}{\partial y^j} = \left(F(0,0) + \frac{\partial^2 F(u,v)}{\partial u\partial v}\alpha\beta + \cdots\right)\frac{\partial\beta}{\partial y^i}\frac{\partial\alpha}{\partial y^j} \ . \tag{1.130}$$

Looking at $(\mu,\nu) = (i,j)$, we have the following contribution to $\mathcal{G}_{ij}(\vec{y})$:

$$g_{ij}(X(\vec{y}))dy^i dy^j = \left(G(0,0) + \frac{\partial^2 G(u,v)}{\partial u\partial v}\alpha\beta + \cdots\right)h_{ij}(\vec{y})dy^i dy^j \ . \tag{1.131}$$

We observe that the leading correction to $\mathcal{G}_{ij}(\vec{y})$ is at the order of $\alpha\beta$. In the expansion of Eq. (1.130), the term $\frac{\partial^2 F(u,v)}{\partial u\partial v}\alpha\beta\frac{\partial\beta}{\partial y^i}\frac{\partial\alpha}{\partial y^j}$ is at the order of $\alpha^2\beta^2$, and hence is negligible. Then we find

$$\begin{aligned} d\sigma^2 &\approx 2F(0,0)\frac{\partial\beta}{\partial y^i}\frac{\partial\alpha}{\partial y^j}dy^i dy^j + \left(G(0,0) + \frac{\partial^2 G(u,v)}{\partial u\partial v}\alpha\beta\right)h_{ij}(\vec{y})dy^i dy^j \\ &= d\sigma_*^2 + 2F(0,0)\frac{\partial\beta}{\partial y^i}\frac{\partial\alpha}{\partial y^j}dy^i dy^j + \frac{\partial^2 G(u,v)}{\partial u\partial v}\alpha\beta h_{ij}(\vec{y})dy^i dy^j \ , \end{aligned} \tag{1.132}$$

where $d\sigma_*^2$ represents the unperturbed intrinsic metric when $(\alpha,\beta) = (0,0)$:

$$d\sigma_*^2 = G(0,0)h_{ij}dy^i dy^j \ . \tag{1.133}$$

To summarize, we have obtained the variation of the intrinsic metric:

$$\delta\mathcal{G}_{ij} = 2F\partial_i\alpha\partial_j\beta + G_{,uv}\alpha\beta h_{ij} \ , \tag{1.134}$$

where F and $G_{,uv}$ are evaluated at $u = v = 0$.

Recall that the first order variation of the determinant is

$$\delta\sqrt{\mathcal{G}} = \frac{1}{2}\sqrt{\mathcal{G}}\mathcal{G}^{ij}\delta\mathcal{G}_{ij} \ . \tag{1.135}$$

Hence the variation of the determinant at Σ_B is given by

$$\delta\sqrt{\mathcal{G}} = \sqrt{\mathcal{G}}\frac{1}{2G}\Big(2F(\partial^j\alpha)(\partial_j\beta) + (d-2)G_{,uv}\alpha\beta\Big) \ . \tag{1.136}$$

Note $\mathcal{G}^{ij} = g^{ij} = \frac{h^{ij}}{F}$ at $u = v = 0$. Observe $\partial_j\Big(\beta(\partial^j\alpha)\sqrt{h}\Big) = (\partial_j\beta)(\partial^j\alpha)\sqrt{h} + \beta\partial_j\Big((\partial^j\alpha)\sqrt{h}\Big)$. After integrating by parts, we obtain

$$\delta\sqrt{\mathcal{G}} = \sqrt{\mathcal{G}}\frac{1}{2G}\Big(-2F\beta(\Delta_{h_{ij}}\alpha) + (d-2)G_{,uv}\alpha\beta\Big) \ . \tag{1.137}$$

Note that total derivatives do not contribute to the surface integral of the Noether charge.

Finally, we derive the shock wave equations of motion. Using the expression of deformed area density Eq. (1.137), we obtain

$$\delta\mathbf{Q}_{\chi,\text{field}} = -(\mathbf{d^{d-2}x})_{uv}\frac{\kappa\sqrt{\mathcal{G}}}{8\pi}\frac{1}{2GF}\Big(-2F\beta(\Delta_{h_{ij}}\alpha) + (d-2)G_{,uv}\alpha\beta\Big) \ , \tag{1.138}$$

and combining it with the variation of the matter Noether charge in Eq. (1.110), we then find that the constraint Eq. (1.121) leads to

$$-\sqrt{\mathcal{G}}\frac{\kappa}{8\pi}\frac{1}{2GF}\Big(-2F(\Delta_{h_{ij}}\alpha) + (d-2)G_{,uv}\alpha\Big)\beta = \sqrt{\mathcal{G}}4p^{(v)}F\delta(x)\kappa\beta \ . \tag{1.139}$$

We can arrange it into the following form

$$\Big((\Delta_{h_{ij}}\alpha) - \frac{(d-2)}{2}\frac{G_{,uv}}{F}\alpha - 32\pi GFp^{(v)}\delta(x)\Big)\beta\kappa\sqrt{\mathcal{G}} = 0 \ . \tag{1.140}$$

Hence we recover the shock wave equation of motion

$$\Delta_{h_{ij}}\alpha(x) - \frac{d-2}{2}\frac{G_{,uv}}{F}\alpha(x) = 32\pi p^{(v)}GF\delta(x) \ . \tag{1.141}$$

1.8 Gauss-Bonnet gravity

The microscopic thermodynamic relation can be generalized to arbitrary covariant theory of gravity. The gravitational shock wave equations in the Gauss-Bonnet gravity have been previously studied for a certain black hole solution in [114] while exhaustive studies have been presented in [70, 69] with rather involved calculations. In this section, we present a simpler derivation of the gravitational shock wave equations of motion in the Gauss-Bonnet gravity based on the Noether charge method.

1.8.1 Noether charge in Gauss-Bonnet gravity

The Lagrangian d-form of the Gauss-Bonnet gravity is

$$\mathbf{L} = \boldsymbol{\epsilon}_{a_1\cdots a_d}\Big(\frac{1}{16\pi}R + \frac{\lambda}{16\pi}(R_{abcd}R^{abcd} - 4R_{ab}R^{ab} + R^2)\Big) \ . \tag{1.142}$$

This yields the Noether charge $(d-2)$-form [74]

$$\mathbf{Q}_{\chi a} = -\boldsymbol{\epsilon}_{ade}\Big(\frac{1}{16\pi}\nabla^d\chi^e + \frac{\lambda}{8\pi}(R\nabla^d\chi^e + 4\nabla^{[f}\chi^{d]}R^e{}_f + R^{defh}\nabla_f\chi_h)\Big) \ . \tag{1.143}$$

We shall focus on a family of the metric of Eq. (1.76). The main task is to evaluate the variation $\delta\mathbf{Q}_\chi$ under $(u,v) = (\beta,0) \to (\beta,\alpha)$ at Σ_B. It suffices to compute the contribution to $\mathbf{Q}_{\chi a}$ from $(d,e) = (u,v)$ in Eq. (1.143)[21]:

$$-(\mathbf{d^{d-2}x})_{uv}\sqrt{\mathcal{G}}\frac{1}{8\pi}\left(1+2\lambda H\right)\nabla^{[u}\chi^{v]} \ , \qquad H \equiv R_{\mu\nu}g^{\mu\nu} - 4R_{uv}g^{uv} - 2R^u{}_{uvu}g^{uv} \ , \tag{1.144}$$

and its variation $\delta\mathbf{Q}_{\chi a}$:

$$-(\mathbf{d^{d-2}x})_{uv}(\delta\sqrt{\mathcal{G}})\frac{1}{8\pi}\left(1+2\lambda H\right)\nabla^{[u}\chi^{v]} - (\mathbf{d^{d-2}x})_{uv}\sqrt{\mathcal{G}}\frac{1}{8\pi}2\lambda(\delta H)\nabla^{[u}\chi^{v]} \ . \tag{1.145}$$

To evaluate H, we need the following non-vanishing Ricci tensors:

$$R_{uv} = \left(-\frac{F_{,uv}}{F} - \frac{d-2}{2}\frac{G_{,uv}}{G}\right) \ , \qquad R_{ij} = R^{(d-2)}_{ij} - \left(\frac{G_{,uv}}{F}\right)h_{ij} \ , \tag{1.146}$$

[21] Note there is a factor of 2 which comes from the contribution of $(d,e) = (v,u)$

and the following Riemann tensor:

$$R^u{}_{uvu} = \partial_v \Gamma^u_{uu} - \partial_u \Gamma^u_{vu} + \Gamma^u_{v\alpha}\Gamma^\alpha_{uu} - \Gamma^u_{u\alpha}\Gamma^\alpha_{vu} = \partial_v \frac{F_{,u}}{F} \simeq \frac{F_{,uv}}{F} \ . \tag{1.147}$$

Putting all these together, we obtain the following simple result

$$H = R^{(d-2)}_{ij} g^{ij} = \frac{R^{(d-2)}}{G} \ , \tag{1.148}$$

where $R^{(d-2)}$ is the scalar curvature computed from h_{ij}.

As for δH, the following observation simplifies the calculation. Let us focus on the variation of the first term $R_{\mu\nu}g^{\mu\nu}$ in H:

$$\delta(R_{\mu\nu}g^{\mu\nu}) = \delta(R_{\mu\nu})g^{\mu\nu} + R_{\mu\nu}\delta(g^{\mu\nu}) \ . \tag{1.149}$$

By utilizing the following relation

$$\delta R^\rho{}_{\mu\lambda\nu} = \nabla_\lambda \delta\Gamma^\rho_{\nu\mu} - \nabla_\nu \delta\Gamma^\rho_{\lambda\mu} \ . \tag{1.150}$$

$\delta(R_{\mu\nu})g^{\mu\nu}$ can be expressed as a total derivative. In a similar manner, we can show the following

$$\delta H = R_{\mu\nu}\delta g^{\mu\nu} - 4R_{uv}\delta g^{uv} - 2R^u{}_{uvu}\delta g^{uv} + (\text{total derivative}) \ . \tag{1.151}$$

Note that total derivatives do not contribute to the surface integral of the Noether charge. By dropping these terms, we obtain

$$\delta H = R^{(d-2)}_{ij}\delta\mathcal{G}^{ij} \ . \tag{1.152}$$

Using $\delta\sqrt{\mathcal{G}} = \frac{1}{2}\sqrt{\mathcal{G}}\mathcal{G}^{ij}\delta\mathcal{G}_{ij} = -\frac{1}{2}\sqrt{\mathcal{G}}\mathcal{G}_{ij}\delta\mathcal{G}^{ij}$, the total variation Eq. (1.145) can be expressed as

$$\delta\mathbf{Q}_{\chi,\text{field}} = (\mathbf{d^{d-2}x})_{uv}\frac{\kappa\sqrt{\mathcal{G}}}{8\pi F}\delta\mathcal{G}^{ij}\left(\frac{1}{2}\mathcal{G}_{ij}\left(1 + 2\lambda\frac{R^{(d-2)}}{G}\right) - 2\lambda R^{(d-2)}_{ij}\right) \ . \tag{1.153}$$

Finally, by making use of Eq. (1.134) and integrating by parts, one can obtain the gravitational shock wave equations of motion in the Gauss-Bonnet gravity.

1.8.2 Isotropic space

So far, our discussions are applicable to generic static spacetime in the Gauss-Bonnet gravity. By specializing in isotropic spaces, one obtains further simplified expression of $\delta\mathbf{Q}_\chi$. Let us consider the following metric

$$ds^2 = 2F(u,v)dudv + r_H^2 h_{ij}dx^i dx^j \ , \tag{1.154}$$

where the metric of h_{ij} depends on $k = 1, 0, -1$ and looks like

$$\begin{aligned} k = 1 \quad &\rightarrow \text{ de Sitter (sphere) ;} \\ k = 0 \quad &\rightarrow \text{ flat space ;} \\ k = -1 \quad &\rightarrow \text{ anti-de Sitter (hyperbolic) .} \end{aligned} \tag{1.155}$$

We then have

$$R^{(d-2)}_{ij} = (d-3)kh_{ij} \ , \qquad R^{(d-2)} = k(d-3)(d-2) \ . \tag{1.156}$$

Plugging these into Eq. (1.153), we obtain

$$\delta\mathbf{Q}_{\chi,\text{field}} = -(\mathbf{d^{d-2}x})_{uv}\frac{\kappa\sqrt{\mathcal{G}}}{8\pi}\frac{1}{2GF}\left(1 + 2\lambda\frac{k(d-3)(d-4)}{r_H^2}\right)\Big(- 2F\beta(\Delta_{h_{ij}}\alpha) + (d-2)G_{,uv}\alpha\beta\Big) \ , \tag{1.157}$$

where $r = r_H$ is the horizon radius. By setting $\lambda_{GB} \equiv \lambda\frac{k(d-3)(d-4)}{r_H^2}$, we see that the Noether charge variation $\delta\mathbf{Q}_{\chi,\text{field}}$ is rescaled by a factor of $(1+2\lambda_{GB})$. This recovers the result from [114]

$$(1+2\lambda_{GB})\Big(\Delta_{h_{ij}}\alpha(x) - \frac{d-2}{2}\frac{G_{,uv}}{F}\alpha(x)\Big) = 32\pi p^{(v)}GF\delta(x) \ . \tag{1.158}$$

1.9 Measuring shock waves

In this section, we present a brief discussion on the connection between OTOCs and the thermodynamic relation. Here we shall focus on the cases where the infalling and outgoing matters consist of scalar fields for simplicity of discussion.

So far, we have studied geometries with single gravitational shock waves running in the v direction and derived the thermodynamic relations by using shifted vector fields in

the u direction. A naturally arising question concerns geometries with two intersecting gravitational shock waves running in the u and v directions. Such geometries can be generated by using shifts $\alpha(x)$ and $\beta(x)$ in the v and u coordinates respectively, and have been explicitly considered by Kiem, Verlinde and Verlinde [79]. Readers might wonder if similar analyses based on the covariant phase space formalism would reveal thermodynamic relations for two gravitational shock waves. Unfortunately, the horizon area deformation induced by two gravitational shock waves are the second-order contributions, and as such, naive application of the covariant phase space formalism does not appear to give rise to non-trivial thermodynamic relations [22].

Nevertheless, the microscopic thermodynamic relation can be generalized to geometries with two shock waves in a straightforward manner. This is due to the fact that, in a geometry with two shock waves, each shock wave needs to satisfy the corresponding gravitational equation of motion, which is identical to the one with a single shock wave [79]. Then the effective Lagrangian for two shock waves can be constructed as follows:

$$L_{\text{shock}} = \text{Area}(\alpha(x), \beta(x)) - \int \alpha(x) T^{(u)}(x) - \int \beta(x) T^{(v)}(x) \ , \tag{1.159}$$

up to a multiplicative factor for the area term [23]. Here $T^{(u)}(x)$ and $T^{(v)}(x)$ are energy-momentum tensor profiles running in the u and v directions respectively. By taking the variation of the Lagrangian, we indeed recover equations of motion for $\alpha(x)$ and $\beta(x)$:

$$\frac{\delta L_{\text{shock}}}{\delta \alpha(x)} = E_{\text{shock}}(\beta(x), T^{(u)}(x)) \ , \qquad \frac{\delta L_{\text{shock}}}{\delta \beta(x)} = E_{\text{shock}}(\alpha(x), T^{(v)}(x)) \ . \tag{1.160}$$

Here one may interpret $\alpha(x)$ and $\beta(x)$ as scalar fields which live on the bifurcating sphere, and the $\text{Area}(\alpha(x), \beta(x))$ as interacting massive scalar fields. Then, $T^{(u)}(x)$ and $T^{(v)}(x)$ can be interpreted as source terms coupled with fields $\alpha(x)$ and $\beta(x)$.

Let us now turn our attention to the problem of measuring gravitational shock waves. Recall that the gravitational shock waves can be probed by out-of-time order correlation (OTOC) functions of the form $\langle V(0)W(t)V(0)W(t)\rangle$ within the AdS/CFT correspondence. When the time separation between two perturbations $V(0)$ and $W(t)$ is larger than the scrambling time, two gravitational shock waves will intersect with each other near the black hole horizon. It can then be approximated that two shock waves are running on the black

[22] We expect that a careful analysis, similar to the one given in this chapter for higher-order matter field variations, will enable us to treat higher-order metric variations in a controlled matter.

[23] The effective Lagrangian similar to Eq. (1.159) was derived by 't Hooft [134] for the Schwarzschild black hole.

hole horizon and intersect with each other. Such an approximation would be valid if the values of u, v at the intersecting location are smaller than the effective shifts α, β.

At the heart of the calculation of OTOCs is the gravitational scattering unitary matrix. In his pioneering work, 't Hooft [134] derived the following scattering unitary matrix for the Schwarzschild black hole (often called the horizon S-matrix) [24]:

$$\mathcal{U} = \exp\left[i \iint d\Omega_1 d\Omega_2 P_{\text{out}}(\Omega_1) f(\Omega_1, \Omega_2) P_{\text{in}}(\Omega_2) \right] , \tag{1.161}$$

where $f(\Omega_1, \Omega_2)$ is the green function for the shock wave equation [25]. Here P_{in} and P_{out} account for the incoming and outgoing energy flux. One can cast the above scattering matrix in an illuminating form which makes the relation to the soft thermodynamics more explicit. Here we consider the limit where two shock waves collide near the horizon. Note that one can identify $P_{\text{out}}(\Omega_1)$ and $P_{\text{in}}(\Omega_2)$ as $T^{(u)}(x)$ and $T^{(v)}(x)$ respectively. Then, one obtains

$$\alpha(x) = \int d\Omega_2 f(\Omega_1, \Omega_2) P_{\text{in}}(\Omega_2), \qquad \beta(x) = \int d\Omega_1 f(\Omega_1, \Omega_2) P_{\text{out}}(\Omega_1) \, . \tag{1.162}$$

Recalling that the equations of motion for $\alpha(x)$ and $\beta(x)$ read

$$\delta\text{Area}(\alpha(x), \beta(x)) = \int \alpha(x) T^{(u)}(x) \, , \qquad \delta\text{Area}(\alpha(x), \beta(x)) = \int \beta(x) T^{(v)}(x) \, , \tag{1.163}$$

one can rewrite the above expression as follows:

$$\mathcal{U} = \exp\Big(i(\text{Area}(\alpha(x), \beta(x)) - \text{Area}(0,0)) \Big) \, , \tag{1.164}$$

up to a multiplicative factor for the area. Since $\text{Area}(0,0)$ is a constant, one may express the scattering unitary matrix simply as

$$\mathcal{U} = \exp\Big(i \cdot \text{Area}(\alpha(x), \beta(x)) \Big) \, . \tag{1.165}$$

This expression suggests that, at the leading order, the decorrelation of OTOCs is given by the area deformation [26]

[24] see [112] also.

[25] Quantum mechanical aspects of the scattering matrix have been further studied in recent years, see [11, 45, 46] for instance.

[26] Direct measurement of OTOCs in many-body quantum systems is challenging with inverse time-evolution, but an alternative formulation of OTOCs using entanglement does not require inverse time-evolution. Various experimental protocols for measuring OTOCs in controlled quantum simulators have been proposed [130, 146, 58, 39, 152], with concrete experimental demonstrations [56, 92, 87, 12]. It is still premature to envision measurement of OTOCs in gravitational systems, but at the conceptual level, the smoothness of the black hole horizon for an infalling observer is an indirect evidence of decorrelation of OTOCs, see [148, 110] for details.

It is worth emphasizing that this expression of the scattering unitary matrix can be generalized to any bifurcating horizon in generic static spacetime in arbitrary covariant theory of gravity [27]. The above expression of the scattering matrix can be viewed as a concrete method of measuring the soft charge, which is realized as the gravitational shock wave. We hope to further expand this observation in the future.

Also note that matter contents render rich additional structures in the scattering matrix, see the table in [134] for instance. It will be interesting to extend our analysis to the cases where the infalling and outgoing matters are not scalar fields.

1.10 Discussions

In this chapter, we derived microscopic thermodynamic relations localized at the near-horizon region, providing a concrete and universal characterization of gravitational shock waves applicable to generic static spacetimes, including those beyond the AdS/CFT correspondence. By extending Wald's covariant phase space formalism to arbitrary-order matter variations, we obtained an explicit formula relating the horizon area deformation—interpreted as a soft charge—to the incoming and outgoing energy fluxes. Specifically, we showed that the gravitational scattering matrix, encapsulated by out-of-time-order correlation functions, is directly related to the horizon area variations induced by intersecting shock waves.

Our result provides a clear physical interpretation of soft hair in gravitational contexts, bridging abstract soft charges and measurable horizon deformations. Moreover, our method generalizes beyond general relativity, illustrated through explicit calculations in Gauss-Bonnet gravity, underscoring the robustness and universality of our approach.

We now turn to the black hole information problem, a central challenge in quantum gravity. The Hayden-Preskill thought experiment presents a sharp framework for understanding how information can be retrieved from a black hole after the scrambling time. In the next chapter, we analyze quantum information retrieval in holographic systems by studying gravitational shock wave geometries, which model backreaction from infalling and outgoing perturbations. This framework allows us to explicitly test recoverability criteria and deepen our understanding of holographic information dynamics.

[27] While OTOCs are concepts primarily studied within the framework of the AdS/CFT correspondence, they can be generically defined as correlation functions of matter fields which were initially located away from the horizon. Hence, we speculate that our observation may be generalized to asymptotically flat spaces as well as the de-Sitter horizon.

Chapter 2

Information recovery and shock waves

2.1 Introduction

One of the most famous puzzles in modern theoretical physics, which has been fundamental in shaping our expectations for a complete theory of quantum gravity, is the celebrated *black hole information problem* [61]. This apparent tension between general relativity and quantum mechanics arises at the semiclassical regime: quantum field theory in curved spacetimes seems to suggest, at the same time, that (i) black holes are thermodynamic objects whose entropy is finite and proportional to their horizon area, as given by the Bekenstein-Hawking formula, and (ii) black holes can evaporate and shrink in size by emitting Hawking radiation that is nearly exactly thermal, and whose entropy seemingly grows monotonically even after exceeding the entropy of the black hole itself. These two seem to naively imply that black hole evaporation could violate unitarity, hinting at fundamental information loss[61] and suggesting that, within effective field theory on a curved spacetime alone, information that falls into a black hole will never be revealed in the Hawking radiation.

The first substantial theoretical challenge to the notion of fundamental information loss came from the discovery of the AdS/CFT correspondence [99, 54, 142]. In this duality, quantum gravity in asymptotically anti-de Sitter (AdS) spacetimes is equivalently described by a conformal field theory (CFT) living on the asymptotic boundary, without dynamical gravity. Consequently, information falling into a black hole in AdS must be preserved,

and black hole dynamics must remain intrinsically unitary, as the complete gravitational physics is encoded in a manifestly unitary boundary theory.

It is now widely accepted that AdS/CFT exemplifies a broader principle known as *holography*, which posits that gravitational physics in various spacetimes—not only AdS-like—can be fully captured by non-gravitational theories defined in fewer dimensions. Leveraging holographic insights, Hayden and Preskill (HP) [64] formulated precise criteria for the recoverability of information falling into black holes, assuming unitary dynamics and a finite black hole Hilbert space dimension determined by the exponential of the horizon area. In the Hayden-Preskill thought experiment, if one can manipulate the Hawking radiation with arbitrarily high precision, then after waiting until past the Page time (the moment at which the entropy of the collected Hawking radiation exceeds the Bekenstein-Hawking entropy of the black hole) [108, 107], any additional information thrown into the black hole is returned to the exterior in the form of Hawking radiation as quickly as it physically can. The key to such quick recovery in the Hayden-Preskill thought experiment is the phenomenon of *quantum information scrambling*, a manifestation of quantum chaos describing how information disperses across strongly coupled many-body quantum systems[67, 81, 121, 119]. Even though the original claims by Hayden and Preskill were derived under the assumption of Haar-random dynamics (which is highly unphysical). Remarkably, the conditions for quick information recovery originally derived under idealized Haar-random dynamics have been observed in strongly coupled, large-N conformal field theories—precisely those theories expected to admit semiclassical holographic duals [115].

The Hayden-Preskill recovery criterion establishes a "pure-boundary" argument for the recoverability of information after the Page time, based on the scrambling properties of the boundary theory. On another front, inspired by insights from AdS/CFT, the most significant recent breakthrough in understanding information recovery from a bulk perspective was the discovery of the Quantum Extremal Surface (QES) formula and entanglement wedge reconstruction [43, 141, 116, 71]. When applied to evaporating black holes in AdS with absorbing boundary conditions, the QES prescription directly yields the Page curve for the von Neumann entropy of Hawking radiation, demonstrating the purification of radiation entropy after the Page time [111, 6]. Furthermore, entanglement wedge reconstruction provides a rigorous criterion identifying the bulk subregion encoded in the collected Hawking radiation. Crucially, after the Page time, the entanglement wedge of the radiation includes the black hole interior, confirming that information falling into the black hole approximately a scrambling time earlier can be explicitly reconstructed from radiation degrees of freedom. Surprisingly, these profound results derive largely from Hawking's original quantum field theory calculations in curved spacetime—methods previously believed capable only of suggesting information loss. The essential new ingredient enabling this bulk

calculation to demonstrate information recoverability is entanglement wedge reconstruction, which is inherently holographic and has no purely semiclassical gravity counterpart, yet precisely reproduces Hayden-Preskill's criteria for information recovery.

Nevertheless, the absence of explicit scrambling dynamics in previous QES calculations, despite their agreement with HP criteria, remains puzzling. A plausible explanation is that the QES formula implicitly encodes scrambling effects as a "black box." Supporting this perspective, minimal surface analyses at the boundary have been shown to become valid precisely when associated out-of-time-order correlators (OTOCs)—direct indicators of scrambling—begin to decay [67]. However, it remains unclear how semiclassical methods, traditionally neglecting gravitational backreaction and thus unable to explicitly reproduce scrambling phenomena like OTOC decay, can nevertheless yield correct predictions consistent with HP-type recoverability.

Recent algebraic perspectives have further illuminated aspects of this puzzle. Witten observed that including specific $O(G_N)$ corrections modifies the algebraic structure of quantum field theory from type III_1 to type II_∞, enabling direct computations of entanglement entropy and extending the QES formula's validity into algebraic frameworks within the semiclassical limit [143, 26, 76]. Still, this algebraic approach inherently neglects gravitational backreaction, emphasizing instead an emergent algebra defined within a strictly semiclassical regime.

Our goal in this work is to bridge these perspectives by explicitly incorporating scrambling dynamics through gravitational backreaction, thus directly testing the connection between scrambling and entanglement wedge reconstruction. It is well-established that in holography, boundary perturbations, in certain regimes, correspond to shockwaves propagating in the bulk geometry, which explicitly encode scrambling dynamics [121, 93]. Here, we demonstrate that analyzing a non-evaporating eternal black hole geometry incorporating gravitational backreaction from incoming and outgoing shockwaves precisely reproduces the HP criteria for information recoverability, thus making explicit the role of scrambling in bulk semiclassical calculations. Our findings constitute a significant step toward unifying entanglement wedge reconstruction, scrambling dynamics, and Hayden-Preskill recoverability criteria within a consistent dynamical framework.

2.1.1 Overview of results

The underlying motivation behind this work is to derive the standard quantum mechanical and information theoretic results on the HP thought experiments from the bulk calculations. In doing so, we hope *not* to ignore the gravitational backreaction from the infalling

and outgoing matter fields that will be induced near the black hole horizon. While a complete study would require us to consider the all the possible effects of backreaction from the existing matter fields, we expect that this will lead to tremendously complex analysis. Instead, we will consider a minimal toy model that includes an infalling quantum state A and a particular outgoing Hawking quanta D as *the only matter degrees of freedom* on the bulk that are present in the model. That is, our toy model will have only three matter quanta, the infalling matter A, the outgoing radiation D to be caught, and its entangled partner $\overline{D}$. Our hope is to analyze this toy model carefully and derive conceptual implications in a universal matter that will remain valid in the full model of quantum gravity.

Even with this drastic simplification, however, consideration of gravitational backreaction for this minimal toy model poses a crucial challenge. This essentially results from the treatment of the colliding shockwaves concerning how to incorporate two coordinate shifts in a consistent manner. To see this, we begin by constructing the bulk geometries that corresponds to the infalling matter and the outgoing radiation respectively. By carefully comparing the results with quantum mechanical predictions, we argue that corresponding bulk geometries are those as depicted in Fig. 2.7. Here, the geometry for the infalling matter simply treat A as a spherically symmetric shell of matter which can be approximated as a gravitational shockwave in the proximity of the black hole horizon. On the other hand, the geometry for the outgoing radiation comes with two entangled gravitational shockwaves $\overline{D}$ and D where D is caught when it reaches the boundary. Note that D's backreaction disappears once it is caught, and this is consistent with the quantum mechanical predictions. Indeed, we demonstrate that other approaches to incorporate backreaction from D fail to obtain consistent results.

We then construct a combined shockwave geometry in which A, D, and $\overline{D}$ all propagate as gravitational shockwaves, and carefully examine how the coordinate shifts should be applied across intersecting null shells. We propose that the backreaction from A must be applied first, followed by that of $\overline{D}$. This ordering ensures consistency across all Cauchy slices and correctly reproduces the entanglement structure of the Hayden-Preskill setup.

Evaluating the generalized entropy associated with various candidate surfaces via the maximin prescription, we recover the quantum information-theoretic prediction that the mutual information $I(\overline{A}, \overline{B}D)$ sharply increases when $M_D \gtrsim M_A$, consistent with the recoverability threshold $n_D \gtrsim n_A$ in the boundary model. We show that this is a direct consequence of the entanglement wedge of $\overline{B}D$ encompassing the infalling quantum state A, and that this reconstruction criterion fails precisely when the shocks do not intersect sufficiently close to the horizon—that is, when the separation between A and D is below the scrambling time.

Moreover, the generalized entropy profile exhibits a local maximum in the strongly scrambling regime, forming a "Python's lunch" structure with barrier height $\Delta S \sim 2S_A$. This is in precise agreement with expectations from circuit complexity arguments, where the decoding cost for retrieving A scales as $\mathcal{C} \sim O(2^{S_A})$. In contrast, in the weakly scrambling regime with $M_D \gg M_A$, the entropy barrier disappears and decoding becomes efficient. This geometric transition captures the crossover to traversable wormhole-style protocols, where decoding can be performed qubit by qubit with low overhead.

We also demonstrate that our treatment resolves the apparent paradox that arises when evaluating $S_{\overline{B}D}$ on a Cauchy slice where D has been caught but the infalling backreaction is not yet visible. Without a proper treatment of the shift ordering, one would conclude that A is not recoverable. Our prescription removes this inconsistency and aligns the bulk calculation with the quantum information-theoretic picture, ensuring that recoverability is immediate upon catching D.

Lastly, we examine the implications of this geometry for the Almheiri-Marolf-Polchinski -Sully thought experiment. We show that the partner mode $\widetilde{D}$ of the outgoing Hawking radiation is reconstructed on $\overline{A}C$, not on $\overline{B}$. In particular, the entanglement wedge of $\overline{A}C$ contains $\widetilde{D}$, while that of $\overline{B}$ does not. This indicates that the interior partner is dynamically re-encoded in a way that depends on the infalling observer, providing a geometric realization of observer-dependent interior reconstruction. The same bulk configuration thus resolves both the information recovery and monogamy paradoxes within a unified framework

Together, these results provide a derivation of the Hayden-Preskill protocol from semiclassical gravity, explicitly incorporating backreaction. The agreement with quantum information theory is exact at the level of mutual information, entanglement wedges, and decoding cost, and does not rely on auxiliary systems or replica wormholes. Our results highlight how the causal structure and shockwave dynamics of AdS gravity encode fine-grained quantum information recovery, and suggest a mechanism by which black hole interiors can remain smooth without violating monogamy or unitarity.

2.2 Hayden-Preskill thought experiment

One concrete approach to addressing the information loss problem is to throw a quantum state into a black hole and see if we can later recover it. In this chapter, we focus on a variant of information loss thought experiments due to Hayden and Preskill (HP) [64] as it captures the most essential moment when the information leaks out from the black hole.

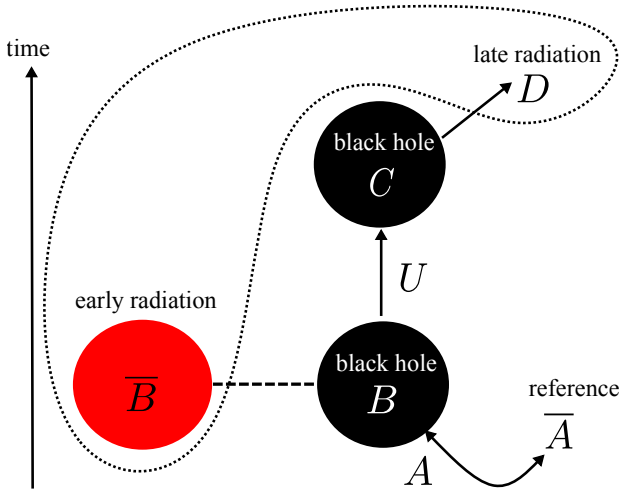


Figure 2.1: Schematics of the HP thought experiment and labelling of Hilbert spaces.

2.2.1 Setup

Consider a black hole B which has already emitted more than a half of its content. This *old black hole* will be maximally entangled with the early Hawking radiation $\overline{B}$. For our purposes, this black hole can be modelled as n_B copies of EPR pairs where n_B corresponds to the coarse-grained entropy of the black hole.

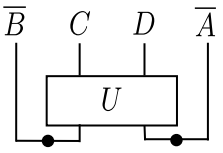
The HP thought experiment proceeds as follows: Alice throws an unknown quantum state into an old black hole, and Bob will reconstruct it by collecting the Hawking radiation which comes out from the black hole. The interesting twist is the assumption that Bob has an access to all the early radiation $\overline{B}$, which has already been emitted. We graphically summarize the setup in Fig. 2.1. Here, A is the Hilbert space for Alice's initial quantum state while B is the black hole and $\overline{B}$ is the early radiation. The black hole time evolves by a unitary operator U, and D is the late Hawking radiation whereas C is the remaining black hole. Bob has an access to both $\overline{B}$ and D (the early and late radiations), aiming to reconstruct the state on A.

Can Bob reconstruct Alice's quantum state? The answer is yes, but surprisingly, Bob requires only a handful of qubits. Namely, if Alice's quantum state is supported on n_A qubit Hilbert space, Bob only needs to collect $n_D = n_A + O(1)$ qubits. Hayden and Preskill demonstrated this result in the case of Haar random U. Subsequent works showed this generically for scrambling dynamics U [151].

For a later purpose, let us recall one useful information theoretic trick. Here, we

introduce the *reference system* $\overline{A}$ which is maximally entangled with the Hilbert space A. Namely, instead of a secret quantum state $|\psi\rangle$ supported on A, we prepare EPR pairs on $A\overline{A}$. The quantum state of our interest is given by

$$|\Psi\rangle_{\overline{B}CD\overline{A}} = U_{BA}|\mathrm{EPR}\rangle_{\overline{B}B}|\mathrm{EPR}\rangle_{A\overline{A}} = \quad , \tag{2.1}$$



where the horizontal line in the diagram represents EPR pairs and black dots represent normalization factors of $\frac{1}{\sqrt{d_B}}$ and $\frac{1}{\sqrt{d_A}}$.

The reference system effectively keeps track of the initial quantum state which was prepared on the Hilbert space A. The crucial point is that, if an unknown quantum state $|\psi\rangle$ is recoverable from the subsystem $\overline{B}D$, then the reference system $\overline{A}$ would be entangled with $\overline{B}D$. Hence, the recoverability of Alice's diary can be quantitatively assessed by computing the mutual information:

$$I(\overline{A}, \overline{B}D) = S_{\overline{A}} + S_{\overline{B}D} - S_{\overline{A}\overline{B}D} \ . \tag{2.2}$$

Namely, $I(\overline{A}, \overline{B}D) \approx 2S_{\overline{A}}$ guarantees the information recoverability, on average over all the input states [67].

2.2.2 Minimal surface calculus

The information recoverability can be demonstrated by explicitly computing (or lower bounding) $I(\overline{A}, \overline{B}D)$ using the Haar integral formula. Here, instead of explicit calculations, we present a heuristic analysis using the *minimal surface* trick for computing entanglement entropies in Haar random states. This essentially says that entanglement entropies in quantum states, built out of Haar random unitary operators, obey the minimal surface rule similar to the Ryu-Takayanagi formula. Namely, letting E be an arbitrary subsystem of $|\Psi\rangle$, we have

$$S_E \approx \min(n_E, 2n - n_E) \ . \tag{2.3}$$

The resemblance to the Ryu-Takayanagi formula in holography hints the parallels between the bulk and boundary calculations, as further elaborated in section 2.5.

Let us compute each term in Eq. (2.2):

$$\begin{aligned} S_{\overline{A}} &= \min\{n_A, n_B + n_C + n_D\} = n_A \ ; \\ S_{\overline{B}D} &\approx \min\{n_B + n_D, n_A + n_C\} \ ; \\ S_{\overline{A}D\overline{B}} &\approx \min\{n_A + n_B + n_D, n_C\} = n_C \ , \end{aligned} \tag{2.4}$$

where we compared the number of cuts in two candidate surfaces as depicted in Fig. 2.2. The second term depends on the sizes of A and D. Using $n = n_A + n_B = n_C + n_D$, we have

$$\min\{n_B + n_D, n_A + n_C\} = 2 \cdot \min\{n_A, n_D\} - n_A + n_C \ . \tag{2.5}$$

Putting these together, we can conclude

$$I(\overline{A}, \overline{B}D) \approx 2 \cdot \min\{n_A, n_D\} \ , \qquad (U\text{: strongly-scrambling}) \ . \tag{2.6}$$

Hence, for $n_D > n_A$, we have $I(\overline{A}, \overline{B}D) \approx 2S_{\overline{A}}$, and the information is recoverable. The Page curve behaviour [108, 107] can be studied by setting $n_A = 0$ in the HP thought experiment. Namely, it simply corresponds to the statement $S_{\overline{B}D} \approx n_C$.

Note that, in the above analysis, the choice of qubits D to be caught is arbitrary as long as $n_D > n_A$. This is an artifact of using Haar random unitary. In the next section, this limitation will be lifted by introducing the notion of scrambling and out-of-time ordered correlation functions.

The argument presented above primarily addresses the recoverability of the initial state, proving the existence of a recovery quantum operation. However, it is important to note that the decoding problem, which pertains to the efficient reconstruction of the initial state on A, is also vital. An algorithm from [151] employs the Grover search algorithm and achieves a decoding complexity that scale proportionally with the input Hilbert space dimension $d_A = 2^{n_A}$. Although this algorithm may not be highly efficient when the input qubit size is polynomially large with respect to $n = S_{\rm BH}$, it is expected that this algorithm performs optimally with respect to $d_A = 2^{n_A}$ when U is strongly scrambling [151].

It will be also useful to study how the information recovery *fails* in non-scrambling cases. While a formal definition of scrambling will be introduced in the next section, one may consider a many-body dynamics U that runs very short period of time as an example. We have $S_{\overline{A}} = n_A$ and $S_{\overline{A}D\overline{B}} = n_C$. As for $S_{\overline{B}D}$, since we are primarily interested in the cases with $n_A, n_D \ll n_B, n_C$, the outgoing radiation D will remain entangled with $\overline{B}$. So, we have

$$S_{\overline{B}D} \approx n_B - n_D \ , \qquad (U\text{: non-scrambling}) \ . \tag{2.7}$$

and $I(\overline{A}, \overline{B}D) \approx 0$. Hence the information is not recoverable.

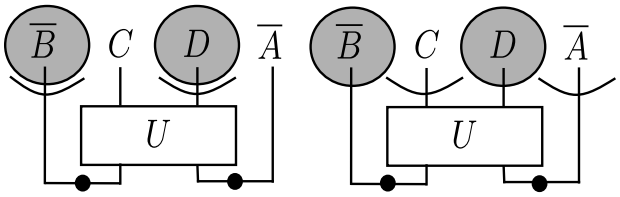
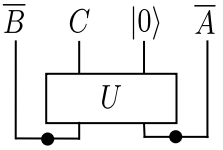


Figure 2.2: Two candidate surfaces, homologous to $\overline{B}D$, with "Areas" being $n_B + n_D$ and $n_C + n_A$ respectively.

2.2.3 Projective measurement

Given recent progress on studies of projective measurements and their consequence on the entanglement structure in many-body quantum systems, it will be also interesting to ask what happens when the outgoing radiation D is projectively measured. Effects of projective measurements on quantum black holes have been studied in [153]. In this case, the quantum state of interest is given by

$$|\Psi^{(\Pi)}\rangle_{\overline{B}C\overline{A}} \propto \Pi_D U_{BA} |\mathrm{EPR}\rangle_{\overline{B}B} |\mathrm{EPR}\rangle_{A\overline{A}} = \quad , \tag{2.8}$$



where $\Pi_D = |0\rangle\langle 0|_D$.

The information recoverability can be addressed by computing $I(\overline{A}, \overline{B})$. Using the minimal surface trick, we have

$$\begin{aligned} S_{\overline{A}} &= \min\{n_A, n_B + n_C\} = n_A \ ; \\ S_{\overline{B}} &\approx \min\{n_B, n_A + n_C\} = \min\{n_D, 2n_A\} - n_A + n_C \ ; \\ S_{\overline{AB}} &= \min\{n_A + n_B, n_C\} = n_C \ , \end{aligned} \tag{2.9}$$

and

$$I(\overline{A}, \overline{B}) \approx \min\{n_D, 2n_A\} \ . \tag{2.10}$$

Hence, if $n_D > 2n_A$, the information is recoverable.

When U is non-scrambling, measuring D will reduce entanglement entropies of $\overline{B}$ as D and $\overline{B}$ would be entangled. Hence we have

$$S_{\overline{B}} \approx n_B - n_D \ , \tag{2.11}$$

and thus $I(\overline{A}, \overline{B}) \approx 0$.

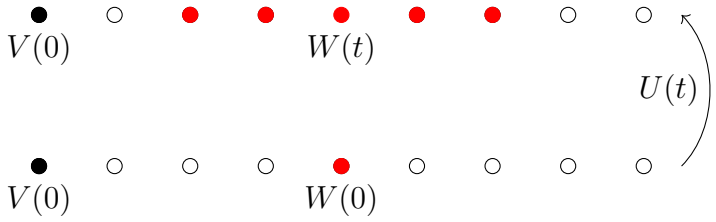


Figure 2.3: Operator growth and scrambling. $V(0)$ and $W(t)$ will eventually overlap, leading to the decay of OTOC $\langle V(0)W(t)V(0)W(t)\rangle$.

2.3 Scrambling and Information Recoverability

We have seen that the information recovery is possible for Haar random U. Subsequent works [67] established that the information recovery remains possible as long as U is *scrambling*.

2.3.1 Out-of-time order correlation

To illustrate the idea, we begin by considering a one-dimensional chain of qubits as in Fig. 2.3. Let W be a local Pauli operator acting on a single qubit. In the Heisenberg picture, the local operator $W(0)$ will eventually grow into some complex operator $W(t) = U^\dagger W U$ that has non-trivial supports on many qubits for large t. In other words, the size of the operator $W(t)$ grows under scrambling dynamics. This generic growth of the operator size is the essence of scrambling, and is indeed what makes the information recovery in the HP thought experiment possible.

The growth of operators can be studied by using out-of-time order correlation (OTOC) functions, which have the following form:

$$\langle V(0)W(t)V(0)W(t)\rangle \equiv \frac{1}{d}\,\mathrm{Tr}\left[V(0)W(t)V(0)W(t)\right]\,, \tag{2.12}$$

where V is another single-qubit Pauli operator which is distant from W. Initially, V and W commute with each other, so OTOC $= 1$. At later times, however, $W(t)$ becomes a big operator, and $V(0)$ and $W(t)$ do not commute with each other anymore. As a result, we tend to have OTOC ≈ 0. Hence, OTOCs measure if $W(t)$ has a non-trivial support on the qubit where $V(0)$ is supported. So, OTOCs can measure the size of the operators.

While the definition of OTOCs involves time-reversal, it can be measured without a time machine. Namely, one can rewrite OTOCs as follows:

$$\text{OTOC} = \langle W(t)|(V^T \otimes V)|W(t)\rangle \ , \tag{2.13}$$

where V^T is a transpose of V. For the EPR state, we have

$$(I \otimes V)|\text{EPR}\rangle = (V^T \otimes I)|\text{EPR}\rangle \ . \tag{2.14}$$

The purturbed state $|W(t)\rangle$ is then defined as

$$|W(t)\rangle \equiv (I \otimes W(t))|\text{EPR}\rangle \ . \tag{2.15}$$

Thus, OTOCs can be represented as two-point correlation functions in $|W(t)\rangle$.

The information recoverability follows directly from scrambling dynamics. Letting A and D be subsystems supporting the infalling quantum state and the outgoing radiation respectively, we assume that U is scrambling with respect to operators on A and D. In particular, we assume

$$\langle V_D(0)W_A(t)V_D(0)W_A(t)\rangle \approx 0 \ , \tag{2.16}$$

when V_D and W_A are non-identity Pauli operators (or any set of complete basis operators) supported on D and A. As shown in [67], values of OTOCs are sufficient to explicitly compute Renyi-2 entropy $S^{(2)}_{\overline{B}D}$. One can then lower bound $I(\overline{A}, \overline{B}D)$ from Renyi-2 quantities, showing $I(\overline{A}, \overline{B}D) \approx 2S_A$ when U is scrambling. Furthermore, one can show that the estimates from the minimal surface treatment remain valid when U is scrambling.

The upshot is that the quantum state on A is recoverable if the outgoing Hawking mode D is scrambling with respect to A in a sense of OTOCs for operators acting on A and D.

Note that previous approaches for deriving the Page curve, coupling the system to an external heat bath, do not offer precise control over which outgoing modes to be collected. Furthermore, as a consequence of introducing a heat bath, further infalling energy flux is created, which makes difficult to ask the recoverability question in a precise form. One of the goals of this chapter is to derive the HP recoverability without introducing a heat bath.

2.3.2 Weakly scrambling regime

We have addressed the information recoverability for two regimes; i) when U is strongly scrambling, and ii) when U is non-scrambling. It turns out that, in the intermediate time

scale where U is not strongly scrambling, the information recovery still remains possible if one takes $n_D \gg n_A$. It is worth analyzing this weakly scrambling regime since the underlying physical mechanism behind the information recoverability and the recoverability is qualitatively distinct from the strongly scrambling case.

A defining feature of strongly scrambling dynamics is the full operator growth where any local operator grows to an $O(n)$-body operator. Instead, we consider a dynamics U that runs shorter than the scrambling time where the operator size becomes large, but is still much smaller than $O(n)$. The well-celebrated Lyapunov growth in a black hole occurs in this time regime.

A weak scrambling dynamics can be modelled by considering an all-to-all coupled system where random two-qubit Haar random gates are applied sequentially on randomly chosen pairs of qubits. Such a random circuit can generate scrambling unitary dynamics after $O(\log n)$ time steps. Here we simply stop the circuit before it gets fully scrambled. In order to further simplify the analysis, we will focus on random Clifford gates and assume that, under the circuit U, any single body Pauli operator will grow to W-body Pauli operator with randomly distributed supports of qubits. Namely, we will work in a regime with $1 \ll W \ll n$ and $1 \ll n_A < n_D \ll n$. Note that the variance of the size can be of order of the size itself since the size growth is exponential, but for our purpose, this approximation suffices.

In order for the information recoverability, a local operator on A needs to grow large enough to overlap with D. This occurs with high probability when

$$W n_D \gtrapprox n \, . \tag{2.17}$$

This criteria can be better understood by considering the inverse time evolution $U^\dagger$. A typical Pauli operator supported on D has size of $O(n_D)$. Its inverse evolution will make it grow to size of $O(W n_D)$. The information recovery becomes possible when a typical Pauli operator on D becomes as large as the whole system size.

In the weakly scrambling regime, there exists an efficient decoding algorithm whose quantum circuit complexity does not increase with d_A. The algorithm is identical to the one presented in [153] for the specific case where $n_A = 1$. This follows from the fact that, in the weakly scrambling regime, local operators acting on each qubit on A grows to $O(W)$-body operators, and these operators do not overlap significantly with each other. Consequently, one can treat the quantum information, prepared on each qubit in A, as more or less independent when recovering it. Note that this algorithm is fundamentally akin to the traversable wormhole, see [47, 98, 103] for details.

2.3.3 Scrambling and shockwaves

In the bulk, scrambling emerges from the fact that finite energy signals at asymptotic infinity will be blueshifted as they approach the black hole horizon. This blueshift will lead to a large backreaction on the geometry and can be described as a gravitational shockwave [38, 37, 68, 121].

Let us illustrate this by focusing on the BTZ black hole in 2+1 bulk dimensions. In the Kruskal coordinates, the unperturbed BTZ metric is given by

$$ds^2 = \frac{1}{(1+uv)^2}\left[-4\ell^2 du dv + R^2(1-uv)^2 d\phi^2\right] \ , \tag{2.18}$$

where $R = \sqrt{8G_N M}\ell$ is the location of the horizon and ℓ is the AdS radius. If we send some matter into the bulk at a boundary time $-t_W$ and boundary energy E, requiring continuous radial evolution (defined by $\sqrt{g_{\phi\phi}}$) imposes

$$R\frac{1-uv}{1+uv} = \widetilde{R}\frac{1-\widetilde{u}\widetilde{v}}{1+\widetilde{u}\widetilde{v}} \ , \tag{2.19}$$

where $\widetilde{u}, \widetilde{v}$ are the coordinates to the future of the shock. Furthermore, $\widetilde{R} = \sqrt{8G_N(M+E)}\ell$ is the radius corresponding to a black hole of mass $M+E$ to account for the infalling matter's energy. For small E/M and large t_w, this leads to a coordinate shift:

$$\widetilde{v} = v + \frac{E}{4M}\exp\left[\frac{2\pi}{\beta}t_w\right]\Theta(u-u_0) \ . \tag{2.20}$$

Figure 2.4 shows this geometry.

To establish a relation to OTOCs, recall that the eternal black hole is dual to the thermofield double state

$$|\text{TFD}\rangle \propto \sum_n e^{-\frac{\beta}{2}E_n}|n\rangle|n\rangle \ , \tag{2.21}$$

which becomes $|\text{EPR}\rangle$ at the infinite temperature limit of $\beta \to 0$. The shockwave can be created by perturbing the boundary with simple operator W, and time-evolving it:

$$|W(t)\rangle = (I \otimes W(t))|\text{TFD}\rangle \ . \tag{2.22}$$

OTOC-like functions can then be probed by two-point correlation functions across the two-sided black hole:

$$\langle V_L \otimes V_R\rangle = \langle W(t)|V_L \otimes V_R|W(t)\rangle \ . \tag{2.23}$$

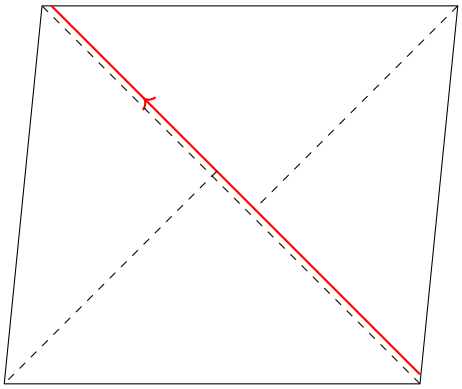


Figure 2.4: The Penrose diagram of the black hole geometry after including the perturbation from asymptotic infinity. The dashed lines are the horizon, and the infalling matter shockwave is in red. The horizon is disjoint because of the coordinate shift.

This correlation function can be estimated to the leading order by recalling

$$\langle \varphi(x)\varphi(y)\rangle \sim e^{-md(x,y)} \,, \tag{2.24}$$

where m is the mass of φ and $d(x,y)$ is the geodesic distance between x, y. With a gravitational shockwave, $d(x,y)$ increases sharply due to the coordinate shift, leading to a sharp decrease in the mutual information between the left and right boundaries [121].

The exponential growth of the shift appears to suggest that, for large t_w, an infalling signal would be sent to the black hole singularity immediately after crossing the gravitational shockwave. It would be, however, more reasonable to think that some quantum gravity effect will suppress such a growth after the scrambling time. Also, whether a sudden coordinate shift is an accurate description of the bulk physics is questionable. It may be more reasonable to think that the shockwave geometry is simply an effective description that matches the phase shift experienced by infalling/outgoing modes asymptotically. Throughout the chapter, we will work on regimes where t_w is at most the scrambling time, so that the coordinate shift is not negligible, but still remains less violent, as in [121].

2.4 Shockwave Geometry for Infalling/Outgoing Matter

In this section, we will derive the shockwave geometries that correspond to throwing a quantum state A and catching a Hawking radiation D. Our strategy is to construct the

candidate geometries, compute entanglement entropies using the HRT formula (namely the Maxmin prescription), and confirm that the results precisely match with quantum mechanical predictions. We will work on the two-sided eternal AdS black hole, but our discussions are applicable to generic maximally entangled black holes.

2.4.1 Quantum extremal surface and maxmin prescription

The Ryu-Takayanagi (RT) formula [116] relates the bulk geometry to the entanglement of the boundary by

$$S(R) = \min_{\chi} \frac{\text{Area}(\chi)}{4G_N} \,, \tag{2.25}$$

where R is a subregion of the boundary at a fixed time and χ is a codimension-2 surface in the bulk that is homologous to R.

The RT formula can be generalized in two ways, and indeed the more general form is key for understanding the HP recovery in the bulk. On the one hand, the area term accounts for the gravitational entropy, but we must also account for the matter entropy in the bulk as well. On the other hand, the RT formula assumes the spacetime in question is static. If the spacetime is dynamical, there may be multiple saddle points to choose between instead of a unique minimum. In that case, the correct χ must be extremal, in the sense that changing $\chi \to \chi + \delta\chi$ leaves the area invariant to first order. If there are multiple extrema, we pick the minimum one. Let E be the bulk region bounded by the RT surface χ, meaning $\partial E = \chi \cup R$. Then, putting the pieces together, we have

$$S(R) = \min \text{ ext}_{\chi} \frac{\text{Area}(\chi)}{4G_N} + S_{bulk}(E) \,. \tag{2.26}$$

Often, this variation is hard to compute in practice. Luckily, there is an easier (but still difficult) variation called the maximin prescription which is equivalent to the Hubeny-Rangamani-Takayanagi (HRT) formula if one only further assumes the quantum focusing conjecture [141]. The maximin prescription is given by

$$S(R) = \max_{\Sigma} \min_{\chi} \frac{\text{Area}(\chi)}{4G_N} + S_{bulk}(E) \,. \tag{2.27}$$

The quantity we are varying is the same, but the variation is different. The procedure is as follows:

(a) Fix a Cauchy slice Σ.

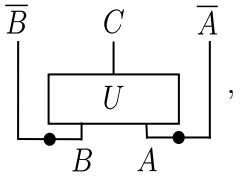
(b) Find the χ_Σ on this Cauchy slice that minimizes $S(R)$.

(c) Among all candidate $\{\chi_\Sigma\}$, the true HRT surface maximizes $S(R)$.

Calculating the generalized entropy of a boundary subregion is not the only purpose that the HRT surface serves. It also demarcates which regions of the bulk can be reconstructed on which boundary subregions. E is often called the *entanglement wedge*, and any operator with support on E can be reconstructed as a (possibly complicated and nonlocal) operator on R.

2.4.2 Infalling quantum state

As a warmup exercise, let us look at the geometry corresponding to an infalling quantum state. Throwing a spherically symmetric quantum state A into a black hole can be modelled as an infalling shockwave which glues two black hole geometries with $M = M_B$ and $M = M_B + M_A$ respectively (Fig. 2.5). This can be verified easily by computing entanglement entropies by the Maxmin prescription. In a quantum circuit language, the state of interest is given by



, (2.28)

where the outgoing radiation D is absent. We assume the scrambling time separation between A and the boundary time of the Cauchy slice. This ensures that the intersection between the Cauchy slice Σ and the shockwave will be close to the horizon.

Let us begin by computing $S_{\overline{B}}$. Since a Cauchy slice Σ goes through two regions, labelled by I and II in Fig. 2.5, we begin by finding candidate generalized entropy for each region.Candidate generalized entropies for the region I and II are given by

$$\begin{aligned} S_{\overline{B}}^{(I)} &= \frac{1}{4G_N}\text{Area}(M_B)\ ; \\ S_{\overline{B}}^{(II)} &= \frac{1}{4G_N}\text{Area}(M_B + M_A) + S_A\ . \end{aligned} \tag{2.29}$$

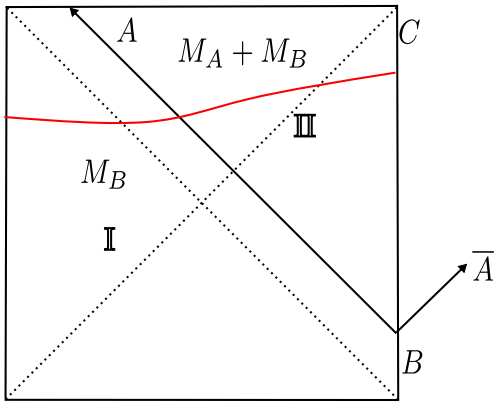


Figure 2.5: Shockwave geometry for an infalling quantum state.

Here, $S_{\overline{B}}^{(II)}$ includes the matter contribution S_A as the candidate surface encompasses A. Also, $S_{\overline{B}}^{(II)}$'s area contribution is approximated by $\frac{1}{4G_N}\text{Area}(M_B + M_A)$ because the shockwave runs close to the horizon. One can verify that there exists a Cauchy slice Σ that contains these two candidate surfaces simultaneously. The remaining task is to find the global maxmin. After looking at possible configurations of Cauchy slices, one can verify that this amounts to considering the minimum of the candidate generalized entropies $S_{\overline{B}}^{(I)}$ and $S_{\overline{B}}^{(II)}$:

$$S_{\overline{B}} = \min\left\{S_{\overline{B}}^{(I)}, S_{\overline{B}}^{(II)}\right\} = \frac{1}{4G_N}\text{Area}(M_B)\ . \tag{2.30}$$

One can repeat a similar analysis to compute S_C. We find the following candidate generalized entropies

$$S_C^{(I)} = \frac{1}{4G_N}\text{Area}(M_B) + S_A\ , \qquad S_C^{(II)} = \frac{1}{4G_N}\text{Area}(M_B + M_A)\ . \tag{2.31}$$

Here we must have $S_C^{(I)} \leq S_C^{(II)}$ in order for the Bekenstein to be obeyed. One may also rely on a standard thermodynamic argument.Hence we have

$$S_C = \min\left\{S_C^{(I)}, S_C^{(II)}\right\} = \frac{1}{4G_N}\text{Area}(M_B) + S_A\ . \tag{2.32}$$

These results are consistent with predictions from Eq. (2.28).

2.4.3 A detour

The horizon radius R and the inverse temperature β of BTZ black holes are given by

$$R^2 = 8G_N M\ell^2, \qquad \beta = \frac{2\pi\ell^2}{R} \tag{2.33}$$

where ℓ is the AdS radius. Suppose that the mass changes to $M + \Delta M$ by an insertion of the shock. For $\Delta M \ll M$, the increase of the horizon radius and the coarse-grained entropy are

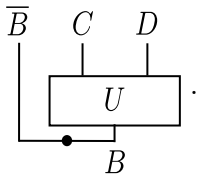
$$\Delta R \approx \frac{\Delta M}{2M} R, \qquad \Delta S_{\rm BH} = \frac{1}{4G_N} 2\pi\Delta R. \tag{2.34}$$

This obeys the first law of the thermodynamics

$$T\Delta S_{\rm BH} = \Delta M. \tag{2.35}$$

2.4.4 Failed approaches for outgoing mode

Next, let us discuss the effect of catching a Hawking radiation D. The state of interest is given by



. (2.36)

In order to illustrate our proposal better, we begin by demonstrating that naive approaches fail to reproduce consistent results.

Catching a Hawking radiation D and removing it from the system effectively reduces the mass of the black hole from $M = M_B = M_C + M_D$ to $M = M_C$. Then one might be tempted to think that this could be modelled as a gravitational shockwave which falls into the black hole and glues black hole geometries with $M = M_C$ and $M = M_C + M_D$ as shown in Fig. 2.6(a). Indeed, this geometry appears to reproduce entanglement entropies consistent with Eq. (2.36). This treatment, however, suffers from a number of inconsistencies. First, this would be a negative energy shockwave. Indeed, a signal from the lower left boundary would reach the right boundary due to the negative coordinate

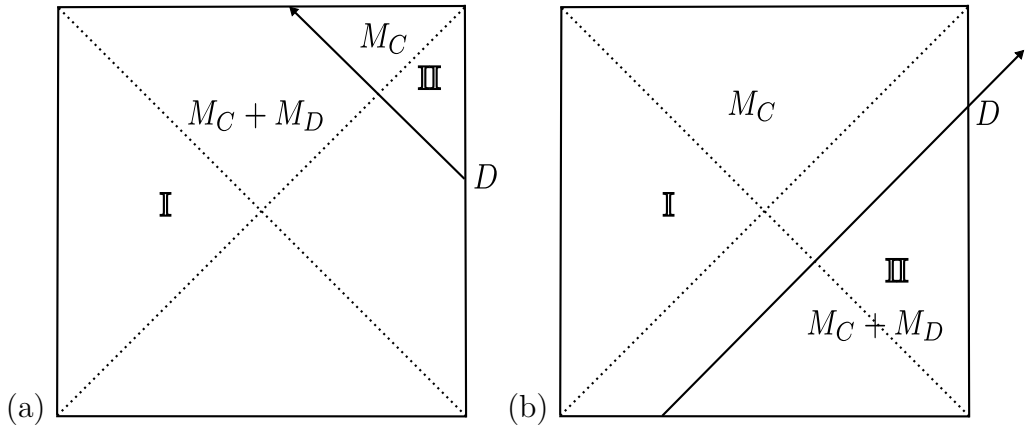


Figure 2.6: Failed approaches for geometries corresponding to catching the Hawking radiation. (a) Sending a negative energy shockwave. (b) Shockwave exiting the asymptotic boundary.

shift, resulting in a traversable wormhole [1]. Second, in the HP thought experiment, the infalling quantum state A must become recoverable as a result of catching the Hawking radiation D. A shockwave as in Fig. 2.6(a), however, will not be able to make any influence on A as shockwaves from A and D will not intersect with each other.

Another approach would be to consider a shockwave that exits the asymptotic boundary at D as shown in Fig. 2.6(b). Note that this geometry can be obtained by turning the geometry of Fig. 2.5 upside down, so it is essentially the time-reversal of throwing a matter into a black hole. This geometry, however, does not reproduce entanglement entropies consistent with quantum mechanical predictions from Eq. (2.36) because from this bulk geometry we would have

$$S_{\overline{B}} = \frac{1}{4G_N}\text{Area}(M_C) \ , \tag{2.37}$$

instead of $S_{\overline{B}} = \frac{1}{4G_N}\text{Area}(M_C + M_D)$.

2.4.5 Outgoing mode

Here, we propose that one needs to include a shockwave propagating on the left hand side which glues black hole geometries with $M = M_C$ and $M = M_C + M_D$. The inclusion of

[1] In this chapter, we do not introduce any coupling between two sides of the black hole, and thus, the black hole must remain non-traversable.

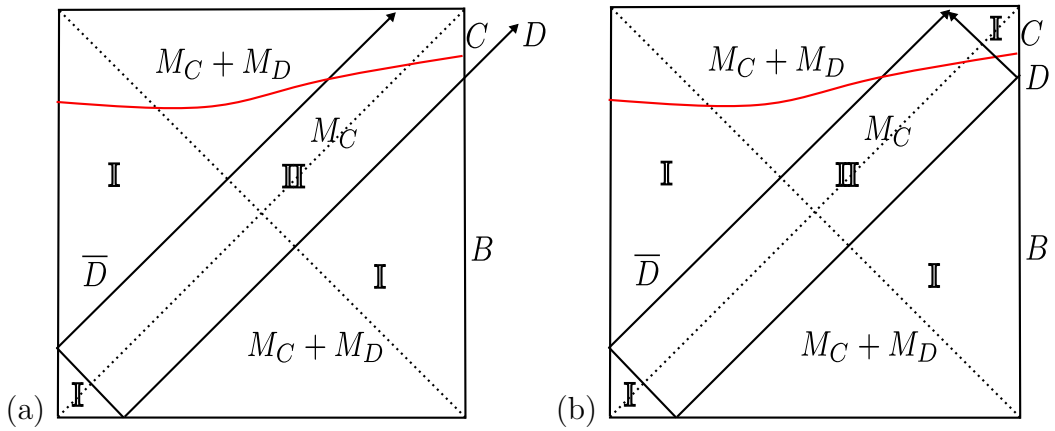


Figure 2.7: (a) Shockwave geometry for catching a Hawking radiation. (b) Shockwave geometry for measuring a Hawking radiation.

such a shockwave can be justified as, from the perspective of the left hand side, the mass of the black hole will remain $M_B = M_C + M_D$. Here, this shockwave on the left hand side corresponds to the entangled partner mode $\overline{D}$ of the Hawking radiation D as depicted in Fig. 2.7(a).

With this geometry with the partner shockwave, let us compute entanglement entropies via the maxmin prescription. The Cauchy slice Σ goes through two regions, labelled by I and II in Fig. 2.7(a). We begin with $S_{\overline{B}}$. We have

$$S_{\overline{B}}^{(I)} = \frac{1}{4G_N}\text{Area}(M_C + M_D) \ , \qquad S_{\overline{B}}^{(II)} = \frac{1}{4G_N}\text{Area}(M_C) + S_D \ . \tag{2.38}$$

Here S_D corresponds to the entanglement entropy of the Hawking radiation, and must satisfy $S_{\overline{B}}^{(I)} = S_{\overline{B}}^{(II)}$. As such, we obtain

$$S_{\overline{B}} = \frac{1}{4G_N}\text{Area}(M_C + M_D) \ . \tag{2.39}$$

Next, let us compute S_C. We have

$$S_C^{(I)} = \frac{1}{4G_N}\text{Area}(M_C + M_D) + S_D \ , \qquad S_C^{(II)} = \frac{1}{4G_N}\text{Area}(M_C) \ , \tag{2.40}$$

and thus,

$$S_C = \frac{1}{4G_N}\text{Area}(M_C) \ . \tag{2.41}$$

These results are consistent with Eq. (2.36).

Our proposal can be also verified by considering the case when the outgoing radiation D is projectively measured. The corresponding geometry is depicted in Fig. 2.7(b). The Cauchy slice Σ goes through three regions: $I \to II \to I$. For $S_{\overline{B}}$, we have

$$S_{\overline{B}}^{(I)} = \frac{1}{4G_N}\text{Area}(M_C + M_D) \ , \qquad S_{\overline{B}}^{(II)} = \frac{1}{4G_N}\text{Area}(M_C) \ . \tag{2.42}$$

Note that $S_{\overline{B}}^{(II)}$ does not contain the matter contribution as the Hawking radiation D, as well as the partner mode $\overline{D}$, have been projected onto a pure state $|0\rangle$. Hence, we have

$$S_{\overline{B}} = \frac{1}{4G_N}\text{Area}(M_C) \ , \tag{2.43}$$

where the entanglement entropy is reduced due to the projective measurement of D by $S_D = \frac{1}{4G_N}\text{Area}(M_D)$. This follows from our assumption that $M_D \ll M_B$, allowing us to use the first-order expansion as discussed in section 2.4.3.

One might worry about that the shockwave geometry in Fig. 2.7 would be running backward in time as a result of catching D, and thus sending a signal to the past. The correct interpretation is to think that D and $\overline{D}$ were present from the beginning, regardless of catching D or not. Namely, these represent gravitational backreaction from the matter fields. In the vacuum solution of quantum field theory on a black hole geometry, one finds that the Hawking pairs are created across the horizon at a constant rate. In order for the full treatment, one needs to consider backreaction from all the outgoing Hawking radiations and their partners.

In this sense, the shockwave geometry in Fig. 2.7 should be viewed as a toy model which assumes that D and $\overline{D}$ are the only matter fields present in this theory. Admittedly, this is a drastic simplification of the full problem. But the fact that this treatment perfectly reproduce predictions from quantum mechanical analysis of the HP thought experiment should be viewed as its justification. We expect that general lessons from this analysis will remain qualitatively valid.

2.5 Information Recovery from Shockwave Geometry

In this section, we discuss the information recoverability via catching the Hawking radiation (*i.e.* the HP thought experiment).

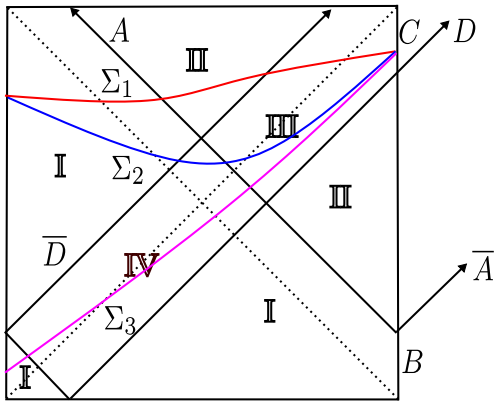


Figure 2.8: Shockwave geometry for the HP recovery via catching the Hawking radiation.

2.5.1 Scrambling case

The corresponding shockwave geometry is depicted in Fig. 2.8 where two kinds of shockwaves, associated with $D, \overline{D}$ and A, coexist. The treatment of two colliding shockwaves needs a careful analysis, and will have an important conceptual implication as we argue in section 2.5.4. For the moment, we ignore subtleties associated with the treatment of two coordinate shifts at the colliding point, but we still assume that each region in the colliding shockwave geometry has a well-defined mass.

We need to evaluate $I(\overline{A}, \overline{B}D)$. Here $S_{\overline{A}}$ and $S_{\overline{A}D\overline{B}} = S_C$ can be easily computed, so we focus on computing $S_{\overline{B}D}$ by looking for the extremal surfaces homologous to the subsystem $\overline{B}D$. There are four kinds of regions of interest, which are labelled by I, II, III and IV. Candidate generalized entropies are

$$\begin{aligned} S_{\overline{B}D}^{(I)} &= \frac{1}{4G_N}\text{Area}(M_B) + S_D \ ; \\ S_{\overline{B}D}^{(II)} &\approx \frac{1}{4G_N}\text{Area}(M_A + M_B) + S_A + S_D \ ; \\ S_{\overline{B}D}^{(III)} &= \frac{1}{4G_N}\text{Area}(M_A + M_B - M_D) + S_A \ ; \\ S_{\overline{B}D}^{(IV)} &= \frac{1}{4G_N}\text{Area}(M_B - M_D) \ . \end{aligned} \tag{2.44}$$

Here, the area contribution in $S_{\overline{B}D}^{(II)}$ is approximated by the area of the black hole horizon for region II due to the scrambling time separation.

We begin with two types of Cauchy slices (Σ_1, Σ_2) where boundary times are located in the future of D and $\overline{D}$ (Fig. 2.8). We will discuss another Cauchy slice, Σ_3, in the later subsection. Consider first Σ_1 which goes through the following regions:

$$\Sigma_1 : I \to II \to III \ . \tag{2.45}$$

Since $S^{(II)}_{\overline{BD}} \geq S^{(I)}_{\overline{BD}}, S^{(III)}_{\overline{BD}}$, we have $S^{(\Sigma_1)}_{\overline{BD}} = \min\left\{S^{(I)}_{\overline{BD}}, S^{(III)}_{\overline{BD}}\right\}$. Next, consider Σ_2 which goes through the following regions:

$$\Sigma_2 : I \to IV \to III \to II \ . \tag{2.46}$$

Here, $S^{(IV)}_{\overline{BD}}$ is the smallest, so we have $S^{(\Sigma_2)}_{\overline{BD}} = S^{(IV)}_{\overline{BD}}$. Taking the maximum of $S^{(\Sigma_1)}_{\overline{BD}}$ and $S^{(\Sigma_2)}_{\overline{BD}}$, we obtain $S_{\overline{BD}} = S^{(\Sigma_1)}_{\overline{BD}}$. Finally, assuming that S_A is nearly maximal for a given mass M_A, we have [2]

$$\begin{aligned} S_{\overline{BD}} &= S^{(I)}_{\overline{BD}} \qquad && M_D \lessapprox M_A \ ; \\ &= S^{(III)}_{\overline{DB}} \qquad && M_D \gtrapprox M_A \ . \end{aligned} \tag{2.47}$$

This allows us to compute $I(\overline{A}, \overline{B}D)$,

$$\begin{aligned} I(\overline{A}, \overline{B}D) = S_{\overline{A}} + S_{\overline{B}D} - S_{\overline{A}D\overline{B}} &\approx 0 \qquad && M_D \lessapprox M_A \ ; \\ &\approx 2S_A \qquad && M_D \gtrapprox M_A \ . \end{aligned} \tag{2.48}$$

Therefore we conclude that the information recovery is possible when $M_D \gtrapprox M_A$. The information recoverability follows also from a stronger hypothesis of entanglement wedge reconstruction, as the HRT surface for $\overline{B}D$ lies in region III, and the infalling mode A is contained in the entanglement wedge. We have thus obtained a result which exactly parallels with the one predicted from scrambling dynamics in the HP thought experiment.

One potential issue would be that, while $S^{(IV)}_{\overline{B}}$ is given by an extremal surface, $S^{(I)}_{\overline{B}}, S^{(II)}_{\overline{B}}, S^{(III)}_{\overline{B}}$ are not. This subtlety results from using shockwaves with delta-function profiles that introduce sudden coordinate shifts. For a more realistic treatment, one can introduce some broadening in shockwaves to avoid sudden increase of generalized entropies when crossing shockwaves as in Fig. 2.9(a). Noting that area terms in $S^{(I)}_{\overline{B}}, S^{(II)}_{\overline{B}}, S^{(III)}_{\overline{B}}$ are already nearly extremal due to the proximity to the bifurcation surface, we expect that compensations from gradual increase/decrease of matter contributions will make these candidate generalized entropies extremal [3].

[2] If we use non-maximal S_A, this reduces to the situation considered in [66] where the input information of entropy S_A is encoded into an object with a coarse-grained entropy $S^{(\text{coarse})}_A > S_A$.

[3] For instance, when moving from region I to region II, the area decreases slightly while we also pick up gradual increase of S_A.

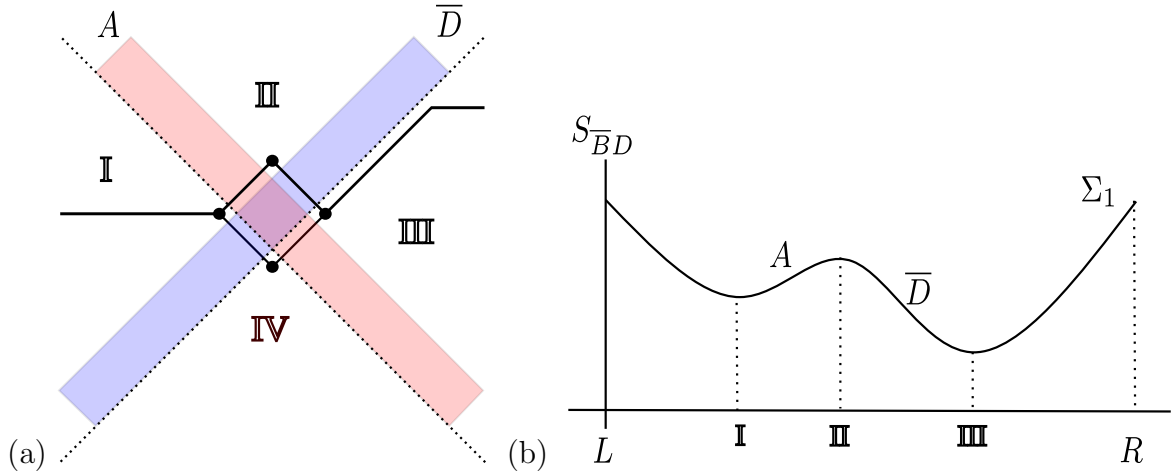


Figure 2.9: (a) Approximate locations of candidate surfaces and two kinds of Cauchy slices. (b) Generalized entropies associated with the Cauchy slice Σ_1. Here, $M_D \geq M_A$ guarantees $S^{(III)}_{\overline{B}D} \leq S^{(I)}_{\overline{B}D}$. The entropy barrier is $S^{(II)}_{\overline{B}D} - S^{(I)}_{\overline{B}D} \approx 2S_A$.

Let us draw a schematic picture of the generalized entropies associated with the Cauchy slice Σ_1 as in Fig. 2.9(b). We find that the true minima (region III) is separated from the left boundary by a local maxima (region II). Such a geometry is often called the Python's lunch, and is associated with high decoding complexity [16]. Namely, it has been conjectured that the decoding complexity is proportional to the exponential of the barrier of generalized entropies. In our setup, we find $S^{(II)}_{\overline{B}D} - S^{(I)}_{\overline{B}D} \approx 2S_A$, suggesting the decoding complexity of $\mathcal{C} \approx 2^{O(S_A)}$. Indeed, the recovery algorithm from [151], which works deterministically in the regime where the time separation of the boundary time of A and D is longer than the scrambling time, has a run time proportional to the Hilbert space size of A, in other words, $\mathcal{C} \approx O(2^{S_A})$.

2.5.2 Non-scrambling case

Now we discuss what would go wrong in the above calculation when A and D are not separated by the scrambling time. A crucial difference is that shockwaves run far from the horizons. Let us revisit candidate generalized entropies associated with the Cauchy slice Σ_1. When moving from the region I to II, the area term decreases further from $\frac{1}{4G_N}\text{Area}(M_B)$ after crossing region I's horizon. As a result, $S^{(I)}_{\overline{B}D}$ will have a much smaller value. Similar observation holds for $S^{(II)}_{\overline{B}D}$ and $S^{(III)}_{\overline{B}D}$. On the other hand, candidate generalized entropies

for the Cauchy slice Σ_2 remain unchanged. As a result, the Cauchy slice Σ_2 will be preferred, and we have

$$S_{\overline{B}D} = S^{(IV)}_{\overline{B}D} = \frac{1}{4G_N}\text{Area}(M_B - M_D)\ . \tag{2.49}$$

This corresponds to the non-scrambling regime, as in Eq. (2.11), where $\overline{B}$ and D remain entangled.

2.5.3 Weakly scrambling regime

Beyond the fully scrambling and non-scrambling limits discussed above, it is also instructive to consider the weakly scrambling regime as discussed in section 2.3.2. Notably, the value of $S_{\overline{B}D}$ depends on how small $S^{(II)}_{\overline{B}D}$ is. Here, we focus on the cases with $M_D \gtrapprox M_A$ and thus $S^{(III)}_{\overline{B}D} \leq S^{(I)}_{\overline{B}D}$.

When $S^{(III)}_{\overline{B}D} \leq S^{(II)}_{\overline{B}D} \leq S^{(I)}_{\overline{B}D}$, we have $S^{(\Sigma_1)}_{\overline{B}D} = S^{(III)}_{\overline{B}D}$, and thus, the information recovery is still possible. The corresponding entropy landscale is depicted in Fig. 2.10. A notable feature of this geometry is that the local maxima in region II disappears since $S^{(II)}_{\overline{B}D}$ becomes smaller due to the decrease of Δt. According to the hypothesis of Python's lunch, this suggests the possibility of an efficient decoding. This is consistent with the decoding complexity of the traversable wormhole algorithms. The algorithms is known to work efficiently in a regime of $n_D \gg n_A$ with intermediate time scale before the scrambling time. Here, we can observe that taking $M_D \gg M_A$ makes the generalized entropy difference $S^{(I)}_{\overline{B}D} - S^{(III)}_{\overline{B}D}$ large. This will make it easier to tune the time separation Δt so that $S^{(III)}_{\overline{B}D} \leq S^{(II)}_{\overline{B}D} \leq S^{(I)}_{\overline{B}D}$. A similar observation has been made in the traversable wormhole algorithms.

When $S^{(II)}_{\overline{B}D} \leq S^{(III)}_{\overline{B}D}$, we have $S^{(\Sigma_1)}_{\overline{B}D} = S^{(II)}_{\overline{B}D}$, and thus, the information recovery is not fully possible. In this case, we have

$$S_{\overline{B}D} = \max\left\{S^{(II)}_{\overline{B}D}, S^{(IV)}_{\overline{B}D}\right\}. \tag{2.50}$$

The case with $S^{(II)}_{\overline{B}D} \leq S^{(IV)}_{\overline{B}D}$ corresponds to the non-scrambling regime where $\overline{B}$ and D remain entangled.

2.5.4 On Σ_3

So far, we have computed $S_{\overline{B}D}$ for a family of Cauchy slices Σ_1, Σ_2. Since unitary time-evolution does not change the entanglement entropy, one naturally expect that the same

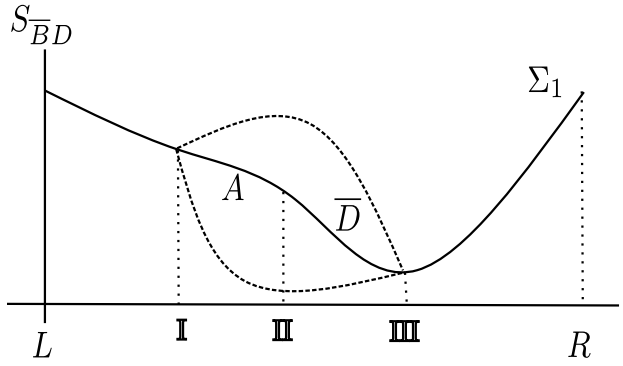


Figure 2.10: Weakly scrambling regime where the traversable wormhole algorithms work efficiently. The Python's lunch disappears by tuning the time separation Δt so that $S^{(III)}_{\overline{B}D} \leq S^{(II)}_{\overline{B}D} \leq S^{(I)}_{\overline{B}D}$. Dotted paths correspond to $\Delta t \gtrapprox t_{\rm scr}$ and $\Delta t \ll t_{\rm scr}$.

result would be obtained when we change the boundary time on the left side by considering another Cauchy slice Σ_3 as in Fig. 2.8.

It turns out, however, that this is not the case. Or, at least, a naive approach fails. Here the Cauchy slice Σ_3 goes through the following regions:

$$\Sigma_3 : I \to IV \to III \ . \tag{2.51}$$

Since $S^{(IV)}_{\overline{B}D}$ is the smaller, one would have

$$S_{\overline{B}D} = S^{(IV)}_{\overline{B}D} \ , \tag{2.52}$$

which implies the failure of information recovery. Hence, we have obtained an inconsistent result by considering the Cauchy slice Σ_3.

The key to resolve this puzzle is to consider the effect of coordinate shifts from an infalling state A carefully. Recall that we have ignored the subtlety concerning the two shifts at the colliding points of two shockwaves. Here we make the following crucial proposal:

- The gravitational backreaction from an infalling quantum state A needs to be considered before considering the one from an outgoing Hawking radiation D.

From the causal perspective, our proposal of considering the shift from A first is natural as throwing A into a black hole occurs earlier than catching the Hawking radiation D. With this proposal, one can resolve the aforementioned puzzle concerning the calculation on Σ_3.

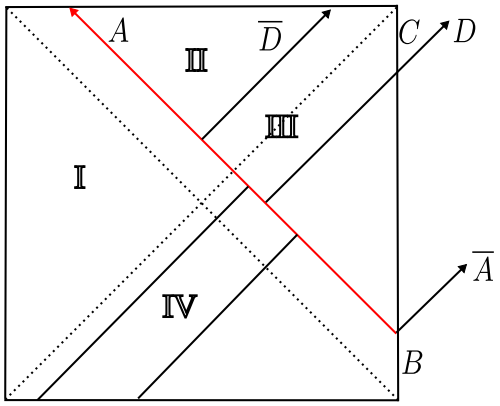


Figure 2.11: The effect of coordinate shift from A's shockwave. The Cauchy slice Σ_3 is not well defined.

Namely, we can observe that the shift along the infalling state A shifts the trajectories of D and $\overline{D}$ as in Fig. 2.11. Then, under the scrambling dynamics, the Cauchy slice Σ_3 simply becomes unphysical as one cannot pick a boundary time before $\overline{D}$ on the left side.

Another independent check of this proposal is to consider the non-scrambling cases. Recall that, with $\Delta t \ll t_{\rm scr}$, the coordinate shift is negligible. Ignoring the gravitational backreaction from A, we obtain

$$S_{\overline{B}D} = S^{(IV)}_{\overline{B}D} \,, \tag{2.53}$$

for Σ_3. This indeed coincides with the result for non-scrambling cases where Σ_3 remains physically valid.

It is worth pointing out that this proposed resolution is consistent with the fact that the expression of the interior partner mode $\overline{D}$ changes due to the backreaction from the infalling mode A as pointed out in [150]. Furthermore, based on these observations, we propose that the shift from A should be introduced first, and then the shift from $\overline{D}$ should be introduced after, concerning the potential subtleties at the colliding point of two shockwaves. We will further elaborate on this in the next section.

2.5.5 Measuring a Hawking radiation

Next, let us discuss the information recovery via projective measurement. There are four kinds of regions of interest, which are labelled by I, II, III and IV, in Fig. 2.12. Applying

the maxmin prescription for each region, candidate generalized entropies are

$$\begin{aligned} S_{\overline{B}}^{(I)} &= \frac{1}{4G_N}\text{Area}(M_B)\ ; \\ S_{\overline{B}}^{(II)} &\approx \frac{1}{4G_N}\text{Area}(M_A + M_B) + S_A\ ; \\ S_{\overline{B}}^{(III)} &= \frac{1}{4G_N}\text{Area}(M_A + M_B - M_D) + S_A\ ; \\ S_{\overline{B}}^{(IV)} &= \frac{1}{4G_N}\text{Area}(M_B - M_D)\ . \end{aligned} \tag{2.54}$$

Consider first Σ_1 which goes through the following regions:

$$\Sigma_1 : I \to II \to III \to II\ . \tag{2.55}$$

Since $S_{\overline{B}}^{(II)} \geq S_{\overline{B}}^{(I)}, S_{\overline{B}}^{(III)}$, we have $S_{\overline{B}}^{(\Sigma_1)} = \min\left\{S_{\overline{B}}^{(I)}, S_{\overline{B}}^{(III)}\right\}$. Next, consider Σ_2 which goes through the following regions:

$$\Sigma_2 : I \to IV \to III \to II\ . \tag{2.56}$$

Here, $S_{\overline{B}}^{(IV)}$ is the smallest, so we have $S_{\overline{B}}^{(\Sigma_2)} = S_{\overline{B}}^{(IV)}$. Taking the maximum of $S_{\overline{B}}^{(\Sigma_1)}$ and $S_{\overline{B}}^{(\Sigma_2)}$, we obtain that $S_{\overline{B}} = S_{\overline{B}}^{(\Sigma_1)}$. Assuming that S_A is nearly maximal, we have

$$\begin{aligned} S_{\overline{B}} &= S_{\overline{B}}^{(I)} \qquad M_D \lessapprox 2M_A \\ &= S_{\overline{B}}^{(III)} \qquad M_D \gtrapprox 2M_A\ , \end{aligned} \tag{2.57}$$

(SL) *check this* recovering the results from section 2.2. When the time separation is much less than the scrambling time, we will have $S_{\overline{B}} = S_{\overline{B}}^{(IV)}$, which is consistent with the non-scrambling case.

To conclude, we discuss what would go wrong if we did not consider the effect of backreaction from infalling and outgoing modes. For spherically symmetric matter, the coordinate shifts simply amounts to the change of the area terms in generalized entropies. Then, ignoring backreaction is essentially the same as ignoring the change of the mass. We

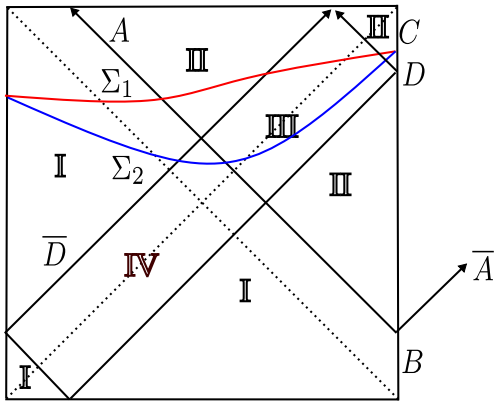


Figure 2.12: Shockwave geometry for the HP recovery via measuring the Hawking radiation.

then would have

$$\begin{aligned} S_{\overline{B}D}^{(I)} &\approx \frac{1}{4G_N}\text{Area}(M)\ ; \\ S_{\overline{B}D}^{(II)} &\approx \frac{1}{4G_N}\text{Area}(M) + S_A\ ; \\ S_{\overline{B}D}^{(III)} &\approx \frac{1}{4G_N}\text{Area}(M) + S_A\ ; \\ S_{\overline{B}D}^{(IV)} &\approx \frac{1}{4G_N}\text{Area}(M)\ , \end{aligned} \tag{2.58}$$

which leads to an incorrect surface, namely $S_{\overline{B}D}^{(I)}$ or $S_{\overline{B}D}^{(IV)}$.

2.6 Interior Partner from Shockwave Geometry

The same setup, used for the HP thought experiment, can be used to study the problem concerning the Hawking interior partner mode, as conceived in the Almheiri-Marolf-Polchinski -Sully (AMPS) thought experiment [7]. In this section, we briefly discuss the implication to the firewall puzzle.

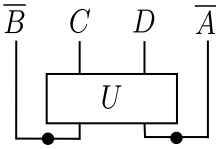
2.6.1 Setup

Again, consider an old black hole B which is maximally entangled with the early Hawking radiation $\overline{B}$. According to the effective field theory calculation, pairs of entangled particles will be created across the horizon. Here we will focus on a particular pair of the outgoing mode D and the interior partner mode $\widetilde{D}$. Since the old black hole is maximally entangled, the outgoing mode D must be entangled with some degrees of freedom in the early radiation $\overline{B}$. Let us denote such partner degrees of freedom by $\overline{D}$. So far, we have found that the outgoing mode D is entangled with $\overline{D}$ which is in the early radiation $\overline{B}$, and also entangled with $\widetilde{D}$ which is inside the black hole. However, this contradicts with the monogamy of entanglement, namely, D cannot be entangled with two things simultaneously.

The AMPS thought experiment is designed to probe a potential violation of the monogamy of entanglement by an infalling observer A. Imagine that Bob carefully distill the mode $\overline{D}$, which will be entangled with D, into multiple qubits, by applying some appropriate quantum operation on the early radiation $\overline{B}$. Alice then jumps into the black hole while carrying these qubits $\overline{D}$ in her pocket. Her goal is to verify that the outgoing mode D is entangled with $\overline{D}$ in her pocket, as well as the interior mode $\widetilde{D}$. Successful crossing of the horizon would mean that Alice could verify the violation of the monogamy of entanglement. Hence, the AMPS thought experiment appears to suggest that the outgoing mode D and the interior mode $\widetilde{D}$ may not be entangled with each other across the horizon, implying the presence of the firewall across the horizon.

One possible resolution of the firewall puzzle can be obtained by considering the effect of backreaction from the infalling observer A. Suppose that an infalling observer A carries finite energy and entropy and jumps into a black hole. Depending on the initial state of the infalling observer, the black hole will respond in a different manner. Again, it is convenient to append the reference system $\overline{A}$ which is entangled with A. Then, the quantum state of our interest is given by

$$|\Psi\rangle_{\overline{B}CD\overline{A}} = U_{BA}|\mathrm{EPR}\rangle_{\overline{B}B}|\mathrm{EPR}\rangle_{A\overline{A}} = \quad , \tag{2.59}$$



which is exactly the same as Eq. (2.1).

In the absence of the infalling observer A, the interior partner mode can be found on the early radiation $\overline{B}$. In the presence of the infalling observer A, the situation changes drastically. The problem of finding the interior partner mode $\widetilde{D}$ can be addressed in an

unambiguous manner by finding degrees of freedom with which the outgoing mode D is entangled. Below we identify the interior partner mode from both boundary and bulk perspectives.

2.6.2 Boundary picture

Here, we demonstrate that the partner mode $\widetilde{D}$ can be found on $\overline{A}C$ by computing $I(D, \overline{A}C)$. We have

$$\begin{aligned} S_D &= n_D \ ; \\ S_{\overline{A}C} &\approx \min\{n_D + n_B, n_A + n_C\} \ ; \\ S_{D\overline{A}C} &= n_B \ , \end{aligned} \tag{2.60}$$

and thus

$$I(D, C\overline{A}) \approx 2 \cdot \min\{n_A, n_D\} \ . \tag{2.61}$$

Hence, for $n_A \geq n_D$, we have $I(D, C\overline{A}) \approx 2S_D$, and D will be entangled with $\overline{A}C$. Equivalently, we have $I(D, \overline{B}) \approx 0$, suggesting that D is decoupled from the early radiation $\overline{B}$. The assumption of $n_A \geq n_D$ can be justified by recalling that A needs to carry $\overline{D}$ in order to perform the AMPS thought experiment.

This analysis suggests that, due to the backreaction from an infalling observer A, the qubits $\overline{D}$ inside the early radiation $\overline{B}$, are no longer entangled with the outgoing mode D, avoiding the violation of the monogamy of entanglement. One crucial implication of the above analysis is that the construction of interior operator is state-independent in a sense that it does not depend on how the black hole was initially entangled with the early radiation $\overline{B}$. In fact, one can easily see that the same construction of the interior partner mode $\widetilde{D}$ works for a young pure state black hole as well, resolving the Marolf-Polchinski version of the firewall puzzle [101].

However, it is still important to keep in mind that the construction of interior operators depends on the initial state of the infalling observer. This suggests that the black hole interior is observer-dependent object. Depending on how you enter the black hole, the entanglement structure is reorganized and observer-dependent interior mode will be dynamically created.

2.6.3 Bulk picture

One can reproduce these results from the bulk perspective by following the argument for the HP thought experiment. The only difference is that one must compute $S_{\overline{A}C}$ for the interior reconstruction problem while one computes $S_{\overline{B}D}$ for the HP recovery. These calculations of course match, but the entanglement wedge reconstruction occurs for modes encompassed by the HRT surface and $\overline{A}C$, instead of $S_{\overline{B}D}$. For $M_A \geq M_D$, the entanglement entropy is given by

$$S_{\overline{A}C} = S^{(I)}_{\overline{A}C} \,, \tag{2.62}$$

and the partner mode $\widetilde{D}$ is inside the entanglement wedge of $\overline{A}C$. Hence, $\widetilde{D}$ can be reconstructed from $\overline{A}C$.

A crucial benefit of this analysis is that the firewall puzzle and the information loss problem can be resolved on an equal footing, at least when $M_A \approx M_D$. Namely, we observe that, under the scrambling dynamics, the Cauchy slice Σ_1 is preferred (regardless of $M_A > M_D$ or $M_A < M_D$), supporting the proposal that the black hole interior mode $\widetilde{D}$ changes dynamically as a result of backreaction from A and is no longer the same as $\overline{D}$. The Cauchy slice Σ_2 would contain the original partner mode $\overline{D}$, but those modes relevant to the firewall puzzle are the ones close to the horizon that requires the scrambling time separation. This observation provides a further counterargument, from the bulk perspective, against previous proposals for resolving the firewall puzzle, which are often bundled under the slogan of ER = EPR. For counterarguments from the boundary perspective, see [148].

2.7 Outlook

Throughout this chapter, we have focused on the cases where infalling matter A and outgoing mode D are spherically symmetric. It will be interesting to consider the cases where A and D are spatially localized modes. For a $1+1$-dimensional BTZ black hole, the locations of A and D may be specified by an angular variable θ. When A and D are spatially separated, say at $\theta = 0$ and $\theta = \pi$ respectively, the scrambling time separation is not sufficient to make $\langle O_A(0)O_D(t)O_A(0)O_D(t)\rangle$ decay [114]. In such a case, the recovery will not be possible even when shockwaves are running close to the horizon.

One useful observation is that the partner mode $\overline{D}$ will not experience the shift from the infalling mode A when $\langle O_A(0)O_D(t)O_A(0)O_D(t)\rangle$ did not decay. Similarly, the null trajectory from A will not experience the shift from $\overline{D}$. This suggests that, when moving

the surface through region I, region II and region III on a Cauchy slice, one can deform the surface without crossing the shockwaves at all. This keeps the area terms almost constant and will prevent us from finding the HRT surfaces in region I and region III.

With this observation in mind, it will be interesting to revisit a previous work [93] by two of the authors. Inspired by a pioneering work by 't Hooft, it has been advocated that the area deformation under two null shifts $(\delta u, \delta v) = (\alpha(x), \beta(x))$ is a quantitative measure of scrambling. Appealing to the above observation, the absence of the area deformation seems to suggest the failure of the information recovery as expected from non-decaying OTOCs. We hope to elaborate on this in the future.

Chapter 3

Noise-strength-adapted approximate quantum codes inspired by machine learning

3.1 Introduction

Having explored the fundamental challenges of information recovery in black hole physics and the role of quantum information scrambling in the previous chapter, we now turn to the practical frontiers of quantum information science. In this chapter, we shift our focus from the theoretical landscape of quantum gravity to the design and discovery of quantum error-correcting codes, leveraging modern machine learning techniques. This transition highlights the versatility of quantum information concepts, bridging foundational questions in physics with emerging applications in quantum technologies.

Large-scale fault-tolerant quantum computers are essential for achieving the theoretically promised advantages over classical machines in solving certain important problems, and quantum error correction (QEC) plays an essential role for enabling that. While much research has focused on QEC codes that enable exact recovery of logical information, namely perfect error correction [123, 51, 105], approximate quantum error-correcting (AQEC) codes, which allow for a certain degree of recovery inaccuracy, can provide more practical and sometimes necessary alternatives with important advantages. For example, AQEC can outperform exact QEC in terms of code rates [91, 30] and facilitate continuous transversal logical gates in covariant codes [65, 44, 144, 86, 145, 156, 96, 95, 85] which would otherwise be prohibited [42].

Despite the importance of AQEC codes, our understanding of them remains limited. In particular, there lacks a systematic framework for designing useful AQEC codes, and most known codes depend on pre-assumed algebraic structures, such as stabilizer formalism [51] or self-complementary quantum codes [125, 88]. In this work, we propose to employ machine learning (ML) to explore AQEC codes beyond conventional algebraic constraints, thereby expanding the range of possibilities. Specifically, we demonstrate how a hybrid quantum-classical learning algorithm leads to the discovery of novel AQEC codes with unique structures and improved performance. This machine learning approach offers several advantages. Its flexible architecture provides a systematic way to search for AQEC codes suitable for any noise model, while also enables the exploration of non-stabilizer codes. Although there have been previous works [20, 77] using variational ML to find QEC codes, little efforts have been devoted to exploring AQEC.

This work makes the following key contributions. We introduce *noise-strength-adapted (NSA)* quantum codes – new families of approximate quantum error-correcting codes whose codewords vary continuously with noise strength, enabling substantial improvements in code performance. These codes are discovered using a hybrid quantum-classical machine learning protocol and then analytically generalized to arbitrary system sizes and bosonic systems. Specifically, we focus on AQEC codes for amplitude damping (AD) noise—a common noise that models energy losses in quantum systems, such as photon losses and spin relaxations in photonic and superconducting systems respectively [25, 28]. Through our learning protocol, we first find two $((4, 1))$ codes capable of approximately correcting a single AD error, with the innovative NSA feature that the code is adaptive to arbitrary AD noise strength (note that Ref. [100] also identifies a different 4-qubit AD code with this feature). This feature allows the possibility of tailoring the codes towards different quantum device architectures to improve performance. It should be noted that no additional information is required in designing these adaptive codes, since the noise strength value is already necessary for decoding AD codes, as demonstrated in [91]. Leveraging the NSA property, our codes outperform the Leung–Nielsen–Chuang–Yamamoto (LNCY) AD code [91] by exhibiting lower Knill–Laflamme condition violation and higher worst-case fidelity. The two new codes discovered from ML are named the self-complementary (SC) code and the pair-complementary (PC) code. The former is an NSA version of the LNCY code, while the latter is an entirely new code. Moreover, the PC code achieves even better performance than SC code, with higher-order loss suppression and increased fidelity. Note that the code found in Ref. [100] through biconvex optimization can also achieve higher fidelity than LNCY code [100], together suggesting that the NSA formalism can consistently improve code performances.

Remarkably, both the SC and PC structures can be systematically generalized into

families of AD codes under the framework of NSA, each capable of correcting a single AD error. Compact codeword forms of these families for arbitrary n can also be obtained. We explicitly compute the fidelities for both families of codes as functions of n and γ, and show that the performance advantage of NSA codes grows with the number of qubits. We also observe that the two families present a trade-off between error-correction performance and code rates. Specifically, for a given n, PC code has a smaller k but achieves a higher fidelity. We note that the NSA SC codes can be generalized further to qudit systems with arbitrary local dimension $q \geq 2$. In addition, our NSA framework can be applied to systems consisting of bosons. One of the simplest examples of a bosonic code, the 0-2-4 binomial code, which can correct a single photon loss [102], can be adapted to an NSA version with improved fidelity. Our results demonstrate that NSA is an advantageous and comprehensive formalism that may significantly benefit the design of a wide variety of codes.

It is important to emphasize that the NSA $(\!(4, 1)\!)$ AD codes originate from our observations in the learning procedures, which are non-stabilizer codes, highlighting the potential of machine learning to identify and refine new codes. Furthermore, we successfully developed an analytical understanding of our NSA codes, enabling systematic generalizations to broader classes of codes.

3.1.1 Organization

The remainder of this chapter is organized as follows: Sec. 3.2 reviews the principles of (approximate) quantum error correction and defines amplitude damping noise. Sec. 3.3 demonstrates how our ML experiments lead to discovering NSA codes, followed by Sec. 3.4 where we present our $(\!(4, 1)\!)$ NSA code findings. In Secs. 3.5, we present recovery protocols for the self-complementary and pair-complementary NSA codes, along with calculations of their worst-case fidelities. In Sec. 3.6, we generalize these results by deriving families of $(\!(n, k)\!)$ NSA codes and an NSA version of the 0-2-4 binomial code. Finally, Sec. 3.7 summarizes our findings and outlines potential directions for future research.

3.2 Preliminaries

3.2.1 Quantum error correction

To perform quantum error correction, we need to encode the logical qubits into a larger physical Hilbert space, so that the encoded information is protected from the noise. Con-

sider a a 2^k-dimensional subspace of an n-qubit Hilbert space as a $((n, k))$ code. This code is considered a perfect QECC against a set of errors $\{E_a\}$ *if and only if* it satisfies the the Knill-Laflamme (KL) condition [83]:

$$\langle\bar{\psi}_\alpha|E_a^\dagger E_b|\bar{\psi}_\beta\rangle = C_{ab}\delta_{\alpha\beta} \ , \tag{3.1}$$

where $\{|\bar{\psi}_\alpha\rangle\}$ forms the basis of the code subspace, and E_a are the Kraus operators for the noise channel. Here, the weight of an error operator is defined as the number of qubits on which it acts nontrivially, for example an opeartor $E_a = \mathrm{id}_1 \otimes \mathrm{id}_2 \otimes \sigma_{x,3} \otimes \mathrm{id}_4$ has weight 1. A code can correct weight-t errors if Eq. 3.1 holds for all error operators with weight up to t, in which case it can also detect errors of weight $2t$.

The KL condition guarantees that there exists a perfect recovery channel $\mathcal{R}$ after the encoded information is contaminated by the noise operators, in other words $\mathcal{R}\circ\mathcal{N}\circ\mathcal{E} = \mathrm{id}_\mathrm{L}$, where $\mathcal{N}$ and $\mathcal{E}$ are the noisy and encoding channels respectively, and id_L denotes the identity channel on the logical subspace, see Fig. 3.1 for an illustration.

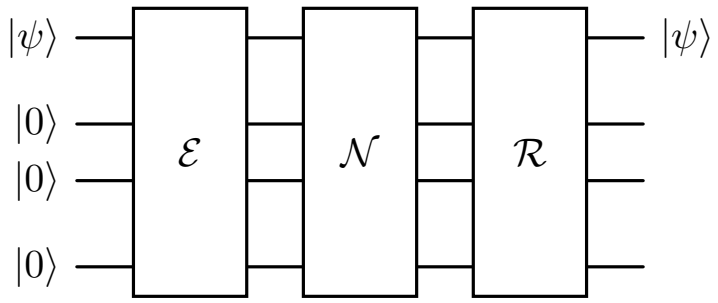


Figure 3.1: Cartoon of quantum error correction.

One of the foundational frameworks for constructing quantum error-correcting codes is the stabilizer formalism, which employs groups of commuting Pauli operators—known as stabilizers—to define a protected subspace. A stabilizer code is specified by its stabilizer group, and the code space is the simultaneous +1 eigenspace of all stabilizer generators. A notable property of stabilizer codes is that the norm of each logical state's components remains *uniform* across the code space. These structural features make stabilizer codes both conceptually elegant and practically powerful, establishing them as the cornerstone of modern fault-tolerant quantum computation.

3.2.2 Approximate quantum error correction

Conversely, when the recovery channel can not be perfect, namely $\mathcal{R} \circ \mathcal{N} \circ \mathcal{E} \approx \mathrm{id}_{\mathrm{L}}$, an approximate quantum error correction code is obtained and Eq. 3.1 can be relaxed to be [10, 104]:

$$\langle \bar{\psi}_\alpha | E_a^\dagger E_b | \bar{\psi}_\beta \rangle = C_{ab} \delta_{\alpha\beta} + \epsilon_{\alpha\beta,ab} \ , \tag{3.2}$$

where $\epsilon_{\alpha\beta,ab}$ are small parameters quantifying the KL condition violation and depend on both the code states and the error operators. Alternative ways to describe the deviations of an AQEC code from an exact one include the subsystem variance [147] and the near-optimal fidelity [155], both of which are related to the violation of the KL condition in Eq. (3.2).

One can also benchmark the performance of an AQEC code by the worst-case fidelity, which we will simply refer to as fidelity later on,

$$\mathcal{F} = \max_{\mathcal{R}} \left\{ \min_{|\bar{\psi}\rangle} \langle \bar{\psi}| \left(\mathcal{R} \circ \mathcal{N}\right)(|\bar{\psi}\rangle\langle\bar{\psi}|) \, |\bar{\psi}\rangle \right\} \ , \tag{3.3}$$

where minimization is taken over the code states $|\bar{\psi}\rangle$ and maximization is taken over all possible recovery channels $\mathcal{R}$. While in general there is no restriction on the recovery channel, we restrict ourselves to $\mathcal{R}$ such that after the recovery there will be no residual logical error. Note that the scaling of $1 - \mathcal{F}$ with respect to the noise strength γ is lower-bounded by the KL condition violation's scaling; a detailed proof is presented in the next subsection, section 3.2.3.

3.2.3 Worst-case fidelity and loss function for AQEC codes

In this subsection, we provide detailed information on the relationship between the worst-case fidelity and the loss function that we discuss in section 3.2.2. Specifically, we will define three types of measures that describe the deviation from the exact QEC: the completely bounded fidelity $\mathcal{F}_c$, the worst-case fidelity $\mathcal{F}$, and the loss function $\mathcal{L}$, here $\mathcal{L} = \mathcal{L}(\epsilon_{\alpha\beta,ab})$ can be regarded as a function of the KL condition violation. We will focus on the scaling of these measures in terms of the noise strength γ, and we prove that the scaling of the loss function $\mathcal{L}$ provides a lower bound for the scaling of $1 - \mathcal{F}$ and $1 - \mathcal{F}_c$, i.e. if $\mathcal{L} = O(\gamma^{t+1})$ where γ is the noise strength, then $1 - \mathcal{F}_c = O(\gamma^{t+1})$ and $1 - \mathcal{F} = O(\gamma^{t+1})$.

We now define these two fidelity measures individually:

Definition 1 (Completely bounded fidelity)**.** For encoding map $\mathcal{E}$ and noise channel $\mathcal{N}$ acting on the physical system, the *Completely bounded fidelity* is defined as

$$\mathcal{F}_c \coloneqq \max_{\mathcal{R}} \min_{\rho} f((\mathcal{R} \circ \mathcal{N} \circ \mathcal{E} \otimes \mathrm{id})(\rho), (\mathrm{id}_L \otimes \mathrm{id})(\rho)) \ , \tag{3.4}$$

where f is the Uhlmann fidelity $f(\rho, \sigma) \coloneqq \mathrm{Tr}\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}$, and id_L denotes the logical identity channel. The optimization involves $\mathcal{R}$, which runs over recovery channels, and ρ, which runs over input states on any extended system.

Definition 2 (Worst-case fidelity)**.** For encoding map $\mathcal{E}$ and noise channel $\mathcal{N}$ acting on the physical system, the *worst-case fidelity* is defined as

$$\mathcal{F} \coloneqq \max_{\mathcal{R}} \min_{\rho} f(\mathcal{R} \circ \mathcal{N} \circ \mathcal{E}(\rho), \rho) \ , \tag{3.5}$$

where f is the Uhlmann fidelity $f(\rho, \sigma) \coloneqq \mathrm{Tr}\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}$, and id_L denotes the logical identity channel. The optimization involves $\mathcal{R}$, which runs over recovery channels, and ρ, which runs over input logical states.

Note that $\mathcal{F}_c \leq \mathcal{F}$, so we just need to prove that if $\mathcal{L} = O(\gamma^{t+1})$ where γ is the noise strength, then $1 - \mathcal{F}_c = O(\gamma^{t+1})$. Our proof is based on the following lemma:

Lemma 1 (Proposition 2 in [20])**.** *Consider an n-qubit noise channel $\mathcal{N}(\rho) = \sum_a E_a \rho E_a^\dagger$, and a quantum error correcting code spanned by code words $\{|\psi_1\rangle, ..., |\psi_K\rangle\}$. Denote the cost function of the basis states as $\mathcal{L}$. Then if the code is ε-correctable under $\mathcal{N}$, i.e. $\sqrt{1 - \mathcal{F}_c} = \varepsilon$, we can bound ε by*

$$\varepsilon \leq K\sqrt{2\mathcal{L}} \ . \tag{3.6}$$

Using this lemma [20], we have

$$\mathcal{F}_c \geq 1 - 2K^2\mathcal{L} \ . \tag{3.7}$$

So if $\mathcal{L} = O(\gamma^{t+1})$ then $1 - \mathcal{F}_c = O(\gamma^{t+1})$ and thus $1 - \mathcal{F} = O(\gamma^{t+1})$.

3.2.4 Amplitude damping noise

In this work, we aim to use machine learning to find AQEC codes suitable for amplitude damping noise, which in a two-level system corresponds to the spontaneous decay from the

excited state to the ground state. The noise channel $\mathcal{N}^A(\rho) = A^0\rho A^{0\dagger} + A^1\rho A^{1\dagger}$ is given by the Kraus operators

$$A^0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix} \; ; \quad A^1 = \begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix} \; , \tag{3.8}$$

where the decay rate γ characterizes the noise strength. A^1 corresponds to a 'jump' from the state $|1\rangle$ to $|0\rangle$, while A^0 reduces the amplitude of $|1\rangle$ by the amount $\sqrt{1-\gamma}$.

We consider the following independent amplitude damping noise model, $\mathcal{N} = \otimes_{i=1}^{n}\mathcal{N}_i^A$, where $\mathcal{N}_i^A$ is the amplitude damping channel acting on the i-th qubit with Kraus operators A^0 and A^1. Note that each qubit is affected independently by its own noise operator.

3.3 Machine learning experiments

In the following, we demonstrate how our machine learning experiments played a key role in discovering NSA codes for amplitude damping noise. While our NSA formulation is motivated by experimental results, its analytical construction remains rigorous and independent of the learning process.

To search for AQEC codes, we employ a variational quantum learning approach, where the bases of the code space are generated by a parametrized quantum circuit. This circuit consists of multiple layers of one- and two-qubit rotation gates with trainable variables. The details of the circuit are provided later in this section. For an AQEC code designed to approximately correct t errors, the KL condition violation is required to scale as $O(\gamma^{t+1})$ [52], where γ denotes the noise strength. To enforce this scaling, we define our loss function in our variational quantum learning as the sum of the L^1-norms of the KL condition violations [20]:

$$\begin{aligned} \mathcal{L} = \sum_{\mu} \left[\sum_{1\leqslant\alpha<\beta\leqslant 2^k} |\langle\bar{\psi}_\alpha|E_\mu|\bar{\psi}_\beta\rangle| \right] \\ + \sum_{\mu} \left[\sum_{\alpha=1}^{2^k} \frac{1}{2} \left| \langle\bar{\psi}_\alpha|E_\mu|\bar{\psi}_\alpha\rangle - \overline{\langle E_\mu\rangle} \right| \right] \; , \end{aligned} \tag{3.9}$$

where $E_\mu = E_a^\dagger E_b$ and each Kraus operator E_a is drawn from the error set $\{E_a\}$. Our experiments focus on a 4-qubit system designed to approximately correct a single amplitude

damping error ($t = 1$). The independent error set consists of operators $E_a = \otimes_{i=1}^{n} A_i^{l_i}$, where at most one qubit experiences a jump error A^1, while the remaining qubits are affected by A^0.

We begin with applying the variational quantum learning protocol to search for the best code states at a fixed noise strength γ_0 by minimizing the loss function defined in Eq. (3.9). The details of the learning procedure are provided in later in this section. After obtaining an optimized code at γ_0, we evaluate its performance across a range of noise strengths γ, as shown in Fig. 3.2(a). In general, for an AQEC code correcting a single AD noise, the loss function is expected to vary smoothly with γ and typically scales as $O(\gamma^2)$, ensuring that the code remains effective for a general strength γ. This expected behavior is exhibited in the $((4, 1))$ LNCY code in [91], as shown in the green line in Fig. 3.2(a).

However, the codes found through our variational learning optimized for a specific noise strength γ_0 exhibit a strikingly different pattern. Not only do they achieve lower losses around γ_0 compared to the LNCY code, but their loss functions also reveal a distinct *non-smooth point* precisely at $\gamma = \gamma_0 = 10^{-1.5}$, as shown in the orange and blue solid lines in Fig. 3.2(a). This observation leads to significant implications: i) the codes learned by our variational protocol must be fundamentally different from the LNCY code; ii) AQEC codes optimized for a specific noise strength can not be realized as a single fixed-parameter code across varying noise regimes. Instead, the optimal code structure itself should vary with γ. This motivates us to introduce a new code construction—noise-strength adapted (NSA) codes—in which the code states are designed to continuously adjust with the noise strength γ.

As shown in Fig. 3.2(a), the machine learning results reveal the presence of two saddle points, indicated in the blue and orange lines. Both saddle points correspond to code structures with reduced loss around γ_0, and they belong to two distinct families of codes. With the new NSA construction, we demonstrate that the loss function for codes with varying code states smoothly changes with γ, as indicated in the dashed lines in Fig. 3.2(a). We demonstrate in the next section that these two families—termed self-complementary and pair-complementary codes—differ in their structural properties (with the naming convention becoming clear from the analytical formulation), and offer better performance over a broad range of noise strength, compared to the conventional non-NSA approximated code.

3.3.1 Details of variational quantum learning

Here, we provide details of our variational quantum learning approach and explain how the properties of the logical states found through machine learning (ML) guided us in

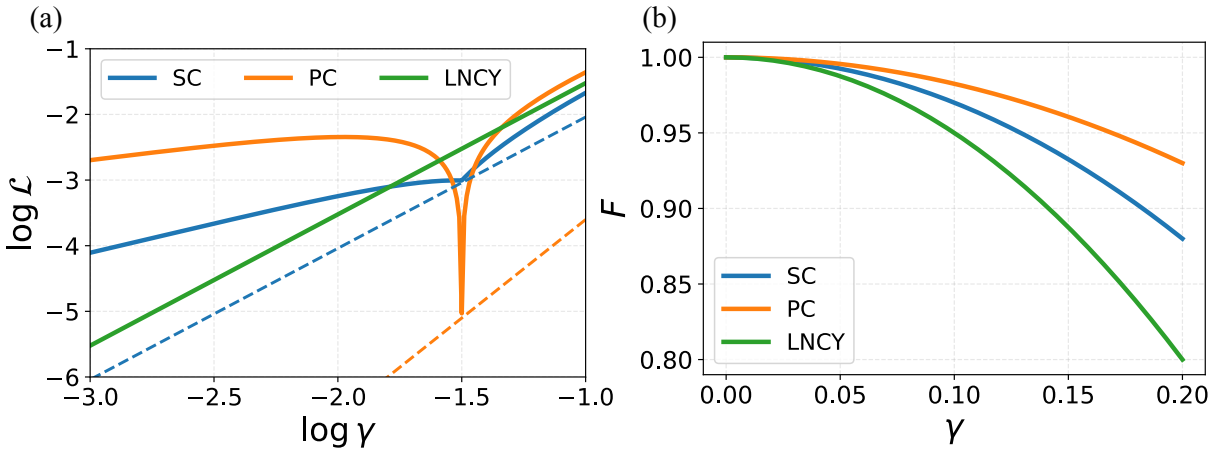


Figure 3.2: (a) Loss $\mathcal{L}$ as a function of γ is shown. The solid green line shows the smooth γ^2 dependence of LNCY code. The solid blue and orange lines show the losses of two codes optimized for $\gamma = 10^{-1.5}$. The blue and orange dashed lines, instead, show the losses for NSA codes with adaptive code states. (b) shows the fidelity plotted as a function of γ for $((4, 1))$ NSA SC, $((4, 1))$ NSA PC, and LNCY codes.

developing our noise-strength adapted (NSA) code.

To identify the code subspace, we employ a parameterized quantum circuit introduced in [20]. This circuit comprises multiple layers of one- and two-qubit rotation gates with trainable parameters, as shown in Fig. 3.3. For an $((n, k))$ code, the first k qubits are initialized as the logical state we want to encode, and the rest of the qubits are initialized as the $|0\rangle$ states. For instance, if $k = 1$, the first qubit are initialized as $|0\rangle$ and $|1\rangle$ for the logical states $|\bar{0}\rangle$ and $|\bar{1}\rangle$ respectively, while all other states are initialized as the $|0\rangle$ states. After initialization, we apply multiple layers of rotation gates, $\{R_x, R_z, R_{zz}\}$, to encode the states. Here, R_x and R_z are single-qubit rotation gates about the x- and z-axes, respectively, while R_{zz} is a two-qubit entangling gate defined as $e^{-i\theta\sigma_i^z\sigma_j^z/2}$. This gate set is universal, as any unitary operation can be decomposed into a combination of these three gates.

Specifically, our circuit consists of three layers of rotation gates. Within each layer, we apply R_x and R_z rotations to each physical qubit and R_{zz} rotations to all possible qubit pairs. Each rotation gate has a trainable rotation angle, resulting in a total of $(3n^2 + 13n)/2$ parameters. In our experiments, we focus on searching for the $((4, 1))$ code using a 50-parameter parameterized quantum circuit.

As discussed earlier in this section an approximate quantum error correcting (AQEC)

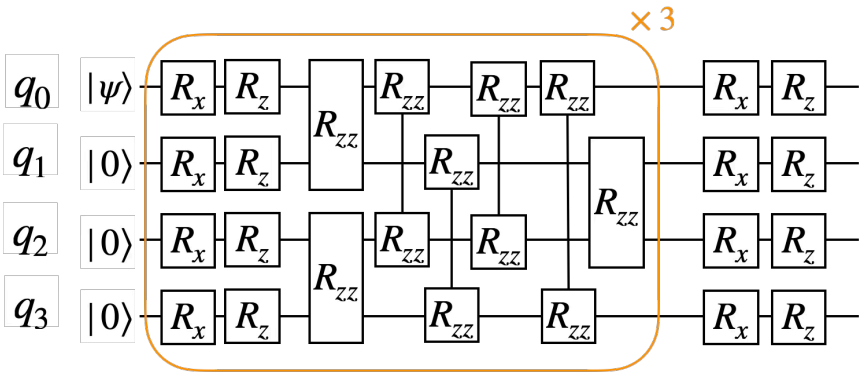


Figure 3.3: Encoding circuit $\mathcal{E}$ with trainable parameters for searching $(\!(4,1)\!)$ code.

code designed to approximately correct $t = 1$ errors requires the violation of the KL condition to scale as $O(\gamma^2)$ [52]. To enforce this scaling, we define the $\mathcal{L}_1$ loss function in our variational quantum learning framework as the sum of the L^1-norms of the KL condition violations. To enhance training stability, we introduce an additional $\mathcal{L}_2$ loss function, also defined as the sum of the L^2-norms of the KL condition violations. The training procedure consists of two stages: we first perform pre-training using the $\mathcal{L}_2$ loss function, followed by fine-tuning with the $\mathcal{L}_1$ loss function [20]. The loss function $\mathcal{L}_1$ is defined above in Eq. 3.9, and the loss function $\mathcal{L}_2$ is defined as follows

$$\mathcal{L}_2 = \sum_{\mu} \left[\sum_{1\leqslant\alpha<\beta\leqslant 2^k} |\langle\bar{\psi}_\alpha|E_\mu|\bar{\psi}_\beta\rangle|^2 \right] + \sum_{\mu} \left[\sum_{\alpha=1}^{2^k} \frac{1}{4} \left| \langle\bar{\psi}_\alpha|E_\mu|\bar{\psi}_\alpha\rangle - \overline{\langle E_\mu\rangle} \right|^2 \right] . \tag{3.10}$$

Here as before, $\{|\bar{\psi}_\alpha\rangle\}$ are encoded states after parameterized quantum circuit in Fig. 3.3 , $E_\mu = E_a^\dagger E_b$, and each Kraus operator is drawn from the independent error set $E_a = \otimes_{i=1}^n A_i^{l_i}$, where at most one qubit experiences a jump error (A^1), while the remaining qubits are affected by A^0.

In our experiments, we used the BFGS method from the SciPy optimization package (scipy.optimize) as our optimizer. Within 20,000 steps, our minimization process successfully identified optimized encoding circuits for the code subspace. As discussed and illustrated in Fig. 3.2, the codes discovered through our variational quantum learning are optimized for a specific noise strength γ_0. However, while these codes outperform known codes [91] within certain regimes around this specific noise strength γ_0, they perform worse in other regimes. This suggests that AQEC codes optimized for a specific noise strength cannot performs optimally across different noise regimes. Instead, the optimal code struc-

ture itself should vary with γ.

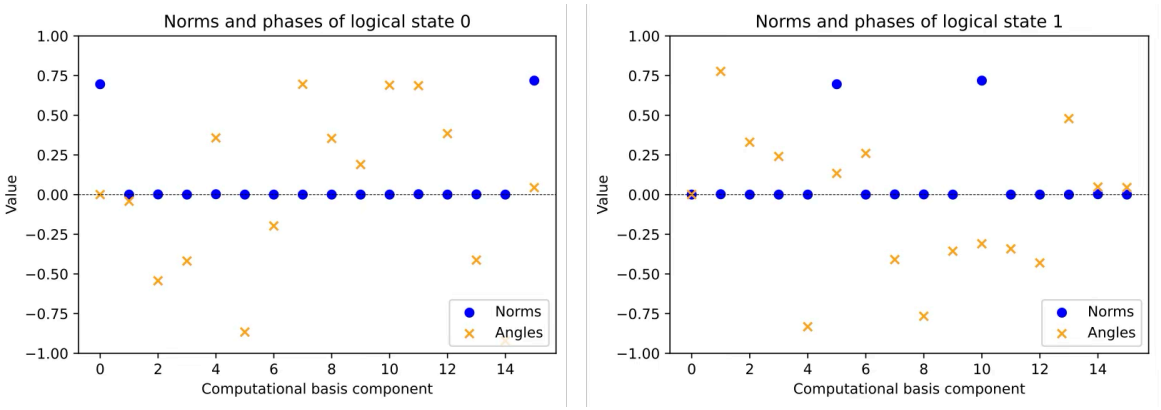


Figure 3.4: Codeword obtained through variational quantum learning, corresponding to the self-complementary NSA code.

To investigate how codewords vary with noise strength γ, we analyzed the norms and angles of codewords obtained through our variational quantum learning method at $\gamma = \gamma_0 = 10^{-1.5}$. The codewords corresponding to the blue line in Fig. 3.2 are illustrated in Fig. 3.4 below. We observe that only two computational basis components have a considerable norm, and their phases appear to be quite random. Upon closer examination of the norms, we find that their differences are proportional to certain factors of γ_0. Therefore, we propose the following ansatz for the codewords:

$$\begin{aligned} |\bar{0}\rangle &= A(\gamma)\,|0000\rangle + \sqrt{1-A(\gamma)^2}\,|1111\rangle \ ; \\ |\bar{1}\rangle &= B(\gamma)\,|0011\rangle + \sqrt{1-B(\gamma)^2}\,|1100\rangle \ . \end{aligned} \tag{3.11}$$

We further compute the L_1 loss of these codewords as a function of the coefficients $A(\gamma)$ and $B(\gamma)$. Our results show that the choice $A(\gamma) = 1/\sqrt{1+(1-\gamma)^{-4}}$ and $B(\gamma) = 1/\sqrt{2}$ yields a minimum of the loss function in Eq. (3.9), leading to the self-complementary NSA code presented in Eq. 3.12.

We perform the same analysis for a second code learned from ML, corresponding to the orange line in Fig. 3.2. The optimized codewords from ML are shown in Fig. 3.5. Similarly, we find that when parameterizing the codeword basis as a function of γ as in Eq. 3.11, we obtain a code of the form given by Eq. 3.13, which yields a minimum of the loss function.

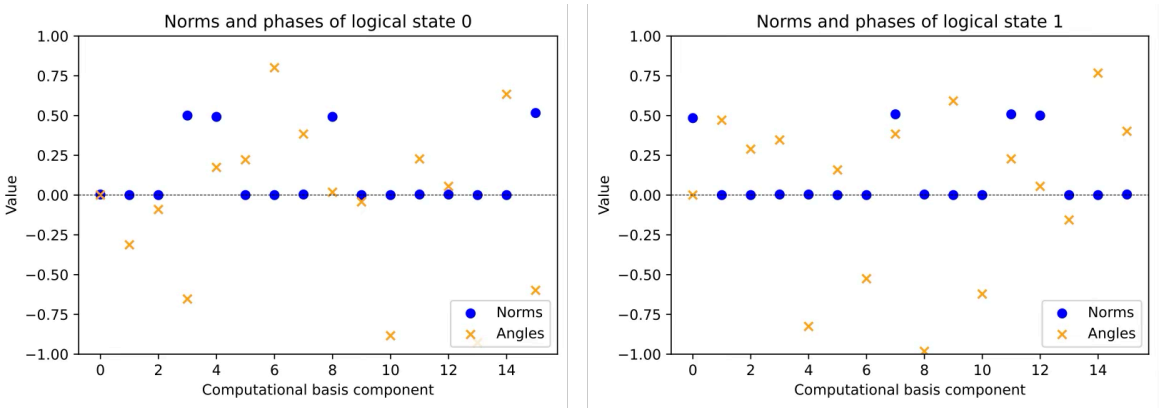


Figure 3.5: Codeword obtained through variational quantum learning, corresponding to the pair-complementary NSA code.

3.4 Basic analytical results

Our machine learning experiment on 4-qubit codes reveals a distinct structured pattern in the code states, which can be captured analytically by a closed-form expression that adapts to the noise strength γ.

We begin by presenting the 4-qubit NSA self-complementary code:

$$|\bar{0}\rangle = \sqrt{\frac{1}{1+(1-\gamma)^{-4}}}(|0000\rangle + (1-\gamma)^{-2}|1111\rangle) \; ;$$
$$|\bar{1}\rangle = \sqrt{\frac{1}{2}}\,(|0011\rangle + |1100\rangle) \; . \tag{3.12}$$

This code differs from LNCY code in that the coefficients in $|\bar{0}\rangle$ depend on γ, whereas the coefficients in both $|\bar{0}\rangle$ and $|\bar{1}\rangle$ in LNCY code are fixed at $1/\sqrt{2}$. In the limit of $\gamma \to 0$, the NSA code reduces to its non-NSA counterpart, the LNCY code.

The loss $\mathcal{L}$, calculated using Eq. (3.9), for the NSA code in Eq. (3.12) is $\gamma^2 + O(\gamma^3)$, which is smaller compared to the LNCY code, $3\gamma^2 + O(\gamma^3)$. Furthermore, without allowing residual logical errors, the fidelity for the NSA code is $1 - 3\gamma^2 + \mathcal{O}(\gamma^3)$, cf. $1 - 5\gamma^2 + \mathcal{O}(\gamma^3)$ for the LNCY code, plotted as the blue and green lines in Fig. 3.2(b) respectively. The difference in the fidelities lies in that the recovery channel distorts the encoded states in LNCY non-NSA code, but such distortion can be absent in our NSA code, as detailed in section 3.5.1. Notably, even in the non-NSA code, where the codewords are independent of γ, knowledge of the noise strength remains crucial for decoding [91].

Next, we introduce another 4-qubit NSA code, termed the pair-complementary (PC) code, which differs from the self-complementary code by having code states consisting of four basis states:

$$
\begin{aligned}
|\bar{0}\rangle = \big((1-\gamma)^{-1}|0011\rangle - (1-\gamma)^{-3/2}|1110\rangle \\
-(1-\gamma)^{-3/2}|1101\rangle + |0000\rangle\big)/N_0 \,; \\
|\bar{1}\rangle = \big((1-\gamma)^{-1}|1100\rangle + (1-\gamma)^{-1/2}|0001\rangle \\
+(1-\gamma)^{-1/2}|0010\rangle + (1-\gamma)^{-2}|1111\rangle\big)/N_1 \,,
\end{aligned} \tag{3.13}
$$

where N_0 and N_1 are normalizations. Notice that the basis states involved in the two logical states are pairwise complementary to each other. The loss for the pair-complementary code is $\gamma^3/4 + O(\gamma^4)$, and achieves higher-order loss suppression compared to the self-complementary codes, as also illustrated by the dashed orange lines in Fig. 3.2(a). Additionally, the fidelity for the pair-complementary code, given by $1 - 7\gamma^2/4 + O(\gamma^3)$ and marked as orange line in Fig. 3.2(b), is larger than the NSA 4-qubit self-complementary code. Notably, this new code has no corresponding non-NSA versions that have been reported in the literature.

Both codes derived from ML experiments can be generalized to arbitrary system size. In later section, we introduce these two families of NSA codes and demonstrate their enhanced performance in terms of fidelity.

3.5 Recovery protocol and fidelity calculation

In this section, we present the recovery protocol and detailed fidelity calculation for the self-complementary and pair-complementary $(\!(4,1)\!)$ NSA codes.

3.5.1 Recovery and worst-case fidelity of $(\!(4,1)\!)$ self-complementary NSA code

Our recovery channel is similar to the non-NSA $(\!(4,1)\!)$ amplitude damping code in [91], and we will showcase that the self-complementary NSA code can obtain a larger fidelity using a similar decoding circuit.

We consider the input qubit state,

$$|\bar{\psi}_{\rm in}\rangle = a|\bar{0}\rangle + b|\bar{1}\rangle \ , \tag{3.14}$$

where the codewords of the 4-qubit NSA AD code are as follows:

$$\begin{aligned} |\bar{0}\rangle &= \left(|0000\rangle + (1-\gamma)^{-2}|1111\rangle\right)/\sqrt{1+(1-\gamma)^{-4}} \ ; \\ |\bar{1}\rangle &= \left(|0011\rangle + |1100\rangle\right)/\sqrt{2} \ . \end{aligned} \tag{3.15}$$

After $t = 1$ amplitude damping errors $A_{0000} = A^0A^0A^0A^0$, $A_{1000} = A^1A^0A^0A^0$, $A_{0100} = A^0A^1A^0A^0$, $A_{0010} = A^0A^0A^1A^0$ and $A_{0001} = A^0A^0A^0A^1$, all possible final states occurring with probabilities at least $O(\gamma)$ are,

$$|\phi_{0000}\rangle = a\left[\frac{|0000\rangle + |1111\rangle}{\sqrt{1+(1-\gamma)^{-4}}}\right] + b\left[\frac{(1-\gamma)\left[|0011\rangle + |1100\rangle\right]}{\sqrt{2}}\right] \ , \tag{3.16}$$

$$|\phi_{1000}\rangle = a\sqrt{\frac{\gamma}{(1-\gamma)(1+(1-\gamma)^{-4})}}|0111\rangle + b\sqrt{\frac{\gamma(1-\gamma)}{2}}|0100\rangle \ , \tag{3.17}$$

$$|\phi_{0100}\rangle = a\sqrt{\frac{\gamma}{(1-\gamma)(1+(1-\gamma)^{-4})}}|1011\rangle + b\sqrt{\frac{\gamma(1-\gamma)}{2}}|1000\rangle \ , \tag{3.18}$$

$$|\phi_{0010}\rangle = a\sqrt{\frac{\gamma}{(1-\gamma)(1+(1-\gamma)^{-4})}}|1101\rangle + b\sqrt{\frac{\gamma(1-\gamma)}{2}}|0001\rangle \ , \tag{3.19}$$

$$|\phi_{0001}\rangle = a\sqrt{\frac{\gamma}{(1-\gamma)(1+(1-\gamma)^{-4})}}|1110\rangle + b\sqrt{\frac{\gamma(1-\gamma)}{2}}|0010\rangle \ . \tag{3.20}$$

The first step of the recovery process is to perform the error syndrome detection as shown in Fig. 3.6, which is similar to the circuit in Fig. 2 in [91]. There could only be three possible outcomes, $(M_2, M_4) = (0,0), (0,1), (1,0)$. If $(M_2, M_4) = (0,0)$, the inferred state will be $|\psi_{0000}\rangle$; if $(M_2, M_4) = (1,0)$, the inferred state will be $|\psi_{1000}\rangle$ or $|\psi_{0100}\rangle$; if $(M_2, M_4) = (0,1)$, the inferred state will be $|\psi_{0010}\rangle$ or $|\psi_{0001}\rangle$; and we discard the output state if $(M_2, M_4) = (1,1)$. In the following, we examine the first three cases separately.

After the syndrome measurement, only the first the third qubits remain. We denote their states as $|q_1\rangle$ and $|q_3\rangle$ respectively. If $(M_2, M_4) = (0,0)$, that implies A_{0000} has occurred, then $|q_1, q_3\rangle$ after measurement becomes,

$$a\left[\frac{|00\rangle + |11\rangle}{\sqrt{1+(1-\gamma)^{-4}}}\right] + b\left[\frac{(1-\gamma)(|01\rangle + |10\rangle)}{\sqrt{2}}\right] \ . \tag{3.21}$$

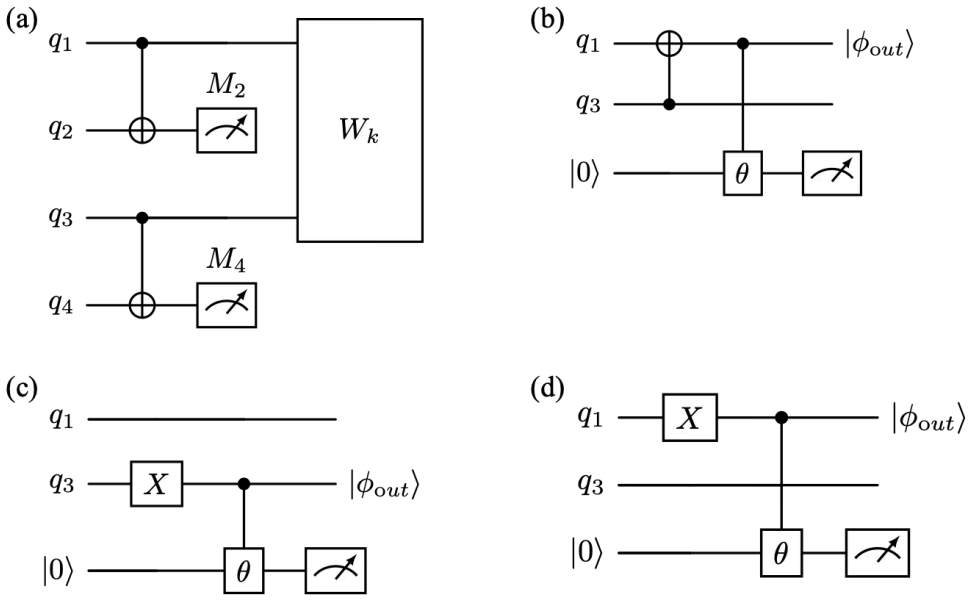


Figure 3.6: **(A)** Circuit for error syndrome detection. The measurement outcomes (M_2, M_4) determine which of the three recovery operations W_k to apply. **(B)**, **(C)**, and **(D)** show the corresponding recovery circuits for measurement results 00, 10, and 01, respectively. The controlled-rotation gates are CR_y gates with rotation angle $2\cos^{-1}\left(\sqrt{\frac{2}{(1-\gamma)^2+(1-\gamma)^{-2}}}\right)$.

Applying a control-not using $|q_3\rangle$ as a control, gives,

$$|q_1 q_3\rangle = \left[(a\sqrt{\frac{2}{1+(1-\gamma)^{-4}}}|0\rangle + b(1-\gamma)|1\rangle) \right] \frac{|0\rangle + |1\rangle}{\sqrt{2}} \ . \tag{3.22}$$

We then connect the system with an ancilla qubit prepared to be $|0\rangle$, and consider a non-unitary operation on the q_1 qubit: $\mathcal{N}(\rho) = N_0 \rho N_0^\dagger + N_1 \rho N_1^\dagger$ where

$$N_0 = |0\rangle\langle 0| + \sqrt{\frac{2}{(1-\gamma)^2 + (1-\gamma)^{-2}}}|1\rangle\langle 1| \ ; \tag{3.23}$$

$$N_1 = \sqrt{1 - \frac{2}{(1-\gamma)^2 + (1-\gamma)^{-2}}}|1\rangle\langle 1| \ . \tag{3.24}$$

This operation is equivalent to applying a controlled rotation gate CR_y—with $|q_1\rangle$ as the control qubit and the ancilla as the target qubit—using a rotation angle of $2\cos^{-1}\left(\sqrt{\frac{2}{(1-\gamma)^2+(1-\gamma)^{-2}}}\right)$. Subsequently, the ancilla qubit is measured. If the measurement outcome on the ancilla is $|0\rangle$, the resulting state is:

$$|\phi_{\text{out}}\rangle = \sqrt{\frac{2}{1+(1-\gamma)^{-4}}}(a|0\rangle + b|1\rangle) \ . \tag{3.25}$$

If $(M_2, M_4) = (1, 0)$, that implies A_{1000} or A_{0100} has occurred, then the final states that we should consider will be $|\phi_{1000}\rangle$ or $|\phi_{0100}\rangle$. In either case, $|q_1 q_3\rangle$ is a product state and the third qubit will be distorted as

$$|q_3\rangle = a\sqrt{\frac{\gamma}{(1-\gamma)(1+(1-\gamma)^{-4})}}|1\rangle + b\sqrt{\frac{\gamma(1-\gamma)}{2}}|0\rangle \ . \tag{3.26}$$

We can also connect the system with an ancilla qubit prepared to be $|0\rangle$, and considering a non-unitary operation $\mathcal{N}(\rho) = N_0 \rho N_0^\dagger + N_1 \rho N_1^\dagger$ where

$$N_0 = |0\rangle\langle 1| + \sqrt{\frac{2}{(1-\gamma)^2 + (1-\gamma)^{-2}}}|1\rangle\langle 0| \ ; \tag{3.27}$$

$$N_1 = \sqrt{1 - \frac{2}{(1-\gamma)^2 + (1-\gamma)^{-2}}}|1\rangle\langle 0| \ . \tag{3.28}$$

This operation consists of applying an Pauli X gate to qubit $|q_3\rangle$, followed by a controlled rotation gate CR_y—with $|q_3\rangle$ as the control qubit and the ancilla as the target qubit—using the same rotation angle as previously described for A_{0000}. Subsequently, the ancilla qubit is measured. If the measurement outcome on the ancilla is $|0\rangle$, the resulting state is:

$$|\phi_{\text{out}}\rangle = \sqrt{\frac{\gamma}{(1-\gamma)+(1-\gamma)^{-3}}}\,[a|0\rangle + b|1\rangle] \ . \tag{3.29}$$

Same decoding process can be performed when error A_{0010} or A_{0001} has occurred, when $(M_2, M_4) = (0, 1)$. Thus the worst case fidelity is

$$\mathcal{F} = \frac{2}{1+(1-\gamma)^{-4}} + \frac{4\gamma}{(1-\gamma)+(1-\gamma)^{-3}} = 1 - 3\gamma^2 + O(\gamma^3) \ . \tag{3.30}$$

In contrary, the fidelity for the non-NSA 4-qubit self-complementary AD code in [91] is $1 - 5\gamma^2 + O(\gamma^3)$. Our NSA code indeed has better fidelity at the order of $O(\gamma^2)$.

3.5.2 Recovery and worst-case fidelity of $(\!(4,1)\!)$ pair-complementary NSA code

In this subsection, we continue to present the recovery protocol and calculate the worst-case fidelity achievable without residual logical errors for the $(\!(4,1)\!)$ pair-complementary NSA code. The key idea of the recovery protocol is similar to the self-complementary code, so we will not repeat the details here. We consider the following codewords,

$$\begin{aligned}
|\bar{0}\rangle &= \frac{1}{N_0}\left((1-\gamma)^{-1}|0011\rangle - (1-\gamma)^{-3/2}|1110\rangle - (1-\gamma)^{-3/2}|1101\rangle + |0000\rangle\right) \ ; \\
|\bar{1}\rangle &= \frac{1}{N_1}\left((1-\gamma)^{-1}|1100\rangle + (1-\gamma)^{-1/2}|0001\rangle + (1-\gamma)^{-1/2}|0010\rangle + (1-\gamma)^{-2}|1111\rangle\right) \ ,
\end{aligned} \tag{3.31}$$

where N_0 and N_1 are normalization factors,

$$N_0 = \sqrt{(1-\gamma)^{-2} + 2(1-\gamma)^{-3} + 1} \ , \quad N_1 = \sqrt{(1-\gamma)^{-2} + 2(1-\gamma)^{-1} + (1-\gamma)^{-4}} \ . \tag{3.32}$$

For an input state $|\bar{\psi}_{\text{in}}\rangle = a\,|\bar{0}\rangle + b\,|\bar{1}\rangle$, the possible output states after $t = 1$ amplitude

damping, occurring with probabilities of order $O(\gamma)$ or higher, are:

$$|\phi_{0000}\rangle = \frac{a}{N_0}\left(|0011\rangle - |1110\rangle - |1101\rangle + |0000\rangle\right) + \frac{b}{N_1}\left(|1100\rangle + |0001\rangle + |0010\rangle + |1111\rangle\right) \ , \tag{3.33}$$

$$|\phi_{1000}\rangle = \sqrt{\frac{\gamma}{1-\gamma}}\left[\frac{-a}{N_0}\left(|0110\rangle + |0101\rangle\right) + \frac{b}{N_1}\left(|0100\rangle + |0111\rangle\right)\right] \ , \tag{3.34}$$

$$|\phi_{0100}\rangle = \sqrt{\frac{\gamma}{1-\gamma}}\left[\frac{-a}{N_0}\left(|1010\rangle + |1001\rangle\right) + \frac{b}{N_1}\left(|1000\rangle + |1011\rangle\right)\right] \ , \tag{3.35}$$

$$|\phi_{0010}\rangle = \sqrt{\frac{\gamma}{1-\gamma}}\left[\frac{a}{N_0}\left(|0001\rangle - |1100\rangle\right) + \frac{b}{N_1}\left(|0000\rangle + |1101\rangle\right)\right] \ , \tag{3.36}$$

$$|\phi_{0001}\rangle = \sqrt{\frac{\gamma}{1-\gamma}}\left[\frac{a}{N_0}\left(|0010\rangle - |1100\rangle\right) + \frac{b}{N_1}\left(|0000\rangle + |1110\rangle\right)\right] \ . \tag{3.37}$$

Note that while orthogonality can assure us to distinguish the errors $A_{0000}, A_{1000}, A_{0100}$, there is a large overlap between the the error words $|\phi_{0010}\rangle$ and $|\phi_{0001}\rangle$, and we cannot distinguish the error A_{0010} and A_{0001} with a projective measurement. However, one can consider the classical mixture of the two error words and finish the error correction in its diagonal basis. More explicitly, consider the following projective measurement:

$$\begin{aligned} P_{0000} =& \frac{1}{4}\left(|0011\rangle - |1110\rangle - |1101\rangle + |0000\rangle\right)\left(\langle 0011| - \langle 1110| - \langle 1101| + \langle 0000|\right) \\ &+ \frac{1}{4}\left(|1100\rangle + |0001\rangle + |0010\rangle + |1111\rangle\right)\left(\langle 1100| + \langle 0001| + \langle 0010| + \langle 1111|\right) \ , \end{aligned} \tag{3.38}$$

$$P_{1000} = \frac{1}{2}\left(|0110\rangle + |0101\rangle\right)\left(\langle 0110| + \langle 0101|\right) + \frac{1}{2}\left(|0100\rangle + |0111\rangle\right)\left(\langle 0100| + \langle 0111|\right) \ , \tag{3.39}$$

$$P_{0100} = \frac{1}{2}\left(|1010\rangle + |1001\rangle\right)\left(\langle 1010| + \langle 1001|\right) + \frac{1}{2}\left(|1000\rangle + |1011\rangle\right)\left(\langle 1000| + \langle 1011|\right) \ , \tag{3.40}$$

$$\begin{aligned} P_S =& \frac{1}{6}\left(|0001\rangle + |0010\rangle - 2\,|1100\rangle\right)\left(\langle 0001| + \langle 0010| - 2\,\langle 1100|\right) \\ &+ \frac{1}{6}\left(|1101\rangle + |1110\rangle + 2\,|0000\rangle\right)\left(\langle 1101| + \langle 1110| + 2\,\langle 0000|\right) \ , \end{aligned} \tag{3.41}$$

$$P_A = \frac{1}{2}\left(|0001\rangle - |0010\rangle\right)\left(\langle 0001| - \langle 0010|\right) + \frac{1}{2}\left(|1101\rangle - |1110\rangle\right)\left(\langle 1101| - \langle 1110|\right) \ . \tag{3.42}$$

Here, we use S and A to denote the 'symmetrized' and 'anti-symmetrized' invariant subspaces. After performing the projection operators, we can further apply some recovery operations based on the projective measurement outcomes. As mentioned before, we choose recovery operations such that there is no residual logical error in the recovered state.

We can now explicitly compute the worst-case fidelity achievable without residual logical errors. The contribution from the error operator P_{0000} is given by

$$\mathcal{F}_0 = 4 \min \left\{ \frac{1}{N_0^2}, \frac{1}{N_1^2} \right\} = 1 - 2\gamma + \frac{1}{4}\gamma^2 + O(\gamma^3) \, . \tag{3.43}$$

The factor of 4 is due to the error state $|\phi_{0000}\rangle$ comprising four terms, and the minimum is derived from the calculation of the worst-case fidelity. The contribution from P_{1000} or P_{0010} is

$$\mathcal{F}_1 = \frac{2\gamma}{1-\gamma} \min \left\{ \frac{1}{N_0^2}, \frac{1}{N_1^2} \right\} = \frac{1}{2}\gamma - \frac{1}{2}\gamma^2 + O(\gamma^3) \, . \tag{3.44}$$

Here we use subscript 0 or 1 to denote the scaling order with respect to the noise strength γ of the code states after the amplitude damping errors.

As P_S and P_A are the projections to the invariant subspace of the error subspace, the output states, despite being affected by the noise channel, remain in a pure state after projective measurement, namely

$$P_S(|\phi_{0010}\rangle \langle\phi_{0010}| + |\phi_{0001}\rangle \langle\phi_{0001}|)P_S = |\phi_S\rangle \langle\phi_S| \ , \tag{3.45}$$

$$P_A(|\phi_{0010}\rangle \langle\phi_{0010}| + |\phi_{0001}\rangle \langle\phi_{0001}|)P_A = |\phi_A\rangle \langle\phi_A| \ , \tag{3.46}$$

where the projected states are

$$|\phi_S\rangle = \sqrt{\frac{\gamma}{1-\gamma}} \left[\frac{a}{\sqrt{2}N_0}(|0001\rangle + |0010\rangle - 2\,|1100\rangle) + \frac{b}{\sqrt{2}N_1}(|1101\rangle + |1110\rangle + 2\,|0000\rangle) \right] \, , \tag{3.47}$$

$$|\phi_A\rangle = \sqrt{\frac{\gamma}{1-\gamma}} \left[\frac{a}{\sqrt{2}N_0}(|0001\rangle - |0010\rangle) + \frac{b}{\sqrt{2}N_1}(|1101\rangle - |1110\rangle) \right] \, . \tag{3.48}$$

Their contributions to the worst-case fidelity are

$$\mathcal{F}_1' = \mathcal{F}_{1,S}' + \mathcal{F}_{1,A}' = (3+1)\frac{\gamma}{1-\gamma} \min \left\{ \frac{1}{N_0^2}, \frac{1}{N_1^2} \right\} = \gamma - \gamma^2 + O(\gamma^3) \, . \tag{3.49}$$

To summarize, we get the worst-case fidelity up to $O(\gamma^3)$,

$$\mathcal{F} = \mathcal{F}_0 + 2\mathcal{F}_1 + \mathcal{F}_1' = 1 - \frac{7}{4}\gamma^2 + O(\gamma^3) \, , \tag{3.50}$$

which is higher than the worst-case fidelity Eq. (3.30) of the $(\!(4,1)\!)$ self-complementary NSA code.

3.5.3 Entanglement fidelity of $(\!(4,1)\!)$ pair-complementary NSA code

Apart from worst-case fidelity, we can also calculate the entanglement fidelity of the $(\!(4,1)\!)$ pair-complementary NSA code. Consider the maximal entangled state:

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}\left(|\bar{0}\rangle|0_r\rangle + |\bar{1}\rangle|1_r\rangle\right) . \tag{3.51}$$

The entanglement fidelity is given by

$$F_{\text{ent}} = \langle\Phi^+|(\mathcal{R}\circ\mathcal{N})\otimes \text{id}_r(|\Phi^+\rangle\langle\Phi^+|)|\Phi^+\rangle . \tag{3.52}$$

Notice that by linearity we can calculate the contribution of the entanglement fidelity for different errors. We will focus on the $t=1$ amplitude damping errors here, the possible output states after $t=1$ amplitude damping, are:

$$\Phi_a^+ = \frac{1}{2}|\phi_a\rangle\langle\phi_a| , \tag{3.53}$$

where $a = 0000, 1000, 0100, 0010, 0001$, and

$$\begin{aligned}
|\phi_{0000}\rangle &= \frac{1}{N_1}\left(|0011\rangle - |1110\rangle - |1101\rangle + |0000\rangle\right)\otimes|0_r\rangle \\
&\quad + \frac{1}{N_2}\left(|1100\rangle + |0001\rangle + |0010\rangle + |1111\rangle\right)\otimes|1_r\rangle , \qquad (3.54)\\
|\phi_{1000}\rangle &= \sqrt{\frac{\gamma}{1-\gamma}}\left[\frac{-1}{N_1}\left(|0110\rangle + |0101\rangle\right)\otimes|0_r\rangle + \frac{1}{N_2}\left(|0100\rangle + |0111\rangle\right)\otimes|1_r\rangle\right] , \qquad (3.55)\\
|\phi_{0100}\rangle &= \sqrt{\frac{\gamma}{1-\gamma}}\left[\frac{-1}{N_1}\left(|1010\rangle + |1001\rangle\right)\otimes|0_r\rangle + \frac{1}{N_2}\left(|1000\rangle + |1011\rangle\right)\otimes|1_r\rangle\right] , \qquad (3.56)\\
|\phi_{0010}\rangle &= \sqrt{\frac{\gamma}{1-\gamma}}\left[\frac{1}{N_1}\left(|0001\rangle - |1100\rangle\right)\otimes|0_r\rangle + \frac{1}{N_2}\left(|0000\rangle + |1101\rangle\right)\otimes|1_r\rangle\right] , \qquad (3.57)\\
|\phi_{0001}\rangle &= \sqrt{\frac{\gamma}{1-\gamma}}\left[\frac{1}{N_1}\left(|0010\rangle - |1100\rangle\right)\otimes|0_r\rangle + \frac{b}{N_2}\left(|0000\rangle + |1110\rangle\right)\otimes|1_r\rangle\right] . \qquad (3.58)
\end{aligned}$$

We apply the projective measurements as described in Eqs. (3.38)–(3.42).

We can now explicitly compute the entanglement fidelity. The contribution from the error operator P_{0000} is given by

$$\mathcal{F}^0_{\text{ent}} = \frac{1}{4}(\frac{2}{N_1} + \frac{2}{N_2})^2 = 1 - 2\gamma + \frac{1}{4}\gamma^2 + O(\gamma^3) \tag{3.59}$$

The contribution from P_{1000} or P_{0010} is

$$\mathcal{F}^1_{\text{ent}} = \frac{\gamma}{4(1-\gamma)}(\frac{\sqrt{2}}{N_1} + \frac{\sqrt{2}}{N_2})^2 = \frac{1}{2}\gamma - \frac{1}{2}\gamma^2 + O(\gamma^3)\ . \tag{3.60}$$

The projected states for P_S and P_A are given by

$$|\phi_S\rangle = \sqrt{\frac{\gamma}{1-\gamma}}\left[\frac{1}{\sqrt{2}N_1}(|0001\rangle + |0010\rangle - 2\,|1100\rangle) \otimes |0_r\rangle + \frac{1}{\sqrt{2}N_2}(|1101\rangle + |1110\rangle + 2\,|0000\rangle) \otimes |1_r\rangle\right]\ , \tag{3.61}$$

$$|\phi_A\rangle = \sqrt{\frac{\gamma}{1-\gamma}}\left[\frac{1}{\sqrt{2}N_1}(|0001\rangle - |0010\rangle) \otimes |0_r\rangle + \frac{1}{\sqrt{2}N_2}(|1101\rangle - |1110\rangle) \otimes |1_r\rangle\right]\ . \tag{3.62}$$

Their contributions to the entanglement fidelity are

$$\mathcal{F}'^1_{\text{ent}} = \mathcal{F}'^{1,S}_{\text{ent}} + \mathcal{F}'^{1,A}_{\text{ent}} = \frac{1}{4}\frac{\gamma}{1-\gamma}\left[(\frac{\sqrt{3}}{N_1} + \frac{\sqrt{3}}{N_2})^2 + (\frac{1}{N_1} + \frac{1}{N_2})^2\right] = \gamma - \gamma^2 + O(\gamma^3)\ . \tag{3.63}$$

To summarize, we get the entanglement fidelity up to $O(\gamma^3)$,

$$\mathcal{F}_{\text{ent}} = \mathcal{F}^0_{\text{ent}} + 2\mathcal{F}^1_{\text{ent}} + \mathcal{F}'^1_{\text{ent}} = 1 - \frac{7}{4}\gamma^2 + O(\gamma^3)\ , \tag{3.64}$$

which is the same as the worst-case fidelity, and better than Liang's code in [100] with an analytical quantum error recovery.

3.6 Generalizations

As observed in the 4-qubit NSA codes, the optimized code structure introduces asymmetry into the traditional self-complementary codes which admit a basis of codewords of the form $|\bar{\psi}_u\rangle = (|u\rangle + |\tilde{u}\rangle)/\sqrt{2}$, where for each qubit i, $\tilde{u}_i = 1 - u_i$. The key insight from the

machine learned self-complementary codes is that allowing different coefficients for $|u\rangle$ and its complement, adjusted based on the noise strength, enhances code performance.

Expanding on this concept, we extend the formalism to general $(\!(n,k)\!)$ NSA SC codes[1], building upon an existing family of $(\!(n,k)\!)$ self-complementary codes designed to correct $t=1$ amplitude damping noise [88], and arrive at the following structure for the codeword:

$$|\bar{\psi}_u\rangle = (1-\gamma)^{-\|u\|/2}|u\rangle + (1-\gamma)^{-\|\tilde{u}\|/2}|\bar{u}\rangle/N_u \ , \tag{3.65}$$

where N_u is the normalization, and for a given bit string u, $\|u\|$ denotes the Hamming weight. Notice that each basis coefficient is a function of the noise strength γ and the basis itself. We then compute the fidelity for the NSA self-complementary codes and find

$$\mathcal{F}_{\rm SC} = 1-(n^2-n)\,\gamma^2/4+\mathcal{O}(\gamma^3) \ . \tag{3.66}$$

We compare it with the non-NSA codes and find that up to $O(\gamma^2)$ NSA codes achieve a higher fidelity by the amount given by $(2n\lfloor n/2\rfloor - 2\lfloor n/2\rfloor^2)\gamma^2/4$. Note that the performance enhancement of NSA codes grows as the system size increases. In addition, the NSA self-complementary code can be readily extended to qudit systems with a local dimension $q\geq 2$, and for further details, including the fidelity calculations, are provided in section 3.6.2.

We now turn to discuss the generalization of pair-complementary NSA codes. Building on our finding of $(\!(4,1)\!)$ NSA pair-complementary code through machine learning, we show that there exists a new family of $(\!(n,k)\!)$ code to correct a single amplitude damping noise. The codeword takes the following form,

$$\begin{aligned} |\bar{\psi}_u\rangle &\sim |u11\rangle + |u00\rangle - |\tilde{u}10\rangle - |\tilde{u}01\rangle \ ; \\ |\bar{\psi}'_u\rangle &\sim |\tilde{u}00\rangle + |\tilde{u}11\rangle + |u01\rangle + |u10\rangle \ , \end{aligned} \tag{3.67}$$

where all coefficients and normalizations are omitted to emphasize the code structure more clearly. The coefficients, which follow the same form as in Eq. (3.65), depend on both γ and the Hamming weight of the basis. Notice that the codewords of this new family is constructed by extending the basis u in the self-complementary codeword with two additional bits. Specifically, the $|\bar{\psi}_u\rangle$ and $|\bar{\psi}'_u\rangle$ pair contains superpositions of all combinations $\{0,1\}^2$ appending after the self-complementary codeword u and its complementary $\tilde{u}$. Hence, we claim that for any given $(\!(n,k)\!)$ self-complementary code, there exists a corresponding

[1]We assume the largest possible k for a given n qubit systems The fidelity calculations conducted later are also under the same assumption.

$((n+2, k+1))$ pair-complementary code that can correct one amplitude damping error; further details of the calculation and the proof of this claim are provided in Section 3.6.3..

We compute the fidelity of NSA pair-complementary codes for general n, and obtain

$$\mathcal{F}_{\mathrm{PC}} = 1 - (n^2 - 3n + 3)\gamma^2/4 + O(\gamma^3) \,. \tag{3.68}$$

We observe that for any given n, pair-complementary NSA codes achieve higher fidelity than the self-complementary ones, at the cost of fewer logical qubits. Specifically, we establish the following inequality,

$$\mathcal{F}_{\mathrm{SC}}[((n,k))] > \mathcal{F}_{\mathrm{PC}}[((n+2,k+1))] > \mathcal{F}_{\mathrm{SC}}[((n+2,k+1))] \tag{3.69}$$

The formalism of NSA can be further extended to the bosonic codes [102]. For the simplest 0-2-4 binomial code, with the code states: $|\bar{0}'\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |4\rangle)$, $|\bar{1}'\rangle = |2\rangle$,

it can be generalized to the NSA version, with $|\bar{0}'\rangle$ becoming

$$|\bar{0}\rangle = \left(|0\rangle + (1-\gamma)^{-2}|4\rangle\right) / \sqrt{1 + (1-\gamma)^{-4}} \,, \tag{3.70}$$

while $|\bar{1}\rangle = |2\rangle$ remains. Note that this code can be reduced from the $((4,1))$ qubit code by summing up the bit strings in each basis. Hence, the behavior of the fidelity is also the same, which is $1 - 3\gamma^2 + \mathcal{O}(\gamma^3)$ for the NSA code, as compared to $1 - 5\gamma^2 + \mathcal{O}(\gamma^3)$ in the non-NSA code.

3.6.1 Fidelity of self-complementary qubit NSA code generalization

In this subsection, we show the details of how to generalize the $((4,1))$ NSA amplitude damping (AD) code to a family of self-complementary NSA AD codes, and compare the fidelity difference between the NSA and non-NSA AD codes for arbitrary n.

As shown in [88, 124], for a self-complementary $((n,k))$ code $C = \mathrm{span}\{(|u\rangle + |\tilde{u}\rangle)/\sqrt{2} \mid u \in S\}$ can correct $t = 1$ amplitude damping errors if and only if the set $S \subset \{0,1\}^n$ satisfies:

(a) If $u \in S$, then $\bar{u} \in S$.

(b) Error spaces from different u must not overlap.

Note that we consider largest possible k given a n in the self-complementary code for the following calculations. A non-NSA self-complementary AD code can be written as,

$$|\bar{\psi}_u\rangle = \frac{1}{\sqrt{2}}\left(|u\rangle + |\tilde{u}\rangle\right) \ , \tag{3.71}$$

where $\tilde{u}_i = 1 - u_i$ for every single qubit site. The AD error that has A_0 acting on every single qubit transforms an input state $|\bar{\psi}_{\rm in}\rangle = \sum_u c_u \left(|u\rangle + |\tilde{u}\rangle\right)/\sqrt{2}$ into

$$\sum_a c_u \left(\sqrt{\frac{(1-\gamma)^{\|u\|}}{2}}|u\rangle + \sqrt{\frac{(1-\gamma)^{\|\tilde{u}\|}}{2}}|\tilde{u}\rangle \right) \ , \tag{3.72}$$

where $\|u\|$ denotes the Hamming weight of the code word basis $|u\rangle$. The contribution to the worst-case fidelity is then dominated by the case where the number of 0s and 1s in the bit string u is balanced. Thus

$$\mathcal{F}_0^{\text{non-NSA}} = \frac{(1-\gamma)^{\lfloor n/2 \rfloor} + (1-\gamma)^{n-\lfloor n/2 \rfloor}}{2} \ . \tag{3.73}$$

Consider an AD error that has A^1 acting only on one qubit i and A^0 acting on the rest of the qubits, denoted as A_{1_i}. It transforms the input state into,

$$\frac{1}{\sqrt{2}}\left(\delta_{1,u_i}(1-\gamma)^{(\|u\|-1)/2}\gamma^{1/2}|u-e_i\rangle + \delta_{1,\tilde{u}_i}(1-\gamma)^{(\|\tilde{u}\|-1)/2}\gamma^{1/2}|\tilde{u}-e_i\rangle\right) \ , \tag{3.74}$$

where e_i denotes a bit-string with 1 at the i-th bit and 0 on the others. Note that after the A^1 noise, only one of the two bases in Eq. (3.74) survives, whose weight takes the minimum value when u contains all zeros or all ones. This gives the contribution of the worst-case fidelity

$$\mathcal{F}_1^{\text{non-NSA}} = \frac{(1-\gamma)^{(n-1)}\,\gamma}{2} \ . \tag{3.75}$$

Hence, the worst-case fidelity is,

$$\mathcal{F}^{\text{non-NSA}} = \mathcal{F}_0^{\text{non-NSA}} + n\mathcal{F}_1^{\text{non-NSA}} = 1 - (n^2 - n + 2n\lfloor n/2 \rfloor - 2\lfloor n/2 \rfloor^2)\gamma^2/4 + \mathcal{O}(\gamma^3) \ . \tag{3.76}$$

Now, we generalize the non-NSA self-complementary codeword to the following NSA self-complementary codeword for arbitrary n,

$$|\bar{\psi}_u\rangle = \frac{(1-\gamma)^{-\|u\|/2}|u\rangle + (1-\gamma)^{-\|\tilde{u}\|/2}|\bar{u}\rangle}{\sqrt{(1-\gamma)^{-\|u\|} + (1-\gamma)^{-\|\tilde{u}\|}}} \ . \tag{3.77}$$

Similarly, consider an input state that is a random superposition of all logical qubits. The AD error A_0 transforms the input state into,

$$\sum_u c_u \frac{|u\rangle + |\tilde{u}\rangle}{\sqrt{(1-\gamma)^{-\|u\|} + (1-\gamma)^{-\|\tilde{u}\|}}} \,. \tag{3.78}$$

Its contribution to the worst-case fidelity is

$$\mathcal{F}_0^{\mathrm{NSA}} = \min_u \left\{ \frac{2}{(1-\gamma)^{-\|u\|} + (1-\gamma)^{-(n-\|u\|)}} \right\} = \frac{2}{1+(1-\gamma)^{-n}} \tag{3.79}$$

Using the mean inequality, we find that when u is the all-zero bit string (or equivalently, the all-one bit string), the minimum can be achieved.

And the AD error A_{1_i} transforms the input state into,

$$\sum_u c_u \, (1-\gamma)^{-1/2} \, \gamma^{1/2} \, \frac{\delta_{1,u_i}|u-e_i\rangle + \delta_{1,\tilde{u}_i}|\tilde{u}-e_i\rangle}{\sqrt{(1-\gamma)^{-\|u\|} + (1-\gamma)^{-\|\tilde{u}\|}}} \,. \tag{3.80}$$

Its contribution to the worst-case fidelity is

$$\mathcal{F}_1^{\mathrm{NSA}} = \min_u \left\{ \frac{\gamma \, (1-\gamma)^{-1}}{(1-\gamma)^{-\|u\|} + (1-\gamma)^{-(n-\|u\|)}} \right\} = \frac{\gamma}{(1-\gamma)^{1-n} + (1-\gamma)} \tag{3.81}$$

As above, using the mean inequality, we find that when u contains all zeros or all ones, the minimum can be achieved.

The total contribution to the worst-case fidelity hence is,

$$\mathcal{F}^{\mathrm{NSA}} = \mathcal{F}_0^{\mathrm{NSA}} + n\mathcal{F}_1^{\mathrm{NSA}} = 1 - (n^2-n)\,\gamma^2/4 + \mathcal{O}(\gamma^3) \,. \tag{3.82}$$

The fidelity for the NSA self-complementary codes is shown to be better than the non-NSA self-complementary codes, and their difference is $(2n\lfloor n/2 \rfloor - 2\lfloor n/2 \rfloor^2)/4$ in the order of γ^2. Notice that the advantage of NSA self-complementary code grows as the qubit number n increases.

3.6.2 Self-complementary qudit NSA code generalization

In this subsection, we discuss the generalization to the NSA self-complementary *qudit* code with local dimension $q \geq 2$. We start with an example of $n = 4$ qutrit codes. Then, we show the definition of the amplitude damping noise for qudit systems, and detail the extension

to the non-NSA self-complementary qudit codes from the non-NSA self-complementary qubit codes. Afterwards, we show how we can generalize the non-NSA self-complementary qudit codes to the NSA self-complementary qudit codes. Along the way, we show how to compute the worst-case fidelities for both non-NSA and NSA self-complementary qudit codes, and argue that NSA self-complementary qudit codes have better performance.

We start with an example of $n = 4$, $q = 3$ amplitude damping code. We first introduce the non-NSA version of the code and then move to the NSA version, and calculate their fidelities respectively. For $q = 3$, the amplitude damping operator takes the form,

$$A_0 = |0\rangle\langle 0| + \sqrt{1-\gamma}|1\rangle\langle 1| + (1-\gamma)|2\rangle\langle 2| \ , \tag{3.83}$$

$$A_1 = \sqrt{\gamma}|0\rangle\langle 1| + \sqrt{2\gamma(1-\gamma)}|1\rangle\langle 2| \ , \tag{3.84}$$

$$A_2 = \gamma|0\rangle\langle 2| \ . \tag{3.85}$$

The non-NSA codewords are the following,

$$\begin{aligned} |\bar{0}\rangle &= \sqrt{\frac{1}{3}}(|0000\rangle + |1111\rangle + |2222\rangle) \ ; \\ |\bar{1}\rangle &= \sqrt{\frac{1}{3}}(|0011\rangle + |1122\rangle + |2200\rangle) \ ; \\ |\bar{2}\rangle &= \sqrt{\frac{1}{3}}(|0022\rangle + |1100\rangle + |2211\rangle) \ . \end{aligned} \tag{3.86}$$

Consider an input state $|\bar{\psi}\rangle = a_0|\bar{0}\rangle + a_1|\bar{1}\rangle + a_2|\bar{2}\rangle$. The output states after error operators

A_{0000}, A_{1000} and A_{2000} respectively are,

$$
\begin{aligned}
|\psi_{0000}\rangle = a_0 & \left[\frac{|0000\rangle + (1-\gamma)^2|1111\rangle + (1-\gamma)^4|2222\rangle}{\sqrt{3}}\right] \\
& + a_1 \left[\frac{[(1-\gamma)|0011\rangle + (1-\gamma)^3|1122\rangle + (1-\gamma)^2|2200\rangle]}{\sqrt{3}}\right] \\
& + a_2 \left[\frac{[(1-\gamma)^2|0022\rangle + (1-\gamma)|1100\rangle + (1-\gamma)^3|2211\rangle]}{\sqrt{3}}\right] ; \qquad (3.87) \\
|\psi_{1000}\rangle = a_0 & \left[\frac{\sqrt{\gamma}(1-\gamma)^{3/2}|0111\rangle + \sqrt{2\gamma}(1-\gamma)^{7/2}|1222\rangle}{\sqrt{3}}\right] \\
& + a_1 \left[\frac{\sqrt{\gamma}(1-\gamma)^{5/2}|0122\rangle + \sqrt{2\gamma}(1-\gamma)^{3/2}|1200\rangle}{\sqrt{3}}\right] \\
& + a_2 \left[\frac{\sqrt{\gamma}(1-\gamma)^{1/2}|0100\rangle + \sqrt{2\gamma}(1-\gamma)^{5/2}|1211\rangle}{\sqrt{3}}\right] ; \qquad (3.88) \\
|\psi_{2000}\rangle = a_0 & \left[\frac{\gamma(1-\gamma)^3|0222\rangle}{\sqrt{3}}\right] \\
& + a_1 \left[\frac{\gamma(1-\gamma)|0200\rangle}{\sqrt{3}}\right] \\
& + a_2 \left[\frac{\gamma(1-\gamma)^2|0211\rangle}{\sqrt{3}}\right] . \qquad (3.89)
\end{aligned}
$$

Note that there is no overlap between the error subspaces, enabling us to distinguish different errors by a projective measurement. The error correction procedure is almost identical to section 3.5.1, and the total worse-case fidelity would be

$$
\begin{aligned}
\mathcal{F}_{4_\text{qutrit}}^{\text{non-NSA SC}} &= ((1-\gamma)^2 + (1-\gamma)^4 + (1-\gamma)^6)/3 + 4 \times (\gamma(1-\gamma)^3 + 2\gamma(1-\gamma)^7)/3 \\
&\quad + 4 \times (\gamma^2(1-\gamma)^6)/3 \\
&= 1 - 14\gamma^2 + O(\gamma^3) . \qquad (3.90)
\end{aligned}
$$

Now, we move on to NSA case. Inspired by the qubit case, we can choose the logical

states to be the following,

$$
\begin{aligned}
|\bar{0}\rangle &= \frac{|0000\rangle + (1-\gamma)^{-2}|1111\rangle + (1-\gamma)^{-4}|2222\rangle}{\sqrt{1+(1-\gamma)^{-4}+(1-\gamma)^{-8}}} \; ; \\
|\bar{1}\rangle &= \frac{(1-\gamma)^{-1}|0011\rangle + (1-\gamma)^{-3}|1122\rangle + (1-\gamma)^{-2}|2200\rangle}{\sqrt{(1-\gamma)^{-2}+(1-\gamma)^{-6}+(1-\gamma)^{-4}}} \; ; \\
|\bar{2}\rangle &= \frac{(1-\gamma)^{-2}|0022\rangle + (1-\gamma)^{-1}|1100\rangle + (1-\gamma)^{-3}|2211\rangle}{\sqrt{(1-\gamma)^{-2}+(1-\gamma)^{-6}+(1-\gamma)^{-4}}} \; .
\end{aligned} \tag{3.91}
$$

After affected by noise $A_{0000},\ A_{1000},\ A_{2000}$,

$$
\begin{aligned}
|\phi_{0000}\rangle =& a_0 \left[\frac{|0000\rangle + |1111\rangle + |2222\rangle}{\sqrt{1+(1-\gamma)^{-4}+(1-\gamma)^{-8}}}\right] \\
&+ a_1 \left[\frac{|0011\rangle + |1122\rangle + |2200\rangle}{\sqrt{(1-\gamma)^{-2}+(1-\gamma)^{-6}+(1-\gamma)^{-4}}}\right] \\
&+ a_2 \left[\frac{|0022\rangle + |1100\rangle + |2211\rangle}{\sqrt{(1-\gamma)^{-2}+(1-\gamma)^{-6}+(1-\gamma)^{-4}}}\right] \; ;
\end{aligned} \tag{3.92}
$$

$$
\begin{aligned}
|\phi_{1000}\rangle =& a_0 \left[\frac{\gamma^{1/2}|0111\rangle + (2\gamma)^{1/2}|1222\rangle}{\sqrt{(1-\gamma)+(1-\gamma)^{-3}+(1-\gamma)^{-7}}}\right] \\
&+ a_1 \left[\frac{\gamma^{1/2}|0122\rangle + (2\gamma)^{1/2}|1200\rangle}{\sqrt{(1-\gamma)^{-1}+(1-\gamma)^{-5}+(1-\gamma)^{-3}}}\right] \\
&+ a_2 \left[\frac{\gamma^{1/2}|0100\rangle + (2\gamma)^{1/2}|1211\rangle}{\sqrt{(1-\gamma)^{-1}+(1-\gamma)^{-5}+(1-\gamma)^{-3}}}\right] \; ;
\end{aligned} \tag{3.93}
$$

$$
\begin{aligned}
|\phi_{2000}\rangle =& a_0 \left[\frac{\gamma\,|0222\rangle}{\sqrt{(1-\gamma)^{2}+(1-\gamma)^{-2}+(1-\gamma)^{-6}}}\right] \\
&+ a_1 \left[\frac{\gamma\,|0200\rangle}{\sqrt{1+(1-\gamma)^{-4}+(1-\gamma)^{-2}}}\right] \\
& a_2 \left[\frac{\gamma\,|0211\rangle}{\sqrt{(1+(1-\gamma)^{-4}+(1-\gamma)^{-2}}}\right] \; .
\end{aligned} \tag{3.94}
$$

The error correction procedure is also very similar to section 3.5.1, and the total worse-case

fidelity would be

$$\begin{aligned}\mathcal{F}_{4_\text{qutrit}}^{\text{NSA SC}} &= \frac{3}{1+(1-\gamma)^{-4}+(1-\gamma)^{-8}} + 4\times\frac{3\gamma}{(1-\gamma)+(1-\gamma)^{-3}+(1-\gamma)^{-7}} \\ &\quad + 4\times\frac{\gamma^2}{(1-\gamma)^2+(1-\gamma)^{-2}+(1-\gamma)^{-6}} \\ &= 1-10\gamma^2+\mathcal{O}(\gamma^3)\ . \end{aligned} \tag{3.95}$$

With the examples of $n=4$ qutrit codes, we now move to the general qudit case. Now, consider the amplitude damping noises for general qudit systems that take the following form,

$$A^l = \sum_{a=l}^{q-1}\sqrt{\binom{a}{l}}\sqrt{(1-\gamma)^{a-l}\gamma^l}\,|a-l\rangle\langle a|\ , \tag{3.96}$$

where l indicates the level of jumps in the AD noises.

Inspired by the non-NSA self-complementary qubit codes, we define a new family of non-NSA self-complementary qudit codes with local dimension q as $\mathfrak{C}_d = \text{span}\{\sum_{a=0}^{q-1}\left|u^{\uparrow_a}\right\rangle/\sqrt{q}|u\in S_d\}$ for $S_d\subset\{0,\cdots,(q-1)\}^n$, where $\left|u^{\uparrow_a}\right\rangle = |(u+a^{\otimes n})\bmod q\rangle$ and the module is taken at every qudit site. Here, we continue to consider weight $t=1$ amplitude damping errors, where there is at most one non-A_0 acting on all qudits. Code $\mathfrak{C}_d$ can correct $t=1$ AD errors if and only if the set S_d satisfies:

(a) if $u\in S_d$, then $u^{\uparrow_a}\in S_d$, $\forall\, a\in\{0,\cdots,(q-1)\}$;

(b) error spaces from different u must not overlap [88, 52].

The codeword takes the form of,

$$|\bar{\psi}_u\rangle = \frac{1}{\sqrt{q}}\sum_{a=0}^{q-1}\left|u^{\uparrow_a}\right\rangle\ . \tag{3.97}$$

Notice that a enumerates each basis in the codeword. Consider the following input state,

$$|\bar{\psi}_{\text{in}}\rangle = \frac{1}{\sqrt{q}}\sum_u c_u\sum_{a=0}^{q-1}|u^{\uparrow_a}\rangle\ . \tag{3.98}$$

An amplitude damping error with only A^0 acting on every single qudit changes the input state to,

$$\frac{1}{\sqrt{q}}\sum_u c_u \sum_{a=0}^{q-1}(1-\gamma)^{\|u^{\uparrow_a}\|/2}|u^{\uparrow_a}\rangle \ . \tag{3.99}$$

Its contribution to the worst-case fidelity is taken from the codeword that has the smallest coefficients. For error of all A_0, the codeword that has the smallest coefficient is the one where $u^{\uparrow_0}$ is half being 0 and the rest being 1.

$$\min_u \left\{ \frac{\sum_{a=0}^{q-1}(1-\gamma)^{\|u^{\uparrow_a}\|}}{q} \right\} = \min_u \left[1 - \frac{1}{q}\left(\sum_{a=0}^{q-1} \|u^{\uparrow_a}\| \right)\gamma + \frac{1}{2q}\left(\sum_{a=0}^{q-1} \left(-\|u^{\uparrow_a}\| + \|u^{\uparrow_a}\|^2 \right) \right)\gamma^2 + \cdots \right] \ . \tag{3.100}$$

Notice that every dit in $u^{\uparrow_a}$ is a permutation of $\{0,1,...,q-1\}$ as a runs from 0 to $q-1$, hence the summation of weight is a constant

$$\sum_a \|u^{\uparrow_a}\| = nq(q-1)/2 \ . \tag{3.101}$$

Thus, finding the minimum of Eq.(3.100) is equivalent to finding the minimum of $\sum_{a=0}^{q-1}\|u^{\uparrow_a}\|^2$. For simplicity of the calculation, we assume that n is divisible by q. By the mean inequality we can minimize the summation of the square by taking $\|u^{\uparrow_a}\|$ to be independent of a, which can be achieved by taking u with each of the digits from 0 to $q-1$ appearing n/q times, thus $\|u^{\uparrow_a}\| = n(q-1)/2$, and we obtain the minimum to be

$$\min_u \sum_a \|u^{\uparrow_a}\|^2 = n^2 q(q-1)^2/4 \ . \tag{3.102}$$

By substituting the minimum quantity into Eq. (3.100), we obtain its contribution to the fidelity,

$$\mathcal{F}_0 = 1 - \frac{1}{2}n(q-1)\gamma + \frac{1}{8}n(q-1)(n(q-1)-2)\gamma^2 + O(\gamma^3) \ . \tag{3.103}$$

Let's now consider the case where an amplitude damping error that evolves jump of $|u^{\uparrow_a} - e_i^{(l)}\rangle\langle u^{\uparrow_a}|$ where $u_i^{\uparrow_a} \geq l$ denotes the i-th dit in the dit string $u^{\uparrow_a}$. We define $e_i^{(l)}$ to be a dit string that is all 0 except that the i-th dit is of value l. It effectively represent the AD error which has A^l is acting on qudit site i and A^0 acting on all the rest qudits. After the error, the input state becomes,

$$\frac{1}{\sqrt{q}}\sum_u c_u \sum_{a=0}^{q-1}\sqrt{\binom{u_i^{\uparrow_a}}{l}}\,(1-\gamma)^{(\|u^{\uparrow_a}\|-l)/2}\gamma^{l/2}\,|u^{\uparrow_a} - e_i^{(l)}\rangle \ . \tag{3.104}$$

Since we assumed that different error spaces do not overlap, we can use projective measurement to distinguish the noise state after any $e_i^{(l)}$. The contribution to the worst-case fidelity comes from a codeword of smallest coefficient, and is,

$$\min_u \left\{ \frac{1}{q} \sum_{a=0}^{q-1} \binom{u_i^{\uparrow_a}}{l} (1-\gamma)^{(\|u^{\uparrow_a}\|-l)} \gamma^l \right\} . \tag{3.105}$$

In the case where $l = 1$,

$$\begin{aligned} \mathcal{F}_1 &= \min_u \left\{ \frac{1}{q} \sum_{a=0}^{q-1} u_i^{\uparrow_a} (1-\gamma)^{(\|u^{\uparrow_a}\|-1)} \gamma \right\} \\ &= \min_u \left\{ \frac{1}{q} \sum_{a=0}^{q-1} u_i^{\uparrow_a} \gamma - \frac{1}{q} \max_u \sum_{a=0}^{q-1} u_i^{\uparrow_a} (\|u^{\uparrow_a}\| - 1)\gamma^2 + O(\gamma^3) \right\} \qquad (3.106) \\ &= \frac{q-1}{2}(\gamma + \gamma^2) - \frac{\gamma^2}{q} \max_u \sum_{a=0}^{q-1} u_i^{\uparrow_a} \|u^{\uparrow_a}\| + O(\gamma^3) . \qquad (3.107) \end{aligned}$$

the minimization is taken when $\sum_{a=0}^{q-1} u_i^{\uparrow_a} \|u^{\uparrow_a}\|$ is taken to be maximal. Note that

$$\sum_{a=0}^{q-1} u_i^{\uparrow_a} \|u^{\uparrow_a}\| = \sum_j \sum_{a=0}^{q-1} u_i^{\uparrow_a} u_j^{\uparrow_a} \leq n \sum_{a=0}^{q-1} a^2 = n \frac{(q-1)q(2q-1)}{6} . \tag{3.108}$$

where by the rearrangement inequality, the maximum is taken when $u_i^{\uparrow_a} = u_j^{\uparrow_a}$ for any i, j and a, i.e. u is a dit string where all dits are the same. Hence,

$$\mathcal{F}_1 = \frac{q-1}{2}\gamma + \frac{(q-1)\,(3 - n(2q-1))}{6}\gamma^2 + O(\gamma^3) . \tag{3.109}$$

In the case where $l = 2$,

$$\mathcal{F}_2 = \sum_{a=2}^{q-1} \binom{a}{2} \gamma^2 = \frac{q(q-1)(q-2)}{6}\gamma^2 + O(\gamma^3) . \tag{3.110}$$

where we get rid of the dependence of u because when we sum over a, $u_i^{\uparrow_a}$ is just a permutation of integers from 0 to $q-1$. Moreover, note that for $l > 2$, its contribution to the worst-case fidelity is already on the order of $O(\gamma^3)$; thus, these terms can be safely neglected.

The total worst-case fidelity for non-NSA code is

$$\mathcal{F}^{\text{non-NSA}} = \mathcal{F}_0 + n\mathcal{F}_1 + n\mathcal{F}_2 = 1 - \frac{1}{24}\left(n(q-1)(2-4q+n(5q-1))\right)\gamma^2 + O(\gamma^3)\ . \quad (3.111)$$

For the generalization to NSA qudit amplitude damping code, we follow the similar analysis above to calculate the worst case fidelity. The general form of the NSA codeword with local dimension q is,

$$|w\rangle = \frac{\sum_{a=0}^{q-1}(1-\gamma)^{(-\|u^{\uparrow a}\|)/2}}{\sqrt{\sum_{a=0}^{q-1}(1-\gamma)^{-\|u^{\uparrow a}\|}}}\,|u^{\uparrow a}\rangle\ , \quad (3.112)$$

For error with only the damping A^0, a general input state becomes,

$$\sum_u c_u \frac{|u^{\uparrow a}\rangle}{\sqrt{\sum_{a=0}^{q-1}(1-\gamma)^{-\|u^{\uparrow a}\|}}}\ , \quad (3.113)$$

and contribution to the worst case fidelity is the minimal coefficient of the codeword,

$$\min_u\left\{\frac{q}{\sum_{a=0}^{q-1}(1-\gamma)^{-\|u^{\uparrow a}\|}}\right\} = \frac{q}{\sum_{a=0}^{q-1}(1-\gamma)^{-a\cdot n}}\ . \quad (3.114)$$

Now consider the amplitude damping errors evolving jumps l. The input state becomes,

$$\sum_u c_u \frac{\sum_{a=0}^{q-1}\sqrt{\binom{u_i^{\uparrow a}}{l}}\,\gamma^{l/2}(1-\gamma)^{-l/2}}{\sqrt{\sum_{a=0}^{q-1}(1-\gamma)^{-\|u^{\uparrow a}\|}}}\,|u^{\uparrow a} - e_i^{(l)}\rangle\ . \quad (3.115)$$

Its contribution to the worst case fidelity is,

$$\min_u\left\{\frac{\sum_{a=0}^{q-1}\binom{u_i^{\uparrow a}}{l}\gamma^l(1-\gamma)^{-l}}{\sum_{a=0}^{q-1}(1-\gamma)^{-\|u^{\uparrow a}\|}}\right\} = \frac{\sum_{a=0}^{q-1}\binom{a}{l}\gamma^l(1-\gamma)^{-l}}{\max_u\sum_{a=0}^{q-1}(1-\gamma)^{-\|u^{\uparrow a}\|}} = \frac{\sum_{a=0}^{q-1}\binom{a}{l}\gamma^l(1-\gamma)^{-l}}{\sum_{a=0}^{q-1}(1-\gamma)^{-a\cdot n}}\ . \quad (3.116)$$

From the rearrangement inequality, the smallest coefficient here corresponds to the codeword in which u is all 0. And the total worst case fidelity is,

$$\begin{aligned}\mathcal{F}^{\text{NSA}} &= \frac{q}{\sum_{a=0}^{q-1}(1-\gamma)^{-a\cdot n}} + n\times\sum_{l=1}^{q-1}\frac{\sum_{a=0}^{q-1}\binom{a}{l}\gamma^l(1-\gamma)^{-l}}{\sum_{a=0}^{q-1}(1-\gamma)^{-a\cdot n}}\\ &= 1 - \frac{1}{12}(q-1)(2q-1)(n^2-n)\gamma^2 + O(\gamma^3)\ . \end{aligned} \quad (3.117)$$

Hence, we find the difference between NSA and non-NSA fidelities,

$$\Delta\mathcal{F} = \frac{1}{24}n^2(q^2-1)\gamma^2 \tag{3.118}$$

Below we display fidelities up to second order in γ for several different q of general n for both NSA and non-NSA self-complementary codes.

q	$F^{\text{NSA SC}}$	$F^{\text{non-NSA SC}}$
3	$1-\frac{5}{6}(n^2-n)\gamma^2$	$1-\frac{1}{12}(14n^2-10n)\gamma^2$
4	$1-\frac{7}{4}(n^2-n)\gamma^2$	$1-\frac{1}{8}(19n^2-14n)\gamma^2$
5	$1-3(n^2-n)\gamma^2$	$1-\frac{1}{6}(24n^2-18n)\gamma^2$
6	$1-\frac{55}{12}(n^2-n)\gamma^2$	$1-\frac{5}{24}(29n^2-22n)\gamma^2$
7	$1-\frac{13}{2}(n^2-n)\gamma^2$	$1-\frac{1}{4}(34n^2-26n)\gamma^2$

3.6.3 Fidelity of pair-complementary qubit NSA code generalization

In this subsection, we show the details of how togeneralize the $(\!(4,1)\!)$ pair-complementary NSA code to a family of pair-complementary NSA codes, and calculate the worst-case fidelity achievable without residual logical errors for arbitrary n.

By inspecting the codewords of the $(\!(4,1)\!)$ pair-complementary NSA code, one observes that the basis states are paired with each other. Thus we consider the codewords of the following form:

$$\begin{aligned}\left|\bar{\psi}_u^{i,j}\right\rangle &\sim |u\rangle + |u-e_i-e_j\rangle - |\tilde{u}+e_i\rangle - |\tilde{u}+e_j\rangle \ ; \\ \left|\bar{\psi}_u'^{i,j}\right\rangle &\sim |u-e_i\rangle + |u-e_j\rangle + |\tilde{u}+e_i+e_j\rangle + |\tilde{u}\rangle \ ,\end{aligned} \tag{3.119}$$

where $e_i \in \{0,1\}^n$ denotes the error vector with 1 at the i-th bit and 0 on the others. To highlight the structure of this code more clearly, we have omitted the coefficients in the equations above. Specifically, each computational basis state $|u\rangle$ should include a factor of $(1-\gamma)^{-\|u\|/2}$ before it, where $\|u\|$ denotes the Hamming weight of the bit-string u.

We then try to find the set of $\{u,i,j\}$ such that the Knill-Laflamme condition can be approximately satisfied. To simplify notation, we define $S_{u,i,j} = \{u, u-e_i, u-e_j, u-e_i-e_j\}$

in the following, and use $\tilde{S}_{u,i,j} = \{\tilde{u}, \tilde{u} + e_i, \tilde{u} + e_j, \tilde{u} + e_i + e_j\}$ to denote its element-wise complement. The pairwise orthogonality condition of codewords then imposes the following constraint:

(C1) $(S_{u,i,j} \cup \tilde{S}_{u,i,j}) \cap (S_{v,k,l} \cup \tilde{S}_{v,k,l}) \neq \varnothing$, iff $u = v$ and $\{i, j\} = \{k, l\}$.

To assure that no confusion between different $\{u, i, j\}$ arise after one damping, we will also require

(C2) Error spaces from different $\{u, i, j\}$ must not overlap.

Since for each triplet $\{u, i, j\}$, the corresponding pair-complementary code words $\left|\bar{\psi}_u^{i,j}\right\rangle$ and $\left|\bar{\psi}_u'^{i,j}\right\rangle$ contains superpositions of all combinations of 0 and 1s on the i-th and j-th qubits, for our condition (C2) it is only relevant to consider code word on the other $n-2$ qubits. Similarily, for another triplet $\{v, k, l\}$ it is only relevant to consider code word on $n-2$ qubits other than the k-th and l-th qubits. To maximize the code rate, it is then natural to pick our code spaces such that for any triplet, $\{u, i, j\}$, the set $\{i, j\}$ are the same. This inspires us to construct the following family of pair-complementary NSA codes.

Proposition 2. *For any $((n, k))$ self-complementary code*

$$\mathfrak{C}_1 = \text{span}\{(|u\rangle + |\tilde{u}\rangle)/\sqrt{2} \mid u \in S\} , \tag{3.120}$$

we can construct a $((n+2, k+1))$ pair-complementary code

$$\mathfrak{C}_2 = \text{span}\{\left|\bar{\psi}_u\right\rangle, \left|\bar{\psi}_u'\right\rangle\} \mid u \in S/\mathbb{Z}_2\} , \tag{3.121}$$

where

$$\begin{aligned} |\bar{\psi}_u\rangle &\sim |u00\rangle + |u11\rangle - |\tilde{u}10\rangle - |\tilde{u}01\rangle ; \\ |\bar{\psi}_u'\rangle &\sim |u01\rangle + |u10\rangle + |\tilde{u}11\rangle + |\tilde{u}00\rangle . \end{aligned} \tag{3.122}$$

Here the quotient set $S/\mathbb{Z}_2$ is defined such that if $u \in S/\mathbb{Z}_2 \subset S$ then $\tilde{u} \notin S/\mathbb{Z}_2$, i.e. the quotient set only contains one element in each complementary pairs u and $\tilde{u}$. $u00$ denotes a bit string starting with u and ending with 00, with similar conventions for the other terms.

Proof. To show that our construction gives an AQEC code for one amplitude damping error, we verify that it satisfies the condition C1 and C2, and demonstrate the error correction process and the worst case fidelity for it.

To verify C1, notice that for $S_{u,i,j} \cup \tilde{S}_{u,i,j}$, the first $n-2$ bits in the bit strings are either u or $\tilde{u}$, thus $(S_{u,i,j} \cup \tilde{S}_{u,i,j}) \cap (S_{v,k,l} \cup \tilde{S}_{v,k,l}) = \varnothing$ iff $u = v$ or $u = \tilde{v}$. Since $u, v \in S/\mathbb{Z}_2$, $\tilde{v} \notin S/\mathbb{Z}_2$, so $u = v$.

To verify C2, consider code states generated from different bit strings u and v, notice that if the errors happen in the last 2 qubits, then since $u \neq v$ and $u \neq \tilde{v}$, the error spaces are guaranteed to not overlap with each other. If the errors do not happen in the last 2 qubits, then the non-overlap requirement is equivalent to saying that for $u, \tilde{u}, v, \tilde{v}$, there is no confusion arises assuming the decay occurs at no more than one qubit.

This is exactly the condition for $\mathfrak{C}_1$ to be an AQEC code for a single AD noise. □

It is worth mentioning that when we take $u = 00$, the above codeword corresponds to the $((4, 1))$ NSA pair-complementary code in Eq. 3.13.

We now demonstrate the error correction process for the pair-complementary code. Consider the following codewords of $((n, k))$ pair-complementary code:

$$\begin{aligned}
\left|\bar{\psi}_u\right\rangle = \frac{1}{N_u}\Big((1-\gamma)^{-\frac{\|u\|}{2}}\left|u00\right\rangle + (1-\gamma)^{-(\frac{\|u\|+2}{2})}\left|u11\right\rangle \\
- (1-\gamma)^{-\frac{n-\|u\|-1}{2}}\left|\tilde{u}10\right\rangle - (1-\gamma)^{-\frac{n-\|u\|-1}{2}}\left|\tilde{u}01\right\rangle\Big)\ ; \\
\left|\bar{\psi}'_u\right\rangle = \frac{1}{N'_u}\Big((1-\gamma)^{-\frac{n-\|u\|}{2}}\left|\tilde{u}11\right\rangle + (1-\gamma)^{-\frac{n-\|u\|-2}{2}}\left|\tilde{u}00\right\rangle \\
+ (1-\gamma)^{-\frac{\|u\|+1}{2}}\left|u01\right\rangle + (1-\gamma)^{-\frac{\|u\|+1}{2}}\left|u10\right\rangle\Big)\ .
\end{aligned} \tag{3.123}$$

Here we use the superscript u to label the pairs of logical states generated by the codeword $(|u\rangle + |\bar{u}\rangle / \sqrt{2})$ of the $((n-2, k-1))$ self-complementary code. And similarly as before, N_u and N'_u are normalization factors, and $\bar{u}$ is the complement of the bit-string u, where for each qubit site i, we have $\tilde{u}_i = 1 - u_i$.

For an input state $\left|\bar{\psi}_{\text{in}}\right\rangle = \sum_u a_u \left|\bar{\psi}_u\right\rangle + b_u \left|\bar{\psi}'_u\right\rangle$, there are three kinds of output states after $t = 1$ amplitude damping, occurring with probabilities of order $O(\gamma)$ or higher. Firstly, an amplitude damping error with only A^0 acting on every single qubit changes the input state to,

$$|\phi_0\rangle = \sum_u \frac{a_u}{N_u}\left(|u00\rangle + |u11\rangle - |\tilde{u}10\rangle - |\tilde{u}01\rangle\right) + \frac{b_u}{N'_u}\left(|\tilde{u}11\rangle + |\tilde{u}00\rangle + |u01\rangle + |u10\rangle\right)\ . \tag{3.124}$$

Secondly, for an amplitude damping error with operator A^1 acting on qubit site i, where $1 \leq i \leq n-2$, and operator A^0 acting on all other qubits, the input state transforms as:

$$|\phi_{1,i}\rangle = \begin{cases} \sqrt{\frac{\gamma}{1-\gamma}} \sum_u \left[\frac{a_u}{N_u}(|u'00\rangle + |u'11\rangle) + \frac{b_u}{N'_u}(|u'01\rangle + |u'10\rangle)\right]\ , \text{ if } u_i = 1\ ; \\ \sqrt{\frac{\gamma}{1-\gamma}} \sum_u \left[\frac{-a_u}{N_u}(|\tilde{u}'10\rangle + |\tilde{u}'01\rangle) + \frac{b_u}{N'_u}(|\tilde{u}'11\rangle + |\tilde{u}'00\rangle)\right]\ , \text{ if } u_i = 0\ . \end{cases} \tag{3.125}$$

Note that this state contains half as many terms as $|\phi_0\rangle$. We denote by u' the bit-string obtained by changing the bit at site i from 1 to 0, and similarly define $\tilde{u}'$.

Finally, for an amplitude damping error with operator A^1 acting on one of the last two qubit sites and operator A^0 acting on all remaining qubits, the resulting output states are:

$$|\phi_{\ldots 10}\rangle = \sqrt{\frac{\gamma}{1-\gamma}} \sum_u \left[\frac{a_u}{N_u} \left(|u01\rangle - |\tilde{u}00\rangle \right) + \frac{b_u}{N'_u} \left(|\tilde{u}01\rangle + |u00\rangle \right) \right] , \tag{3.126}$$

$$|\phi_{\ldots 01}\rangle = \sqrt{\frac{\gamma}{1-\gamma}} \sum_u \left[\frac{a_u}{N_u} \left(|u10\rangle - |\tilde{u}00\rangle \right) + \frac{b_u}{N'_u} \left(|\tilde{u}10\rangle + |u00\rangle \right) \right] . \tag{3.127}$$

Note that, similarly to the second case discussed above, these states contain half as many terms as $|\phi_0\rangle$. Furthermore, as discussed in Section 3.5.2, there is a large overlap between the error words $|\phi_{\ldots 10}\rangle$ and $|\phi_{\ldots 01}\rangle$. We therefore consider the classical mixture of these two error states and adopt the following set of projective measurements:

$$\begin{aligned} P_0 =& \frac{1}{4} \left(|u00\rangle - |\tilde{u}10\rangle - |\tilde{u}01\rangle + |u11\rangle \right) \left(\langle u00| - \langle \tilde{u}10| - \langle \tilde{u}01| + \langle u11| \right) \\ &+ \frac{1}{4} \left(|\tilde{u}11\rangle + |u01\rangle + |u10\rangle + |\tilde{u}00\rangle \right) \left(\langle \tilde{u}11| + \langle u01| + \langle u10| + \langle \tilde{u}00| \right) , \end{aligned} \tag{3.128}$$

$$P_{1,i} = \begin{cases} \frac{1}{2} \left(|u'00\rangle + |u'11\rangle \right) \left(\langle u'00| + \langle u'11| \right) + \frac{1}{2} \left(|u'01\rangle + |u'10\rangle \right) \left(\langle u'01| + \langle u'10| \right) , & \text{if } u_i = 0 ; \\ \frac{1}{2} \left(|\tilde{u}'10\rangle + |\tilde{u}'01\rangle \right) \left(\langle \tilde{u}'10| + \langle \tilde{u}'01| \right) + \frac{1}{2} \left(|\tilde{u}'11\rangle + |\tilde{u}'00\rangle \right) \left(\langle \tilde{u}'11| + \langle \tilde{u}'00| \right) , & \text{if } u_i = 1 , \end{cases} \tag{3.129}$$

$$\begin{aligned} P_S =& \frac{1}{6} \left(|u01\rangle + |u10\rangle - 2\,|\tilde{u}00\rangle \right) \left(\langle u01| + \langle u10| - 2\,\langle \tilde{u}00| \right) \\ &+ \frac{1}{6} \left(|\tilde{u}01\rangle + |\tilde{u}10\rangle + 2\,|u00\rangle \right) \left(\langle \tilde{u}01| + \langle \bar{u}10| + 2\,\langle u00| \right) , \end{aligned} \tag{3.130}$$

$$P_A = \frac{1}{2} \left(|u01\rangle - |u10\rangle \right) \left(\langle u01| - \langle u10| \right) + \frac{1}{2} \left(|\tilde{u}01\rangle - |\tilde{u}10\rangle \right) \left(\langle \tilde{u}01| - \langle \tilde{u}10| \right) . \tag{3.131}$$

We continue to calculate the worst-case fidelity achievable without residual logical errors. To ensure clarity, we introduce the following notation. For simplicity of calculation, we denote the logical states by $|\bar{\psi}_\lambda\rangle$, labeled by index λ, which have the form:

$$|\bar{\psi}_\lambda\rangle = \frac{1}{N_\lambda} \sum_\alpha \pm (1-\gamma)^{-\|\lambda_\alpha\|/2} \, |\lambda_\alpha\rangle . \tag{3.132}$$

Here, the index α labels the individual computational-basis components of the logical state, N_λ is the normalization factor, and some bases carry alternating signs. If the noisy

states are projectively measured to be in the subspace of P_0, after the error correction, the contribution to the fidelity is given by

$$\mathcal{F}_0 = 4 \min_{\lambda} N_{\lambda}^{-2}, \tag{3.133}$$

which can be expanded in a Taylor series up to order $O(\gamma^2)$.

$$\begin{aligned} N_{\lambda}^{-2} &= \left[\sum_{\alpha}(1-\gamma)^{-\|\lambda_\alpha\|}\right]^{-1} = \left[4 + \left(\sum_{\alpha}\|\lambda_\alpha\|\right)\gamma + \frac{1}{2}\left(\sum_{\alpha}\|\lambda_\alpha\| + \sum_{\alpha}(\|\lambda_\alpha\|)^2\right)\gamma^2 + \cdots\right]^{-1} \\ &= \frac{1}{4}\left[1 - \frac{\sum_\alpha \|\lambda_\alpha\|}{4}\gamma + \left(\frac{\left(\sum_\alpha \|\lambda_\alpha\|\right)^2}{16} - \frac{\sum_\alpha \left(\|\lambda_\alpha\| + (\|\lambda_\alpha|)^2\right)}{8}\right)\gamma^2\right] + \cdots . \end{aligned} \tag{3.134}$$

Note that from the structure of the pair-complementary codewords given in Eq. (3.123), it follows that $\sum_\alpha \|\lambda_\alpha\| = 2n$ for any basis logical state. Therefore, the leading contribution to worst-case fidelity is

$$\mathcal{F}_0 = 1 - \frac{n}{2}\gamma + \left(\frac{n^2}{4} - \frac{2n + \max_\lambda \left[\sum_\alpha (\|\lambda_\alpha|)^2\right]}{8}\right)\gamma^2 + O(\gamma^3) . \tag{3.135}$$

Note that within each pair of the logical states in the pair-complementary codes, i.e. consider Eq. (3.31) for a specific u, the sum of squared Hamming weights in each computational basis is the same, i.e. the quantity $\sum_\alpha (\|\lambda_\alpha\|)^2$ is equal. Thus to find the maximum, we only need to identify the bit-string u that maximizes the following quantity $S(u)$,

$$S(u) = \sum_{\alpha}(\|\lambda_\alpha\|)^2 = \left[(\|u\|)^2 + (\|u\| + 2)^2 + 2(n - \|u\| - 1)^2\right] . \tag{3.136}$$

This maximum occurs when u is the all-zero bit-string (or equivalently, the all-one bit-string), resulting in the maximum value:

$$\max_u S(u) = 2^2 + 2(n-1)^2 = 2n^2 - 4n + 6 . \tag{3.137}$$

Thus, we arrive at

$$\mathcal{F}_0 = 1 - \frac{n}{2}\gamma + \frac{n-3}{4}\gamma^2 + O(\gamma^3) . \tag{3.138}$$

We can similarly obtain the contribution from the projection $P_{1,i}$:

$$\mathcal{F}_1 = \frac{2\gamma}{1-\gamma} \min_{\lambda} N_{\lambda}^{-2} = \frac{\gamma}{2} + (1 - \frac{n}{2})\frac{\gamma^2}{2} + O(\gamma^3) . \tag{3.139}$$

Finally, we can determine the contributions from the projections P_S and P_A:

$$\mathcal{F}_1' = \mathcal{F}_{1,S}' + \mathcal{F}_{1,A}' = (3+1)\cdot\frac{\gamma}{1-\gamma}\min_{\lambda} N_\lambda^{-2} = \gamma + (1-\frac{n}{2})\gamma^2 + O(\gamma^3)\ . \tag{3.140}$$

Summarizing these results, we obtain the worst-case fidelity up to the order $O(\gamma^3)$:

$$\mathcal{F} = F_0 + (n-2)\mathcal{F}_1 + \mathcal{F}_1' = 1 - \frac{n^2-3n+3}{4}\gamma^2 + O(\gamma^3)\ , \tag{3.141}$$

which is higher than the worst-case fidelity Eq. (3.82) of the $((n,k))$ self-complementary NSA code.

3.6.4 Recovery and worst-case fidelity of binomial NSA codes

The binomial code has been proposed to use a single bosonic mode for boson loss and gain errors. In this subsection, we show that one can have an NSA generalization of the binomial code with improved fidelity. Since we focus on a single atom loss error in this chapter, we consider the simplest version of the binomial code, i.e. the 0-2-4 code, which can correct the amplitude damping noise with $\ell \leq 1$, where the Kraus operators are as follows:

$$A^\ell = \left(\frac{\gamma}{1-\gamma}\right)^{\ell/2}\frac{a^l}{\sqrt{\ell!}}(1-\gamma)^{\hat{n}/2}\ . \tag{3.142}$$

The code states are as follows:

$$\begin{aligned} |\bar{0}\rangle &= \frac{1}{\sqrt{2}}(|0\rangle + |4\rangle)\ ; \\ |\bar{1}\rangle &= |2\rangle\ . \end{aligned} \tag{3.143}$$

Note that this can be thought of as being reduced from the $((4,1))_2$ qubit code in [91] by summing up the bit strings in each basis. Following our NSA generalization of the 4-qubit self-complementary code, we can generalize the 0-2-4 code to the NSA version, with the code states:

$$|\bar{0}'\rangle = \sqrt{\frac{1}{1+(1-\gamma)^{-4}}}\left(|0\rangle + (1-\gamma)^{-2}|4\rangle\right)\ ,\quad |\bar{1}'\rangle = |2\rangle\ . \tag{3.144}$$

The recovery and fidelity of the (non-)NSA 0-2-4 code is also similar to their (non-)NSA 4-qubit counterpart, which we state below for concreteness.

For the non-NSA 0-2-4 code, we consider the input qubit state,

$$|\bar{\psi}_{\text{in}}\rangle = a|\bar{0}\rangle + b|\bar{1}\rangle \ , \tag{3.145}$$

after the amplitude damping noise A^0 or A^1 all possible final states will be:

$$|\phi_0\rangle = a\frac{|0\rangle + (1-\gamma)^2|4\rangle}{\sqrt{2}} + b(1-\gamma)|2\rangle \ ; \tag{3.146}$$

$$|\phi_1\rangle = \sqrt{\frac{2\gamma}{1-\gamma}}(a(1-\gamma)^2|3\rangle + b(1-\gamma)|1\rangle) \ . \tag{3.147}$$

The syndrome measurement could be done by a parity measurement of the boson number, and the fidelity after the recovery will be

$$\mathcal{F} = \min\left\{\frac{1+(1-\gamma)^4}{2}, (1-\gamma)^2\right\} + \frac{2\gamma}{1-\gamma}\min\{(1-\gamma)^4, (1-\gamma)^2\} = 1-5\gamma^2+O(\gamma^3) \tag{3.148}$$

For the NSA 0-2-4 code, we consider the input qubit state,

$$|\bar{\psi}_{\text{in}}\rangle = a|\bar{0}'\rangle + b|\bar{1}'\rangle \ , \tag{3.149}$$

after the amplitude damping noise A^0 or A^1 all possible final states will be:

$$|\phi_0\rangle = a\frac{|0\rangle + |4\rangle}{\sqrt{1+(1-\gamma)^{-4}}} + b(1-\gamma)|2\rangle \ , \tag{3.150}$$

$$|\phi_1\rangle = \sqrt{\frac{2\gamma}{1-\gamma}}(\sqrt{\frac{2}{1+(1-\gamma)^{-4}}}a|3\rangle + b(1-\gamma)|1\rangle) \ . \tag{3.151}$$

The syndrome measurement could also be done by a parity measurement of the boson number, and the fidelity after the recovery will be

$$\mathcal{F} = \min\left\{\frac{2}{1+(1-\gamma)^{-4}}, (1-\gamma)^2\right\} + \frac{2\gamma}{1-\gamma}\min\left\{\frac{2}{1+(1-\gamma)^{-4}}, (1-\gamma)^2\right\} = 1-3\gamma^2+O(\gamma^3) \ , \tag{3.152}$$

which is higher than the non-NSA counterpart as in Eq. (3.148).

3.7 Discussions

Building on observations from machine learning methods, we formalized a novel framework for AQEC codes featuring noise strength adaptation, which enables advantages over

conventional codes. Specifically, by employing a hybrid quantum-classical learning algorithm, we identified two $(\!(4,1)\!)$ noise-strength-adapted AQEC codes tailored for amplitude damping noise, with performance surpassing its previously known non-NSA counterpart. The NSA framework generalizes to two families of codes for arbitrary system size, showcasing its scalability. Moreover, it extends to AQEC for single-photon loss through the NSA version of the 0-2-4 binomial code. These codes are all non-stabilizer codes.

Our NSA-based codes optimize code performance by directly incorporating the actual noise strength into the design process, opening a new avenue for code design. Our findings show that NSA codes consistently outperform their non-NSA counterparts, highlighting the advantages of adapting codes to specific noise environments. This suggests that quantum error correction and fault tolerance procedures may significantly benefit from noise-strength adaptation. Future investigations could explore how the NSA principle may extend to other noise models or hybridize with existing QEC paradigms to enhance performance under real-world conditions. Furthermore, loss functions based on different measures [155, 100] can be employed, which may lead to other structures of NSA codes and inspire new AQEC studies.

In this study, the enhanced performance of our NSA codes relies on the constraint that no residual logical errors remain after correction. Nevertheless, we anticipate that our NSA framework can also be effectively applied to scenarios where this constraint is not strictly necessary. Additionally, we generalized the NSA self-complementary code to qudit systems and expect similar generalizations to be feasible for pair-complementary codes.

Beyond NSA codes, our ML-based methodology provides a systematic approach for exploring quantum error correction codes. In particular, for codes with well-defined algebraic structures—such as stabilizer codes— machine learning methods, including reinforcement learning, could be effective in identifying new codes [106, 129, 154]. By applying these techniques to structured families of quantum codes, future research may uncover novel error correction strategies that would be difficult to find through conventional analytical approaches.

References

[1] S. L. Adler. Einstein gravity as a symmetry-breaking effect in quantum field theory. *Rev. Mod. Phys.*, 54:729, 1982. [Erratum: Rev.Mod.Phys. 55, 837 (1983)].

[2] A. Aggarwal, A. Castro, and S. Detournay. Warped symmetries of the kerr black hole. *JHEP*, 01:016, 2020.

[3] P. Aichelburg and R. Sexl. On the gravitational field of a massless particle. *Gen. Rel. Grav.*, 2:303–312, 1971.

[4] M. Alishahiha, A. Davody, A. Naseh, and S. F. Taghavi. On butterfly effect in higher derivative gravities. *JHEP*, 11:032, 2016.

[5] A. Almheiri, X. Dong, and D. Harlow. Bulk locality and quantum error correction in ads/cft. *JHEP*, 04:163, 2015.

[6] A. Almheiri, N. Engelhardt, D. Marolf, and H. Maxfield. The entropy of bulk quantum fields and the entanglement wedge of an evaporating black hole. *JHEP*, 12:063, 2019.

[7] Ahmed Almheiri, Donald Marolf, Joseph Polchinski, and James Sully. Black Holes: Complementarity or Firewalls? *JHEP*, 02:062, 2013.

[8] A. Averin. Entropy counting from a schwarzschild/cft correspondence and soft hair. *Phys. Rev. D*, 101(4):046024, 2020.

[9] Y. BenTov and J. Swearngin. Gravitational shockwaves on rotating black holes. *Gen. Rel. Grav.*, 51:25, 2019.

[10] Cédric Bény and Ognyan Oreshkov. General Conditions for Approximate Quantum Error Correction and Near-Optimal Recovery Channels. *Phys. Rev. Lett.*, 104(12):120501, March 2010.

[11] P. Betzios, N. Gaddam, and O. Papadoulaki. The black hole s-matrix from quantum mechanics. *JHEP*, 11:131, 2016.

[12] M. S. Blok, V. V. Ramasesh, T. Schuster, K. O'Brien, J. M. Kreikebaum, D. Dahlen, A. Morvan, B. Yoshida, N. Y. Yao, and I. Siddiqi. Quantum information scrambling on a superconducting qutrit processor. *Phys. Rev. X*, 11:021010, 2021.

[13] Dolev Bluvstein, Simon J. Evered, Alexandra A. Geim, Sophie H. Li, Hengyun Zhou, Tom Manovitz, Sepehr Ebadi, Madelyn Cain, Marcin Kalinowski, Dominik Hangleiter, J. Pablo Bonilla Ataides, Nishad Maskara, Iris Cong, Xun Gao, Pedro Sales Rodriguez, Thomas Karolyshyn, Giulia Semeghini, Michael J. Gullans, Markus Greiner, Vladan Vuletić, and Mikhail D. Lukin. Logical quantum processor based on reconfigurable atom arrays. *Nature*, 626(7997):58–65, December 2023.

[14] H. Bondi, M. van der Burg, and A. Metzner. Gravitational waves in general relativity. 7. waves from axisymmetric isolated systems. *Proc. Roy. Soc. Lond. A*, A269:21–52, 1962.

[15] Fernando G. S. L. Brandão, Elizabeth Crosson, M. Burak Şahinoğlu, and John Bowen. Quantum error correcting codes in eigenstates of translation-invariant spin chains. *Phys. Rev. Lett.*, 123:110502, Sep 2019.

[16] Adam R. Brown, Hrant Gharibyan, Geoff Penington, and Leonard Susskind. The Python's Lunch: geometric obstructions to decoding Hawking radiation. *JHEP*, 08:121, 2020.

[17] L. S. Brown and J. C. Collins. Dimensional renormalization of scalar field theory in curved space-time. *Annals Phys.*, 130:215, 1980.

[18] R.-G. Cai and J. B. Griffiths. Null particle solutions in three-dimensional (anti-)de sitter spaces. *J. Math. Phys.*, 40:3465–3475, 1999.

[19] M. Campanelli and C. Lousto. Exact gravitational shock wave solution of higher order theories. *Phys. Rev. D*, 54:3854–3860, 1996.

[20] Chenfeng Cao, Chao Zhang, Zipeng Wu, Markus Grassl, and Bei Zeng. Quantum variational learning for quantum error-correcting codes. *Quantum*, 6:828, October 2022.

[21] S. Carlip. Black hole entropy from conformal field theory in any dimension. *Phys. Rev. Lett.*, 82:2828–2831, 1999.

[22] S. Carlip. Entropy from conformal field theory at killing horizons. *Class. Quant. Grav.*, 16:3327–3348, 1999.

[23] S. Carlip. Black hole entropy from bondi-metzner-sachs symmetry at the horizon. *Phys. Rev. Lett.*, 120(10):101301, 2018.

[24] A. Castro, A. Maloney, and A. Strominger. Hidden conformal symmetry of the kerr black hole. *Phys. Rev. D*, 82:024008, 2010.

[25] N. Cerf, Gerd Leuchs, and E. Polzik. *Quantum Information With Continuous Variables of Atoms and Light*. Imperial College Press, 02 2007.

[26] Venkatesa Chandrasekaran, Geoff Penington, and Edward Witten. Large N algebras and generalized entropy. *JHEP*, 04:009, 2023.

[27] L.-Q. Chen, W. Z. Chua, S. Liu, A. J. Speranza, and B. d. S. L. Torres. Virasoro hair and entropy for axisymmetric killing horizons. *Phys. Rev. Lett.*, 125:241302, 2020.

[28] Luca Chirolli and Guido Burkard. Decoherence in solid-state qubits. *Advances in Physics*, 57(3):225–285, 2008.

[29] G. Compère and A. Fiorucci. Advanced lectures on general relativity, 2018.

[30] Claude Crépeau, Daniel Gottesman, and Adam D Smith. Approximate quantum error-correcting codes and secret sharing schemes. In *Eurocrypt*, volume 3494, pages 285–301. Springer, 2005.

[31] C. Crnkovic and E. Witten. Covariant description of canonical formalism in geometrical theories. In *Three Hundred Years of Gravitation*, pages 676–684. Cambridge University Press, 1987.

[32] X. Dong, D. Harlow, and A. C. Wall. Reconstruction of bulk operators within the entanglement wedge in gauge-gravity duality. *Phys. Rev. Lett.*, 117:021601, 2016.

[33] L. Donnay, G. Giribet, H. A. Gonzalez, and M. Pino. Supertranslations and superrotations at the black hole horizon. *Phys. Rev. Lett.*, 116(9):091101, 2016.

[34] L. Donnay, G. Giribet, H. A. González, and M. Pino. Extended symmetries at the black hole horizon. *JHEP*, 09:100, 2016.

[35] L. Donnay, G. Giribet, H. A. González, and A. Puhm. Black hole memory effect. *Phys. Rev. D*, 98:124016, 2018.

[36] W. Donnelly and L. Freidel. Local subsystems in gauge theory and gravity. *JHEP*, 09:102, 2016.

[37] T. Dray and G. 't Hooft. The effect of spherical shells of matter on the schwarzschild black hole. *Commun. Math. Phys.*, 99:613–625, 1985.

[38] T. Dray and G. 't Hooft. The gravitational shock wave of a massless particle. *Nucl. Phys. B*, 253:173–188, 1985.

[39] J. Dressel, J. González Alonso, M. Waegell, and N. Yunger Halpern. Strengthening weak measurements of qubit out-of-time-order correlators. *Phys. Rev. A*, 98(1):012132, 2018.

[40] Sourav Dutta, Debjyoti Biswas, and Prabha Mandayam. Noise-adapted qudit codes for amplitude-damping noise. *arXiv e-prints*, page arXiv:2406.02444, June 2024.

[41] Sourav Dutta, Aditya Jain, and Prabha Mandayam. Smallest quantum codes for amplitude damping noise. *arXiv e-prints*, page arXiv:2410.00155, September 2024.

[42] Bryan Eastin and Emanuel Knill. Restrictions on transversal encoded quantum gate sets. *Phys. Rev. Lett.*, 102:110502, Mar 2009.

[43] Netta Engelhardt and Aron C. Wall. Quantum Extremal Surfaces: Holographic Entanglement Entropy beyond the Classical Regime. *JHEP*, 01:073, 2015.

[44] Philippe Faist, Sepehr Nezami, Victor V. Albert, Grant Salton, Fernando Pastawski, Patrick Hayden, and John Preskill. Continuous symmetries and approximate quantum error correction. *Phys. Rev. X*, 10:041018, Oct 2020.

[45] Nava Gaddam and Nico Groenenboom. Soft graviton exchange and the information paradox. *Phys. Rev. D*, 109(2):026007, 2024.

[46] Nava Gaddam, Nico Groenenboom, and Gerard 't Hooft. Quantum gravity on the black hole horizon. *JHEP*, 01:023, 2022.

[47] Ping Gao, Daniel Louis Jafferis, and Aron C. Wall. Traversable Wormholes via a Double Trace Deformation. *JHEP*, 12:151, 2017.

[48] H. Geng. Some information theoretic aspects of de-sitter holography. *JHEP*, 02:005, 2020.

[49] H. Geng. Non-local entanglement and fast scrambling in de-sitter holography. *Ann Phys*, 426:168402, 2021.

[50] S. B. Giddings and V. S. Rychkov. Black holes from colliding wavepackets. *Phys. Rev. D*, 70:104026, 2004.

[51] Daniel Gottesman. *Stabilizer Codes and Quantum Error Correction*. PhD thesis, California Institute of Technology, 1997.

[52] Markus Grassl, Linghang Kong, Zhaohui Wei, Zhang-Qi Yin, and Bei Zeng. Quantum Error-Correcting Codes for Qudit Amplitude Damping. *arXiv e-prints*, page arXiv:1509.06829, September 2015.

[53] Nikodem Grzesiak, Reinhold Blümel, Kenneth Wright, Kristin M. Beck, Neal C. Pisenti, Ming Li, Vandiver Chaplin, Jason M. Amini, Shantanu Debnath, Jwo-Sy Chen, and Yunseong Nam. Efficient arbitrary simultaneously entangling gates on a trapped-ion quantum computer. *Nature Communications*, 11(1):2963, June 2020.

[54] S. S. Gubser, Igor R. Klebanov, and Alexander M. Polyakov. Gauge theory correlators from noncritical string theory. *Phys. Lett. B*, 428:105–114, 1998.

[55] S. S. Gubser, S. S. Pufu, and A. Yarom. Entropy production in collisions of gravitational shock waves and of heavy ions. *Phys. Rev. D*, 78:066014, 2008.

[56] M. Gärttner, J. G. Bohnet, A. Safavi-Naini, M. L. Wall, J. J. Bollinger, and A. M. Rey. Measuring out-of-time-order correlations and multiple quantum spectra in a trapped-ion quantum magnet. *Nature Physics*, 13:781–786, 2017.

[57] S. Haco, S. W. Hawking, M. J. Perry, and A. Strominger. Black hole entropy and soft hair. *JHEP*, 12:098, 2018.

[58] N. Yunger Halpern, B. Swingle, and J. Dressel. Quasiprobability behind the out-of-time-ordered correlator. *Phys. Rev. A*, 97:042105, 2018.

[59] D. Harlow and J. q. Wu. Covariant phase space with boundaries. *JHEP*, 10:146, 2020.

[60] Daniel Harlow and Hirosi Ooguri. Symmetries in quantum field theory and quantum gravity. *Communications in Mathematical Physics*, 383:1669–1804, 2021.

[61] S. W. Hawking. Breakdown of Predictability in Gravitational Collapse. *Phys. Rev. D*, 14:2460–2473, 1976.

[62] S. W. Hawking, M. J. Perry, and A. Strominger. Soft hair on black holes. *Phys. Rev. Lett.*, 116:231301, 2016.

[63] S. W. Hawking, M. J. Perry, and A. Strominger. Superrotation charge and supertranslation hair on black holes. *JHEP*, 05:161, 2017.

[64] P. Hayden and J. Preskill. Black holes as mirrors: quantum information in random subsystems. *JHEP*, 09:120, 2007.

[65] Patrick Hayden, Sepehr Nezami, Sandu Popescu, and Grant Salton. Error correction of quantum reference frame information. *PRX Quantum*, 2:010326, Feb 2021.

[66] Patrick Hayden and Geoffrey Penington. Learning the Alpha-bits of Black Holes. *JHEP*, 12:007, 2019.

[67] P. Hosur, X.-L. Qi, D. A. Roberts, and B. Yoshida. Chaos in quantum channels. *JHEP*, 02:004, 2016.

[68] M. Hotta and M. Tanaka. Shock wave geometry with nonvanishing cosmological constant. *Class. Quant. Grav.*, 10:307–314, 1993.

[69] W.-H. Huang. Butterfly velocity in quadratic gravity. *Class. Quant. Grav.*, 35(19):195004, 2018.

[70] W.-H. Huang. Holographic butterfly velocities in brane geometry and einstein-gauss-bonnet gravity with matters. *Phys. Rev. D*, 97:066020, 2018.

[71] Veronika E. Hubeny, Mukund Rangamani, and Tadashi Takayanagi. A Covariant holographic entanglement entropy proposal. *JHEP*, 07:062, 2007.

[72] J. i. Koga. Asymptotic symmetries on killing horizons. *Phys. Rev. D*, 64:124012, 2001.

[73] Mohsin Iqbal, Nathanan Tantivasadakarn, Ruben Verresen, Sara L. Campbell, Joan M. Dreiling, Caroline Figgatt, John P. Gaebler, Jacob Johansen, Michael Mills, Steven A. Moses, Juan M. Pino, Anthony Ransford, Mary Rowe, Peter Siegfried, Russell P. Stutz, Michael Foss-Feig, Ashvin Vishwanath, and Henrik Dreyer. Non-abelian topological order and anyons on a trapped-ion processor. *Nature*, 626(7999):505–511, February 2024.

[74] V. Iyer and R. M. Wald. Some properties of noether charge and a proposal for dynamical black hole entropy. *Phys. Rev. D*, 50:846–864, 1994.

[75] T. Jacobson. Black hole radiation in the presence of a short distance cutoff. *Phys. Rev. D*, 48:728–741, 1993.

[76] Kristan Jensen, Jonathan Sorce, and Antony J. Speranza. Generalized entropy for general subregions in quantum gravity. *JHEP*, 12:020, 2023.

[77] Peter D. Johnson, Jonathan Romero, Jonathan Olson, Yudong Cao, and Alán Aspuru-Guzik. Qvector: an algorithm for device-tailored quantum error correction, 2017.

[78] I. Khavkine. Covariant phase space, constraints, gauge and the peierls formula. *Int. J. Mod. Phys. A*, 29:1430009, 2014.

[79] Y. Kiem, H. L. Verlinde, and E. P. Verlinde. Black hole horizons and complementarity. *Phys. Rev. D*, 52:7053–7065, 1995.

[80] J. Kirklin. Localisation of soft charges, and thermodynamics of softly hairy black holes. *Class. Quant. Grav.*, 35(17):175010, 2018.

[81] A. Kitaev. Talks at kitp. Unpublished.

[82] A. Kitaev and S. J. Suh. The soft mode in the sachdev-ye-kitaev model and its gravity dual. *JHEP*, 05:183, 2018.

[83] Emanuel Knill and Raymond Laflamme. Theory of quantum error-correcting codes. *Phys. Rev. A*, 55:900–911, Feb 1997.

[84] M. Koloğlu, P. Kravchuk, D. Simmons-Duffin, and A. Zhiboedov. Shocks, superconvergence, and a stringy equivalence principle. *J. High Energy Phys.*, 2020(11):96, 2020.

[85] Linghang Kong and Zi-Wen Liu. Near-optimal covariant quantum error-correcting codes from random unitaries with symmetries. *PRX Quantum*, 3(2):020314, 2022.

[86] Aleksander Kubica and Rafał Demkowicz-Dobrzański. Using quantum metrological bounds in quantum error correction: A simple proof of the approximate eastin-knill theorem. *Phys. Rev. Lett.*, 126:150503, Apr 2021.

[87] K. A. Landsman, C. Figgatt, T. Schuster, N. M. Linke, B. Yoshida, N. Y. Yao, and C. Monroe. Verified quantum information scrambling. *Nature*, 567:61–65, 2019.

[88] Ruitian Lang and Peter W. Shor. Nonadditive quantum error correcting codes adapted to the ampltitude damping channel. *arXiv e-prints*, page arXiv:0712.2586, December 2007.

[89] Martin Larocca, Supanut Thanasilp, Samson Wang, Kunal Sharma, Jacob Biamonte, Patrick J. Coles, Lukasz Cincio, Jarrod R. McClean, Zoë Holmes, and M. Cerezo. A review of barren plateaus in variational quantum computing, 2024.

[90] J. Lee and R. M. Wald. Local symmetries and constraints. *J. Math. Phys.*, 31:725–743, 1990.

[91] Debbie W. Leung, M. A. Nielsen, Isaac L. Chuang, and Yoshihisa Yamamoto. Approximate quantum error correction can lead to better codes. *Physical Review A*, 56(4):2567–2573, October 1997.

[92] J. Li, R. Fan, H. Wang, B. Ye, B. Zeng, H. Zhai, X. Peng, and J. Du. Measuring out-of-time-order correlators on a nuclear magnetic resonance quantum simulator. *Phys. Rev. X*, 7:031011, 2017.

[93] Shuwei Liu and Beni Yoshida. Soft thermodynamics of gravitational shock wave. *Phys. Rev. D*, 105(2):026003, 2022.

[94] Shuwei Liu, Shiyu Zhou, Zi-Wen Liu, and Jinmin Yi. Noise-strength-adapted approximate quantum codes inspired by machine learning, 3 2025.

[95] Zi-Wen Liu and Sisi Zhou. Quantum error correction meets continuous symmetries: fundamental trade-offs and case studies.

[96] Zi-Wen Liu and Sisi Zhou. Approximate symmetries and quantum error correction. *npj Quantum Information*, 9(1):119, November 2023.

[97] D. Lovelock. The einstein tensor and its generalizations. *J. Math. Phys.*, 12:498–501, 1971.

[98] Juan Maldacena, Douglas Stanford, and Zhenbin Yang. Diving into traversable wormholes. *Fortsch. Phys.*, 65(5):1700034, 2017.

[99] Juan Martin Maldacena. The Large N limit of superconformal field theories and supergravity. *Adv. Theor. Math. Phys.*, 2:231–252, 1998.

[100] Xuanhui Mao, Qian Xu, and Liang Jiang. Optimized four-qubit quantum error correcting code for amplitude damping channel.

[101] Donald Marolf and Joseph Polchinski. Gauge/Gravity Duality and the Black Hole Interior. *Phys. Rev. Lett.*, 111:171301, 2013.

[102] Marios H. Michael, Matti Silveri, R. T. Brierley, Victor V. Albert, Juha Salmilehto, Liang Jiang, and S. M. Girvin. New Class of Quantum Error-Correcting Codes for a Bosonic Mode. *Physical Review X*, 6(3):031006, July 2016.

[103] Sepehr Nezami, Henry W. Lin, Adam R. Brown, Hrant Gharibyan, Stefan Leichenauer, Grant Salton, Leonard Susskind, Brian Swingle, and Michael Walter. Quantum Gravity in the Lab. II. Teleportation by Size and Traversable Wormholes. *PRX Quantum*, 4(1):010321, 2023.

[104] Hui Khoon Ng and Prabha Mandayam. Simple approach to approximate quantum error correction based on the transpose channel. *Phys. Rev. A*, 81:062342, Jun 2010.

[105] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[106] Jan Olle, Remmy Zen, Matteo Puviani, and Florian Marquardt. Simultaneous discovery of quantum error correction codes and encoders with a noise-aware reinforcement learning agent. *npj Quantum Information*, 10(1):126, December 2024.

[107] D. N. Page. Information in black hole radiation. *Phys. Rev. Lett.*, 71:3743–3746, 1993.

[108] Don N. Page. Average entropy of a subsystem. *Phys. Rev. Lett.*, 71:1291–1294, 1993.

[109] F. Pastawski, B. Yoshida, D. Harlow, and J. Preskill. Holographic quantum error-correcting codes: toy models for the bulk/boundary correspondence. *JHEP*, 06:149, 2015.

[110] Sabrina Pasterski and Herman Verlinde. HPS meets AMPS: how soft hair dissolves the firewall. *JHEP*, 09:099, 2021.

[111] G. Penington. Entanglement wedge reconstruction and the information paradox. *JHEP*, 09:002, 2020.

[112] J. Polchinski. Chaos in the black hole s-matrix, 2015.

[113] D. A. Roberts, D. Stanford, and A. Streicher. Operator growth in the syk model. *JHEP*, 06:122, 2018.

[114] D. A. Roberts, D. Stanford, and L. Susskind. Localized shocks. *JHEP*, 03:051, 2015.

[115] Daniel A. Roberts and Douglas Stanford. Two-dimensional conformal field theory and the butterfly effect. *Phys. Rev. Lett.*, 115(13):131603, 2015.

[116] Shinsei Ryu and Tadashi Takayanagi. Holographic derivation of entanglement entropy from AdS/CFT. *Phys. Rev. Lett.*, 96:181602, 2006.

[117] R. Sachs. Asymptotic symmetries in gravitational theory. *Phys. Rev.*, 128:2851–2864, 1962.

[118] K. Schoutens, H. L. Verlinde, and E. P. Verlinde. Black hole evaporation and quantum gravity. In *Quantum Aspects of Black Holes*, 1994.

[119] Yasuhiro Sekino and Leonard Susskind. Fast Scramblers. *JHEP*, 10:065, 2008.

[120] K. Sfetsos. On gravitational shock waves in curved space-times. *Nucl. Phys. B*, 436:721–745, 1995.

[121] S. H. Shenker and D. Stanford. Black holes and the butterfly effect. *JHEP*, 03:067, 2014.

[122] S. H. Shenker and D. Stanford. Stringy effects in scrambling. *JHEP*, 05:132, 2015.

[123] Peter W. Shor. Scheme for reducing decoherence in quantum computer memory. *Phys. Rev. A*, 52:R2493–R2496, Oct 1995.

[124] Peter W. Shor, Graeme Smith, John A. Smolin, and Bei Zeng. High performance single-error-correcting quantum codes for amplitude damping. *arXiv e-prints*, page arXiv:0907.5149, July 2009.

[125] John A. Smolin, Graeme Smith, and Stephanie Wehner. Simple family of nonadditive quantum codes. *Phys. Rev. Lett.*, 99:130505, Sep 2007.

[126] A. J. Speranza. Local phase space and edge modes for diffeomorphism-invariant theories. *JHEP*, 02:021, 2018.

[127] K. Stelle. Renormalization of higher derivative quantum gravity. *Phys. Rev. D*, 16:953–969, 1977.

[128] A. Strominger and A. Zhiboedov. Gravitational memory, bms supertranslations and soft theorems. *JHEP*, 01:086, 2016.

[129] Vincent Paul Su, ChunJun Cao, Hong-Ye Hu, Yariv Yanay, Charles Tahan, and Brian Swingle. Discovery of optimal quantum error correcting codes via reinforcement learning.

[130] B. Swingle, G. Bentsen, M. Schleier-Smith, and P. Hayden. Measuring the scrambling of quantum information. *Phys. Rev. A*, 94:040302, 2016.

[131] G. 't Hooft. Graviton dominance in ultrahigh-energy scattering. *Phys. Lett. B*, 198:61–63, 1987.

[132] G. 't Hooft. Strings from gravity. *Phys. Scripta T*, 15:143, 1987.

[133] G. 't Hooft. The black hole interpretation of string theory. *Nucl. Phys. B*, 335:138–154, 1990.

[134] G. 't Hooft. The scattering matrix approach for the quantum black hole: An overview. *Int. J. Mod. Phys. A*, 11:4623–4688, 1996.

[135] E. Verlinde and H. Verlinde. Black hole entanglement and quantum error correction. *JHEP*, 10:107, 2013.

[136] H. L. Verlinde and E. P. Verlinde. Scattering at planckian energies. *Nucl. Phys. B*, 371:246–268, 1992.

[137] W. Vleeshouwers. On bms-transformations and the shock wave s-matrix, 2018. Master thesis.

[138] R. M. Wald. On identically closed forms locally constructed from a field. *J. Math. Phys.*, 31:2378–2384, 1990.

[139] R. M. Wald. Black hole entropy is the noether charge. *Phys. Rev. D*, 48:3427–3431, 1993.

[140] R. M. Wald and A. Zoupas. A general definition of 'conserved quantities' in general relativity and other theories of gravity. *Phys. Rev. D*, 61:084027, 2000.

[141] Aron C. Wall. Maximin Surfaces, and the Strong Subadditivity of the Covariant Holographic Entanglement Entropy. *Class. Quant. Grav.*, 31(22):225007, 2014.

[142] Edward Witten. Anti de Sitter space and holography. *Adv. Theor. Math. Phys.*, 2:253–291, 1998.

[143] Edward Witten. Gravity and the crossed product. *JHEP*, 10:008, 2022.

[144] Mischa P. Woods and Álvaro M. Alhambra. Continuous groups of transversal gates for quantum error correcting codes from finite clock reference frames. *Quantum*, 4:245, March 2020.

[145] Yuxiang Yang, Yin Mo, Joseph M. Renes, Giulio Chiribella, and Mischa P. Woods. Optimal universal quantum error correction via bounded reference frames. *Phys. Rev. Res.*, 4:023107, May 2022.

[146] N. Y. Yao, F. Grusdt, B. Swingle, M. D. Lukin, D. M. Stamper-Kurn, J. E. Moore, and E. A. Demler. Interferometric approach to probing fast scrambling, 2016.

[147] Jinmin Yi, Weicheng Ye, Daniel Gottesman, and Zi-Wen Liu. Complexity and order in approximate quantum error-correcting codes. *Nature Physics*, 20(11):1798–1803, September 2024.

[148] B. Yoshida. Firewalls vs. scrambling. *JHEP*, 10:132, 2019.

[149] B. Yoshida. Soft mode and interior operator in the hayden-preskill thought experiment. *Phys. Rev. D*, 100:086001, 2019.

[150] B. Yoshida. Observer-dependent black hole interior from operator collision. *Phys. Rev. D*, 103:046004, 2021.

[151] B. Yoshida and A. Kitaev. Efficient decoding for the hayden-preskill protocol.

[152] B. Yoshida and N. Y. Yao. Disentangling scrambling and decoherence via quantum teleportation. *Phys. Rev. X*, 9:011006, 2019.

[153] Beni Yoshida. Projective measurement of black holes, 3 2022.

[154] Austin Yubo He and Zi-Wen Liu. Discovering highly efficient low-weight quantum error-correcting codes with reinforcement learning. *arXiv e-prints*, page arXiv:2502.14372, February 2025.

[155] Guo Zheng, Wenhao He, Gideon Lee, and Liang Jiang. Near-optimal performance of quantum error correction codes. *Phys. Rev. Lett.*, 132:250602, Jun 2024.

[156] Sisi Zhou, Zi-Wen Liu, and Liang Jiang. New perspectives on covariant quantum error correction. *Quantum*, 5:521, August 2021.