University of California
Santa Barbara

Traversable Wormholes, The Positivity of Negativity, and How to Gauge Your Tensor Network

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Physics

by

Sean Anthony McBride

Committee in charge:

Professor Xi Dong, Chair
Professor Donald Marolf
Professor Andrea Young

June 2024

The Dissertation of Sean Anthony McBride is approved.

Professor Donald Marolf

Professor Andrea Young

Professor Xi Dong, Committee Chair

June 2024

Traversable Wormholes, The Positivity of Negativity,

and How to Gauge Your Tensor Network

by

Sean Anthony McBride

Mishuk gotal'u meshuroke, pako kyore.

Pressure makes gems, ease makes decay.

—Mandalorian proverb

Acknowledgements

When I was in high school, I was big into the Oscars. In private moments I often fantasized about what my acceptance speech for Best Actor would look like and who I would thank. I guess this will have to do.

This thesis is for the following:

for my advisor, Xi Dong. Too often, I would rush into physics with reckless abandon, emerging with sloppy calculations and half-baked conclusions. I'm indebted to Xi's careful nature for teaching me to think slower, for showing when to simplify and when to complicate, and for finding when to follow the math instead of following your nose. Xi's kindness extends beyond physics; he's never once failed to assist with bureaucratics and would always lend an ear when I asked. I feel fortunate to have had him righting the ship these last six years;

for Don Marolf. Don was my advisor when I arrived at UCSB, after a chance phone call led to a generous offer for summer work. Don guided me through my first publication with pedagogical aplomb, and has continued to educate me and the rest of the high energy community as a professor-cum-Gravity-Lunch-question-answerer;

for my other collaborators Fernando Iniguez, Wayne Weng, Pratik Rath, and Jiuci Xu. I've never written a paper alone, nor have I yet tried, and who would want to when such brilliant people are a door's knock away? I thank all of them for their insights, creativity, and determination;

for Mark van Raamsdonk and Gordon Semenoff, for taking a chance on me with a postdoc at the University of British Columbia, allowing me to continue this crazy journey in academia. Let's hope this dedication ages well;

for Gary Horowitz, a fantastic educator and fantasticer human being. Gary led the charge of gravity lunch for all of grad school, filling the silence with interesting questions

and updates on the broader physics world. He's truly earned the title President of Gravity;

for David Berenstein, who gives great life advice in addition to great, though less understandable, physics advice;

for all the institutes, universities, and resorts which have fostered my growth as a physicist with schools, seminars, and tea times: the Erwin Schrödinger Institute at the University of Vienna, the Galileo Galilei Institute, the Centro de Ciencias de Benasque Pedro Pascal, the IAS, Princeton, Caltech, Berkeley, and UCSC;

for the UCSB postdocs, who from my first year to my last have regaled me with physics knowledge, some of which may have stuck: Henry Maxfield, Joaquin Turiaci, Alexey Milekhin, Amir Tajdini, Kwinten Fransen, Eugenia Colafrancesci, and Maciej Kolanowski;

for the amazing faculty, postdocs, and grad students from far and wide I've had the good luck to interact with at conferences, dinners, and bars: Daniel Jafferis, Kristan Jensen, Jonah Kudler-Flam, Liz Wildenhain, and especially Bob Knighton and Neeraj Tata;

for David Grabovsky. David's unmatched search for rigor complements my own search for the lack thereof, and it's been great having him as an AdS_3 companion when the world has been oh-so four-dimensional;

for Jesse Held, who was never afraid to tell me I'm wrong and has mastered the art of humming two notes to set off a singing frenzy. Here's to inappropriate giggles and boundary conditions;

for Robinson Mancilla, who never fails to peacefully intrude with an interesting question. His refusal to conform to orthodoxy is a Herculean task which he somehow makes look quotidian;

for the rest of the UCSB High Energy Theory community: Brianna Grado-White,

Sergio Hernandez Cuenca, Molly Kaplan, George Hulsey, Xiaoyi Liu, Zi-Yue Wang, Chih-Hung Wu, Zhencheng Wang, Henry Leung, Diandian Wang, and Hiroki Kawai. You've made running HEPJC and bumbling about in the sixth floor lounge a genuinely rewarding intellectual experience. I'm eternally grateful that I had a group so ready to be subjected to my latest rants, and who would rant back to me in kind;

for Tom Hartman, my first theoretical physics mentor, who gleefully (in his own way) led me into the world of high energy physics, in a way where I always felt like an equal when I clearly was not. I can only aspire to write papers as correct yet austerely sardonic as his;

for Ivan Bazarov, who granted me my first real physics research opportunity, then gave it to me again after I dipped out, providing me with a whirlwind four years learning how a real lab operates;

for Luca Cultrera, the wizard of photocathode production, whose faith in an unmotivated freshman taught me how to manage labs, people, and collaborations;

for Liam McAllister, whose physics pedagogy is second to none. I'd be a fish out of water in elecromagnetism, QFT, and string theory without his mélange of mathematical rigor, physical intuition, and blackboard clarity;

for Bob Kaita, who basically gave me free reign of the Princeton Plasma Physics Labs once a week as long as I hit some metal bars with a big hammer;

for Jing Li and Debasis Banerjee, who gave me my first shot at science in a dingy Rutgers chemistry lab: mixing solvents and solutes, baking vials, gassing myself with dibutyl ether, the works;

for Jennifer Martino, who graciously let me pursue my own physics interests in high school in defiance of what I imagine were codified guidelines that students should do assigned homework;

for Ellen LeBlanc, for a similar exercise involving seven unruly calculus students;

for Charles Christoe, for shouting, "Sometimes shit happens!", laying the bedrock for years spent toiling away on engineering projects and in dingy labs;

for Linda Grunthaner, for laying the groundwork for any good engineering education with hand-drawn sketches and strict safety regulations. I never would have started any form of serious scientific research without her training, and I'm grateful for all the times I reconnected with her and with High Tech over the years;

for Kevin Zhou, for showing me the vastness of math and physics as a wee lad, and (as I found out while writing this) for Clark Lyons for emailing me a copy of David Morin's *Introduction to Classical Mechanics*, kickstarting a decades-long love affair;

for MIT OpenCourseWare, Adrian Banner's *The Calculus Lifesaver*, and Linus Pauling's *General Chemistry*, which taught me the value of a library card and the nobility of free education;

for Katherine Lin. Five years ago she mailed me a shovel for digging up penis fish, which just about sums up our friendship. Sorry (not sorry) for farting in your face when we were fourteen, and thank you for opening my mind to a new modality of art every time our paths cross;

for the Cornell Quiz Bowl Team: Amith Punyala, Erial Zheng, James Razumoff, Irene Lin, and Brock Mapes. Thanks for all the seven hour road trips, dingy motels, questionable lakehouse stays, and for being test subjects for Sean McBride's Wild Ride;

for my Swole Bros Craig Pellegrino and Molly Wolfson, whose friendship is too immeasurable to fit in these margins, but for now I thank for all the 7 AM gym sessions, nights watching Seinfeld and playing (read: getting beaten at) Mario Kart, and being the people I turn to first for basically anything. It's tough to find one friend you won't get sick of seeing every day for months on end; I'm so happy to have found two;

for Jenny Smith, for the uncountable Starbucks runs, *Dune* viewings, launch alerts, and dragon fruit educational pamphlets;

for Alex Potts, the first friend I ever made in Santa Barbara. Thanks for the garage gym sessions and education on all things heavy metal and Ohian;

for Franny Setti, the Italian Stallion. Who ever thought that a haircut could change someone's life? Cheers for the Mustang rides, summertime espressos, and rooftop DJ sessions;

for Ari Kaplan, always climbing and only occasionally falling. Thanks for the gymnastics, beach parties, and your stoic sincerity;

for Dima Prokopovich, a true and honest friend, generous with his space, food, and Dr. Manhattan. Thanks for making my life a little more vinyl as I constantly endeavor to make your life a little more electronic;

for Jordan Sinibaldi, who despite my best efforts keeps outsmarting me. Thanks for the late night strolls to Elsies and the racing sessions. I hope one day you'll call me Doctor but I'm not counting on it;

for Sean Beaks, my partner in crime for an unreasonable amount of hobbies, from SCUBA to surfing to climbing to psytrance to ambiguous bass. Not a day goes by in your company without taking in something new;

for Sara Bashore, whose boundless creativity never ceases to inspire me and add much-needed spontaneity to my continued existence. Thanks for the crafts, the snacks, the supremely late hangouts, and for being unfailingly yourself;

for Dillon Cislo, who epitomizes a friend I may not see again for years but the minute we reconnect we'll stay up till 4 AM listening to the Grateful Dead;

for more grad school friends: Hind Al Ali, Shane Bechtel, Mirek Brandt, Tikal Catena, Joe Crowley, Nick Cuccia, Chris Dare, Keegan Downham, Rachel Green, Greta Koumarianou, Madeleine Leibovitch, Eric Liu, Caitlin Patterson, and Malik Tuerkoen;

for more friends: Franny Alcaide, Max Aronow, Jay Casanova, Jackie Castro, Casey Lloyd, Oliver Palmieri, and Bryce West;

for Jared Pagett, my SCUBA sensei. I never would have found this lifelong hobby without his encouragement and leadership. This is for all the 50 degree dives with 3 foot vis, early morning surface swims, joking about pee valves, and Catalina burgers;

for Doug Klug. Doug showed me just how deep the diving world goes, both literally and figuratively. His generosity with his time both for my SCUBA education and organizing his Dives with Doug are unmatched, and I feel fortunate every time I'm in his company.

for the rest of the Paradise Dive Club, a welcome respite from a technology-weary world, where fifty people can gather just to chat about how cool the ocean is, and where all the old heads give me advice on drysuits, photography, and life aboard the Spectre and the Raptor;

for Camilo de Lusa, WookPack, IV Underground, and all the other DJs who, with questionable legality, overtook clearings on Ellwood Beach and East Camino Cielo to deliver free entertainment till the waking dawn;

for Brick Barn Wine Estate, where I've tasted more times than I care to admit, and for Renegade Wines, an unreasonably cozy place to explore a millenia-old tradition;

for all the beds and couches at Little Italy, at Trap House, at Abrego, at Sareaks', at De La Vina which offered me respite in times of need;

for Lighthouse, Dean, Cajé, and Java Station, which fed my caffeine addiction (and often my lox addiction) while I whittled away at a paper or more likely an impromptu chess game;

for that German girl walking through campus at midnight, that Lebanese friend-of-a-friend, Kalon and his wife, and Tom at that fire on East Beach. They don't remember me (and clearly I barely remember them) but I've found more meaning in their fleeting acquaintance than many other core grad school memories;

for Qiao and Tim, my aunt and uncle, without whom I couldn't have invited myself

to Berkeley and Santa Cruz on such a regular basis nor had such a massive stockpile of beans. Thanks for your constant warmth and generosity;

for Santa Barbara, from the mountains to the oceans, from the beaches to the hot springs, from the boulders to the mesas;

for Michael Sipiora, my spiritual mentor. For six years he's informed my life philosophy, not just listening to my woes but educating me on the grand traditions of existentialism, phenomenology, and humanism, with countless reading materials and sage quotes. From politics to academia to Disneyland, I'd emerge from his office with a new and improved perspective on life. I can safely say no one person has had more of an influence on my life than him, with two obvious exceptions.

Last but certainly not least, this thesis is for my parents, Chris and Denise. I've often felt like I'm the sum of their best parts: my mom's gift for gab and love of math, my dad's sense of humor and unnending curiosity. They both valued my education above all else, and were unwavering in their support of my getting a Ph.D from the day I expressed an interest in physics. Through every trial and tribulation they were there, and I thank them both from the bottom of my heart.

Curriculum Vitæ

Sean Anthony McBride

Education

2024	Ph.D. in Physics, University of California, Santa Barbara.
2021	M.A. in Physics, University of California, Santa Barbara.
2018	B.S. in Physics, Cornell University

Publications

X. Dong, **S. McBride**, and W. W. Weng, "Holographic Tensor Networks with Bulk Gauge Symmetries," *JHEP* **02** (2024) 222, arXiv:2309.06436 [hep-th]

S. McBride and F. Iniguez, "Entanglement Negativity Transitions in Chaotic Eigenstates," arXiv:2303.00018 [hep-th]

X. Dong, **S. McBride**, and W. W. Weng, "Replica Wormholes and Holographic Entanglement Negativity," *JHEP* **06** (2022) 094, arXiv:2110.11947 [hep-th]

D. Marolf and **S. McBride**, "Simple Perturbatively Traversable Wormholes from Bulk Fermions," *JHEP* **11** (2019) 037, arXiv:1908.03998 [hep-th]

Abstract

Traversable Wormholes, The Positivity of Negativity,

and How to Gauge Your Tensor Network

by

Sean Anthony McBride

This thesis reports on aspects of semiclassical gravity with an eye towards holography.

Chapter 2 introduces a perturbatively traversable wormhole in a particular four-dimensional quotient spacetime, where traversability is ensured by the Casimir energy of bulk fermions. We compute the fermionic contribution to the integrated null stress-energy tensor and find hints that traversability holds at all times.

Chapters 3 and 4 report on computations of entanglement negativity, a multipartite entanglement measure which distinguishes between classical and quantum correlations. We compute holographic entanglement negativity in a toy model of Jackiw-Teitelboim gravity with end-of-the-world branes, finding a rich phase structure which includes replica symmetry breaking in a large region of phase space. We also compute entanglement negativity in a toy model of chaotic eigenstates, finding some qualitative agreement with phase transitions in the holographic model.

Chapter 5 describes a modification to random tensor networks to incorporate bulk gauge symmetries which we term the "gauged random tensor network." We find an area operator valued in the center of the gauged random tensor network bulk algebra which more closely resembles the area operator provided by the quantum-corrected Ryu-Takayanagi formula.

Contents

Chapter 1

Introduction

We begin this thesis with a definition:

Definition 0: Physics is a predictive framework.

Given a sufficient set of initial conditions for a physical system, one should be able to output the state of that system at any later time. This idealistic picture of physics has no basis in reality, however, as we are constrained by human matters like detector sensitivities in the infrared and the ultraviolet. Thus, any laws of physics we use can only be experimentally verified in an effective sense [1–4].

We hope there exists, in some dark corner of the forest of all knowledge, an absolutely predictive Theory of Everything, expressible in the rigorous *lingua franca* of mathematics. Our only criterion is for this theory to consistently incorporate all observable phenomena. As of the writing of this thesis, this task boils down to incorporating quantum mechanics and general relativity.

Quantum mechanics, and its big brother quantum field theory (QFT), is the theory of the very small. Precision tests of quantum electrodynamics have measured the anomalous magnetic moment of the electron to one part in 10^{-13} [5, 6], and violations of the principles

of quantum mechanics continue to be ruled out by Bell tests [7, 8]. There are certainly many puzzles in the Standard Model of particle physics, but many of these can be resolved by the addition of new gauge fields or matter multiplets without tinkering with the underlying mathematics. This body of evidence suggests that unitarity, in particular the unitary time evolution of quantum mechanics, should be held sacrosanct for the Theory of Everything.

General relativity (GR), on the other hand, is the theory of the very large. When we say "general relativity," we mean that the spacetime metric $g_{\mu\nu}$ is upgraded to a dynamical field, described at low energies by the Einstein-Hilbert action[1]

$$S_{EH} = \frac{1}{16\pi G_N} \int d^4x \sqrt{-g} R. \tag{1.1}$$

This theory is also robust to naïve interference. Tests for violations of the equivalence principle have yielded upper bounds of one part in 10^{15} [9]. Preliminary LIGO tests in the strong gravity regime of black hole mergers have shown no deviation from classical expectations [10], demonstrating that semiclassical gravity (AKA quantum field theory on a curved background) works at experimentally realizable energies. Once again, a century of modern physics has failed to falsify dynamical gravity as a phenomenological truth, so we tack it onto the list.

Unlike the Standard Model, which for all its flaws is a renormalizable quantum field theory, GR as an effective field theory (EFT) is nonrenormalizable, with divergences

[1] Higher derivative corrections to this action should, by the totalitarian principle approach to particle physics (or at least the totalitarian principle I was indoctrinated with), be included in this action suppressed by some string/Planck scale but don't affect the punchline of dynamical gravity.

arising at the energy scale set by the gravitational coupling, the Planck energy[2]

$$E_P = \sqrt{\frac{\hbar c^5}{G_N}} \approx 1.22 \times 10^{19} \text{ GeV}. \tag{1.2}$$

Let's pause to contextualize this energy scale. The Large Hadron Collider's center-of-mass energy currently caps off at 13.6 TeV, and the proposed Future Circular Collider, which would optimistically start collecting data when my great-grandchildren are born, would operate at around 100 TeV. With luck, these mechanical ouroboroi might find an oasis in the current particle physics desert, but can't directly inform us about where to look for modifications to GR. Even inflationary observables, like primordial fluctuations in the CMB power spectrum [12], are too coarse-grained to be useful in the hunt for a UV completion.[3]

We're thus stuck with a theory of gravity which both defies our current mathematical framework and plausibly evades any experimental detection. This tension is precisely why we attack the Theory of Everything from a theoretical vantage point. The rest of this introduction will go over some theories, loosely defined, of quantum gravity, that is frameworks which incorporate gravity as an EFT consistently without abandoning unitarity. This isn't the only approach. We could abandon unitarity and go forth with, say, a unified theory with stochastic dynamics [13], but this is in conflict, albeit a purely philosophical one, with Definition 0! Often we denote our subfield Quantum Gravity™, rather than the arXiV-standardized high energy theoretical physics `[hep-th]`, to emphasize what we want and what we're not willing to lose.

[2]For a more in-depth explanation of this fact, see [11].

[3]This is not to say there don't exist stringy models of inflation, just that the time scales involved (10^{-36} to 10^{-32} seconds) are orders of magnitude separated from the Planck time $t_P = 5.39 \times 10^{-44}$ s, so it's highly unlikely they can elucidate the quantization of gravity.

1.1 String Theory

String theory arose in the 1960s as a happy accident, although one likes to think that, had Veneziano not been reading the right textbook on his maritime return from Israel, kids of today would still be asking what string theory is.[4]

At zeroth order, string theory describes a QFT living on a two-dimensional worldsheet swept out by a one-dimensional string embedded in a target space. One can map the dynamics of the worldsheet into an EFT living on the target space background, and for certain choices of worldsheet theory and target space dimension weakly coupled gravity emerges. Worldsheet unitarity is an important ingredient in this recipe: anomaly cancellation on the worldsheet fixes the equations of motion in the target space. String theory is a goldmine for high energy theorists; by virtue of the fact that the worldsheet theory is a renormalizable QFT sans gravity, mathematical rigor emerges from the darkness, bringing with it what looks like a viable UV theory of quantum gravity.

This is close but no cigar. Aside from some exceptionally simple models [16, 17], scattering amplitudes in string theory must be dealt with perturbatively in the string coupling g_s on a fixed target space. For this reason, string theory is said to be background dependent, which for a full theory of quantum gravity is a big no-no. A nonperturbative or background independent formulation, whether by the name of M-theory or string field theory, is necessary to fully grasp the grandeur of string theory, but this is still on the horizon.

None of the topics in this thesis, by design, rely on the framework of string theory. While there are hints that string theory is the only way of UV completing an effective Einstein-Hilbert action with matter [18, 19], we'll never really know for sure without a rigorous mathematical proof (unlikely) or direct observation of Planck-scale phenomena

[4] The most useful reference for learning string theory remains [14, 15]. I was taught if you understand the glossary you understand string theory; this is still a work in progress.

(impossible). We leave those questions to future work.

1.2 The Black Hole Information Paradox

As we've discussed, the energy scale at which gravity ceases to be a good EFT is roughly the same scale at which we expect our description of spacetime as a smooth background to break down. Thus two competing, well-defined theories of quantum gravity are experimentally indistinguishable, and a strict empiricist might balk at the whole program of quantum gravity and go back to their lab, content with the Wilsonian notion that EFTs are sufficient.

Enter the black hole information paradox. Black holes classically have a nondecreasing horizon area, thermodynamically motivating Bekenstein to conjecture a black hole entropy proportional to the area of its event horizon [20]. Following a heroic calculation by Hawking [21], this entropy was fixed to be

$$S_{BH} = \frac{k_B c^3}{\hbar} \frac{A}{4G_N}. \tag{1.3}$$

With this entropy came a temperature, showing that black holes emit a thermal gas of radiation, shrinking and eventually completely evaporating. The fact that black holes evaporate doesn't necessarily obfuscate the treatment of their area as entropy, as entropy can locally decrease as long as a system's total entropy is nondecreasing. The real paradox arises because the radiation is thermal.

A black hole formed from collapse of matter in a pure state should evolve unitarily to a cloud of radiation in the same pure state. Hawking's calculation instead shows that *any* pure state forming a black hole evolves to the same thermal state. The information contained in the initial shell of matter is gone, replaced with an informationless mess

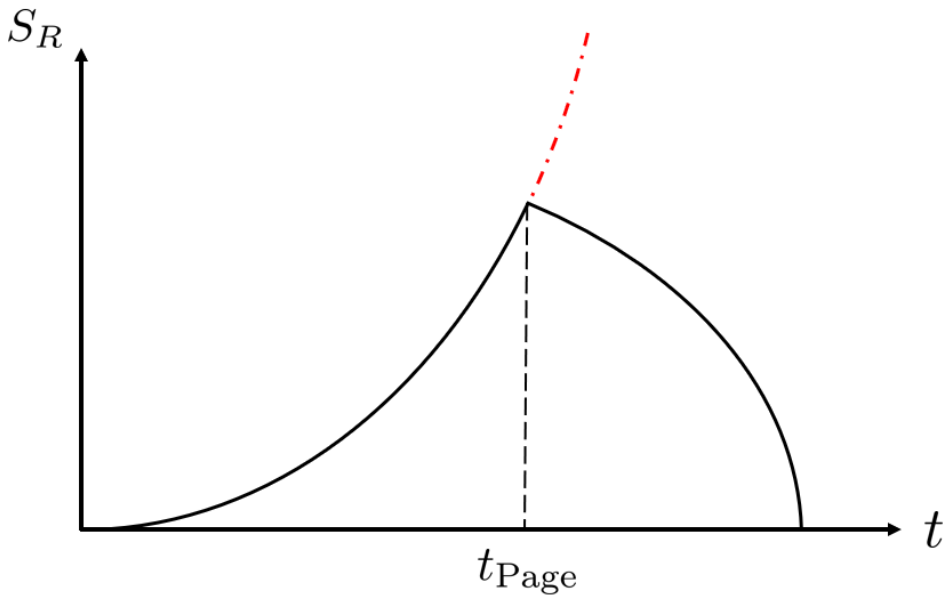


Figure 1.1: The Page curve for an evaporating black hole. The red dotted line denotes the result of Hawking's calculation, which is inconsistent with unitary time evolution.

composed mostly of photons and gravitons.

This confusion is formalized by the Page curve [22]. For a pure state in a Hilbert space of dimension $D = d_R d_B \gg 1$, a random density matrix of rank $d_R < d_B$ obtained via tracing out part of the system will have von Neumann entropy [23]

$$S_R = \log\left(\min\left(d_R, d_B\right)\right) + \mathcal{O}\left(\frac{d_R}{d_B}\right). \tag{1.4}$$

In our scenario, d_R is the dimension of the radiation Hilbert space, and d_B is the dimension of the black hole Hilbert space. If the black hole starts in a pure state, an observer collecting the radiation should see its entropy increase, max out at the Page time[5], and decrease to zero. Hawking's calculation, on the other hand, shows the radiation entropy to be monotonically increasing for the entire evaporation life cycle, which is only half of the story.

[5]This time was treated precisely in [24] and is slightly longer than half of the black hole's total evaporation time.

This paradox is on a completely different footing than the nonrenormalizability of gravity as an effective field theory. A supermassive black hole at the Page time, say Sagittarius A* after $\sim 10^{87}$ years, is still supermassive, with a horizon scale many orders of magnitude larger than the Planck length. On a practical level this isn't strictly speaking falsifiable for astrophysical black holes due to the time scales involved, but it's a semiclassical problem which should have a semiclassical solution. Attempts to evade the black hole information paradox via complementarity [25] fell before the AMPS firewall paradox [26], which we also won't discuss but is resolved via similar techniques to the ones we'll discuss.[6]

1.3 AdS/CFT

The anti-de Sitter/conformal field theory correspondence (AdS/CFT), like most good conjectures in theoretical physics, posits a brevitic equivalence [32–34]:

$$Z_{\text{AdS}} = Z_{\text{CFT}}. \tag{1.5}$$

Also like most conjectures, only one of these terms is actually well-defined. Z_{AdS} is the partition function of a theory of quantum gravity (usually string theory) on a $(d+1)$-dimensional asymptotically AdS background ("the bulk"). Z_{CFT} we understand perfectly well as the partition function of a conformal field theory on a d-dimensional manifold, taken to be the conformal boundary of the bulk spacetime manifold. That the dual quantum field theory lives on a manifold of one fewer dimension is a sharp statement of the holographic principle [35, 36]; just as black hole entropy follows an area law instead of a volume law, gravity itself can be recast as a gravity-less pure boundary

[6] This firewall paradox is not to be confused with the typical state firewall paradox [27], which has a different flavor of potential resolution [28–31].

Prescription	$S(\rho_R) = \cdots$	Euclidean Derivation
Ryu-Takayanagi (RT)	$\min_{\gamma_R} \frac{A(\gamma_R)}{4G_N}$	[40]
Hubeny-Rangamani-Takayangi (HRT) [41]	$\operatorname{ext}_{\gamma_R} \frac{A(\gamma_R)}{4G_N}$	[42]
Faulkner-Lewkowycz-Maldacena (FLM) [43]	$\min_{\gamma_R} \left[\frac{A(\gamma_R)}{4G_N}\right] + S_{\text{bulk}}(\rho_r)$	N/A
Quantum Extremal Surface (QES) [44]	$\operatorname{ext}_{r} \left[\frac{A(\partial r)}{4G_N} + S_{\text{bulk}}(\rho_r)\right]$	[45]

Table 1.1: Holographic entropy formulae. By convention, the bulk subregion bounded by the minimal surface and R, the entanglement wedge, is denoted r.

theory. AdS/CFT has passed numerous tests [37], so Z_{CFT} is sometimes viewed as a *definition* for quantum gravity with asympotically AdS boundary conditions, even though the correspondence has only been demonstrated in any generality for perturbative string theory on fixed backgrounds.

Stated as (1.5), AdS/CFT immediately resolves the black hole information paradox: if quantum gravity really has the dynamics of a unitary quantum system, then black hole evaporation is manifestly unitary! This would have had dramatic consequences, if not for the small problem that a "proof" of AdS/CFT necessitates a rigorous definition of Z_{AdS} via some method besides holography.

A sharp statement of the black hole information paradox in AdS/CFT arose via the Ryu-Takayanagi (RT) formula [38, 39], which conjectured an equivalence between the entanglement entropy of a boundary subregion and boundary-anchored minimal area surfaces in the bulk. For a boundary subregion R with a bulk minimal surface γ_R homologous to R, the von Neumann entropy for the boundary state ρ_R is

$$S(\rho_R) = \frac{A(\gamma_R)}{4G_N}. \tag{1.6}$$

This can be seen as a generalization of the Bekenstein-Hawking entropy for spacetimes without horizons (but with asymptotic boundaries).

The RT formula was later extended to include both time-dependent spacetimes (which upgrades minimization to extremization) and perturbative corrections in G_N such as bulk matter/graviton entanglement entropy to match Bekenstein's generalized entropy [46]. We include a list of these holographic entropy formulae and their associated derivations in Table 1.1. The quantum extremal surface prescription is notable because it's correct to all orders in perturbation theory in G_N. Morally speaking, it treats the area term and the bulk entropy on equal footing such that renormalization of $S_{\text{bulk}}(\rho_r)$ is counteracted by renormalization of Newton's constant.

These formulae were all originally proven using Euclidean methods. Amazingly, these Euclidean derivations are couched firmly in the language of semiclassical gravity, without reference to a boundary dual[7] Here we sketch the derivation of the RT formula.

The von Neumann entropy is difficult to compute on its own, but can be obtained via analytic continuation from the Rényi entropy [52]

$$S(\rho) = \lim_{n\to 1} \frac{1}{1-n} \operatorname{Tr}\left(\rho^n\right). \tag{1.7}$$

What does it mean to compute $\operatorname{Tr}(\rho^n)$ using semiclassical gravity? Ignoring matter for the moment, at lowest order it means to compute the leading saddle to the putative gravitational path integral

$$Z_{\text{grav}} = \int \mathcal{D}g e^{-S_{EH}[g]}, \tag{1.8}$$

with n-fold "replicated" boundary conditions on the spacetime metric. Assuming this path integral is dominated by a single saddle for all integer Rényi index $n > 1$, one can continue the saddle point action to $n = 1$ to obtain a coarse-grained von Neumann entropy. This proves to be equivalent to computing the action on the original, unreplicated

[7] Recently the generalized entropy was shown to be a well-defined entropy of a bulk Type II von Neumann algebra [47–51], constituting a separate Lorentzian derivation.

saddle with the addition of a conical defect, which in the $n \to 1$ limit becomes massless but gives a contribution to the action proportional to its area. The above derivations relied on a choice of a single saddle which respects a $\mathbb{Z}_n$ replica symmetry, such that these two methods are in fact the same.

1.4 A Unitary Page Curve

The modification to this flavor of derivation to obtain a Page curve turns out to be surprisingly simple. The homology condition for extremal surfaces constrains at least part of the surface to be anchored to the boundary. Further restricting to a connected surface yields an entropy with the same pathologies pointed out by Hawking. Via several explicit calculations in AdS/CFT with absorbing boundary conditions [53, 54], it was shown that around the Page time a new disconnected extremal surface appears, part of which lies roughly a Planck length inside the horizon. No wonder it was so hard to find!

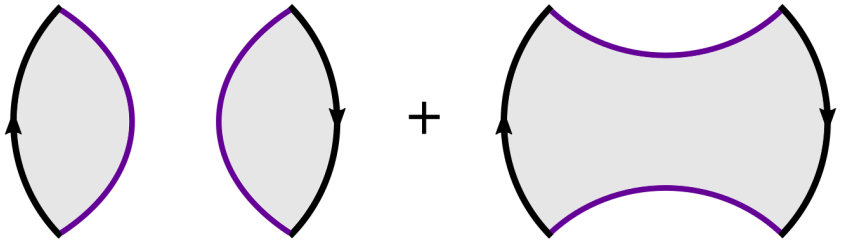
The inclusion of extremal surfaces unanchored to the boundary motivated the so-called "island rule," [55] such that the extremization includes disconnected bulk entanglement wedges. This is the last modification to the Ryu-Takayanagi formula necessary to demonstrate unitarity.[8] The presence of an island implies that, at late times, the degrees of freedom in the black hole are entangled with the early radiation quanta, in a manner not easily explained using quantum field theory on a fixed background [59].

Just as the myriad extremal surface formulas required justification via Euclidean methods, so too did these islands. On Thanksgiving Day 2019[9], two papers appeared with the Euclidean picture, one from groups at Cornell and the IAS [60] (the "East Coast Model") and one from the Stanford group [61] (the "West Coast Model"), both

[8] There are further corrections one can impose for particular incompressible bulk states [56, 57] or for Rényi entropies with $n < 1$ [58], but for our modest historical overview islands alone will suffice.

[9] I joked at Gravity Lunch that reading these papers in the bathroom would give us a good excuse to sneak away from our families during dinner. It did not land.

with the same conclusion: one must include "replica wormholes," connected Euclidean solutions, when computing gravitational path integrals with replicated boundary conditions. This phenomenon is most easily illustrated in the calculation of the second Rényi entropy $S_2(\rho_R) = \mathrm{Tr}(\rho_R^2)$. Here we have two disconnected asymptotic boundaries, so schematically the relevant geometries are:



The disconnected "Hawking saddle" contributes to the first slope of the Page curve, while the connected replica wormhole saddle dominates at later time, eventually bringing the entropy back to zero (or saturating in the case of an eternal black hole). The inclusion of spacetime wormholes into the saddle point calculation of Rényi entropy produces a Page curve consistent with unitarity, but clashes with the orthodoxy of AdS/CFT: if a quantum mechanical dual lived on each boundary, correlation functions should factorize into pieces supported on the individual connected boundary components.

This has resulted in the elevation of the information paradox to the so-called "factorization problem": replica wormholes seem to conflict with the shockingly basic notion that two quantum systems on disconnected manifolds are uncorrelated. We won't attempt a resolution of the factorization problem in this thesis; indeed, its resolution seems to require input beyond semiclassical gravity [62–65].[10] All we'll endeavor to do is use the framework of semiclassical gravity to sharpen some corners of the holographic correspondence. Let's get to it.

[10]Many would argue this is also true for the "real" information paradox, which can only be resolved by counting individual black hole microstates and computing the bulk von Neumann entropy directly

1.5 Permissions and Attributions

1. Chapter 2 is the result of a collaboration with Don Marolf [66], published on arXiV and in the Journal of High Energy Physics.

2. Chapter 3 is the result of a collaboration with Xi Dong and Wayne Weng [67], published on arXiV and in the Journal of High Energy Physics.

3. Chapter 4 is the result of a collaboration with Fernando Iniguez [68], published on arXiV.

4. Chapter 5 is the result of a collaboration with Xi Dong and Wayne Weng [69], published on arXiV and in the Journal of High Energy Physics.

Chapter 2

Perturbatively Traversable Wormholes from Bulk Fermions

2.1 Introduction

Wormholes have long been a source of fascination both in the scientific literature [70–74] and in science fiction [75] as a potential tool for producing superluminal travel. In classical general relativity, wormholes are nontraversable due to constraints on causality from the null energy condition (NEC), which implies the topological censorship theorems [76, 77]. With the assumption of global hyperbolicity, these theorems require causal curves that start and end at the boundary to be deformable to the boundary in both asymptotically flat and asymptotically anti-de Sitter contexts[1]. Adding quantum corrections allows fluctuations that violate the null energy condition, though the topological censorship theorems continue to apply in contexts where the integrated null energy along causal curves is nonnegative so that the averaged null energy condition (ANEC) holds.

[1] Traversable wormholes can be constructed if one drops the requirement of global hyperbolicity, e.g. by introducing NUT charge [78, 79].

More generally, however, one might expect the ANEC to hold only along achronal curves; i.e., along only the fastest causal curves connecting two events [80]. For example, in spacetimes with noncontractible closed spacelike curves, the Casmir effect causes the ANEC to fail. Indeed, it is this achronal ANEC (AANEC) that follows from the generalized second law [81]. Traversable wormholes are thus allowed, so long as it takes longer to travel through the wormhole than to go around. Similar conclusions follow from requiring boundary causality in the context of AdS/CFT [82].

Consistent with such expectations, traversable wormholes were constructed by Gao, Jafferis, and Wall (GJW) [83] by introducing a time-dependent coupling between the two otherwise disconnected asymptotically AdS_3 boundaries. This nonlocal coupling induces a bulk perturbation of a scalar quantum field whose back-reaction allows causal curves to run between the two boundaries. The averaged null energy along these trajectories is negative, but the nonlocal boundary coupling transmits signals from one boundary to the other more quickly than they could travel through the wormhole. In this sense the AANEC is satisfied. Similar setups have been studied for AdS_2 and dual SYK models [84–86] and for rotating AdS_3 wormholes in [87]. Furthermore, since such traversable wormholes provide a holographic dual of certain quantum teleportation protocols, they are of broader interest in connection with the ER = EPR [88, 89] and GR = QM [90] conjectures.

Nonlocal boundary couplings of the sort used in GJW were expected to model more general backgrounds in which causal curves can travel from one wormhole mouth to the other, e.g., when both mouths are embedded in the same asymptotically flat or asymptotically AdS region of spacetime. This expectation was recently verified by two complementary constructions. In [91], a nearly AdS_2 approximation was used to construct a time-independent asymptotically flat four-dimensional wormhole. This approach allowed [91] to address many nonperturbative issues. The contrasting approach of [92] (see also

[93]) used a perturbative framework to argue that standard local quantum fields on a broad class of such classical wormhole geometries have Hartle-Hawking-like states whose stress-energy back-reacts on the geometry to render the wormholes traversable.

The class of backgrounds M considered by [92] were of the form $M = \widetilde{M}/\mathbb{Z}_2$, where the covering space $\widetilde{M}$ contains black holes with Killing horizons and well-defined Hartle-Hawking states for any quantum fields. This context includes many examples with familiar wormhole topologies having fundamental group $\pi_1 = \mathbb{Z}$, though it also includes what one might call torsion wormholes with e.g. $\pi_1 = \mathbb{Z}_2$. For linear quantum fields, the sign of the back-reaction, and thus whether or not it makes the wormhole traversable, is controlled by a choice of periodic or anti-periodic boundary conditions under the action of this $\mathbb{Z}_2$. For scalar fields, for instance, it depends on whether the $\mathbb{Z}_2$ isometry J maps $\phi(x)$ to $\phi(Jx)$ or to $-\phi(Jx)$, and thus whether ϕ satisfies periodic or anti-periodic boundary conditions on the quotient spacetime M. Here we follow [92] in using the term wormhole to refer to any setting where curves both starting and ending at the boundary cannot be smoothly deformed to the boundary. Such curves would classically be forbidden from being causal by the aforementioned topological censorship results of [76, 77].

In certain asymptotically AdS_3 examples of such torsion wormholes, [92] was able to exactly compute quantum stress-energy tensor expectation values for general free bulk scalar fields. The study of more general quantum fields is clearly of interest, though in curved spacetime computations involving higher spin fields can lead to significant technical complications. Here we take a first step in this direction by computing the quantum back-reaction from Weyl fermions of any mass m in the torsion wormholes of [92]. Again, we find that stress-energy tensor expectation values can be computed exactly.

The outline of this chapter is as follows. In section 2.2, we review the construction of $\mathbb{Z}_2$ wormholes from [92] and find geometries useful for our consideration of spinor fields.

In the relevant case the spacetime M is a $\mathbb{Z}_2$ quotient of the rotating Bañados-Teitelboim-Zanelli black hole [94, 95] (rBTZ) times S^1, so that $M = (\text{rBTZ} \times S^1)/\mathbb{Z}_2$. We flesh out the details of this case in section 2.3, providing an analytic expression for the null stress-energy tensor on the horizon of M for a spinor field of arbitrary mass. The above quotient breaks rotational symmetry, and we find the sign of the integrated null energy to generally depend on the BTZ angular coordinate ϕ. But the average is nonzero when the black hole rotates, and is negative with the appropriate choice of periodicity. Following [92] we compute $T\langle\Delta V\rangle$, where T is the black hole temperature and $-\langle\Delta V\rangle$ measures the expectation value of the time advance governing traversability of the wormhole. As in the scalar case, we find $T\langle\Delta V\rangle$ to be independent of T for all fermion masses, suggesting that $\langle\Delta V\rangle$ diverges as $T \to 0$ and that bulk spinors alone would suffice to yield an eternally traversable wormhole in that limit. We conclude in section 2.4 by discussing the extension of our results to other higher spin particles which, like the spinor, have exactly soluble propagators in AdS_d. Working our way all the way up to spin-2 would allow understanding of the effect of linearized gravitons on such wormholes; this remains a goal for future work.

2.2 Preliminaries and Review

We begin in section 2.2.1 with a brief review of certain asymptotically AdS (or asymptotically AdS $\times X$) $\mathbb{Z}_2$ wormholes studied in [92]. We then recall in section 2.2.2 how the stress-energy tensor of quantum scalar fields can be computed exactly in these backgrounds. The extension of such computations to fermions is outlined in section 2.2.3, which in particular describes further properties required for fermions to yield nontrivial results. The example from [92] satisfying these properties is called the Kaluza-Klein zero-brane orbifold (KKZBO) spacetime and is discussed in detail in section 2.2.4.

2.2.1 Review of AdS $\mathbb{Z}_2$ Wormholes

The exactly solvable models of [92] involved quantum fields in a Hartle-Hawking-like state propagating on $\mathbb{Z}_2$ quotients $M = \widetilde{M}/\mathbb{Z}_2$ of BTZ and BTZ $\times S^1$. This setting is useful as the Killing symmetry of BTZ preserves the BTZ Hartle-Hawking state. As a result, in that state on the covering space $\widetilde{M}$ the symmetry requires the expected null-null component of the stress-energy tensor $\langle T_{kk} \rangle_{\widetilde{M}}$ to vanish on the BTZ horizon for any quantum field. This symmetry is then broken by the $\mathbb{Z}_2$ quotient, so on the physical spacetime M the horizon expectation value $\langle T_{kk} \rangle_M$ can be nonzero. Nevertheless, the simplicity of BTZ can be used to provide an analytic expression for $\langle T_{kk} \rangle_M$. The integrated $\langle T_{kk} \rangle_M$ on the horizon can then be evaluated numerically and combined with first order perturbation theory to compute the back-reaction. We begin by reviewing simple examples of the above geometries and techniques. In this section we set the angular momentum of the black hole to zero and consider only nonrotating BTZ.

Without rotation, the BTZ metric can be written in the form

$$ds^2 = \frac{1}{(1+UV)^2}\left(-4\ell^2 dU dV + r_+^2 (1-UV)^2 d\phi^2\right). \tag{2.1}$$

Here ℓ is the AdS length scale (which we will often take to be one) and $2\pi r_+$ is the length of the BTZ horizon. We use Kruskal-like coordinates (U, V, ϕ) where U and V parameterize the null directions. The metric (2.1) is global AdS_3 when ϕ takes values in $(-\infty, \infty)$, but gives the global BTZ black hole when one makes the identification $\phi \sim \phi + 2\pi$ [95].

A variety of interesting spacetimes can be constructed as quotients of (2.1) or of tensor products with some other simple factor X. For example, the $\mathbb{R}P^2$ geon [96] is a quotient of (2.1) under the isometry $J_1 : (U, V, \phi) \mapsto (V, U, \phi + \pi)$; see figure 2.1 below. Note that J_1^2 is the identity and that the quotient BTZ$/J_1$ contains a noncontractible cycle

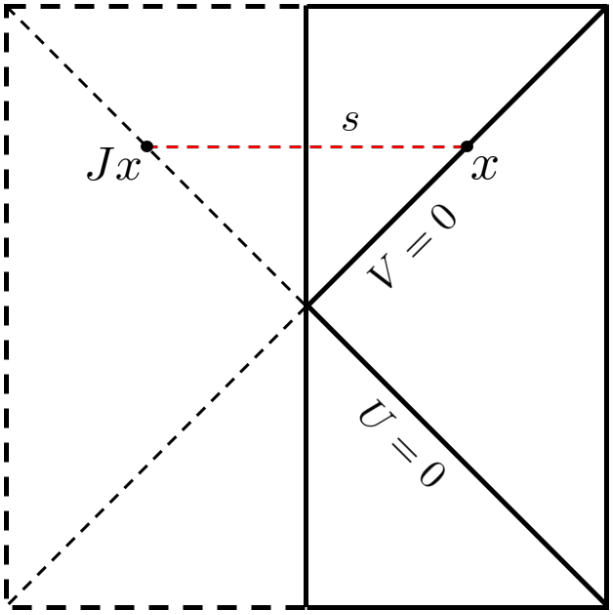


Figure 2.1: Conformal diagram of the $\mathbb{R}P^2$ geon, KKEOW, and KKZBO geometries, with null coordinates (U, V) indicated. The left region (dashed lines) is the image of the right (solid lines) under the $\mathbb{Z}_2$ isometry J which exchanges $U \leftrightarrow V$, so that the right alone may be used to represent the quotient. The desired isometries J also act on S^1 factors (just S^1 for the $\mathbb{R}P^2$ geon and $S^1 \times S^1$ for the KKEOW and KKZBO geometries) not shown in the figure. The geodesic proper distance s between a point x on the horizon and its image Jx is nonzero and spacelike in all cases, though for $U = V = 0$ this is a result of the additional action of J on suppressed angles ϕ and/or θ not shown in this figure.

with $\mathbb{Z}_2$ homotopy that is not deformable to the boundary and can be represented by the closed curve $\{\phi \in [0, \pi]\}$ for any $U(\phi) = V(\phi)$. Another interesting related spacetime is the Kaluza-Klein end-of-the-world brane geometry (hereafter KKEOW) $(\mathrm{BTZ} \times\ S^1)/J_2$, where the isometry J_2 now acts on the angle θ associated with the internal S^1 as well as on the BTZ factor. In particular, $J_2 : (U, V, \phi, \theta) \mapsto (V, U, \phi, \theta + \pi)$. Here the quotient is smooth but Kaluza-Klein reduction on S^1 gives a singular spacetime with what may be called an end-of-the-world brane at $U = V$. Again, J_2^2 is the identity and the quotient contains a noncontractible cycle with $\mathbb{Z}_2$ homotopy that is not deformable to the boundary and can be represented by the closed curve $\{\theta \in [0, \pi],\ \phi = \text{constant}\}$ for any $U(\theta, \phi) = V(\theta, \phi)$.

Both of the above quotients are nonorientable, but an orientable spacetime (BTZ $\times$ $S^1)/J_3$ can be obtained by allowing the isometry J_3 to act on both θ and ϕ. In particular, [92] studied the Kaluza-Klein zero-brane orbifold (KKZBO) given by $J_3 : (U, V, \phi, \theta) \mapsto (V, U, -\phi, \theta + \pi)$. As before, the quotient is smooth but singularities arise when it is Kaluza-Klein reduced on the S^1. These singularities are localized at $U = V$ with $\phi = 0, \pi$ and may be called zero-brane orbifolds as they represent pointlike defects in the resulting three-dimensional spacetime. Once again, J_3^2 is the identity and the quotient contains a noncontractible cycle with $\mathbb{Z}_2$ homotopy that is not deformable to the boundary and is now associated with the zero branes. It can be represented by the closed curves $\{\theta \in [0, \pi]\}$ at $\phi = 0$ or π for any $U(\theta) = V(\theta)$. The KKZBO is more complicated than the above examples as it breaks translational symmetry in ϕ. However, as noted in [92], it has the advantage of extending to rotating black holes where both J_1 and J_2 cease to be isometries. In particular, the KKZBO admits an interesting extreme limit in which back-reaction from quantum fields can become large. As reviewed in section 2.2.3 below, even without rotation, the fact that the KKZBO is orientable makes it a much more interesting context in which to study fermions.

We are interested in finding the quantity $\langle \Delta V \rangle$, the expected null time delay of a geodesic starting at $U = -\infty$ and ending at $U = \infty$ induced by back-reaction from quantum fields. A negative value $\langle \Delta V \rangle < 0$ indicates that the wormhole becomes traversable, as a null geodesic fired from the left boundary in the distant past then arrives at the right boundary at a finite future time with (on average) coordinate $V = \langle \Delta V \rangle$. As shown in [83], in the linearized approximation and in the presence of rotational symmetry, $\langle \Delta V \rangle$ in nonrotating BTZ (or BTZ $\times X$) is related to the integrated null stress-energy tensor

along the horizon through

$$\langle \Delta V \rangle = 4\pi G_N \int_0^\infty dU \, \langle T_{kk} \rangle \, . \tag{2.2}$$

Here $T_{kk} \equiv T_{\mu\nu} k^\mu k^\nu$ for a null horizon generator $k^\mu \partial_\mu = \partial_U$. The extension to rotating BTZ and to broken rotational symmetry will be reviewed below in section 2.2.4 below.

As noted in the introduction, symmetry requires $\langle T_{kk} \rangle$ to vanish in the Hartle-Hawking state on BTZ or BTZ$\times X$, but we will in fact compute $\langle T_{kk} \rangle$ in the quotient $(\text{BTZ} \times X)/\mathbb{Z}_2$ where the symmetry is broken. At linear order in perturbation theory we may then lift the resulting $\langle T_{kk} \rangle$ back to the (BTZ $\times$ X) covering space, compute back-reaction on the metric using (2.2), and take the quotient of the result to obtain a self-consistent semi-classical solution.

The stress-energy tensor of a scalar field ϕ takes the form

$$T_{\mu\nu} = \partial_\mu \phi \partial_\nu \phi - \frac{1}{2} g_{\mu\nu} g^{\rho\sigma} \partial_\rho \phi \partial_\sigma \phi - \frac{1}{2} g_{\mu\nu} m^2 \phi^2. \tag{2.3}$$

Contracting with the null vectors eliminates all but the first term, so $\langle T_{kk} \rangle$ is a coincident limit of derivatives acting on the two-point function $\langle \phi\phi \rangle$. The pointwise contribution to the null stress-energy tensor can thus be determined by derivatives of the scalar Green's function. Since $k^\mu \partial_\mu = \partial_U$, we may write

$$\langle T_{kk}(x) \rangle = \lim_{x \to x'} \langle \partial_U \phi(x) \partial_U \phi(x') \rangle \tag{2.4}$$

where the computation is to be done in the Hartle-Hawking state on the quotient geometry M.

As reviewed in [92], for linear quantum fields correlations functions in this state may be computed using the method of images. In particular, one may identify the quantum

field $\phi(x)$ on the quotient M with an appropriately scaled linear combination of the scalar field $\widetilde{\phi}(x)$ on the covering space $\widetilde{M}$ and its image under the relevant isometry J:

$$\phi(x) \equiv \frac{1}{\sqrt{2}} \left(\widetilde{\phi}(x) \pm \widetilde{\phi}(Jx) \right). \tag{2.5}$$

The choice of sign in (2.5) corresponds to a choice of periodic (+) or anti-periodic (−) boundary conditions on the $\mathbb{Z}_2$ cycle of the quotient.

The stress-energy tensor (2.3) is thus a sum of four terms involving coincident limits of two point functions involving all possible pairings of $\widetilde{\phi}(x)$ and $\widetilde{\phi}(Jx)$. Terms involving coincident points in the covering space are proportional to the stress-energy tensor in the Hartle-Hawking state on $\widetilde{M}$ and thus vanish by the symmetry noted above. Thus all potentially divergent terms vanish and $\langle T_{kk} \rangle_M$ is explicitly finite. Nonzero contributions come only from the two cross terms. These contribute equally and give

$$\langle T_{kk}(x) \rangle_M = \pm \langle \partial_U \phi(x) \partial_U \phi(Jx) \rangle_{\widetilde{M}}. \tag{2.6}$$

We can now see that creating a traversable wormhole is simple, as unless some coincidence or symmetry imposes $\langle T_{kk} \rangle_M = 0$, one will find $\langle T_{kk} \rangle_M$ to be negative for one choice of periodic or anti-periodic scalars. As verified by detailed computations in [92], the correct choice turns out to be periodic boundary conditions for the $\mathbb{R}P^2$ geon, the KKEOW spacetime, and the KKZBO geometry.

2.2.2 Explicit Calculation

We now review further explicit scalar results from [92]. The scalar two-point function on empty AdS_3 is known exactly [97]. It may be written in the form

$$G_{AdS_3}(Z) = \frac{1}{4\pi}\left(Z^2-1\right)^{-1/2}\left(Z+\left(Z^2-1\right)^{1/2}\right)^{1-\Delta}, \tag{2.7}$$

where $Z \equiv 1+\sigma(x,x')/\ell^2$, $\sigma(x,x')$ is the half squared-geodesic distance in an associated four dimensional embedding space $\mathbb{R}^{2,2}$, and Δ is given by the same formula as the scaling dimension of the associated operator in any dual CFT [37],

$$\Delta = 1 \pm \sqrt{1+m^2\ell^2} \tag{2.8}$$

where the $\pm$ denotes a choice of boundary conditions at AdS infinity. We will always take the $(+)$ boundary condition below, as this choice is always consistent with unitarity. However, the $(-)$ choice can also be of interest; see [92] for associated results in the context of the above $\mathbb{Z}_2$ wormholes.

From [92], the half-squared geodesic distance in $\mathbb{R}^{2,2}$ in our Kruskal-like coordinates is

$$\begin{aligned}\sigma(x,x') = \frac{\ell^2}{(1+UV)(1+U'V')}\Bigg[&(UV-1)(U'V'-1)\cosh\left(\frac{r_+\left(\phi-\phi'\right)}{\ell}\right)\\ &-(1+UV)(1+U'V')+2(UV'+VU')\Bigg].\end{aligned} \tag{2.9}$$

We can use the fact that BTZ is a quotient space of AdS_3 to recast the BTZ Green's function as a sum over images in the AdS_3 covering space such that

$$G_{BTZ}(Z) = \sum_{n\in\mathbb{Z}} G_{AdS_3}\left(Z_n\right) \tag{2.10}$$

where $Z_n \equiv 1 + \sigma(x, x'_n)/\ell^2$ and $\phi'_n = \phi' + 2\pi n$. The additional isometry to produce the $\mathbb{R}P^2$ geon means that we are concerned with cases where $\phi' = \phi + \pi$, so the sum over images is a sum over $\phi' = \phi + (2n+1)\pi$ with $U' = V = 0$ and $V' = U$.

By substituting the expression for the embedding space geodesic distance, we can calculate G_{AdS_3}, and by extension the two-point function on the $\mathbb{R}P^2$ geon. The full expression for $\langle T_{kk} \rangle$ is quite complicated, so we recall from [92] only the result in the limit $r_+/\ell \to 0$. Using $\langle T_{kk} \rangle_n$ to denote the contribution to $\langle T_{kk} \rangle$ from the nth term in the sum (2.10), for periodic scalars one finds in this limit the n-independent result

$$\langle T_{kk} \rangle_n = \frac{\left(1 + 2U(U + \sqrt{1+U^2})\right)^{1-\Delta}}{32\pi U^3 (1+U^2)^{5/2}} \Big[1 + \left(2U\sqrt{1+U^2} + 8U^3\sqrt{1+U^2}\right)(\Delta - 1) + 4U^4(2 - 2\Delta + \Delta^2) + U^2(6 - 8\Delta + 4\Delta^2)\Big]. \tag{2.11}$$

While the sum of (2.11) over n clearly diverges, the result (2.11) will still prove useful in section 2.3. For general r_+, numerically integrating the full expression for periodic scalars yields a strictly negative result for each $\langle T_{kk} \rangle_n$, so the back-reaction does indeed render the wormhole traversable.

The calculation for the KKEOW geometry is slightly more involved. For any given four-dimensional mass m, Kaluza-Klein (KK) reduction on the S^1 produces a tower of massive three-dimensional fields with effective masses given by

$$m_{eff}\ell = \sqrt{m^2\ell^2 + \left(\frac{\ell}{R_{S^1}}\right)^2 p^2}, \tag{2.12}$$

where $p \in \mathbb{Z}$ is the mode number on the internal S^1 with radius R_{S^1}. Since the three-dimensional two-point function (2.10) is known exactly, it is useful to write the full four-dimensional stress-energy tensor as a sum over p of contributions $\langle T_{kk} \rangle_{n,p}$ from the n^{th} term in (2.10) and the p^{th} mode on the S^1.

Such contributions include a factor of of $e^{ip\pi} = (-1)^p$ from the action of the isometry on θ. These alternating signs play an important role, as the $n = 0$ contributions $\langle T_{kk} \rangle_{n=0,p}$ all have nonintegrable singularities at $U = 0$. Indeed, such terms are independent of r_+ and give

$$\langle T_{kk} \rangle_{n=0,p} = \frac{1}{32\pi U^3} + \frac{3 - 8\Delta + 4\Delta^2}{64\pi U} + \frac{-2\Delta + 3\Delta^2 - \Delta^3}{6\pi} + O(U). \tag{2.13}$$

Since the four-dimensional state is Hadamard and the quotient is smooth, the full stress-energy tensor can have only integrable singularities and the nonintegrable terms in (2.13) must cancel when summed over p. For numerical calculations, we can simply choose some large N and sum over modes with $|p| \leq N$ if we also impose a cutoff at small U to avoid possible issues from incomplete cancellations of such terms at finite N; see [92] for details[2]. Again, the integrated $\langle T_{kk} \rangle$ is negative for all periodic scalars of the type discussed above.

The KKZBO computations are similar but the integrated $\langle T_{kk} \rangle$ now depends on ϕ. For $\phi \neq 0, \pi$ each term $\langle T_{kk} \rangle_{n,p}$ is finite and continuous. However, at $\phi = 0, \pi$ the action of J_3 coincides with the KKEOW actions of J_3. As a result, the expressions for $\langle T_{kk} \rangle_{n,p}$ coincide there as well. In particular, corresponding care is required for the $n = 0$ modes. For the simple scalars discussed here, periodic boundary conditions make the integrated $\langle T_{kk} \rangle$ negative along all horizon generators, though [92] also found more complicated examples where the sign of $\langle \Delta V \rangle$ varies with ϕ.

[2]It was also shown in [92] that the explicit terms above vanish when summed over p using Dirichlet η regularization for the $1/U^3$ and $1/U$ terms and Abel summation for the constant term. One may thus rewrite $\langle T_{kk} \rangle$ as a sum of explicitly finite terms.

2.2.3 Subtleties of Spinor Representations

We now consider how the above constructions should be generalized to accomodate bulk fermion fields. To begin, recall that spin groups admit two inequivalent fundamental representations in odd dimensional spacetimes. One may think of this choice as arising from two distinct possible definitions of $\gamma^{d-1} = \pm i^{-\frac{d-3}{2}} \prod_{i=0}^{d-2} \gamma^i$, both of which form a representation of the Clifford algebra $\{\gamma^a, \gamma^b\} = 2\eta^{ab}$ [15], or as choice of

$$\prod_{i=0}^{d-1} \gamma^i = \pm \mathbb{I}_{2^{(d-1)/2}}. \tag{2.14}$$

We shall simply denote the two choices as γ^a_A and γ^a_B below, with the understanding that the sign on the right-hand-side of 2.14 is $+$ in the A representation and $-$ in the B representation, with $\gamma^a_A = -\gamma^a_B$.

Since the two spinor representations differ by the sign of all γ^μ, they are related by the action of the three-dimensional parity operator, which we may take to be any orientation-reversing isometry of 3D Minkowski space. As a result, on a nonorientable spacetime like the $\mathbb{R}P^2$ geon, spinors can be well-defined only when both representations are present. In particular, our theory on the geon must contains two spinor fields ψ_A and ψ_B, each corresponding to a different representation, but which are exchanged when one traverses a noncontractible curve that reverses orientation. While it is inconsistent to have a single fermion corresponding to either representation alone, this AB doublet of fermions yields a well-defined theory. In this context, we may use the Lagrangian [98]

$$\mathcal{L} = \det(e) \left[\overline{\psi_A}(i\gamma^\mu_A D_\mu - m)\psi_A + \overline{\psi_B}(i\gamma^\mu_B D_\mu - m)\psi_B\right] \tag{2.15}$$

where $\det(e)$ is the vielbein determinant and $\gamma^\mu_{A,B}$ denote the gamma matrices for each fermion representation. In terms of a method-of-images construction paralleling that for

scalars in [92] in the scalar case, the Lagrangian on the BTZ covering space must also take the form (2.15).

Let us now consider the analogue of the scalar relation (2.5). For fields of nonzero spin, adding the field operators at distinct points x and Jx requires one to first find some way to identify the corresponding two tangent spaces. It is natural to use the isometry J and, as discussed in section 2.3 below, any isometry can be extended to a map $\hat{J}$ taking spinors at x to spinors at Jx. In particular, the spinor field $\hat{J}\widetilde{\psi}$ evaluated at x is an operator built from $\widetilde{\psi}(Jx)$. We may thus write

$$\psi(x) \equiv \frac{1}{\sqrt{2}} \left(\widetilde{\psi}(x) + \left[j\widetilde{\psi} \right] \right) (x). \tag{2.16}$$

Since the equation of motion is linear, it is preserved by $\hat{J}$ if $\hat{J}$ preserves the vielbein and the γ^a. We will use these conditions in section 2.2.4 to define the appropriate extension $\hat{J}$ of the isometry J. The ansatz (2.16) will then satisfy the equation of motion on the quotient M when $\widetilde{\psi}$ satisfies the corresponding equation on the covering space $\widetilde{M}$. One may also check that canonical normalization of $\widetilde{\psi}$ gives canonical normalization of ψ. Thus (2.16) is the desired method-of-images ansatz. As for scalars, correlation functions in spinor Hartle-Hawking states on M and $\widetilde{M}$ will again be related by (2.16).

We now consider the implications for stress-energy tensors on the $\mathbb{R}P^2$ geon, which we represent as BTZ$/J_1$. The important point here is that J_1 reverses orientation and must thus interchange ψ_A and ψ_B. In other words, in this context we may write (2.16) more explicitly as

$$\psi_A(x) \equiv \frac{1}{\sqrt{2}} \left(\widetilde{\psi}_A(x) + \left[\hat{J}_1 \widetilde{\psi}_B \right] (x) \right), \quad \psi_B(x) \equiv \frac{1}{\sqrt{2}} \left(\widetilde{\psi}_B(x) + \left[\hat{J}_1 \widetilde{\psi}_A \right] (x) \right), \tag{2.17}$$

where $\left[J_1 \widetilde{\psi}_B \right] (x)$ is built from the operator $\widetilde{\psi}_B(J_1 x)$. Since the Lagrangian (2.15) con-

tains no interactions, the full stress-energy tensor is of course a sum of separately conserved stress-energy tensors for the A and B fields. Let us consider first the expectation value of the A stress-energy tensor, which is a quadratic composite operator much like the stress-energy tensor of a scalar field. As in the scalar case, using the first line in (2.17) generates four terms. The term involving coincident points at x gives the A stress-energy tensor on the covering space which vanishes by the same symmetry described above for scalars, and the term involving coincident points at $J_1 x$ gives the similarly-vanishing B stress-energy tensor. The final two terms are cross-terms built from the correlators $\langle \psi_A(x)\psi_B(J_1 x)\rangle$ and $\langle \psi_B(J_1 x)\psi_A(x)\rangle$ in the Hartle-Hawking state on the covering space $\widetilde{M}$. The the lack of an interaction term in (2.15) means that such AB cross-correlators vanish identically. So despite the breaking of the Killing symmetry on the $\mathbb{R}P^2$ geon, the A stress-energy tensor continues to vanish on the horizon, as does the B stress-energy tensor by the same argument. The shift ΔV thus remains zero and first-order back-reaction does not render the wormhole traversable.

One might ask if fermionic back-reaction is more interesting on the KKEOW spacetime. There the full spacetime is four-dimensional, so there is only one representation of the Clifford algebra and it is invariant under orientation-reversal. Thus we need only consider a single Weyl fermion. However, as for the scalar case one may proceed by dimensional reduction to three dimensions, where one obtains a set of uncoupled free fields that come in pairs like the fields AB discussed above. In particular, the AB fields in each doublet are related by reversal of orientation on the three-dimensional (orbifold) base space. In terms of each pair, the discussion proceeds precisely as for the $\mathbb{R}P^2$ geon above, and the full stress-energy tensor again vanishes. From the four dimensional perspective the point is that even-dimensional spinors admit a conserved notion of chirality, and that the two chiralities are exchanged by the orientation-reversing isometry J_2. The cross-terms in the stress-energy tensor are then built from two-point functions between

Weyl spinor components of opposite chirality, which vanish in Hartle-Hawking state.

The upshot of this discussion is that in any $\mathbb{Z}_2$ quotient of a spacetime with a Killing symmetry (in any spacetime dimension) the contribution of spinors to expected stress-energy tensors on the horizon will vanish unless the quotient is orientable. For this reason we focus on the orientable KKZBO spacetime in the remainder of this work. The generalization of our discussion thus far to this spacetime is reviewed in section 2.2.4 below.

2.2.4 The KKZBO Spacetime and Back-reaction

As mentioned in section 2.2.1, the KKZBO spacetime can be defined in the presence of rotation. In particular, it admits an extreme limit where [92] found that back-reaction from matter sources becomes large. Perturbations that render the wormhole traversable may thus naturally create an eternally traversable wormhole in this limit in agreement with [86, 91]. We briefly review these results here for later use in studying back-reaction from the stress-energy tensor of our Weyl fermions.

To begin, recall that the rotating BTZ (rBTZ) metric in our Kruskal-like coordinates (U, V, ϕ) takes the form

$$ds^2 = \frac{1}{(1+UV)^2}(-4\ell^2 dUdV + 4\ell r_-(UdV - VdU)d\phi + \left[r_+^2(1-UV)^2 + 4UVr_-^2\right] d\phi^2). \tag{2.18}$$

We are interested in the Cartesian product of (2.18) with an S^1 of radius R_{S^1}, and thus with line element $R_{S^1}^2 d\theta^2$. Defining

$$J_3 : (U, V, \phi, \theta) \to (V, U, -\phi, \theta + \pi) \tag{2.19}$$

as in section 2.2.1, it is clear that the (U, V, ϕ) part of J_3 preserves (2.18) and the θ part

preserves the metric on S^1.

To understand the implications of the $\langle T_{kk}\rangle_{\text{KKZBO}}$ that we will compute in section 2.3, we must understand the back-reactions of such a source on the geometry. As in the case without spin, one may lift the source $\langle T_{kk}\rangle_M$ computed on the quotient $M = \text{KKZBO}$ to the covering space $\widetilde{M} = \text{rBTZ} \times S^1$, compute the associated back-reaction on $\widetilde{M} = \text{rBTZ} \times S^1$, and then quotient the result again by J_3. This lift and quotient procedure gives the same result as computing back-reaction directly on the KKZBO spacetime. And since we preserve rotational symmetry on the internal S^1, the problem can be Kaluza-Klein reduced to studying back-reaction on three-dimensional rBTZ. Indeed, section 2.3 will directly compute the effective three-dimensional stress-energy $\langle T_{kk}\rangle^{\text{3d}}_{\text{KKZBO}}$, which is just the integral over the internal S^1 of $\langle T_{kk}\rangle_{\text{KKZBO}}$.

We thus require the generalization of (2.2) to rotating BTZ and to sources that break rotational symmetry. As shown in [92], the correct result is

$$\Delta V = \frac{1}{4\ell^2}\int_0^\infty dU h_{kk} = \frac{2\pi G_N}{\ell^2}\int_{-\infty}^{\infty}\int_{-\pi}^{\pi} d\phi' dU H(\phi-\phi')\,\langle T_{kk}\rangle^{3d}(\phi'), \tag{2.20}$$

where the Green's function $H(\phi-\phi')$ is usefully described as a sum over Fourier modes of the form

$$H(\phi-\phi') = \sum_q e^{iq(\phi-\phi')}H_q, \quad H_q = \frac{1}{2\pi}\frac{2\ell^2 r_+^2}{r_+^2 - r_-^2 - 2iqr_- + \ell^2 q^2}. \tag{2.21}$$

Perhaps the most interesting feature of (2.21) is that the zero-mode $H_{q=0} = \frac{\ell^2 r_+^2}{\pi\left(r_+^2 - r_-^2\right)}$, diverges in the extreme limit $r_+ \to r_-$. As noted in [92], this invalidates our perturbative analysis for r_+ very close to r_-, but it also suggests that a full nonperturbative analysis could produce a static, eternally traversable wormhole. We will thus be most interested in cases with $r_+ - r_-$ smaller than any classical scale, but with r_+ still far enough from r_-

that our perturbative treatment remains valid. Variations with ϕ are then a subleading effect as the $q \neq 0$ modes in (2.21) remain finite at $r_+ = r_-$. It is thus useful to focus on the average of the time delay (2.20) over ϕ. Since the temperature of rotating BTZ is

$$T = \frac{r_+^2 - r_-^2}{2\pi r_+ \ell^2} = \frac{2r_+}{H_0}, \tag{2.22}$$

we may average $\langle \Delta V \rangle$ over the ϕ-circle to write

$$T \langle \Delta V \rangle_{\text{average}} = \frac{4G_N r_+}{\ell^2} \int_0^\infty \int_{-\pi/2}^{\pi/2} d\phi dU \, \langle T_{kk} \rangle^{3d}, \tag{2.23}$$

where the factor of 4 is associated with making use of symemtries to change the limits of integration relative to those in (2.20). In the scalar case, the numerics in [92] found the extreme limit of (2.23) to be approximately independent of r_+. We will find similar behavior below.

2.3 Spinors and Stress-Energy Tensors on the KKZBO Spacetime

We now turn to the fermion details for stress-energy tensors on the KKZBO. We begin in section 2.3.1 below by extending the KKZBO isometry J_3 of rBTZ $\times$ S^1 to act on spinor fields. Section 2.3.2 then sets up the calculation of the desired stress-energy tensor components and section 2.3.3 studies a simplifying limit in preparation for more complete numerical calculations in section 2.3.4.

2.3.1 Spinor Extensions of the J_3 Isometry

Any isometry has a natural action on tensor fields via the associated diffeomorphism. But spinor fields are not tensors, and are typically defined by attaching an internal tangent space to each point in spacetime. This is done by choosing a vielbein e^a_μ, which has a spacetime index $\mu = (U, V, \phi, \theta)$ and an internal index $a = (0, 1, 2, 3)$. Choosing the metric on the internal space to be $\eta_{ab} = \mathrm{diag}(-1, 1, 1, 1)$, one may use any vielbein that satisfies $e^a_\mu e^b_\nu \eta_{ab} = g_{\mu\nu}$. Any two such vielbeins are related by an internal $O(3, 1)$ gauge transformation. For rBTZ $\times S^1$ we take

$$e^a_\mu = \frac{1}{1+UV}\begin{pmatrix} \ell & \ell & 0 & 0 \\ \ell & -\ell & 0 & 0 \\ -r_-(U-V) & -r_-(U+V) & r_+(1-UV) & 0 \\ 0 & 0 & 0 & R_{S^1}\,(1+UV) \end{pmatrix}, \tag{2.24}$$

where a labels the columns and μ labels the rows.

The natural action of the isometry J_3 on e^a_μ is given by treating e^a_μ as a spacetime vector field for each a. Thus J_3 acts on spacetime indices μ and the point at which the field is evaluated but has no further action on the internal index a. As a result, in addition to acting on the arguments (U, V, ϕ, θ) in (2.24), it also exchanges the U and V rows and changes the sign of all entries in the ϕ row. We note that this combined operation is *not* a symmetry of the vielbein.

However, as noted in section 2.2.3, it would be more useful for our method-of-images construction to have an operation $\hat{J}_3$ that leaves the vielbein invariant. Since the isometry J_3 preserves the metric, the vielbeins $J_3 e^a_\mu$ and e^a_μ can differ only by an internal $O(3, 1)$ transformation. Choosing the right such transformation j (such that $(je)^a_\mu = j^a_b e^b_\mu$) thus allows us to define a vielbein-preserving $\hat{J}_3 = J_3 \circ j$ as desired. It is easy to check that

this is the case for

$$j_a^b = \begin{pmatrix} 1 & & & \\ & -1 & & \\ & & -1 & \\ & & & 1 \end{pmatrix} \tag{2.25}$$

defines a $\hat{J}_3$ that preserves the vielbein (2.24).

To describe the corresponding action on spinors, one should view (2.25) as the action of an $O(3,1)$ transformation on covectors. There is then a corresponding action $\widetilde{j}$ of this transformation on spinors, up to a sign to be discussed below associated with $O(3,1)$ being the double cover of the associated spin group, defined by requiring that j leave invariant the four-dimensional Clifford algebra $\{\Gamma^a\}$ that satisfies $\{\Gamma^a, \Gamma^b\} = 2\eta^{ab}$ and defines the four-dimensional spinor representation. The spinor-space matrix $\widetilde{j}$ must satisfy

$$\Gamma^a = [j(\Gamma)]^a = (\widetilde{j})^{-1} j_b^a \Gamma^b \widetilde{j}. \tag{2.26}$$

Noting that

$$j_a^b \Gamma^a : (\Gamma^0, \Gamma^1, \Gamma^2, \Gamma^3) \to (\Gamma^0, -\Gamma^1, -\Gamma^2, \Gamma^3), \tag{2.27}$$

and setting $\widetilde{j} = i\Gamma^1\Gamma^2 = (\widetilde{j})^{-1} = \widetilde{j}^\dagger$ one finds $(\widetilde{j})^{-1}\Gamma^a\widetilde{j} = j_b^a\Gamma^b = (j)^{-1}{}_b^a\Gamma^b$ so that (2.26) is satisfied as desired.

Since we will use Kaluza-Klein reduction on the S^1 to express the four-dimensional fields in terms of three-dimensional fields on a BTZ orbifold, it is useful to express $\widetilde{j}$ in terms of the three-dimensional gamma matrices γ^a for $a = (0,1,2)$. Since each three-dimensional spinor representation is invariant under the action of any Γ^a, we can choose a basis for the four-dimensional representation in which each Γ^a takes a block diagonal

form $\begin{pmatrix} \gamma^a_A & 0 \\ 0 & \gamma^a_B \end{pmatrix}$, where the subscripts denote the two spinor representations labelled A and B in section 2.2.3. In either representation we thus find

$$\widetilde{j} = i\gamma^1\gamma^2. \tag{2.28}$$

The fact that the action of $\hat{J}_3$ on spinors is ambiguous up to an overall sign implies that there are two natural notions of periodic spinors on the quotient space: those defined by spinors on rBTZ$\times S^1$ that are invariant under $\hat{J}_3$ and those defined by spinors invariant under $-\hat{J}_3$. As a result, any use of the terms periodic and anti-periodic spinors is generally a choice of convention as these terms can become well-defined only after making an arbitrary choice of this sign[3]. Our convention in this work will be to use $\hat{J}_3$ as defined by (2.28). In contrast, consider the ϕ-translation used to construct rBTZ as a quotient of AdS_3. This latter isometry already preserves the vielbein (2.24), so we may take the corresponding extra action j on spinor indices to be trivial.

2.3.2 Computing $\langle T_{kk} \rangle$

For bosonic fields one defines the source for the Einstein equations by

$$T_{\mu\nu} = \frac{-2}{\sqrt{g}} \frac{\delta S_{\text{matter}}}{\delta g^{\mu\nu}}. \tag{2.29}$$

[3]This case is like what one often sees in flat space, where the discrete isometry is a special case of a continuous isometry. Continuous isometries lift uniquely from $O(d-1,1)$ to the associated spin group and can be used to define natural notions of periodic and anti-periodic spinors.

For spinor fields, as in (2.15), the Lagrangian is best written in terms of the vielbein so that this source is defined via

$$T_{\mu\nu} = -\frac{1}{\det(e)} \frac{\delta S_{\rm matter}}{\delta e^{\lambda}_{a}} \left(\delta^{\lambda}_{\mu} e_{a\nu} + \delta^{\lambda}_{\nu} e_{a\mu} \right). \tag{2.30}$$

Note that (2.30) reduces to (2.29) when the vielbein appears in the action only through $g^{\mu\nu} = e^{\mu}_{a} \eta^{ab} e^{\nu}_{b}$. Evaluating (2.30) for a spinor field yields the Belinfante stress-energy tensor [99, 100]

$$T_{\mu\nu} = \frac{i}{2} \left[\overline{\psi} (\gamma_{(\mu} D_{\nu)} \psi) + (\overline{D}_{(\nu} \overline{\psi}) \gamma_{\mu)} \psi \right], \tag{2.31}$$

where $D_{\mu} \equiv \partial_{\mu} + \frac{1}{2} \omega^{ab}_{\mu} \Sigma_{ab}$ is the covariant derivative with spin connection ω^{ab}_{μ} and $\overline{D_{\nu}} \equiv \partial_{\mu} - \frac{1}{2} \omega^{ab}_{\mu} \Sigma_{ab}$. This tensor is both symmetric and conserved.

Having understood the action $\hat{J}_3$ on spinors of the appropriate rBTZ $\times S^1$ isometry J_3 in section 2.3.1 above, we can use the method of images to compute the expectation value $\langle T_{kk} \rangle$ much as for scalars. The critical relation is

$$\psi(x) \equiv \frac{1}{\sqrt{2}} \left(\widetilde{\psi}(x) + \left[\hat{J}_3 \widetilde{\psi} \right](x) \right), \tag{2.32}$$

where we have chosen signs such that (2.32) is a periodic spinor under our $\hat{J}_3$. Recall also that $\left[\hat{J}_3 \widetilde{\psi} \right](x) = j \widetilde{\psi}(J_3 x)$. Inserting this into (2.31) and taking expectation values in the rBTZ $\times S^1$ Hartle-Hawking state again yields 4 terms. Two of these are the expectation values of $\langle T_{kk} \rangle$ on rBTZ$\times S^1$ at x and at $J_3 x$, which for x on the horizon must again vanish by symmetry in the Hartle-Hawking state. We thus need only compute the remaining cross-terms associated with distinct points $x, J_3 x$ in the covering space. As for scalars, in the full four-dimensional stress-energy tensor such terms are manifestly nonsingular for all x.

However, again as for scalars it will be useful to Kaluza-Klein reduce our spinor to a

tower of spinors on AdS_3 by decomposing ψ into Fourier modes $e^{ip\theta}$ on the internal S^1 (we assume periodicity on this circle). The resulting three-dimensional spinors have masses that are again given by (2.12). As noted earlier, this reduction yields three-dimensional spinors in both representations.

For each p we may write the associated 3D stress-energy tensor $\langle T_{kk} \rangle_p$ on the horizon in terms of the 3D spinor propagator $S_{BTZ}{}^{\alpha}_{\beta'}(x, x') = \langle \psi^{\alpha}(x) \bar{\psi}_{\beta'}(x') \rangle_{BTZ}$ for a fermion of the appropriate effective mass and choice of spinor representation in the BTZ Hartle-Hawking state

$$\langle T_{\mu\nu} \rangle_p = (-1)^p \frac{i}{2} \lim_{x \to x'} \sum_{A,B} \mathrm{Tr} \left[\gamma_{(\mu} D_{\nu)} S_{BTZ}(x, J_3 x') \widetilde{j} + \widetilde{j} \overline{D_{(\nu'}} S_{BTZ}(J_3 x, x') \gamma_{\mu)} \right], \quad (2.33)$$

where $D_{\nu'}$ denotes a covariant derivative on the second argument that acts on the associated spinor indices and $\Sigma_{A,B}$ indicates that the right-hand-side adds togeher the contributions from the two three-dimensional representations. The null-null component of this stress-energy tensor can be rewritten

$$\langle T_{kk} \rangle_p = (-1)^{p+1} \sum_{A,B} \mathrm{Im} \{ \mathrm{Tr} \left[\widetilde{j} \not{k} D_k S_{BTZ}(x, J_3 x) \right] \}. \quad (2.34)$$

The propagator S_{BTZ} is further given by an image sum $S_{\mathrm{BTZ}} = \sum_{n \in \mathbb{Z}} S_{\mathrm{AdS}_3}$ over propagators in AdS_3.

Using the maximal symmetry of this spacetime, the AdS_3 propagators can be written

$$S_{\mathrm{AdS}_3}(x, x') = [\alpha(s) + \not{\eta} \beta(s)] \Lambda(x, x')] \quad (2.35)$$

in terms of the spinor parallel propagator Λ along the geodesic connecting x and x', and two functions $\alpha(s), \beta(s)$ of the associated geodesic distance s, and the tangent vector

$n^\mu = \frac{\partial}{\partial x^\mu} s(x, x')$ to this geodesic at x. This construction and the relevant formulae are reviewed in appendix B, following the same procedure as used by [101] in the Euclidean case. In particular, the explicit form of $\alpha(s)$ is given by combining (B.14), (B.17), and (B.19), whence $\beta(s)$ then follows from the second line of (B.9).

However, it remains to find an expression for the spinor parallel propagator Λ. This of course depends on our choice of $SO(2,1)$ gauge, and thus on our choice of vielbein. Rather than compute the result directly for the choice (2.24) and the relevant geodesics, it is simpler to proceed by noting that, for each n and x, the contribution to (2.33) must be invariant under the combined action of AdS_3 isometries, diffeomorphisms, and internal $SO(2,1)$ gauge transformations and using such transformations to separately simplify the computations for each n, x. In particular, we may fix an auxiliary (say, nonrotating) BTZ coordinate system on AdS_3 and then use AdS_3 isometries to map the geodesic segment running from x to the relevant image point x' onto the bifurcation surface of the (auxiliary) BTZ horizon. Using the (nonrotating version of the) vielbein (2.24), one then finds the spinor parallel propagator Λ to be trivial along such geodesics, with $\Lambda(s) = \mathbb{1}$ for all s.

To proceed further, it is useful to note that the quotient $\widetilde{M} = (\text{rBTZ} \times S^1)/J_3$ can also be generated by taking the quotient of AdS_3 under the group generated by both an rBTZ ϕ-translation (for which $AdS_3/\mathbb{Z} = \text{rBTZ}$) and a π rotation in AdS_3 global coordinates. In terms of the standard embedding coordinates reviewed in appendix A, this is a rotation in the (X_1, X_2) plane. For each n, we may choosing the lift of x to AdS_3 so that the geodesic from x to the relevant image point x' intersects the axis of this rotation at the midpoint of the geodesic, and we may then choose the isometry moving the geodesic to the bifurcation surface of our auxiliary BTZ coordinates to be just a rotation around the same axis followed by a translation along it. In particular, we may choose this AdS_3 isometry to preserve the relevant axis so that the action $\widetilde{j}$ on spinor indices

of our extended isometry $\hat{J}_3$ is unchanged. Noting that (2.25) holds in our auxiliary nonrotating BTZ $SO(2,1)$ frame as well as in the physical one, the expression (2.28) for $\widetilde{j}$ must continue to hold in this frame as well. Computing each contribution to (2.34) in the associated auxiliary frame and summing over n and the choice of representations then yields

$$\langle T_{kk} \rangle_p (x) = -4(-1)^p \sum_{n \in \mathbb{Z}} (\varphi_n)_\mu k^\mu \, (n_\mu k^\mu) \left[\left(\frac{d}{ds_n} - \frac{1}{2\ell} \coth \frac{s_n}{2\ell} \right) \beta(s_n) \right], \tag{2.36}$$

where $(\varphi_n)_\mu$ is the unit-normalized vector at x whose lift to AdS_3 points along the infinitesimal generator of rotations of X^1 into X^2 when the geodesic from x to its image point x' is lifted to AdS_3 so as to intersect the associated rotation axis at its midpoint, and where s_n is the geodesic distance in AdS_3 between x and x'. In particular, both s_n and φ_n depend on n.

As reviewed in appendix A, in terms of our rBTZ coordinates the explicit form of the geodesic distance between two points x, x' in AdS_3 can be written

$$\begin{aligned} s(x,x') = \ell \cosh^{-1} \Bigg(\frac{1}{(UV+1)(U'V'+1)} \Bigg[(UV-1)(U'V'-1) \cosh \left(\frac{r_+(\phi-\phi')}{\ell} \right) + \\ 2(UV'+VU') \cosh \left(\frac{r_-(\phi-\phi')}{\ell} \right) + 2(VU'-UV') \sinh \left(\frac{r_-(\phi-\phi')}{\ell} \right) \Bigg] \Bigg). \end{aligned} \tag{2.37}$$

Note that choosing x on the horizon and x' as above imposes $U' = V = 0$, $V' = U$, and $\phi' = -\phi$, simplifying the result to

$$\begin{aligned} s(U,\phi) &= \ell \cosh^{-1} \left(\cosh \left(\frac{2r_+\phi}{\ell} \right) + 2U^2 \exp \left(-\frac{2r_-\phi)}{\ell} \right) \right) \\ &\equiv \ell \cosh^{-1} \left(1 + \frac{2\rho^2}{\ell^2} \right) \end{aligned} \tag{2.38}$$

where $\rho = \ell\sqrt{\sinh^2\left(\frac{r_+\phi}{\ell}\right) + U^2\exp\left(-\frac{2r_-\phi}{\ell}\right)}$ is the radial coordinate defined by either x or x' in a global AdS_3 coordinate system whose rotation axis orthogonally intersects the midpoint of the geodesic from x to x'; in other words, $2\pi\rho$ is the circumference of the circle defined by rotating either x or x' about this axis.

2.3.3 General Features

Since the general form of $\langle T_{kk}\rangle$ is quite complicated, we will compute the details of stress-energy tensor profiles and the associated back-reaction on the metric numerically in section 2.3.4 below. However, it is useful to first discuss certain general features of (2.36).

In the scalar case, [92] found the expressions to simplify greatly in the limit $r_+(\phi - \phi')/\ell, r_-(\phi-\phi')\ell \to 0$, which in particular holds for the $n = 0$ term near $\phi = 0$. This is even more true in our case, as $\varphi_\mu k^\mu$ vanishes in this limit. For $n = 0$ we find

$$\langle T_{kkn=0,p} = \frac{Ue^{-3r_-\phi}\sinh(r_+\phi)\left(\rho+\sqrt{1+\rho^2}\right)^{-2m_3(p)}}{8\pi\rho^5\left(1+\rho^2\right)^2}$$
$$\left[3+\left(6+4m_3^2(p)\right)\rho^2+\left(3+4m_3^2(p)\right)\rho^4+6m_3(p)\rho\sqrt{1+\rho^2}+8m_3(p)\rho^3\sqrt{1+\rho^2}\right] \tag{2.39}$$

with ρ defined as before and where we have set ℓ to 1.

The expression (2.39) is singular at $\rho = 0$, or $U = \phi = 0$. It is useful to understand this singularity since, as described above, the four-dimensional stress-energy tensor can have at most an integrable singularity at this point. For $p \neq 0$ and $r_+ \neq 0$ the terms $\langle T_{kk}\rangle_{n=0,p}$ are finite, so any nonintegrable singularity must cancel when (2.39) is summed over p. This is precisely what occurs in the scalar case studied in [92].

However, in our case the singularity in (2.39) is separately integrable for each p.

Writing $e^{-\frac{r_-\phi}{\ell}}U = \rho\cos\theta$ and $\sinh\frac{r_+\phi}{\ell} = \rho\sin\theta$ yields

$$\langle T_{kk}\rangle_{n=0,p} = \frac{3\sin\theta\cos\theta}{8\pi\rho^3}\left(1 - 2\frac{r_-}{r_+}\cos\theta\rho + O(\rho^2)\right) \tag{2.40}$$

at small ρ. The result simply vanishes for any U at $\phi = 0$ ($\sin\theta = 0$) for any U, or for any ϕ at $U = 0$ ($\cos\theta = 0$). Integrating over U at fixed ϕ thus raises no issues. Integrating over both ϕ and U also yields a finite result since $\int d\theta \sin\theta\cos\theta = 0 = \int d\theta \sin\theta\cos\theta$, so that integral of the explicit terms in (2.40) also vanish, and all other terms give finite results due to the fact that the measure $dUd\phi \propto \rho d\rho d\phi(1+O(\rho^2))$ supplies an additional factor of ρ.

The expression (2.39) has many similarities to the scalar expression (2.11). One might expect a particularly simple relation between the two in the large mass limit where occupation numbers are small and quantum effects are suppressed so that the choice of bosonic vs. fermionic statistics is unimportant. But the kinematic structure of the expressions remains different in that limit, associated with the nontrivial action $\widetilde{j}$ of the isometry $\hat{J}_3$ on fermionic indices. In particular, at large m_3 the spinor result (2.39) yields

$$\langle T_{kk}\rangle_{n=0,\psi} = \frac{U\sinh(r_+\phi)\, e^{-3r_-\phi}\left(\rho + \sqrt{1+\rho^2}\right)^{-2m_3}}{2\pi\rho^3\left(1+\rho^2\right)} m_3^2 + \mathcal{O}(m_3), \tag{2.41}$$

while $\Delta \to 1 + m$ in the bosonic expressions (2.11) gives

$$\langle T_{kk}\rangle_{n=0,\phi} = \frac{U^2 e^{-4r_-\phi}\left(\rho + \sqrt{1+\rho^2}\right)^{-2m}}{8\pi\rho^3\left(1+\rho^2\right)^{3/2}} m^2 + \mathcal{O}(m). \tag{2.42}$$

The fact that the denominators differ by a factor of 4 may be ascribed to the fact that we consider a single real scalar and a four-component spinor (from the four-dimensional

point of view). However, the other discrepancies reflect the difference in kinematics. We find similar differences in the limit of large U with m fixed, which yields

$$\frac{\langle T_{kk}\rangle_{n=0,\psi}}{\langle T_{kk}\rangle_{n=0,\phi}} = 4\sinh(r_+\phi) + \mathcal{O}(1/U^2). \tag{2.43}$$

It is interesting that the above expressions for our fermion field all change signs under $\phi \to -\phi$. This stands in marked contrast to the scalar results whose signs are generally ϕ-independent[4]. This odd-parity behavior turns out to arise from the nonrotating limit $r_-(\phi - \phi')\ell \to 0$ regardless of whether a similar limit is taken for r_+. Indeed, for nonrotating BTZ there is no preferred sign of the unit-vector φ_μ and the symmetry of nonrotating BTZ under $\phi \to -\phi$ requires $(\varphi_n)_\mu k^\mu \to -(\varphi_n)_\mu k^\mu$. Since in that case all other factors in $\langle T_{kk}\rangle$ are even, the integrated stress-energy becomes an odd function of ϕ and the average over the full horizon must vanish. But this symmetry is broken by rotation and, indeed, we will find below that for nonzero angular velocity the average of $\langle T_{kk}\rangle$ over the full horizon is nonzero.

2.3.4 Numerical Results

It remains to study $\langle T_{kk}\rangle$ and $\int dU \langle T_{kk}\rangle$ in detail. For this task we resort to numerics and follow the same basic strategy as in [92]. We will phrase our results in terms of the dimensionless quantity $\ell \langle T_{kk}\rangle \propto \Delta V$. As above, our results are for the effective three-dimensional Kaluza-Klein-reduced stress-energy tensor. We impose a cutoff N on the number of Kaluza-Klein modes over which we sum and also regulate the functions near the (integrable) singularity at $U = 0, \phi = 0$. We choose our cutoffs such that our answers do not change significantly when these cutoffs are altered.

[4]Though with specially engineered boundary conditions [92] found cases where the sign of the integrated scalar stress-energy depends on ϕ as well.

In particular, our numerical expressions are computed via

$$\langle T_{kk}\rangle_{numerical} = \sum_{p=-N}^{N} (-1)^p f(U, m(p)) + (-1)^{N+1} f(U, m(N+1)) \tag{2.44}$$

where we have added an extra term so as to sum over an even number of terms, N of which have an additional sign change. In computing $\left(\int dU \langle T_{kk}\rangle\right)_{\text{numerical}}$ we integrate the stress-energy tensor only over $|U| > \epsilon$. We will interpolate between the origin and $U = \epsilon$ with a linear approximation for the spinor and a constant for the scalar, as shown in Figure 2.3 and integrate the interpolating function for $|U| < \epsilon$. We choose our spinors to be periodic under $\hat{J}_3$ such that we get a positive overall contribution to the stress-energy tensor.

Some results for the dimensionless quantity $\ell \langle T_{kk}\rangle$ are shown in figures 2.2 and 2.3. Figure 2.2 shows the contributions to the stress-energy tensor for a spinor and scalar of the same mass at a particular value of ϕ. Notably, the spinor contribution is everywhere positive, as opposed to the scalar contribution which changes sign. Figure 2.3 shows the details of the interpolation of $\langle T_{kk}\rangle$ for small U. As the spinor stress-energy tensor vanishes at $U = 0$, a linear interpolation from the origin was used, while a constant interpolation was used for the scalar case.

We also numerically calculate the contributions to $\int dU \langle T_{kk}\rangle$ at extremality in figure 2.4 for various values of ϕ and $r_+ = r_-$. As opposed to the scalar contribution, the spinor contribution generically changes sign at different ϕ for a given r_+. In the scalar case, this phenomenon must be engineered [92] by requiring some KK modes to have the $(-)$ boundary condition in (2.8). Figure 2.5 plots the integral of $\langle T_{kk}\rangle$ over both U and ϕ at extremality as a function of the radius $r_+ = r_-$. The physical importance of this quantity is that, as discussed in section 2.2.4, it is proportional to $T \langle V\rangle_{\text{average}}$ (see (2.23)). As in [92], we find numerically that this function is independent of r_+. We thus find a large

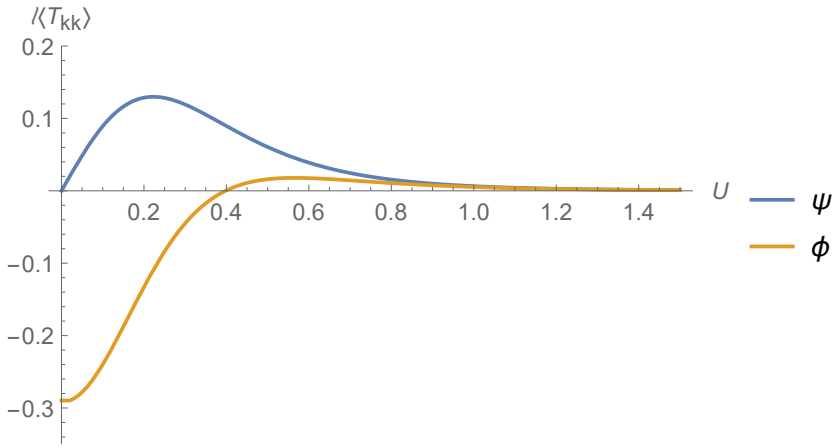


Figure 2.2: Plot of $\ell \langle T_{kk} \rangle$ vs. U with $k^\mu \partial_\mu = \partial_U$ for a spinor field ψ and a scalar field ϕ. Both particles have mass $m = 1$ in units of inverse ℓ_{AdS}. We have chosen $r_+ = 1$ and $r_- = 1/2$ in the same units, as well as $\phi = 0.1$. Here the cutoff $\epsilon = 0.02$, the sum over BTZ images has been done to $n = 3$, and the sum over KK modes has been performed to $N = 50$, with $\ell / R_{S^1} = \sqrt{10}$.

average time advance $\langle V \rangle_{\text{average}} \propto 1/T$ as $T \to 0$ in the extreme limit $r_- \to r_+$, suggesting that a nonperturbative treatment may lead to an eternally traversable wormhole as in [92], at least in the presence of some large parameter that controls quantum fluctuations relative to the mean. Figure 2.6 compares the relative size of this quantity at extremality for one four-dimensional complex scalar field and four four dimensional real scalar fields, which as mentioned before, have the same number of degrees of freedom. For all values of the three-dimensional mass m, the spinor contribution to the stress-energy tensor is less than that of the equivalent scalar fields, owing to cancellations from the spinor kinematic structure.

2.4 Conclusion

We have studied the contributions of bulk spinors to the integrated null stress-energy tensor of a $\mathbb{Z}_2$ quotient of rBTZ $\times$ S^1 known as the Kaluza-Klein zero-brane orbifold

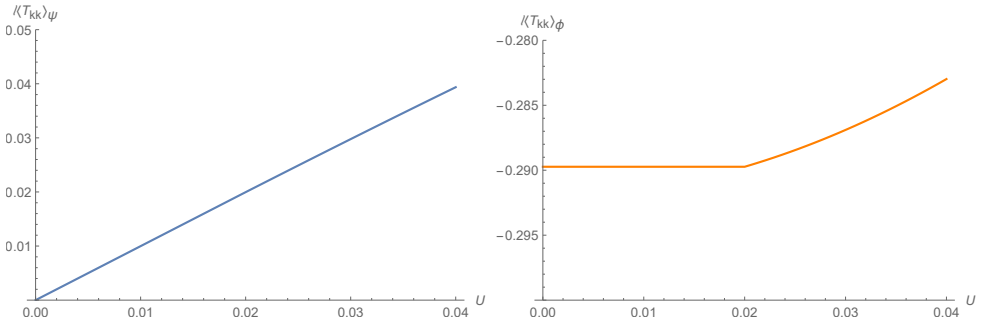


Figure 2.3: Details of the interpolation between the origin and $U = \epsilon = 0.02$ for the spinor and scalar in Figure 2.2.

(KKZBO). As in the scalar case, the $\mathbb{Z}_2$ quotient is associated with a sign that controls the periodicity of the bulk field as well as the overall sign in the null stress-energy $\langle T_{kk} \rangle$ on the horizon.

The fact that spinor fields carry Lorentz indices leads to notable differences from the scalar case. The spinor $\langle T_{kk} \rangle$ includes an overall factor of $\varphi_\mu k^\mu$, where φ is a vector field defined by the $\mathbb{Z}_2$ quotient operation. This $\varphi_\mu k^\mu$ is an odd function of ϕ for nonrotating BTZ, so in that case the spinor $\langle T_{kk} \rangle$ is odd as well and integrates to zero over the ϕ-circle. As a result, at least in the limit of large black hole radius where the Green's function (2.21) relating $\langle T_{kk} \rangle$ to the generator-dependent time delay $\langle \Delta V \rangle$ becomes short-ranged, without rotation one finds the wormhole to be traversable when entered across one half of the ϕ-circle (say for $\phi \in (0, \pi)$) but to remain nontraversable when entered across the other half-circle.

This overall factor of $\varphi_\mu k^\mu$ is associated with the fact that the simplicity of the model in four dimensions means that both possible three-dimensional representations of the Clifford algebra enter with equal weight. Terms that for each representation are not proportional to $\varphi_\mu k^\mu$ have opposite signs in the two representations and cancel. As a result, a more chiral construction that caused the two representations to enter on a less equal footing would not have this property.

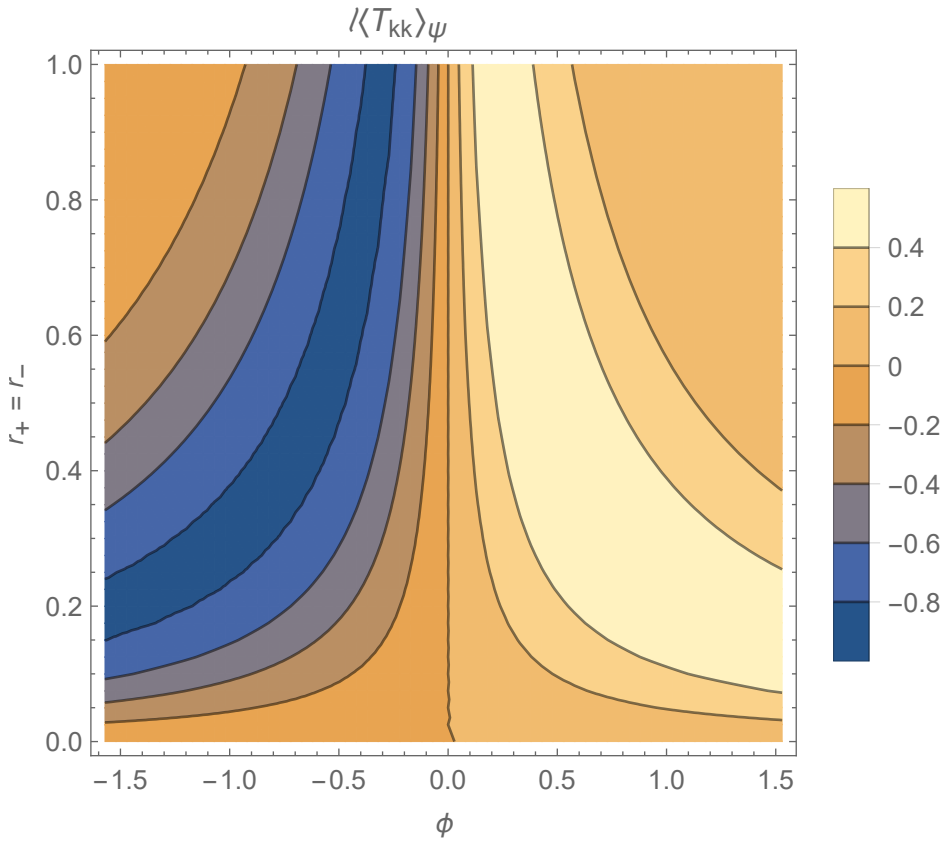


Figure 2.4: Plot of $\ell \langle T_{kk} \rangle$ at extremality for a spinor ψ at various values of $r_+ = r_-$ and ϕ. The four dimensional mass $m = 1$ and $\ell/R_{S^1} = \sqrt{10}$. The cutoff $\epsilon = 0.1$, the sum over Kaluza-Klein modes was performed to $N = 20$, and the sum over BTZ images was performed to $n = 3$.

While the above $\phi \to -\phi$ (anti-)symmetry is broken at nonzero angular velocity, the sign of $\langle T_{kk} \rangle$ still varies with ϕ and the average over the ϕ-circle is correspondingly reduced relative to the scalar case. But choosing the aforementioned sign correctly still makes the average negative (and thus also $\langle V \rangle_{\text{average}}$). As in the scalar case [92], the quantity $T \langle V \rangle_{\text{average}} < 0$ is nonzero at $T = 0$, so that the time delay $- \langle V \rangle_{\text{average}} \propto 1/T$ becomes large. At least in the presence of some large parameter that controls quantum fluctuations relative to the mean, this suggests that a nonperturbative treatment may lead to an eternally traversable wormhole as in [91]. It also interesting that, again as in

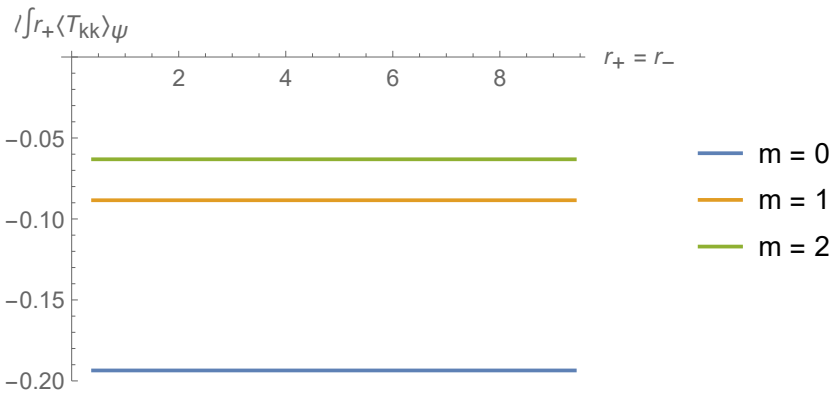


Figure 2.5: Plot of the quantity $\ell \int_0^\infty \int_{-\pi/2}^{\pi/2} d\phi dU r_+ \langle T_{kk} \rangle$ at extremality for $m = (0, 1, 2)$ and $\ell/R_{S_1} = 10$. The cutoff $\epsilon = 0.01$, the sum over Kaluza-Klein modes modes was performed to $N = 200$, and the sum over BTZ images was performed to $n = 3$. This quantity appears constant for all values of r_+, up to corrections at small r_+ for the small value of N. Relevant values of the four-dimensional mass m are indicated above.

the scalar case studied in [92], numerical results suggest $T \langle V \rangle_{\text{average}}$ to be independent of $r_+ = r_-$ at extremality, though with values noticeably smaller than in the scalar case (see figure 2.6) due to partial cancellations associated with the variation in sign with ϕ described above.

Such results in particular make clear that, even though extreme BTZ preserves certain supersymmetries, models with $N = 1$ bulk supersymmetry do not lead to special cancellations in $\langle T_{kk} \rangle$ between bosons and fermions on the extreme KKZBO spacetime. This lack of cancellations is especially manifest at large mass m where AdS bulk supersymmetry relates fermions and bosons of nearly equal masses. It should not be a surprise since the nonvanishing of $\langle T_{kk} \rangle_{\text{KKZBO}}$ is manifestly due to the breaking of the BTZ Killing symmetry by the $\mathbb{Z}_2$ quotient used to build KKZBO from extreme BTZ. Since this Killing symmetry is part of the supersymmetry algebra on extreme BTZ, the quotient must break supersymmetry at the same level.

A natural next step for further investigation would be to study fields of even higher

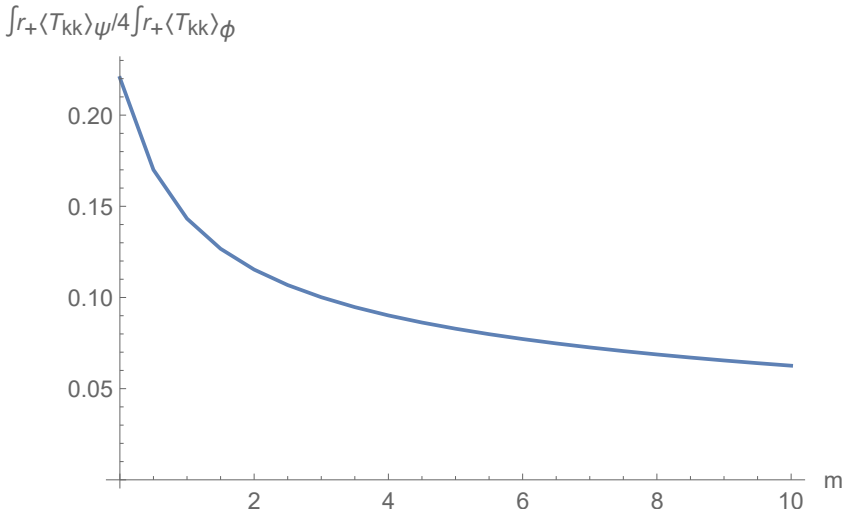


Figure 2.6: Plot of the ratio of the quantity $\int_0^\infty \int_{-\pi/2}^{\pi/2} d\phi dU r_+ \langle T_{kk} \rangle$ at extremality for one four-dimensional complex spinor field and four four-dimensional real scalar fields of the same four-dimensional mass m. $\ell/R_{S_1} = 10$ with a cutoff $\epsilon = 0.01$. The sum over Kaluza-Klein modes was performed to $N = 1000$ and the sum over BTZ images was performed to $n = 5$. The spinor contribution to the integrated stress-energy tensor is always significantly less than that of the scalars with an equivalent number of degrees of freedom.

spin, and in particular to better understand what kinematic factors might arise in such cases. Expressions for the vector propagator for both massive and massless particles on AdS_3 were given in [102] and can be used to compute the stress-energy of such fields on the KKZBO spacetime in direct analogy to the computations performed here. Further extensions to spin-3/2 and spin-2 fields would then allow one to study gravitino and graviton back-reaction to our wormhole geometry. It would also be interesting to study the back-reaction of signals passing through our wormholes as was done for GJW wormholes in [84, 87, 103, 104] to understand how differences from the scalar case above affect limits on the amount of information that can be transmitted.

Chapter 3

Replica Wormholes and Holographic Entanglement Negativity

3.1 Introduction

Replica wormholes have played an important role in recent progress on solving the black hole information problem [60, 61]. These wormholes appear as nontrivial saddle points that could dominate gravitational path integrals with replicated boundary conditions. Their appearance leads to nontrivial "island" contributions in the quantum extremal surface (QES) formula for gravitational entropy [44, 45, 53, 54].

So far most of the discussion has been centered on the von Neumann entropy. While obtaining the von Neumann entropy is a good first step, we need more detailed information about the quantum state — such as more general measures of entanglement — to fully solve the black hole information problem.

In this chapter, we take a first step towards understanding the structure of entanglement in an evaporating black hole and its Hawking radiation by studying entanglement negativity and its Rényi generalizations in a couple of toy models. Just as the von Neu-

mann entropy is a measure of quantum entanglement in pure states, the negativity is an important measure of entanglement in generally mixed states. Therefore, the negativity provides an interesting probe in diagnosing the structure of multipartite entanglement in systems such as an evaporating black hole.

To understand negativity intuitively, consider a general state on two subsystems that is described by a density matrix, and take its partial transpose on the second subsystem. The partially transposed density matrix could have negative eigenvalues, and the degree to which the eigenvalues are negative is characterized by the negativity and logarithmic negativity. Both of these negativity measures are entanglement monotones, and the logarithmic negativity provides an upper bound on the distillable entanglement [105–107]. Negativity has been discussed in a number of interesting prior works [108–130].

We now give a short summary of this chapter.

In Section 4.3, we review the definition and properties of the negativity and its Rényi generalizations. In Section 3.3, we start our study of negativity in a toy model of an evaporating black hole in Jackiw-Teitelboim (JT) gravity with an end-of-the-world (EOW) brane. This is a slight generalization of the model studied in [61], with the system describing the Hawking radiation divided into two subsystems so as to study negativity.

As we tune the parameters of the model, we find a rich phase diagram for the negativities consisting of four phases (see Figure 3.4). Each of the four phases is dominated by a saddle-point geometry of JT gravity (or a set of saddle points). For a black hole before the Page time, we find a phase dominated by a totally disconnected geometry, whereas after the Page time, we find three distinct phases depending on how we divide the radiation system into two subsystems: the first phase is dominated by a cyclically connected geometry (which is the replica wormhole of [61]), the second dominated by an "anti-cyclically" connected geometry, and the third dominated by pairwise connected geometries that are in one-to-one correspondence with noncrossing pairings. These pairwise connected

geometries are new replica wormholes that spontaneously break the replica symmetry. Their appearance agrees with the general discussions on holographic negativity in [128].

In Sections 3.4 and 4.4, we study the behavior of negativities near the transitions between the four phases. Near these phase transitions, more geometries than the four types described earlier could dominate the gravitational path integral for Rényi negativities, and we need to sum over them. In order to obtain the negativity and logarithmic negativity (as well as related negativity measures such as the partially transposed entropy [118, 128]), we need to analytically continue in the replica number. We achieve this by using the resolvent for the partially transposed density matrix to find its eigenvalue distribution (which we call the "negativity spectrum"). To calculate this "negativity resolvent," we organize the sum over geometries into a Schwinger-Dyson equation, which is similar to the method used in [61]. We develop this method for negativity in Section 3.4 and apply it to both a microcanonical ensemble and canonical ensemble in Section 4.4.

When the black hole is in a microcanonical ensemble, the Schwinger-Dyson equation simplifies into a cubic equation for the negativity resolvent, leading to concrete results for the negativities near all phase transitions. This is similar to the case of a random mixed state studied in [127].

When the black hole is in a canonical ensemble, the gravitational calculation is technically more difficult. As a result, we study each of the phase transitions separately, for we only need to sum over a subset of geometries near each transition. Near the transition between the "disconnected" phase and "pairwise" phase, the Schwinger-Dyson equation again simplifies, this time into a quadratic equation for the negativity resolvent. Near the transition between the "cyclic" phase and "pairwise" phase, it is difficult to solve the Schwinger-Dyson equation exactly, but we solve it approximately in the semiclassical, or $\beta \to 0$, limit. From its solution, we find that the negativity spectrum near the phase transition consists of two branches, each of which is approximately a shifted thermal

spectrum with a cutoff. One branch consists of positive eigenvalues, and the other has negative eigenvalues. From this we find enhanced corrections to various negativity measures near the phase transition. In particular, a quantity known as the refined Rényi-2 negativity receives an $\mathcal{O}(1/\sqrt{\beta})$ correction, similar to the enhanced corrections to the von Neumann entropy at the Page transition [61, 131, 132], whereas other negativity measures such as the logarithmic negativity and the partially transposed entropy exhibit $\mathcal{O}(1/\beta)$ corrections, similar to what happens to Rényi entropies S_n with $n < 1$.

Moving beyond the JT gravity model, we study in Section 3.6 the behavior of negativities in a topological model of 2-dimensional gravity with EOW branes. This is a slight generalization of the model of [62], where we again divide the radiation system into two subsystems to study negativity. We find the situation to be very similar to the case of a microcanonical ensemble in JT gravity described earlier. In particular, the Schwinger-Dyson equation again simplifies into a cubic equation for the negativity resolvent, leading to concrete results for the negativities.

We end with some concluding remarks in Section 3.7 and several appendices. In Appendix C, we derive the set of dominant geometries in each of the phases and near phase transitions. In Appendix D, we provide a more detailed analysis near the transition between the cyclic phase and pairwise phase in the canonical ensemble. In Appendix E, we study the Rényi entropies near the Page transition in a similar fashion and show that they exhibit corrections analogous to the corrections to the negativities near the phase transition.

Related works appeared recently and have some partial overlap with our results on the study of negativity in JT gravity in the microcanonical ensemble [133] and the canonical ensemble [130, 134].

3.2 Entanglement Negativity and Rényi Negativities

The motivation for entanglement negativity comes from the Peres-Horodecki criterion [135, 136], also known as the PPT (positive partial transpose) criterion for mixed states, which we review here. Consider a mixed state ρ_{AB} defined on the product Hilbert space $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$. A state is separable if it can be written as

$$\rho_{AB} = \sum_{i=1}^{k} p_i \rho_A^i \otimes \rho_B^i, \qquad \sum_{i=1}^{k} p_i = 1 \tag{3.1}$$

for states ρ_A^i and ρ_B^i on $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively. Separable states are classical mixtures of product states and thus do not contain quantum entanglement; inseparable states are said to be entangled.

We denote the algebra of operators on $\mathcal{H}_i$ by $\mathcal{A}_i$, and the space of linear maps from $\mathcal{A}_A$ to $\mathcal{A}_B$ by $\mathcal{L}(\mathcal{A}_A, \mathcal{A}_B)$. A map $\Lambda \in \mathcal{L}(\mathcal{A}_A, \mathcal{A}_B)$ is said to be positive if

$$\Lambda : \mathcal{A}_A \to \mathcal{A}_B \tag{3.2}$$

maps positive operators to positive operators, and is completely positive if for all non-negative integer n,

$$\Lambda_n \equiv \Lambda \otimes \mathbb{I} : \mathcal{A}_A \otimes \mathcal{M}_n \to \mathcal{A}_B \otimes \mathcal{M}_n \tag{3.3}$$

is positive, where $\mathcal{M}_n$ denotes the algebra of $n \times n$ complex matrices. For separable states, this condition is clearly satisfied when Λ is a positive map, as $(\Lambda \otimes \mathbb{I})(\rho_A \otimes \rho_B) = (\Lambda \rho_A) \otimes \rho_B \geq 0$. For inseparable states, this no longer holds in general, so a good diagnostic of entanglement would be a positive but not completely positive map, such that entangled states would have negative eigenvalues under the action of $\Lambda \otimes \mathbb{I}$.

The partial transpose is such a positive but not completely positive map. Consider

a bipartite system AB with an orthonormal basis $\{|a\rangle\}$ on A and $\{|b\rangle\}$ on B. Given a density matrix ρ_{AB} on AB, we define the partially transposed density matrix as[1]

$$\left\langle a,b|\rho_{AB}^{T_B}|a',b'\right\rangle = \langle a,b'|\rho_{AB}|a',b\rangle\,. \tag{3.4}$$

Acting on a reduced density matrix on B, the partial tranpose becomes the usual transpose which preserves the eigenvalues of the original reduced density matrix and is therefore a positive map. Acting on the full density matrix is not guaranteed to preserve positivity. As an example, take an EPR pair of two qubits A and B. The partial transpose of its density matrix has eigenvalues $\{\frac{1}{2},\frac{1}{2},\frac{1}{2},-\frac{1}{2}\}$. We therefore see that the partial transpose can be a useful tool for differentiating between separable and inseparable states.[2]

The entanglement negativity $\mathcal{N}(\rho)$ is defined as the sum of the absolute values of the negative eigenvalues of this partially transposed density matrix and can be variably written as

$$\mathcal{N}(\rho_{AB}) = \frac{||\rho_{AB}^{T_B}||_1 - 1}{2} = \sum_i \frac{|\lambda_i| - \lambda_i}{2} = \sum_{i:\lambda_i<0} |\lambda_i|. \tag{3.5}$$

Here $||X||_1 \equiv \mathrm{Tr}\,|X| = \mathrm{Tr}\,\sqrt{X^\dagger X}$ is the Schatten 1-norm of a matrix X. We see why negativity is such an appealing entanglement measure, as it is computed directly from a trace, as opposed to a variational principle in the case of other entanglement measures. Note that as we are taking a trace, it does not matter which subsystem we take the partial trace over, so choosing $\rho_{AB}^{T_B}$ instead of $\rho_{AB}^{T_A}$ is merely a convention. The logarithmic

[1]For reasons that will become clear shortly, in later sections we will rename the subsystems A and B as R_1 and R_2, and the partial transposition T_B is thus called T_2.

[2]For 2×2 and 2×3 matrices, the PPT criterion is both necessary and sufficient for the state to be separable. For systems of general dimension it is not sufficient, as bound entangled states have positive semidefinite partial transpose and therefore require further entanglement criteria to be distinguished from separable states.

negativity is similarly defined by

$$\mathcal{E}(\rho_{AB}) = \log\left(\sum_i |\lambda_i|\right) = \log\big(2\mathcal{N}(\rho_{AB}) + 1\big). \tag{3.6}$$

The logarithmic negativity is an upper bound on the distillable entanglement, i.e., the asymptotic number of EPR pairs that can be extracted from a set of identically prepared ρ_{AB} with local operations and classical communication (LOCC).

We can also write a Rényi version of negativity via

$$\mathcal{N}_n(\rho_{AB}) = \mathrm{Tr}\left[\left(\rho_{AB}^{T_B}\right)^n\right]. \tag{3.7}$$

There is a subtlety in the analytic continuation of the Rényi negativity. As the negativity is defined by the absolute value of the eigenvalues of the partially transposed density matrix and the Rényi negativity is defined without an absolute value, we need to define different analytic continuations for even and odd n such that

$$\begin{aligned}\mathcal{N}_{2m}^{(\mathrm{even})} &= \sum_i |\lambda_i|^{2m},\\ \mathcal{N}_{2m-1}^{(\mathrm{odd})} &= \sum_i \mathrm{sgn}(\lambda_i)|\lambda_i|^{2m-1}.\end{aligned} \tag{3.8}$$

The logarithmic negativity is then obtained from the even analytic condition:

$$\mathcal{E}(\rho_{AB}) = \lim_{m\to 1/2} \log \mathcal{N}_{2m}^{(\mathrm{even})}\left(\rho_{AB}\right). \tag{3.9}$$

In this chapter, we will also be interested in a generalization of the Rényi negativities

termed refined Rényi negativities, which are given by

$$S^{T_B(n,\text{even})}(\rho_{AB}) = -n^2\partial_n\left(\frac{1}{n}\log\mathcal{N}_n^{(\text{even})}\right),$$
$$S^{T_B(n,\text{odd})}(\rho_{AB}) = -n^2\partial_n\left(\frac{1}{n}\log\mathcal{N}_n^{(\text{odd})}\right). \tag{3.10}$$

The refined Rényi negativities are inspired by the refined Rényi entropies defined in [137]. In particular, we will be interested in two measures that descend from the refined Rényi negativities. The first is the partially transposed entropy $S^{T_B}(\rho_{AB})$ of [118, 128], defined as the $m \to 1$ limit of the refined odd Rényi negativity:

$$S^{T_B}(\rho_{AB}) = -\frac{1}{2}\lim_{m\to 1}\partial_m \log\mathcal{N}_{2m-1}^{(\text{odd})} = -\sum_i \lambda_i \log|\lambda_i|. \tag{3.11}$$

S^{T_B} is so named in analogy with the von Neumann entropy. The other measure is the refined Rényi-2 negativity $S^{T_B(2,\text{even})}(\rho_{AB})$, which can be written as

$$S^{T_B(2,\text{even})}(\rho_{AB}) = -\lim_{m\to 1} m^2\partial_m\left(\frac{1}{m}\log\mathcal{N}_{2m}^{(\text{even})}\right) = -\sum_i \frac{\lambda_i^2}{\sum_j\lambda_j^2}\log\left(\frac{\lambda_i^2}{\sum_j\lambda_j^2}\right). \tag{3.12}$$

We will refer to this quantity as $S^{T_B(2)}$ for short. It is equivalent to the von Neumann entropy of $\left(\rho_{AB}^{T_B}\right)^2 / \operatorname{Tr}\left[\left(\rho_{AB}^{T_B}\right)^2\right]$.

3.3 The Model and Four Phases

3.3.1 JT gravity with EOW branes

We start by reviewing the simple model of black hole evaporation studied in [61] (see also [138, 139]). This model consists of a black hole in JT gravity, decorated with an

end-of-the-world (EOW) brane with tension μ. The action is given by

$$I = I_{\text{JT}} + \mu \int_{\text{brane}} dy, \tag{3.13}$$

with the JT action being

$$I_{\text{JT}} = -\frac{S_0}{2\pi}\left[\frac{1}{2}\int_{\mathcal{M}} R + \int_{\partial\mathcal{M}} K\right] - \left[\frac{1}{2}\int_{\mathcal{M}} \phi(R+2) + \int_{\partial\mathcal{M}} \phi K\right]. \tag{3.14}$$

We have set $G_N = 1$, though it can be restored by sending the inverse temperature $\beta \to \beta G_N$. The parameter S_0 can be thought of as the zero temperature entropy of an eternal two-dimensional black hole. The EOW brane is endowed with k orthonormal states, or "flavors," which are entangled with an auxiliary reference system R. The states on the brane can be thought of as describing the interior partners of the early Hawking radiation R, so by increasing k we can probe later regimes of an "evaporating" black hole.

The entangled state of the black hole system B and the "radiation" R can be written as

$$|\Psi\rangle = \frac{1}{\sqrt{k}} \sum_{i=1}^{k} |\psi_i\rangle_B |i\rangle_R . \tag{3.15}$$

The density matrix of the R subsystem is therefore

$$\rho_R = \frac{1}{k} \sum_{i,j=1}^{k} |j\rangle \langle i|_R \langle \psi_i | \psi_j \rangle_B. \tag{3.16}$$

The inner product $\langle \psi_i | \psi_j \rangle_B$ is given by a gravitational path integral with standard Dirichlet boundary conditions on an asymptotic boundary interval and Neumann boundary

conditions on the EOW branes:

$$ds^2|_{\partial\mathcal{M}} = \frac{1}{\epsilon^2}d\tau^2, \quad \phi = \frac{1}{\epsilon}, \quad \epsilon \to 0$$

$$\partial_n\phi|_{\text{brane}} = \mu, \quad K = 0 \tag{3.17}$$

as well as specifying the brane states i and j at the endpoints of the EOW brane. As was shown in [61], while naively $\langle\psi_i|\psi_j\rangle \propto \delta_{ij}$, this should be understood as an expectation value in an ensemble, and wormhole contributions indicate exponentially small fluctuations of the inner product. We illustrate the boundary conditions for the matrix elements of ρ_R as follows:

$\langle\psi_i|\psi_j\rangle =$  (3.18)

The solid black line denotes an asymptotic boundary interval for the gravitational path integral, while the blue dashed lines are index lines that impose boundary conditions for the brane states. Computing $\text{Tr}\,(\rho_R)$ means contracting the open index lines and summing over all possible geometries respecting the boundary conditions (3.17).

To study negativity, we need to consider a bipartite mixed state. To that end, we split the radiation system into two subsystems $\mathcal{H}_R = \mathcal{H}_{R_1} \otimes \mathcal{H}_{R_2}$ consisting of k_1 and k_2 states, respectively, such that $k = k_1 k_2$ and

$$\rho_{R_1R_2} = \frac{1}{k}\sum_{i_1,i_2=1}^{k_1}\sum_{j_1,j_2=1}^{k_2}|i_1,j_1\rangle\langle i_2,j_2|\langle\psi_{i_2,j_2}|\psi_{i_1,j_1}\rangle. \tag{3.19}$$

We will refer to this partitioned density matrix as ρ_R from now on. We define our partially

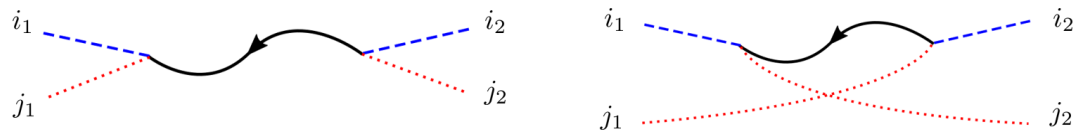


Figure 3.1: Boundary conditions for $\rho_{R_1R_2}$ and $\rho^{T_2}_{R_1R_2}$. Blue (dashed) lines denote states in R_1, and red (dotted) lines denotes states in R_2. If we take a trace, these two boundary conditions are equivalent.

transposed density matrix as the partial transpose over R_2, i.e.,

$$\rho^{T_{R_2}}_{R_1R_2} = \frac{1}{k}\sum_{i_1,i_2=1}^{k_1}\sum_{j_1,j_2=1}^{k_2} |i_1,j_2\rangle\langle i_2,j_1|\langle\psi_{i_2,j_2}|\psi_{i_1,j_1}\rangle. \tag{3.20}$$

We will use the shorthand $\rho^{T_2}_R$ moving forward. This partial transpose affects the boundary conditions for our path integral by swapping the brane flavor index lines corresponding to states in $\mathcal{H}_{R_2}$. The resulting boundary conditions are illustrated in Figure 3.1.

3.3.2 Dominant Saddles

As in any calculation with a gravitational path integral, our first task is to identify the saddle-point geometries which obey the given boundary conditions and sum over them with the appropriate weight. As our goal is to compute Rényi negativities $\mathrm{Tr}\left[\left(\rho^{T_2}_R\right)^n\right]$, our boundary conditions will consist of n copies of the boundary conditions illustrated on the right of Figure 3.1, with matching brane flavor indices contracted. The set of all classical saddles consists of oriented two-dimensional surfaces which end on the asymptotic boundaries and EOW branes, possibly connecting two or more boundaries.

As our gravitational action (3.13) is independent of brane flavor, we can factorize the flavor contributions so that

$$Z_{\text{saddle}} = Z_{\text{grav}} f(k_1,k_2) \tag{3.21}$$

for some function f of the brane Hilbert space dimensions k_1 and k_2. The gravitational

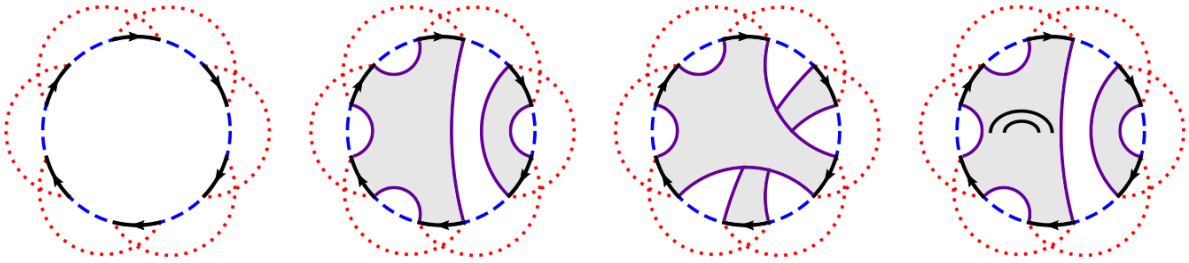


Figure 3.2: Boundary conditions and some possible classical geometries that contribute to $\mathrm{Tr}\left[(\rho_R^{T_2})^6\right]$. Index lines that run along EOW branes are not shown for visual clarity. Left: The partially transposed boundary conditions from Figure 3.1 with brane flavor indices contracted. Center left and right: The geometries consist of disjoint unions of disks in noncrossing or crossing configurations. Right: The geometry has a single handle and will be suppressed by a factor of e^{-2S_0} relative to the first geometry.

partition function Z_n for a surface connecting n boundaries depends on the Euler characteristic χ of the surface in the schematic form

$$Z_n \sim e^{S_0 \chi}. \tag{3.22}$$

The contribution of a surface with genus $g \geq 1$ is therefore suppressed by e^{-2gS_0} for large S_0. This means that the only classical geometries we need to consider are disks or disjoint unions of disks, and Z_{grav} is a product of disk partition functions Z_n. We will therefore assume $S_0 \gg 1$ throughout the rest of the chapter. We illustrate some examples of these disk geometries as well as a higher genus geometry in Figure 3.2.

More precisely, for a disk connecting n boundaries in JT gravity we have

$$Z_n = e^{S_0} \int_0^\infty ds \rho(s) y(s)^n, \quad y(s) \equiv e^{-\frac{\beta s^2}{2}} 2^{1-2\mu} \left| \Gamma\left(\mu - \frac{1}{2} + is\right) \right|^2 \tag{3.23}$$

where $\rho(s) = \frac{s}{2\pi^2} \sinh(2\pi s)$ is the disk density of states[3] in JT gravity. In order to recover

[3]The density of states is more typically written in the $E = s^2$ energy basis such that $\rho(E) = \frac{1}{4\pi^2} \sinh\left(2\pi\sqrt{E}\right)$. Here s can be thought of as an entropy, and is the more natural variable for our purposes.

(3.22), we take $S_0 \gg 1$ while keeping other parameters fixed. We emphasize this is a schematic approximation that should only be used to motivate the pertinent saddles for our problem; in general, there are parametric corrections from the full expression of Z_n which will be discussed in more detail in Sections 3.4 and 4.4.

Since we are ignoring higher genus surfaces, the sum over geometries with n replicated asymptotic boundaries is equivalent to a sum over elements of S_n, the permutation group on n elements. In particular, the sum takes the form

$$\mathrm{Tr}\left[\left(\rho_R^{T_2}\right)^n\right] = \frac{1}{(kZ_1)^n}\sum_{g\in S_n}\left(\prod_{i=1}^{\chi(g)} Z_{|c_i(g)|}\right) k_1^{\chi(g^{-1}X)} k_2^{\chi(g^{-1}X^{-1})} \tag{3.24}$$

$$\sim \frac{1}{(ke^{S_0})^n}\sum_{g\in S_n}\left(e^{S_0}\right)^{\chi(g)} k_1^{\chi(g^{-1}X)} k_2^{\chi(g^{-1}X^{-1})}, \tag{3.25}$$

where $\chi(g)$ is the number of disjoint cycles of the permutation g, $|c_i(g)|$ is the length of the i-th disjoint cycle of g, and X (X^{-1}) is the (anti-)cyclic permutation of length n. Unless otherwise specified, we will take $k, e^{S_0} \gg 1$. Note

$$\chi(\mathbb{1}) = n, \quad \chi(X) = \chi(X^{-1}) = 1. \tag{3.26}$$

The sums (3.25) and (3.24) over elements of the permutation group has a simple geometric interpretation. The permutation g determines how the asymptotic boundaries are connected by EOW branes, while the powers of k_1 and k_2 count the number of index loops. The totally disconnected geometry corresponds to $g = \mathbb{1}$, while the totally connected geometry corresponds to $g = X$. What does the $g = X^{-1}$ geometry look like? We show examples of these three geometries in Figure 3.3. Two of these geometries, the disconnected and cyclic geometries, belong in the class of noncrossing diagrams discussed by [61]. The third, the anti-cyclic geometry, is in some sense equivalent to the cyclic

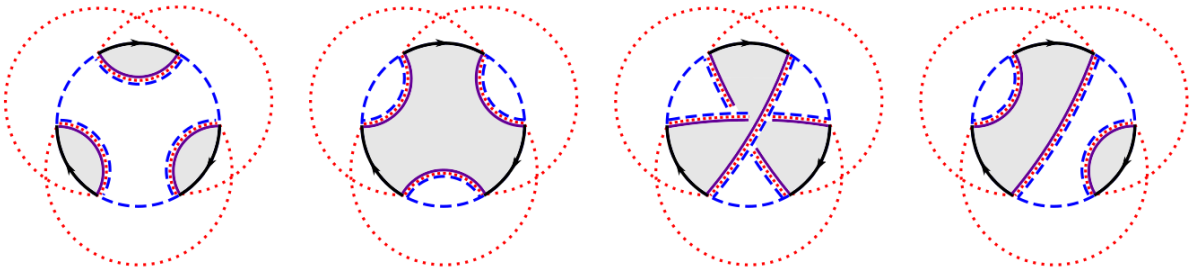


Figure 3.3: The four classes of geometries which dominate the Rényi negativity calculation. In order, they are the disconnected ($g = \mathbb{1}$), cyclic ($g = X$), anti-cyclic ($g = X^{-1}$), and pairwise ($g = \tau$) geometries. The pairwise geometries spontaneously break replica symmetry. The black lines are oriented asymptotic boundaries, the purple lines are EOW branes, the blue (dashed) lines denote k_1 index loops, and the red (dotted) lines denote k_2 index loops.

geometry if one reverses the orientation of the boundary, or equivalently if one exchanges k_1 and k_2. This statement will be explained in more detail in Section 3.4.2.

One might naively guess that the anti-cyclic geometry would never dominate the Rényi negativity, for the same reasons as in [61] where crossing partitions in the calculation of the Rényi entropy were suppressed by factors of $1/k^2$. In fact this geometry dominates in a very large parameter regime: as we show in Appendix C, we have the following phases dominated by the corresponding permutation g:

$$\begin{aligned}
&\text{Totally disconnected:} && e^{S_0} \gg k_1 k_2 && \to && g = \mathbb{1} \\
&\text{Cyclically connected:} && k_1 \gg k_2 e^{S_0} && \to && g = X \\
&\text{Anti-cyclically connected:} && k_2 \gg k_1 e^{S_0} && \to && g = X^{-1} \\
&\text{Pairwise connected:} && k_1 k_2 \gg e^{S_0}, && \to && g = \tau \\
& && e^{-S_0} \ll k_1/k_2 \ll e^{S_0}
\end{aligned} \tag{3.27}$$

To see this intuitively, we calculate the contributions to $\mathrm{Tr}\left[\left(\rho_R^{T_2}\right)^n\right]$ from these geometries. Consider the totally disconnected phase dominated by $g = \mathbb{1}$. We have $\chi(g) = n$

and $\chi(g^{-1}X) = \chi(g^{-1}X^{-1}) = 1$, so this diagram contributes schematically

$$g = \mathbb{1} \quad \Rightarrow \quad k \left(e^{S_0}\right)^n \tag{3.28}$$

to the sum in (3.25) as in [61]. It is then unsurprising that this dominates in the parameter regime $e^{S_0} \gg k$, since it is the unique diagram which maximizes the power of e^{S_0}. Note that the contribution only depends on k, and not k_1 and k_2 individually.

Now, consider the cyclically connected phase with $g = X$. Then $\chi(g) = 1$, $\chi(g^{-1}X) = \chi(\mathbb{1}) = n$, and

$$\chi(g^{-1}X^{-1}) = \chi(X^{-2}) = f(n) \equiv \begin{cases} 1, & n \text{ odd}, \\ 2, & n \text{ even}. \end{cases} \tag{3.29}$$

Hence, the cyclic diagrams contribute schematically

$$g = X \quad \Rightarrow \quad e^{S_0} k_1^n k_2^{f(n)} \tag{3.30}$$

to the sum in (3.25). This configuration maximizes the power of k_1, and therefore it is expected to become important in the parameter regimes where k_1 is comparably large. In fact, as we prove in Appendix C, it is the unique dominant diagram in the regime $k_1 \gg k_2 e^{S_0}$. Note that compared to the Rényi entropy calculation in [61] (which can be recovered by setting $k_2 = 1$ here), the cyclic geometry is suppressed by $1/k_2^{n-f(n)}$. We will also show this diagrammatically in Section 3.4.

The anti-cyclically connected phase with $g = X^{-1}$ is similar, so we will not go through the analysis. In the end, the anti-cyclic diagrams contribute schematically

$$g = X^{-1} \quad \Rightarrow \quad e^{S_0} k_1^{f(n)} k_2^n \tag{3.31}$$

to the sum (3.25). It is thus expected to become important in the parameter regimes where k_2 is comparably large, and we can prove that they are the unique dominant diagrams when $k_2 \gg k_2 e^{S_0}$.

Finally, there is one additional class of dominant geometries we should consider: the pairwise connected phase with $g = \tau$. As we show in Appendix C, these diagrams dominate in a fourth regime satisfying both $k_1 k_2 \gg e^{S_0}$ and $e^{-S_0} \ll k_1/k_2 \ll e^{S_0}$, and they are the only diagrams aside from the disconnected, cyclically connected, and anti-cyclically connected diagrams that can dominate in a large regime of the parameter space. These geometries are in one-to-one correspondence with the set of permutations τ known as noncrossing pairings. For even n, a pairwise connected geometry is constructed by choosing an element in τ, for example $(12)(34)\cdots(n-1,n)$, and connecting paired asymptotic boundaries by two-boundary wormholes. For odd n, the geometries are given by a similar noncrossing pairings of boundaries, plus a single one-boundary connected component. We show an example of such a geometry in Figure 3.3. It is evident that such geometries spontaneously break the replica symmetry.

As we show in Appendix C, each pairwise connected geometry contributes schematically

$$g = \tau \implies \left(e^{S_0}\right)^{\lceil \frac{n}{2} \rceil} k^{\lfloor \frac{n}{2} \rfloor + 1} \tag{3.32}$$

to the sum in (3.25), where $\lceil \frac{n}{2} \rceil$ and $\lfloor \frac{n}{2} \rfloor$ denote the ceiling and floor function, respectively. A pairwise connected diagram in some sense puts k_1, k_2, and e^{S_0} on the most equal footing by maximizing the sum of the three exponents in (3.25). As with the disconnected geometry, the contribution of a pairwise connected geometry only depends on k, and not k_1 and k_2 individually.

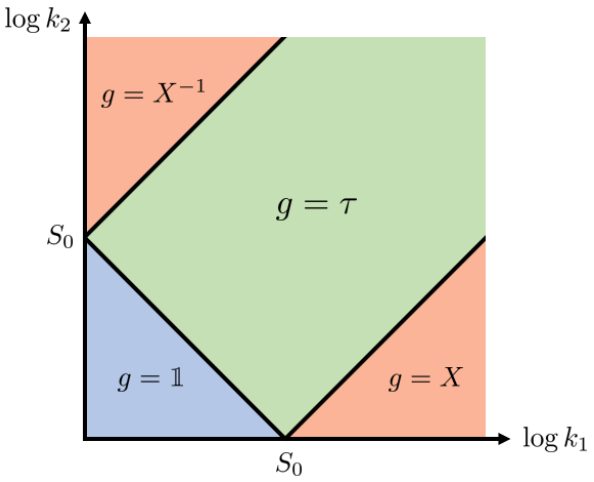


Figure 3.4: The phase diagram for entanglement negativity. The four phases are labeled by the permutations corresponding to their dominant geometries.

Negativities in Dominant Phases				
g	$\mathbb{I}$	X	X^{-1}	τ
$\mathcal{N}_{2m}^{(\text{even})}$	$\dfrac{1}{k^{2m-1}}$	$\dfrac{Z_{2m}}{k_2^{2m-2}Z_1^{2m}}$	$\dfrac{Z_{2m}}{k_1^{2m-2}Z_1^{2m}}$	$\dfrac{C_m Z_2^m}{k^{m-1}Z_1^{2m}}$
$\mathcal{N}_{2m-1}^{(\text{odd})}$	$\dfrac{1}{k^{2m-2}}$	$\dfrac{Z_{2m-1}}{k_2^{2m-2}Z_1^{2m-1}}$	$\dfrac{Z_{2m-1}}{k_1^{2m-2}Z_1^{2m-1}}$	$\dfrac{(2m-1)C_{m-1}Z_2^{m-1}}{k^{m-1}Z_1^{2m-2}}$
$\mathcal{E}$	0	$\log k_2$	$\log k_1$	$\dfrac{1}{2}(\log k - S_0) + \log\dfrac{8}{3\pi}$
S^{T_2}	$\log k$	$\log k_2 + S_0$	$\log k_1 + S_0$	$\dfrac{1}{2}(\log k + S_0) - \dfrac{1}{2}$
$S^{T_2(2)}$	$\log k$	$2\log k_2 + S_0$	$2\log k_1 + S_0$	$\log k - \dfrac{1}{2}$

Table 3.1: Top two rows: Rényi negativities in the four dominant phases labeled by the corresponding permutation g. Bottom three rows: schematic values of three special limits of the analytic continued Rényi negativities (where we have used $Z_n \sim e^{S_0}$). Here $C_m = \frac{1}{m+1}\binom{2m}{m}$ is the Catalan number which gives the number of noncrossing pairings.

3.3.3 Contributions to Negativities

Having worked out the dominant geometries in the four phases and how they contribute to the Rényi negativities schematically (i.e., using $Z_n \sim e^{S_0}$), we now write their contributions exactly using (3.24). This calculation is straightforward to do, for both even and odd n. We then analytically continue the resulting Rényi negativities and find the values of three special limits that we defined in Section 4.3: the logarithmic negativity $\mathcal{E}$, partially transposed entropy S^{T_2}, and refined Rényi-2 negativity $S^{T_2(2)}$. We collect these results for all four phases in Table 3.1.

From these results, we find a phase diagram for negativity, which we show in Figure 3.4. Unlike the von Neumann entropy which only has a single phase transition at $k \sim e^{S_0}$, we see that there are two distinct types of phase transitions for negativity, one from the disconnected phase to the pairwise phase and one from the pairwise phase to the cyclic phase. The transition from the pairwise phase to the anti-cyclic phase is similar to the pairwise-to-cyclic transition under the exchange $k_1 \leftrightarrow k_2$.

3.4 Resolvent Equation for Partial Transpose

Having analyzed the negativity measures deep within each of the four phases, we now turn our attention to the behavior of negativities near the phase transitions. Generally speaking, more geometries than the four types studied in the previous section could dominate the Rényi negativities near a phase transition, and we need to sum over them. We would then need to analytically continue the resulting sum to find special limits such as the logarithmic negativity. This is technically difficult to do directly.

Instead, we study the resolvent for the partially transposed density matrix $\rho_R^{T_2}$, which we refer to as the negativity resolvent or simply the resolvent. From this resolvent, we then extract the negativity spectrum, i.e., the eigenvalue distribution of $\rho_R^{T_2}$. This allows

us to calculate the Rényi negativities and their special limits.

In this section, we derive a self-consistent equation for calculating the resolvent. As we will show, the sum over dominant diagrams in a "planar" regime reduces to a Schwinger-Dyson equation, which can be resummed. This allows us to write down a closed-form equation for the resolvent.

The negativity resolvent is defined in terms of the partially transposed density matrix as[4]

$$R\left(\lambda\right)=\mathrm{Tr}\left(\frac{1}{\lambda\mathbb{I}-\rho_R^{T_2}}\right). \tag{3.33}$$

From the resolvent, the eigenvalue spectrum for $\rho_R^{T_2}$, which we will denote by $D(\lambda)$, can be obtained by taking the discontinuity across the real axis as follows

$$D(\lambda)=\lim_{\epsilon\to 0^+}\frac{1}{2\pi i}\big(R(\lambda-i\epsilon)-R(\lambda+i\epsilon)\big). \tag{3.34}$$

From this, we can compute the Rényi negativities via

$$\mathcal{N}_{2m}^{(\mathrm{even})}=\int d\lambda\, D\left(\lambda\right)|\lambda|^{2m}, \tag{3.35}$$

$$\mathcal{N}_{2m-1}^{(\mathrm{odd})}=\int d\lambda\, D\left(\lambda\right)\mathrm{sgn}\left(\lambda\right)|\lambda|^{2m-1}, \tag{3.36}$$

from which all other negativity measures we consider can be derived.

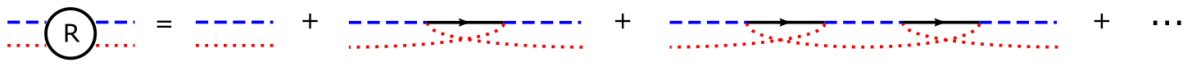
It will be useful to consider (3.33) in the following matrix form:

$$R_{j_1j_2}^{i_1i_2}(\lambda)=\left(\frac{1}{\lambda\mathbb{I}-\rho_R^{T_2}}\right)_{j_1j_2}^{i_1i_2} \tag{3.37}$$

$$=\frac{\delta^{i_1i_2}\delta_{j_1j_2}}{\lambda}+\sum_{n=1}^{\infty}\frac{1}{\lambda^{n+1}}\Big(\left(\rho_R^{T_2}\right)^n\Big)_{j_1j_2}^{i_1i_2}, \tag{3.38}$$

[4]The resolvent $R\left(\lambda\right)$ should not be confused with the subsystem R describing the Hawking radiation.

where we denote the R_1 subsystem by upper i-type indices and R_2 by lower j-type indices. In the last line, we have expanded the expression in a formal power series in $1/\lambda$. Each term in the series is given by the n-replicated density matrix $\left(\rho_R^{T_2}\right)^n$, which defines a boundary condition with n asymptotic boundaries with the brane indices contracted:



where the blue (upper) dashed lines denote i-type index lines and red (lower) dotted lines denote j-type index lines. Each pair of blue/red index lines gives a factor of $1/\lambda$ and each asymptotic boundary gives a factor of $1/(kZ_1)$ coming from the normalization of the density matrix. Note that for up to two boundaries, the boundary conditions are the same after taking a final trace, with or without partial transpose.

In general, the gravitational path integral can be performed as a sum over bulk geometries satisfying the boundary conditions. In the JT model we introduced in Section 3.3, higher genus corrections are highly suppressed by factors of $e^{-S_0} \ll 1$, so we only need to consider disjoint unions of disk geometries connecting any number of asymptotic boundaries. The path integral for the disk can be performed exactly including quantum corrections and is given by (3.23). For n asymptotic boundaries, these disjoint unions of disk geometries are in one-to-one correspondence with elements of the permutation group S_n. As we show in Lemma 11 and Corollary 12 of Appendix C, the only geometries that can possibly dominate in the limit $e^{S_0} \gg 1$ are the planar and anti-planar diagrams, which correspond to certain subsets of permutations in S_n. The term "anti-planar" will be explained in detail in Section 3.4.2 and in Appendix C. In fact, large regions of the phase diagram are dominated by either the planar or anti-planar geometries, which we will call the planar and anti-planar regimes. As we will now show, in each of these two regimes the resulting sum over diagrams can be resummed via a Schwinger-Dyson

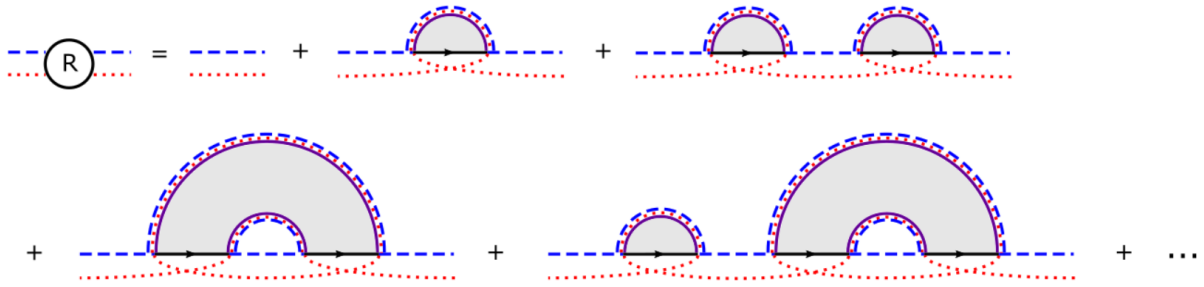
equation.

3.4.1 Planar Regime

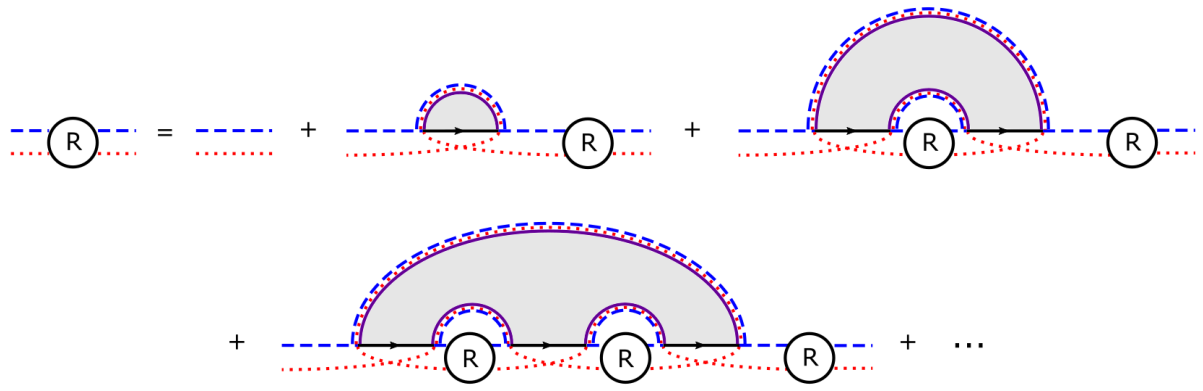
We now derive a resolvent equation in the parameter regime $k_2 \ll k_1 e^{S_0}$. For reasons that will become clear shortly, we call it the planar regime.

Our strategy is to keep only the subset of geometries which have a possibility of dominating. As we outlined in Section 3.3.2, the phases in this regime away from phase transitions are defined by the disconnected, cyclic, and pairwise geometries. Closer to phase transitions, we expect more generic geometries which "interpolate" between these three types of geometries to have a chance of dominating. Indeed, as we show in Lemma 11 of Appendix C, the geometries which can dominate are precisely the planar diagrams, which correspond to the noncrossing partitions, i.e., the permutation group elements $g \in S_n$ lying on a geodesic between $\mathbb{1}$ and X.

We can write the sum over planar geometries diagrammatically as



This sum can be recast as a Schwinger-Dyson equation. Diagrammatically, we have



or, as an equation,

$$R^{i_1 i_2}_{j_1 j_2} = \frac{\delta^{i_1 i_2}\delta_{j_1 j_2}}{\lambda} + \frac{1}{\lambda}\sum_{n=1}^{\infty} \frac{Z_n}{(kZ_1)^n} \widetilde{R}_{k_{n-1}k_1}\widetilde{R}_{j_1 k_2}\widetilde{R}_{k_1 k_3}\cdots\widetilde{R}_{k_{n-3}k_{n-1}} R^{i_1 i_2}_{k_{n-2} j_2}, \tag{3.39}$$

where we have defined the partial trace of the resolvent matrix over the R_1 subsystem: $\widetilde{R}_{j_1 j_2} \equiv \sum_{i=1}^{k_1} R^{ii}_{j_1 j_2}$, and repeated indices are summed over[5]. Note that for $n = 1$ the product of resolvents is simply $R^{i_1 i_2}_{j_1 j_2}$ and for $n = 2$ it is $\widetilde{R}_{k_1 k_1} R^{i_1 i_2}_{j_1 j_2}$. As is evident from the diagrammatics, the i-type indices denoting R_1 form simple self-contractions on all but the last resolvent, while the j-type indices denoting R_2 form a complicated set of contractions. Fortunately, this equation can be solved iteratively: starting with the leading solution $R^{i_1 i_2}_{j_1 j_2} = \delta^{i_1 i_2}\delta_{j_1 j_2}/\lambda + \mathcal{O}\left(1/\lambda^2\right)$, self-consistency then requires that $R^{i_1 i_2}_{j_1 j_2} \propto \delta^{i_1 i_2}\delta_{j_1 j_2}$ to all orders. As we will see, this allows us to rewrite the complicated product of resolvents as the following simple expression

$$\widetilde{R}_{k_{n-1}k_1}\cdots\widetilde{R}_{k_{n-3}k_{n-1}} R^{i_1 i_2}_{k_{n-2} j_2} = \begin{cases} \left(\frac{R}{k_2}\right)^{n-1} R^{i_1 i_2}_{j_1 j_2}, & n \text{ odd}, \\ k_2 \left(\frac{R}{k_2}\right)^{n-1} R^{i_1 i_2}_{j_1 j_2}, & n \text{ even}. \end{cases} \tag{3.40}$$

[5]Note that these repeated indices $k_1, k_2, \cdots, k_{n-1}$ are j-type indices which should not be confused with the parameters k_1, k_2 that count the number of EOW brane states.

Let us explain this in more detail. In general, the behavior for n odd and even are different so we will need to treat these cases separately. To illustrate the simplification for the odd case, we first consider the contribution at $n = 3$, which is the first nontrivial diagram under the partial transpose. We can write

$$\begin{aligned}\widetilde{R}_{k_2k_1}\widetilde{R}_{j_1k_2}R^{i_1i_2}_{k_1j_2} &= \left(\widetilde{R}_{j_1j_1}\right)^2 R^{i_1i_2}_{j_1j_1}\delta_{k_2k_1}\delta_{j_1k_2}\delta_{k_1j_2} \\ &= \left(\widetilde{R}_{j_1j_1}\right)^2 R^{i_1i_2}_{j_1j_1}\delta_{j_1j_2} \\ &= \left(\widetilde{R}_{j_1j_1}\right)^2 R^{i_1i_2}_{j_1j_2} \end{aligned} \tag{3.41}$$

where no summation on j_1 is implied. Now, recalling the full trace $R = \sum_j \widetilde{R}_{jj}$, we find $\widetilde{R}_{jj} = R/k_2$ and we can therefore write

$$\widetilde{R}_{k_2k_1}\widetilde{R}_{j_1k_2}R^{i_1i_2}_{k_1j_2} = \left(\frac{R}{k_2}\right)^2 R^{i_1i_2}_{j_1j_2}. \tag{3.42}$$

We can understand the even case by looking at the first nontrivial contribution at $n = 4$. By a similar analysis as above, we have

$$\begin{aligned}\widetilde{R}_{k_3k_1}\widetilde{R}_{j_1k_2}\widetilde{R}_{k_1k_3}R^{i_1i_2}_{k_2j_2} &= \left(\widetilde{R}_{j_1j_1}\right)^3 R^{i_1i_2}_{j_1j_1}\delta_{k_3k_1}\delta_{j_1k_2}\delta_{k_1k_3}\delta_{k_2j_2} \\ &= k_2\left(\widetilde{R}_{j_1j_1}\right)^3 R^{i_1i_2}_{j_1j_2} \\ &= k_2\left(\frac{R}{k_2}\right)^3 R^{i_1i_2}_{j_1j_2}. \end{aligned} \tag{3.43}$$

Note the additional factor of k_2 compared to $n = 3$ due to the closed index loop formed from the first and third resolvent factors.

More generally, one can show that the even case always has a single index loop and the odd case has no index loops, leading to (3.40). Using this, we can rewrite the

Schwinger-Dyson equation (3.39) as

$$\lambda R^{i_1 i_2}_{j_1 j_2} = \delta^{i_1 i_2}\delta_{j_1 j_2} + k_2 \sum_{m=1}^{\infty} \frac{Z_{2m-1}}{(kk_2 Z_1)^{2m-1}} R^{2m-2} R^{i_1 i_2}_{j_1 j_2} + k_2^2 \sum_{m=1}^{\infty} \frac{Z_{2m}}{(kk_2 Z_1)^{2m}} R^{2m-1} R^{i_1 i_2}_{j_1 j_2}.$$

Taking the full trace, we find

$$\lambda R = k + k_2 \sum_{m=1}^{\infty} \frac{Z_{2m-1} R^{2m-1}}{(kk_2 Z_1)^{2m-1}} + k_2^2 \sum_{m=1}^{\infty} \frac{Z_{2m} R^{2m}}{(kk_2 Z_1)^{2m}}. \tag{3.44}$$

The gravitational partition function of the n-boundary totally connected geometry is given by (3.23), which we repeat here:

$$Z_n = e^{S_0} \int_0^{\infty} ds \rho(s) y(s)^n. \tag{3.45}$$

Since this depends on n only through $y(s)^n$, the sum over n becomes a geometric series and (3.44) can be resummed into

$$\lambda R = k + k_2^2 e^{S_0} \int_0^{\infty} ds \rho(s) \frac{w(s) R (k + w(s) R)}{k^2 k_2^2 - w(s)^2 R^2}, \tag{3.46}$$

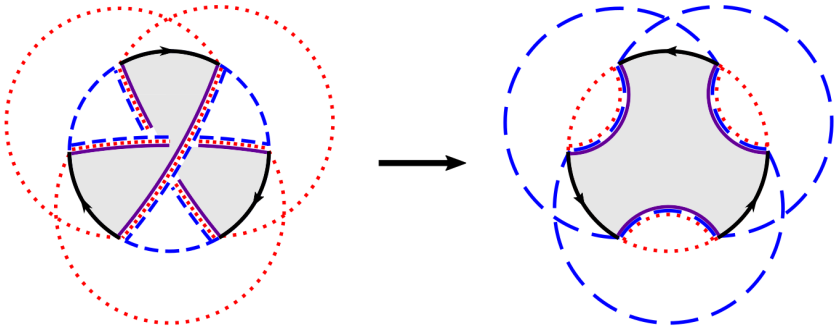
where $w(s) \equiv y(s)/Z_1$. As a consistency check, when $k_2 = 1$ this reduces to the resolvent equation for the original (untransposed) density matrix derived in [61].

3.4.2 Anti-Planar Regime

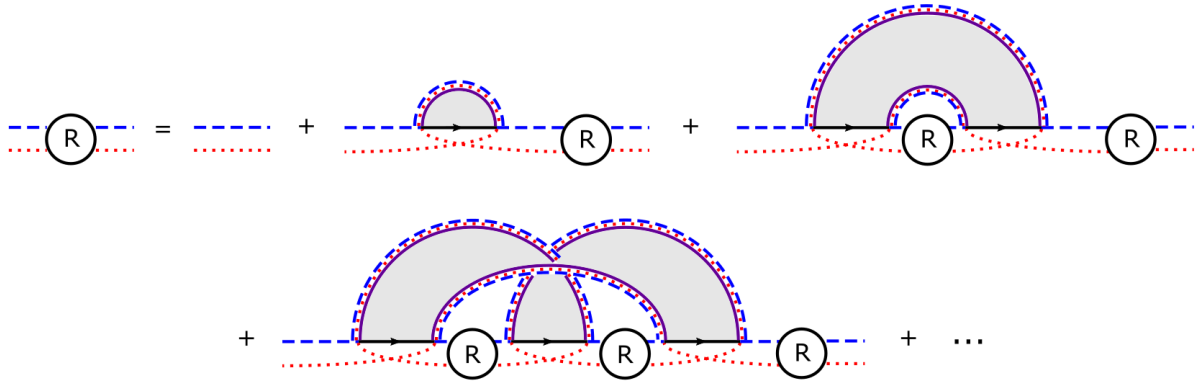
We now consider a different parameter regime $k_1 \ll k_2 e^{S_0}$. For reasons that will become clear shortly, we call it the anti-planar regime. This anti-planar regime has a large overlap with the planar regime; the overlap region is $e^{-S_0} \ll k_1/k_2 \ll e^{S_0}$. Together the two regimes cover the entire parameter space.

Once again, we will keep only the subset of geometries which have a possibility of

dominating. As we outlined in Section 3.3.2, the phases in this regime away from phase transitions are defined by the disconnected, anti-cyclic, and pairwise geometries. Closer to phase transitions, we expect geometries which interpolate between these geometries to have a chance of dominating. Indeed, as we show in Corollary 12 of Appendix C, the geometries which can dominate are precisely the set of anti-planar diagrams, which are in one-to-one correspondence with permutations lying on a geodesic between $\mathbb{1}$ and X^{-1}. An example of an anti-planar geometry is the anti-cyclic geometry; two other (perhaps less obvious) examples are the disconnected and pairwise geometries. Geometrically, anti-planar diagrams are precisely those that become planar diagrams after reversing the orientation of the asymptotic boundaries:



The sum over anti-planar diagrams can similarly be recast as a Schwinger-Dyson equation. Diagrammatically, we have



or, as an equation,

$$R^{i_1 i_2}_{j_1 j_2} = \frac{\delta^{i_1 i_2}\delta_{j_1 j_2}}{\lambda} + \frac{1}{\lambda}\sum_{n=1}^{\infty}\frac{Z_n}{(kZ_1)^n}\widetilde{R}^{k_{n-1}k_1}\widetilde{R}^{i_1 k_2}\widetilde{R}^{k_1 k_3}\cdots\widetilde{R}^{k_{n-3}k_{n-1}}R^{k_{n-2}i_2}_{j_1 j_2}, \tag{3.47}$$

where we have defined the partial trace over the R_2 subsystem: $\widetilde{R}^{i_1 i_2} \equiv \sum_{j=1}^{k_2} R^{i_1 i_2}_{jj}$. Note that the $n = 1$ and $n = 2$ terms are the same as in the planar case. Compared to the planar case, the j-type indices denoting R_2 now form simple self-contractions on all but the last resolvent, while the i-type indices denoting R_1 form the complicated set of contractions. In other words, the i-type indices now play the role of j-type indices in the planar regime, and vice-versa. As before, we can use the fact that $R^{i_1 i_2}_{j_1 j_2} \propto \delta^{i_1 i_2}\delta_{j_1 j_2}$ to all orders to rewrite the complicated product of resolvents as

$$\widetilde{R}^{k_{n-1}k_1}\cdots\widetilde{R}^{k_{n-3}k_{n-1}}R^{k_{n-2}i_2}_{j_1 j_2} = \begin{cases} \left(\frac{R}{k_1}\right)^{n-1} R^{i_1 i_2}_{j_1 j_2}, & n \text{ odd}, \\ k_1\left(\frac{R}{k_1}\right)^{n-1} R^{i_1 i_2}_{j_1 j_2}, & n \text{ even}. \end{cases} \tag{3.48}$$

Using this, we can rewrite the Schwinger-Dyson equation (3.47) as

$$\lambda R^{i_1 i_2}_{j_1 j_2} = \delta^{i_1 i_2}\delta_{j_1 j_2} + k_1\sum_{m=1}^{\infty}\frac{Z_{2m-1}}{(kk_1Z_1)^{2m-1}}R^{2m-2}R^{i_1 i_2}_{j_1 j_2} + k_1^2\sum_{m=1}^{\infty}\frac{Z_{2m}}{(kk_1Z_1)^{2m}}R^{2m-1}R^{i_1 i_2}_{j_1 j_2}.$$

Taking the full trace, we find

$$\lambda R = k + k_1\sum_{m=1}^{\infty}\frac{Z_{2m-1}R^{2m-1}}{(kk_1Z_1)^{2m-1}} + k_1^2\sum_{m=1}^{\infty}\frac{Z_{2m}R^{2m}}{(kk_1Z_1)^{2m}}. \tag{3.49}$$

Finally, using (3.23), we resum (3.49) into

$$\lambda R = k + k_1^2 e^{S_0}\int_0^{\infty} ds\rho(s)\frac{w(s)R(k + w(s)R)}{k^2k_1^2 - w(s)^2R^2}. \tag{3.50}$$

As expected, this is simply the resolvent equation in the planar regime (3.46) with the

exchange $k_1 \leftrightarrow k_2$.

3.5 Negativity Spectrum Near Phase Transitions

Having derived the resolvent equation for the partial transpose, we now solve it to find the negativity spectrum near phase transitions.

For each negativity measure, we will be interested in a neighborhood near one of the phase transitions as we tune the relative sizes of our parameters, and we will analyze the corrections to the negativity measures listed in Table 3.1. Generically these corrections take the form of fluctuations about a fixed saddle point in the gravitational path integral. Near a phase transition, however, multiple saddles are competing for dominance, so enhanced corrections, i.e. corrections larger than those for any individual saddle, provide additional information about the entanglement structure of the system.

3.5.1 Microcanonical Ensemble

Before studying the negativity spectrum in more detail, let us take a brief detour and consider the situation where the black hole is in a microcanonical ensemble in the JT model. In this case, we restrict to some small energy window $[s, s + \Delta s]$. We write

$$e^{\mathbf{S}} = \boldsymbol{\rho}(s)\Delta s, \qquad \mathbf{Z}_n = \boldsymbol{\rho}(s)y(s)^n\Delta s, \qquad \mathbf{w}(s) = \frac{y(s)}{\mathbf{Z}_1} = e^{-\mathbf{S}}, \tag{3.51}$$

where $\boldsymbol{\rho}(s) \equiv e^{S_0}\rho(s)$ is the density of states. In this case, it can be shown that the sum over geometries for the Rényi negativities (3.24) is given by

$$\mathrm{Tr}\left(\rho_R^{T_2}\right)^n = \frac{1}{(ke^{\mathbf{S}})^n}\sum_{g\in S_n}\left(e^{\mathbf{S}}\right)^{\chi(g)} k_1^{\chi(g^{-1}X)}k_2^{\chi(g^{-1}X^{-1})}. \tag{3.52}$$

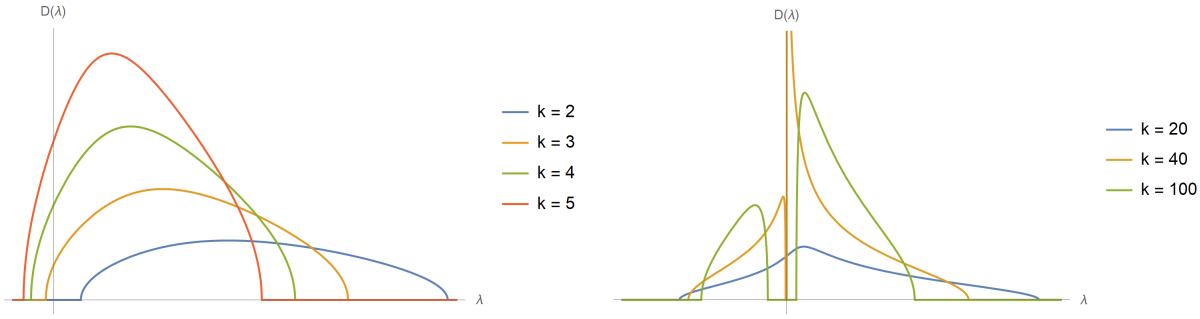


Figure 3.5: Negativity spectrum in the microcanonical ensemble calculated from the cubic resolvent equation (3.53). We fix $e^{\mathbf{S}} = 10$, $k_2 = 2$, and move in a horizontal line in the phase diagram by tuning k. The disconnected-to-pairwise transition (left) occurs at $k = e^{\mathbf{S}} = 10$, though we only plot to $k = 5$ for visual clarity, as the qualitative behavior is the same. The pairwise-to-cyclic transition (right) occurs at $k = k_2^2 e^{\mathbf{S}} = 40$.

We recognize this as the n^{th} Rényi negativity of a Wishart matrix with $e^{\mathbf{S}}$ degrees of freedom. This implies that $\rho_R^{T_2}$ in a microcanonical ensemble can be thought of as the partial transpose of a random matrix drawn from the Wishart distribution. A similar expression for the moments of the partial transpose of a random mixed state was derived in [127]. The similarity between the microcanonical JT model and a random mixed state was also noted in [133].

In the planar regime $k_1 \gg k_2 e^{-S_0}$, the resolvent equation (3.46) becomes

$$R^3 + \left(\frac{e^{\mathbf{S}} k_2^2 - k}{\lambda}\right) R^2 + e^{2\mathbf{S}} k k_2^2 \left(\frac{1}{\lambda} - k\right) R + \frac{e^{2\mathbf{S}} k^3 k_2^2}{\lambda} = 0. \tag{3.53}$$

This matches the resolvent equation derived in [127] for a random mixed state under appropriate rescaling of variables.[6] As was shown there, a closed-form solution to this cubic equation for R can be found, and leads to concrete results for the negativity spectrum

[6] To match to the resolvent equation in [127], define the rescaled variables $z = k_2 e^{\mathbf{S}} \lambda$ and $G = e^{-\mathbf{S}} R / k k_2$ so that (3.53) becomes

$$z G^3 + (\beta - 1) G^2 + (\alpha - z) G + 1 = 0, \tag{3.54}$$

where $\alpha = e^{\mathbf{S}}/k_1$ and $\beta = k_2 e^{\mathbf{S}}/k_1$. This cubic equation was earlier noted in the context of free probability theory [140].

and various negativity measures. The resolvent for the anti-planar regime $k_2 \gg k_1 e^{-S_0}$ can be obtained by $k_1 \leftrightarrow k_2$.

We plot the eigenvalue density in the microcanonical ensemble for various parameter values in Figure 3.5. The spectrum is approximately a Wigner semicircle distribution in the disconnected phase, continues to be connected in the pairwise phase, develops singularities at the pairwise-to-cyclic transition, and has two branches in the cyclic phase, where it is well approximated by the difference of two disjoint Marchenko-Pastur distributions.

3.5.2 Canonical Ensemble: Disconnected-Pairwise Transition

The transitions that involve the cyclic and anti-cyclic phases are complicated, as they involve a sum over diagrams with pieces connecting more than two asymptotic boundaries. Here, we will focus on the transition between the totally disconnected phase and the pairwise connected phase. The disconnected phase involves single-boundary diagrams, while the pairwise phase involves pairwise connected wormholes (plus a single disconnected piece for odd n).

The disconnected-pairwise transition happens within the large overlap $e^{-S_0} \ll k_1/k_2 \ll e^{S_0}$ between the planar regime and the anti-planar regime[7]. Therefore, the dominant geometries are those that are simultaneously planar and anti-planar. As we show in Appendix C, these geometries are disjoint, noncrossing unions of single-boundary disks and pairwise connected wormholes. This result is the content of Lemma 13 in Appendix C.

Intuitively, these dominant geometries interpolate between the disconnected and pairwise geometries. Geometries with pieces connecting more than two asymptotic boundaries are parametrically suppressed. As such, the resolvent equation (3.44) truncates at

[7]In other words, we stay away from the two triple points on the phase diagram, as we would need to analyze transitions to the (anti-)cyclic phase there.

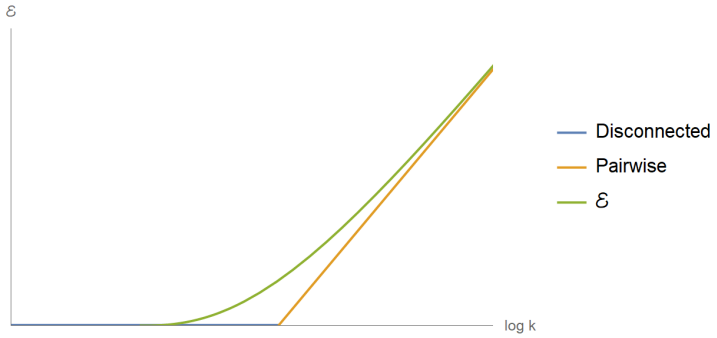


Figure 3.6: Logarithmic negativity near the disconnected-to-pairwise transition with $Z_2/Z_1^2 = e^{-5}$, along with the naive answers in the dominant regions: $\mathcal{E}_{\text{disconnected}} = 0$ and $\mathcal{E}_{\text{pairwise}} = \frac{1}{2}\left(\log k + \log \frac{Z_2}{Z_1^2}\right) + \log \frac{8}{3\pi}$, where $\log \frac{8}{3\pi}$ is an $\mathcal{O}(1)$ term arising from the analytic continuation of the Catalan number in $\mathcal{N}_{2m}^{(\text{even})}$ to $m = 1/2$ [128].

the quadratic order:

$$\lambda R = k + \frac{R}{k} + \frac{Z_2 R^2}{(kZ_1)^2}. \tag{3.55}$$

This quadratic equation can be solved analytically giving the resolvent and eigenvalue density as

$$\begin{aligned} R(\lambda) &= \frac{2k}{A^2}\left(\lambda - \frac{1}{k} - \sqrt{\lambda - \frac{1}{k} + A}\,\sqrt{\lambda - \frac{1}{k} - A}\right), \\ D(\lambda) &= \frac{2k}{\pi A^2}\sqrt{A^2 - \left(\lambda - \frac{1}{k}\right)^2}, \end{aligned} \tag{3.56}$$

where $A^2 \equiv 4Z_2/(kZ_1^2)$. Thus the eigenvalue density is a Wigner semicircle distribution supported on $\lambda \in \left[-A + \frac{1}{k}, A + \frac{1}{k}\right]$. For $k \leq 1/A$, $D(\lambda)$ only has support on $\lambda \geq 0$ and the negativity vanishes; for $k > 1/A$, $D(\lambda)$ has support on $\lambda < 0$ and we find the negativity is nonvanishing. The phase transition thus occurs at $k = 1/A \sim e^{S_0}$, as expected from the schematic analysis in Section 3.3.2.

We can write an explicit expression for the logarithmic negativity using (3.6) and

(3.56):

$$\mathcal{E} = \log \int_{-\infty}^{\infty} d\lambda D(\lambda)|\lambda| \tag{3.57}$$

$$= \log \left[\frac{2}{3\pi} \left(\frac{\sqrt{A^2k^2-1}\,(2A^2k^2+1)}{A^2k^2} + 3\csc^{-1}(Ak) \right) \right] \Theta\left(k - \frac{1}{A}\right). \tag{3.58}$$

We plot this result in Figure 3.6, along with the naive answers for logarithmic negativity in the dominant phases.

How large is the correction at the transition? It is easy to verify that it is $\mathcal{O}(1)$. There are no enhanced corrections here, as we are working in a regime where higher order terms in the Schwinger-Dyson equation are parametrically suppressed. The logarithmic negativity, along with all other negativity measures, never receives contributions from geometries containing pieces with $Z_{n>2}$, so corrections are $\mathcal{O}(1)$.

3.5.3 Canonical Ensemble: Cyclic-Pairwise Transition

In this subsection we will study the richer phase transition between the pairwise phase and the cyclic phase.[8] Our computation is inspired by that of Appendix F of [61].

First, let us define some useful values of s. In the semiclassical $\beta \ll 1$ and large brane mass $\mu \gg 1/\beta$ limits, we have

$$\rho(s)y(s)^n \sim \frac{s}{2\pi^2} y(0)^n e^{2\pi s - n\beta s^2/2}. \tag{3.59}$$

This means that the integral that defines Z_n in (3.23) can be well approximated by the

[8] The pairwise to anti-cyclic transition follows from the calculation in this subsection by exchanging $k_1 \leftrightarrow k_2$.

saddle point located at

$$s^{(n)} = \frac{2\pi}{n\beta} + \mathcal{O}(1)$$
$$\Rightarrow \log Z_n \approx S_0 + \frac{2\pi^2}{n\beta} + \mathcal{O}(\log \beta). \tag{3.60}$$

Throughout, we take our parameters S_0, k_1, and k_2 to be large before taking the semi-classical limit, such that e.g. $\log Z_n \approx S_0$ as $S_0 \gg 1/\beta$. We will also need to define s_k, the value of s for which

$$k = k_2^2 e^{S_0} \int_0^{s_k} ds \rho(s). \tag{3.61}$$

We can approximate s_k by

$$s_k \approx \frac{1}{2\pi} \log\left(\frac{k}{k_2^2}\right) - S_0 + \mathcal{O}(1). \tag{3.62}$$

Note that here we are considering the values of k and k_2 at transition, so the particular values of s_k we are interested in will depend on the details of the negativity measure we are computing. In our schematic analysis where $Z_n \sim e^{S_0}$, we derived the location of the phase transition between the cyclic and pairwise phase and found that it was independent of n. However, taking into account dependence on β, the Z_n's are distinct for different n, which leads to n-dependent transition points. The Rényi negativities in the cyclic and pairwise phases are given by the contributions of the dominant geometries in each phase (see Table 3.1), and coincide at transition. Equating their contributions at the transition, we find, up to factors $\mathcal{O}(1)$ in β,

$$\mathcal{N}_{2m}^{(\text{even})} = \frac{Z_{2m}}{k_2^{2m-2} Z_1^{2m}} = \frac{Z_2^m}{k^{m-1} Z_1^{2m}},$$
$$\mathcal{N}_{2m-1}^{(\text{odd})} = \frac{Z_{2m-1}}{k_2^{2m-2} Z_1^{2m-1}} = \frac{Z_2^{m-1}}{k^{m-1} Z_1^{2m-2}}. \tag{3.63}$$

In terms of the approximation (3.60), we can solve for $\log(k/k_2^2)$ at transition to obtain

$$\begin{aligned}\text{Even:}\quad \log\left(\frac{k}{k_2^2}\right) &= \log\left(\frac{Z_2^m}{Z_{2m}}\right)^{\frac{1}{m-1}} = S_0 + \left(1+\frac{1}{m}\right)\frac{\pi^2}{\beta} + \mathcal{O}(\log\beta)\\ \text{Odd:}\quad \log\left(\frac{k}{k_2^2}\right) &= \log\left(\frac{Z_2^{m-1}Z_1}{Z_{2m-1}}\right)^{\frac{1}{m-1}} = S_0 + \left(1+\frac{4}{2m-1}\right)\frac{\pi^2}{\beta} + \mathcal{O}(\log\beta)\,. \end{aligned} \tag{3.64}$$

From this, we can solve for s_k at the transition using (3.62) to obtain

$$\begin{aligned} s_k^{(n,\text{ even})} &\approx \frac{\pi}{2\beta}\left(1+\frac{2}{n}\right) + \mathcal{O}(\log\beta)\\ s_k^{(n,\text{ odd})} &\approx \frac{\pi}{2\beta}\left(1+\frac{4}{n}\right) + \mathcal{O}(\log\beta)\,. \end{aligned} \tag{3.65}$$

As expected, the transition point depends on n. In particular, it is $\mathcal{O}(1/\beta)$ at leading order and bounded below as a function of n.

For this phase transition, we will fix k and tune k_2. In the phase diagram, this corresponds to moving along a line between the upper left corner and the lower right corner. We need to consider diagrams with pieces made of an arbitrary number of boundaries, and we can restrict ourselves to planar diagrams, as anti-planar diagrams are suppressed by factors of k_2/k_1 relative to their planar counterparts. The resolvent equation is again (3.50):

$$\lambda R = k + k_2^2 e^{S_0}\int_0^\infty ds\rho(s)\frac{w(s)R(k+w(s)R)}{k^2k_2^2 - w(s)^2R^2}. \tag{3.66}$$

We are going to split this integral at some transition s_t such that

$$\lambda R = k + k_2^2 e^{S_0}\int_0^{s_t} ds\rho(s)\frac{w(s)R(k+w(s)R)}{k^2k_2^2 - w(s)^2R^2} + k_2^2 e^{S_0}\int_{s_t}^\infty ds\rho(s)\frac{w(s)R(k+w(s)R)}{k^2k_2^2 - w(s)^2R^2}. \tag{3.67}$$

We rewrite this simple step to emphasize that no approximations have been used yet.

We are now going to use a set of three assumptions:

1. $w(s_t)R \ll kk_2$.

2. $k_2^2 e^{S_0} \int_0^{s_t} ds\rho(s) \frac{w(s)R(k+w(s)R)}{k^2k_2^2 - w(s)^2R^2} \ll k$.

3. $s_t = s_k - \kappa$, where κ is $\mathcal{O}(1)$ but large.

These assumptions are justified in detail in Appendix D, where we show that the resulting simplifications to the resolvent equation give a self-consistent treatment of the problem. For now we will take these as facts and proceed. The first approximation allows us to simplify the final term in (3.67) such that

$$\lambda R \approx k + k_2^2 e^{S_0} \int_0^{s_t} ds\rho(s) \frac{w(s)R(k+w(s)R)}{k^2k_2^2 - w(s)^2R^2} + \frac{e^{S_0}}{k} \int_{s_t}^{\infty} ds\rho(s)w(s)R, \tag{3.68}$$

where we have dropped an R^2 term from the last integral because it can be shown to be much smaller than the leading term k using Assumptions 1 and 3. We define the coefficient of R in the last term to be the constant λ_0, given by

$$\lambda_0 \equiv \frac{e^{S_0}}{k} \int_{s_t}^{\infty} ds\rho(s)w(s). \tag{3.69}$$

We can now write the resolvent equation as

$$(\lambda - \lambda_0)R \approx k + k_2^2 e^{S_0} \int_0^{s_t} ds\rho(s) \frac{w(s)R(k+w(s)R)}{k^2k_2^2 - w(s)^2R^2}. \tag{3.70}$$

Now we turn to Assumption 2, which allows us to treat the second term above as a perturbation to the zeroth order solution

$$R \approx R_0 = \frac{k}{\lambda - \lambda_0}. \tag{3.71}$$

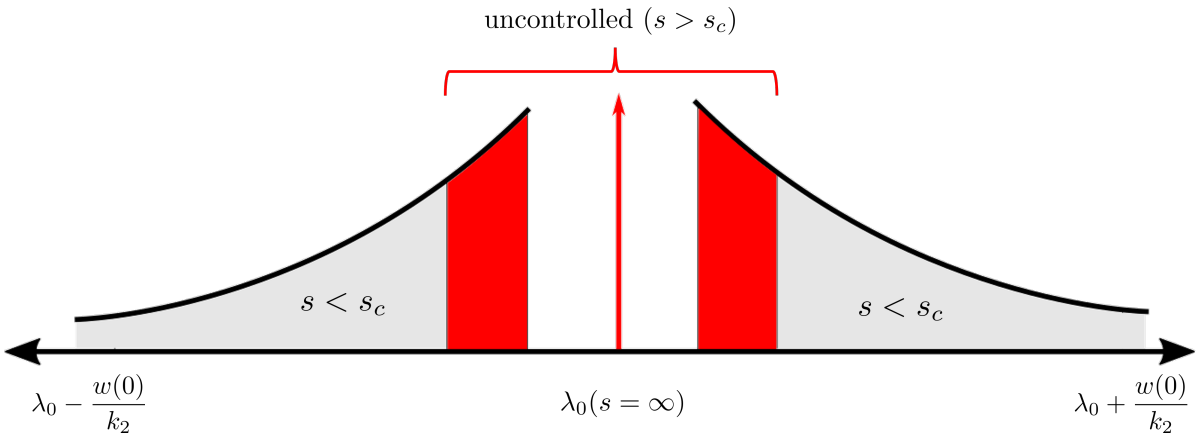


Figure 3.7: A rough sketch (not to scale) of the negativity spectrum near the cyclic–to-pairwise transition.

Plugging this solution back into (3.70), we obtain the first order iterated solution

$$\begin{aligned} R_1 &\approx \frac{k}{\lambda - \lambda_0} + \frac{k_2^2 e^{S_0}}{\lambda - \lambda_0} \int_0^{s_t} ds \rho(s) \frac{w(s) R_0 (k + w(s) R_0)}{k^2 k_2^2 - w(s)^2 R_0^2} \\ &\approx \frac{k}{\lambda - \lambda_0} + e^{S_0} \int_0^{s_t} ds \frac{1}{\lambda - \lambda_0} \frac{\rho(s) w(s) (\lambda - \lambda_0 + w(s))}{(\lambda - \lambda_0)^2 - \left(\frac{w(s)}{k_2}\right)^2}. \end{aligned} \tag{3.72}$$

Now we can find the discontinuity in this expression and extract $D(\lambda)$. There are three contributions to the spectrum: a simple pole at $\lambda = \lambda_0$, and a pair of branch cuts given by the poles at $\lambda = \lambda_0 \pm w(s)/k_2$ in the integrand. We obtain

$$\begin{aligned} D(\lambda) &= \#\delta(\lambda - \lambda_0) \\ &+ e^{S_0} \int_0^{s_t} ds \rho(s) \left[\frac{k_2(k_2+1)}{2} \delta\left(\lambda - \lambda_0 - \frac{w(s)}{k_2}\right) + \frac{k_2(k_2-1)}{2} \delta\left(\lambda - \lambda_0 + \frac{w(s)}{k_2}\right) \right]. \end{aligned} \tag{3.73}$$

Let us pause for a second to unpack this equation. The spectrum consists of a delta function located at λ_0 from the simple pole and two regions of nonzero eigenvalue density from the integrated delta functions. We plot a sketch of this eigenvalue density in Figure 3.7. There are two distinct regions with nonzero eigenvalue density, similar to

the spectrum in the microcanonical ensemble. One point we emphasize in Appendix D is the presence of a "controlled" region $0 < s < s_c$ in which our assumptions hold and an "uncontrolled" region $s > s_c$ where we claim ignorance about the spectrum. In terms of the eigenvalues, this corresponds to an ignorance in the spectrum for a region

$$\lambda \in \left[\lambda_0 - \frac{w(s_c)}{k_2}, \lambda_0 + \frac{w(s_c)}{k_2}\right]. \tag{3.74}$$

We show that our ignorance about the uncontrolled region leads to at most $\mathcal{O}(1)$ multiplicative corrections to the Rényi negativities, or $\mathcal{O}(1)$ additive corrections to $\mathcal{E}$, S^{T_2}, and $S^{T_2(2)}$, due to the constraint that the total number of eigenvalues must be k.

Clearly, λ_0 lies within this uncontrolled region, so we should not take seriously the presence of the delta function at λ_0. In fact, we will now show that this delta function vanishes if we extend the upper limit of the integral from s_t to s_k in (3.73) (which only affects the uncontrolled region and therefore causes a small error). Our density matrix has a total of k eigenvalues and is unit normalized, which translates into conditions on the zeroth and first moments of $D(\lambda)$, namely

$$\int_{-\infty}^{\infty} d\lambda D(\lambda) = k, \quad \int_{-\infty}^{\infty} d\lambda D(\lambda)\lambda = 1. \tag{3.75}$$

We see that these conditions are satisfied by (3.73) if we replace s_t by s_k, compensating for the fact that $s_t = s_k - \kappa$ by sending the coefficient of the $\delta(\lambda - \lambda_0)$ piece to zero. We conclude that a good approximation for the spectrum of the partially transposed density matrix at transition is given by

$$D(\lambda) = e^{S_0} \int_0^{s_k} ds\rho(s) \left[\frac{k_2(k_2+1)}{2}\delta\left(\lambda - \lambda_0 - \frac{w(s)}{k_2}\right) + \frac{k_2(k_2-1)}{2}\delta\left(\lambda - \lambda_0 + \frac{w(s)}{k_2}\right)\right] \tag{3.76}$$

In order to simplify our calculations for negativity, we would like that the delta function branch cuts were purely positive or negative, which is equivalent to the condition $\lambda_0 < w(s_k)/k_2$. By definition, λ_0 is bounded above by $1/k$, and at transition we have $k/k_2^2 = e^{S_0+2\pi s_k}$. In the semiclassical limit, we can use the approximation

$$w(s_k) \approx \frac{e^{-\beta s_k^2/2}}{Z_1} \approx e^{-S_0-\beta s_k^2/2-2\pi^2/\beta}. \tag{3.77}$$

Our condition on the branch cuts becomes

$$\frac{kw(s_k)}{k_2} > 1 \quad \Rightarrow \quad k_2 e^{S_0+2\pi s_k} w(s_k) \approx k_2 e^{2\pi s_k-\beta s_k^2/2-2\pi^2/\beta} \sim k_2 e^{C/\beta} \gg 1 \tag{3.78}$$

as $s_k \sim 1/\beta$ for a generic n. This is satisfied even if the unknown order one constant is negative under our previous assumption that we take our counting parameters to be large before taking small β, such that $\log k_2 \gg \frac{1}{\beta}$.

Now that we have the spectrum at transition, we are ready to calculate the corrections to any negativity measure we want! Let us start with the logarithmic negativity, which has a transition located at

$$s_k^{(1,\text{even})} = \frac{3\pi}{2\beta}. \tag{3.79}$$

We find

$$\begin{aligned}
\mathcal{E} &= \log \int_{-\infty}^{\infty} d\lambda D(\lambda)|\lambda| \\
&= \log e^{S_0} \int_0^{s_k} ds\rho(s) \left[\frac{k_2(k_2+1)}{2} \left(\lambda_0 + \frac{w(s)}{k_2} \right) + \frac{k_2(k_2-1)}{2} \left(\frac{w(s)}{k_2} - \lambda_0 \right) \right] \\
&= \log \left(\frac{1}{k_2} + k_2 e^{S_0} \int_0^{s_k} ds\rho(s) w(s) \right).
\end{aligned} \tag{3.80}$$

As $s_k < s^{(1)}$ for logarithmic negativity, we have approximated $\lambda_0 \approx 1/k$, and the final in-

tegral is well approximated by its maximum value $\rho(s_k)w(s_k)$. The logarithmic negativity is then

$$\begin{aligned}\mathcal{E} &\approx \log\left(\frac{1}{k_2} + k_2 e^{S_0}\rho(s_k)w(s_k)\right)\\ &\approx \log\left(\frac{1}{k_2} + \frac{k w(s_k)}{k_2}\right)\\ &\approx \log k_2 - \frac{\pi^2}{8\beta}.\end{aligned} \tag{3.81}$$

In the second line we used our previous approximation (3.62) for s_k, and in the third line we used (3.78). As we see, the logarithmic negativity experiences an $\mathcal{O}(1/\beta)$ correction to the naive answer $\mathcal{E} = \log k_2$.

Where do we expect $\mathcal{O}(1/\sqrt{\beta})$ corrections? In the case of even Rényi negativities, this happens at

$$s_k^{(n,\text{even})} = s^{(n)} \Rightarrow n = 2, \quad s^{(2)} = \frac{\pi}{\beta}. \tag{3.82}$$

We can check this explicitly for a negativity measure descending from the even analytic continuation. The simplest such measure, the Rényi-2 negativity $\mathcal{N}_2^{(\text{even})}$, is related to the second Rényi entropy S_2 by

$$\mathcal{N}_2^{(\text{even})} = e^{-S_2} \tag{3.83}$$

and therefore comes with $\mathcal{O}(1)$ corrections. We instead turn to the refined Rényi-2 negativity $S^{T_2(2)}$, defined in (3.12). We can read off the naive answer for $S^{T_2(2)}$ from Table 3.1, again using the approximation (3.60). We find

$$S^{T_2(2)} \approx 2\log k_2 + S_0 + \frac{2\pi^2}{\beta}. \tag{3.84}$$

To compute $S^{T_2(2)}$, we first need to compute $\sum_i \lambda_i^2 = \mathcal{N}_2^{(\text{even})}$. At transition, the naive

answer is given by Table 3.1, where $\mathcal{N}_2^{(\text{even})} = Z_2/Z_1^2$. However, using (3.76), we find

$$\begin{aligned}\mathcal{N}_2^{(\text{even})} &= \int_{-\infty}^{\infty} d\lambda D(\lambda)\lambda^2 \\ &= e^{S_0}\int_0^{s_k} ds\rho(s)\left(k_2^2\lambda_0^2 + 2\lambda_0 w(s) + w(s)^2\right)\end{aligned} \tag{3.85}$$

Again, $s_k < s^{(1)}$, so $\lambda_0 \approx 1/k$. This integral gives

$$\begin{aligned}\mathcal{N}_2^{(\text{even})} &= \frac{1}{k} + \frac{2e^{S_0}\rho(s_k)w(s_k)}{k} + \frac{Z_2}{2Z_1^2} \\ &\approx \frac{1}{k} + \frac{2e^{-\pi^2/2\beta}}{k} + \frac{Z_2}{2Z_1^2}.\end{aligned} \tag{3.86}$$

In the limit $k \gg e^{S_0}$, we can safely ignore the first two terms, and the Rényi-2 negativity becomes

$$\mathcal{N}_2^{(\text{even})} \approx \frac{Z_2}{2Z_1^2} \approx \frac{\sqrt{\pi}}{4}\beta^{3/2}e^{-S_0-3\pi^2/\beta}. \tag{3.87}$$

The factor of $1/2$ out front may seem like a problem, as we do not reproduce the naive answer for $\mathcal{N}_2^{(\text{even})}$. However, as the refined Rényi negativities are functions of $\log\mathcal{N}_n$, this factor will only contribute an $\mathcal{O}(1)$ difference from the true answer for $S^{T_2(2)}$, and we are safe in using this approximation. The refined Rényi-2 negativity is therefore given by

$$\begin{aligned}S^{T_2(2)} &= -\int_{-\infty}^{\infty} d\lambda D(\lambda)\frac{\lambda^2}{\mathcal{N}_2^{(\text{even})}}\log\frac{\lambda^2}{\mathcal{N}_2^{(\text{even})}} \\ &= -\frac{e^{S_0}}{\mathcal{N}_2^{(\text{even})}}\int_0^{s_k} ds\rho(s)\left(\frac{k_2(k_2+1)}{2}\left(\lambda_0 + \frac{w(s)}{k_2}\right)^2\log\left(\lambda_0 + \frac{w(s)}{k_2}\right)^2\right. \\ &\quad \left. + \frac{k_2(k_2-1)}{2}\left(\lambda_0 - \frac{w(s)}{k_2}\right)^2\log\left(\lambda_0 - \frac{w(s)}{k_2}\right)^2\right) \\ &\quad + \log\mathcal{N}_2^{(\text{even})}\end{aligned} \tag{3.88}$$

Again, as $\lambda_0 \approx 1/k$, the dominant contribution to this integral will come from the $w(s)^2$

term, which is where the enhanced transition should come in. Previously, using (3.59), we showed that an integral of the form $\rho(s)w(s)^n$ can be approximated by a sharply peaked Gaussian with mean $s^{(n)}$ and standard deviation $1/\sqrt{n\beta}$, up to normalization. We use these simplifications to obtain

$$
\begin{aligned}
&-\frac{e^{S_0}}{\mathcal{N}_2^{(\text{even})}}\int_0^{s_k} ds\rho(s)w(s)^2\log\frac{w(s)^2}{k_2^2} \qquad (3.89)\\
&=-\frac{e^{S_0}}{\mathcal{N}_2^{(\text{even})}}\int_0^{s_k} ds\rho(s)w(s)^2\left(\log\frac{w(s^{(2)})^2}{k_2^2}+\log\frac{w(s)^2}{w(s^{(2)})^2}\right)\\
&=-\log\left(\frac{w(s^{(2)})^2}{k_2^2}\right)-\frac{e^{S_0}}{\mathcal{N}_2^{(\text{even})}}\int_0^{s_k} ds\left(\rho(s)w(s)^2\log\frac{w(s)^2}{w(s^{(2)})^2}\right)\\
&\approx-\log\left(\frac{w(s^{(2)})^2}{k_2^2}\right)-2\sqrt{\frac{\beta}{\pi}}\int_0^{s^{(2)}} dse^{-\beta\left(s-s^{(2)}\right)^2}\beta\left(\left(s^{(2)}\right)^2-s^2\right)\\
&\approx 2\log k_2+2S_0+\frac{5\pi^2}{\beta}-\sqrt{\frac{4\pi}{\beta}}+\mathcal{O}(1) \qquad (3.90)
\end{aligned}
$$

Our final expression for $S^{T_2(2)}$ is therefore

$$
S^{T_2(2)}=2\log k_2+S_0+\frac{2\pi^2}{\beta}-\sqrt{\frac{4\pi}{\beta}} \qquad (3.91)
$$

confirming that there is an $\mathcal{O}(1/\sqrt{\beta})$ correction at transition.

The fact that the refined Rényi-2 negativity experiences this particular correction is not surprising due to its close connection to von Neumann entropies. It is known that von Neumann entropies receive $\mathcal{O}(1/\sqrt{\beta})$ or $\mathcal{O}(1/\sqrt{G_N})$ corrections at the Page transition [61, 131, 132], which can be explained using a diagonal approximation with respect to a basis of fixed-area states [131, 132, 141]. It was shown that the refined Rényi-2 negativity can be written in holography as the sum of the von Neumann entropies of R_1 and R_2 in the state $\rho^2_{R_1R_2}$ (once properly normalized) [128].[9] Therefore, the $\mathcal{O}(1/\sqrt{\beta})$

[9]This can be understood in terms of two cosmic branes homologous to R_1 and R_2, respectively, in the gravity dual of the even Rényi negativity. These cosmic branes arise from a $\mathbb{Z}_{n/2}$ quotient and therefore

correction that we find in the refined Rényi-2 negativity can similarly be explained using the diagonal approximation.

As we show in Appendix E, the Rényi entropy S_n with $n < 1$ experiences $\mathcal{O}(1/\beta)$ corrections in the model of [61], as there too we are computing an entanglement measure with $s_k^{(n)} < s^{(n)}$. In other words, the Rényi index of both the logarithmic negativity and the Rényi entropy with $n < 1$ is below some "critical" Rényi index at which there exist $\mathcal{O}(1/\sqrt{G})$ corrections.

For measures descending from odd Rényi negativity, we might not expect $\mathcal{O}(1/\sqrt{\beta})$ corrections, as we never have $s_k^{(n,\text{odd})} = s^{(n)}$. However, we may still expect some enhanced corrections for some negativity measures in this case. The partially transposed entropy S^{T_2} is one such measure. As $s_k^{(1,\text{odd})} = 5\pi/2\beta$, the naive answer for S^{T_2} is given by

$$S^{T_2} = \log k_2 + S_0 + \frac{4\pi^2}{\beta}. \tag{3.92}$$

Our approximation gives

$$\begin{aligned} S^{T_2} &= -\int_{-\infty}^{\infty} d\lambda D(\lambda)\lambda \log|\lambda| \\ &= -e^{S_0}\int_0^{s_k} ds\rho(s)\left(\frac{k_2(k_2+1)}{2}\left(\lambda_0+\frac{w(s)}{k_2}\right)\log\left(\lambda_0+\frac{w(s)}{k_2}\right)\right. \\ &\quad \left.+\frac{k_2(k_2-1)}{2}\left(\lambda_0-\frac{w(s)}{k_2}\right)\log\left(\frac{w(s)}{k_2}-\lambda_0\right)\right). \end{aligned} \tag{3.93}$$

There is however a subtlety here. The dominant contribution to this integral no longer comes solely from the $w(s)$ term. This can be seen by expanding (3.93) using

$$\log\left(\frac{w(s)}{k_2}\pm\lambda_0\right) \approx \log\frac{w(s)}{k_2} \pm \frac{\lambda_0 k_2}{w(s)}. \tag{3.94}$$

become tensionless in the $n \to 2$ limit, which is similar to the case of the von Neumann entropy.

From this we obtain

$$\begin{aligned} S^{T_2} &= -e^{S_0}\int_0^{s_k} ds\rho(s)\left(k_2^2\left(\lambda_0+\lambda_0\log\frac{w(s)}{k_2}\right)+k_2\left(\frac{\lambda_0^2k_2}{w(s)}+\frac{w(s)}{k_2}\log\frac{w(s)}{k_2}\right)\right) \\ &\approx -e^{S_0}\int_0^{s_k} ds\rho(s)\left(k_2^2\lambda_0+w(s)\right)\log\frac{w(s)}{k_2}. \end{aligned} \tag{3.95}$$

Treating the $\log\frac{w(s)}{k_2}$ as negligible compared to the exponential $\rho(s)$, the two terms in the integrand are of the same order, so we should keep them both. If we look at the naive transition point $s_k = 5\pi/2\beta > s^{(1)}$, we find

$$\begin{aligned} S^{T_2} &\approx -e^{S_0}\int_0^{s_k}\rho(s)w(s)\log\frac{w(s)}{k_2} \\ &\approx -\log\frac{w(s_k)}{k_2} \\ &\approx \log k_2 + S_0 + \frac{4\pi^2}{\beta}, \end{aligned} \tag{3.96}$$

and we would conclude that the correction is $\mathcal{O}(1)$. However, if we were to find the largest correction to this quantity, we would look not at the naive transition, but at the point where we might find $\mathcal{O}(1/\sqrt{\beta})$ corrections, at $s_k = s^{(1)}$. At this point $\lambda_0 = 1/2k$ and we capture half of the Gaussian $\rho(s)w(s)$, so we have

$$\begin{aligned} S^{T_2} &\approx -\frac{1}{2}\log\frac{w(s^{(1)})}{k_2} - \sqrt{\frac{\beta}{2\pi}}\int_0^{s^{(1)}} e^{-\beta(s-s^{(1)})^2/2}\frac{\beta}{2}\log\frac{w(s)}{k_2} \\ &\approx -\log\frac{w(s^{(1)})}{k_2} - \sqrt{\frac{\beta}{2\pi}}\int_0^{s^{(1)}} e^{-\beta(s-s^{(1)})^2/2}\frac{\beta}{2}\left(\left(s^{(1)}\right)^2 - s^2\right) \\ &\approx \log k_2 + S_0 + \frac{4\pi^2}{\beta} - \sqrt{\frac{2\pi}{\beta}}. \end{aligned} \tag{3.97}$$

This looks the same as the naive answer with a $\mathcal{O}(1/\sqrt{\beta})$ correction. However, as we are working at fixed k, we should really be writing everything in terms of k and e^{S_0} using

(3.62), in which case our naive and corrected S^{T_2}'s are

$$\begin{aligned} &\text{Naive:} && S^{T_2} = \frac{1}{2}\left(\log k - S_0\right) + \frac{3\pi^2}{2\beta} \\ &\text{Corrected:} && S^{T_2} = \frac{1}{2}\left(\log k - S_0\right) + \frac{2\pi^2}{\beta}, \end{aligned} \tag{3.98}$$

and we find an $\mathcal{O}(1/\beta)$ correction.

3.6 Topological Model with EOW Branes

Having studied a toy model of an evaporating black hole in JT gravity, we will now consider entanglement negativity in the context of the topological model of Marolf-Maxfield [62], including dynamical end-of-the-world (EOW) branes. This is a theory of topological two-dimensional gravity in which spacetimes are two-dimensional manifolds endowed only with orientation. In contrast to the JT model of Section 3.3, there are no metric or dilaton degrees of freedom and EOW brane boundaries can be generated dynamically.

The action for the topological model is given by

$$S_{\text{top}} = -S_0\chi(M) - S_\partial|\partial M|, \tag{3.99}$$

where S_0 is some arbitrary parameter, $\chi(M)$ is the Euler characteristic of the (possibly disconnected) manifold M, and $|\partial M|$ counts the number of boundaries. $S_\partial|\partial M|$ is a nonlocal term that we put in by hand to ensure reflection positivity. As shown in [62], the simplest choice which results in reflection positivity is $S_\partial = S_0$.[10] The action then

[10] Other valid choices are $S_\partial = S_0 + \log m$ for any positive integer m or $S_\partial > S_0 + \log k$. Any of these choices give a discrete spectrum for the operator $\hat{Z}$.

becomes

$$S_{\text{top}} = -S_0\widetilde{\chi} \tag{3.100}$$

where $\widetilde{\chi} = 2 - 2g$ for any manifold, with or without boundary. To make contact with black hole evaporation, we can extend the model to include EOW branes, which can take one of k "flavors". Since we are interested in studying negativity, we will allow the branes to be labeled by a set of two of flavor indices $\{i, j\}$, where $i \in \{1, \ldots, k_1\}$, $j \in \{1, \ldots, k_2\}$, and such that $k = k_1 k_2$. This is exactly analogous to our construction in Section 3.3 and is a slight generalization of the model in [62].

There are three distinct types of boundaries allowed by the theory. The first are circular asymptotically AdS boundaries denoted by Z. These boundaries are associated with an operator $\widehat{Z}$ which acts on the baby universe Hilbert space $\mathcal{H}_{\text{BU}}$ and creates a Z boundary. Second, there are boundary conditions which we denote by $(\psi_{i_2 j_2}, \psi_{i_1 j_1})$ composed of an oriented interval of asymptotically AdS boundaries with endpoints labeled by flavor indices $\{i_1, j_1\}$ and $\{i_2, j_2\}$. The diagram that describes this is the same as in (3.18). These boundaries are associated with an operator $\widehat{(\psi_{i_2 j_2}, \psi_{i_1 j_1})}$ on $\mathcal{H}_{\text{BU}}$. Finally, there are circular EOW brane boundaries labeled by an arbitrary flavor index $\{i, j\}$, independent of all boundary conditions. These brane boundaries can be dynamically generated as additional boundaries when performing the gravitational path integral.

Let us consider the simplest quantity one can compute with the gravitational path integral of this theory, namely the partition function associated to a single connected component of spacetime with some number of asymptotic boundaries. The gravitational path integral demands that we sum over all such manifolds with arbitrary genus, weighted by $e^{-S_0\widetilde{\chi}}$. Additionally, since the EOW branes are dynamical, there is the possibility of the gravitational path integral generating an arbitrary number of closed brane boundaries, each of which contribute a factor of k and are mutually indistinguishable. The parti-

tion function for a single connected component with some fixed number of asymptotic boundaries is therefore

$$\lambda \equiv \sum_{g=0}^{\infty} \sum_{m=0}^{\infty} e^{(2-2g)S_0} \frac{k^m}{m!} = \frac{e^{2S_0}}{1-e^{-2S_0}} e^k. \tag{3.101}$$

More generally, one can consider amplitudes

$$\left\langle Z^m(\psi_{i'_1 j'_1}, \psi_{i_1 j_1}) \cdots (\psi_{i'_n j'_n}, \psi_{i_n j_n}) \right\rangle \equiv \left\langle \mathrm{NB} \middle| \widehat{Z}^m(\widehat{\psi_{i'_1 j'_1}, \psi_{i_1 j_1}}) \cdots (\widehat{\psi_{i'_n j'_n}, \psi_{i_n j_n}}) \middle| \mathrm{NB} \right\rangle \tag{3.102}$$

which are computed using the gravitational path integral by summing over all (possibly disconnected) manifolds with boundary conditions specified by m circular boundaries Z and n oriented intervals $(\psi_{i'j'}, \psi_{ij})$ with endpoints labeled by the corresponding flavor indices and connected to oriented brane boundaries labeled with matching flavors. The brackets in (3.102) can be interpreted the expectation value of the corresponding operators in the no-boundary state $\big|\mathrm{NB}\big\rangle \in \mathcal{H}_{\mathrm{BU}}$. In what follows, we will assume that the no-boundary state is unit normalized, $\big\langle \mathrm{NB} \big| \mathrm{NB} \big\rangle = 1$.[11]

Let us now proceed to the calculation of negativity in this model. In analogy to Section 3.3, we can define the (unnormalized) density matrix

$$\rho = \sum_{i_1, i_2 = 1}^{k_1} \sum_{j_1, j_2 = 1}^{k_2} |i_1, j_1\rangle\langle i_2, j_2| (\psi_{i_2 j_2}, \psi_{i_1 j_1}), \tag{3.103}$$

which plays the role of the state of the Hawking radiation. We are interested in studying the Rényi negativities $\mathcal{N}_n = \mathrm{Tr}\left[\left(\rho^{T_2}\right)^n\right]$ of this density matrix. The most straightforward method is to use the moment generating function of $\rho^{i_1 i_2}_{j_1 j_2} \equiv (\psi_{i_2 j_2}, \psi_{i_1 j_1})$, which one can

[11] In [62], the no-boundary state has inner product $\big\langle \mathrm{NB} \big| \mathrm{NB} \big\rangle = e^{\lambda}$ and represents the sum over arbitrary numbers of closed universes. This normalization enters as a universal prefactor in all amplitudes we compute, so we can choose to normalize it to one.

show is given by

$$\left\langle \exp\left(\sum_{i_1,i_2=1}^{k_1}\sum_{j_1,j_2=1}^{k_2} t_{j_1 j_2}^{i_1 i_2}\rho_{j_1 j_2}^{i_1 i_2}\right)\right\rangle = e^{-\lambda}\exp\left[\lambda\det\left(\mathbb{I}-t\right)^{-1}\right], \tag{3.104}$$

where t can be thought of as the $k\times k$ matrix with entries $t_{j_1 j_2}^{i_1 i_2}$ by treating $\{i,j\}$ as a single index of size k and $\mathbb{I}$ is the $k\times k$ identity matrix . This is a slight generalization of the result derived in [62]. In principle, one can compute all moments of $\rho_{j_1 j_2}^{i_1 i_2}$, and hence all Rényi negativities, by taking appropriate partial derivatives of (3.104) with respect to $t_{j_1 j_2}^{i_1 i_2}$.

However, there is a shortcut that we will now describe. As shown in [62], the spectrum of the operator $\widehat{Z}$ takes values in $\mathbb{N}$. One can derive the distribution for $\rho_{j_1 j_2}^{i_1 i_2}$ in a fixed $Z=d\in\mathbb{N}$ sector by Taylor expanding (3.104) in λ:

$$\begin{aligned}\left\langle \exp\left(\sum_{i_1,i_2=1}^{k_1}\sum_{j_1,j_2=1}^{k_2} t_{j_1 j_2}^{i_1 i_2}\rho_{j_1 j_2}^{i_1 i_2}\right)\right\rangle &= \sum_{d=0}^{\infty} p_d(\lambda)\left\langle \exp\left(\sum_{i_1,i_2=1}^{k_1}\sum_{j_1,j_2=1}^{k_2} t_{j_1 j_2}^{i_1 i_2}\rho_{j_1 j_2}^{i_1 i_2}\right)\right\rangle_{Z=d}\\ \implies \left\langle \exp\left(\sum_{i_1,i_2=1}^{k_1}\sum_{j_1,j_2=1}^{k_2} t_{j_1 j_2}^{i_1 i_2}\rho_{j_1 j_2}^{i_1 i_2}\right)\right\rangle_{Z=d} &= \det\left(I-t\right)^{-d},\end{aligned} \tag{3.105}$$

where $p_d(\lambda)=e^{-\lambda}\frac{\lambda^d}{d!}$ is a Poisson distribution. We can recognize (3.105) as the moment generating function for a Wishart distribution with d degrees of freedom, and coincides with the distribution of a random mixed state and the microcanonical JT model in Section 3.5.1. Thus, we can immediately write down the Rényi negativities in a fixed $Z=d$ sector

$$\left\langle \mathrm{Tr}\left(\rho_R^{T_2}\right)^n\right\rangle_{Z=d} = \sum_{g\in S_n} d^{\chi(g)} k_1^{\chi(g^{-1}X)} k_2^{\chi(g^{-1}X^{-1})}, \tag{3.106}$$

which matches the answer for the microcanonical ensemble (3.52) with d playing the role of $e^{\mathbf{S}}$. The results for the negativity spectrum obtained in Sections 3.3.2 and 3.5.1 there-

fore apply with this replacement. However, since d does not correspond to the partition function on some manifold, it is difficult to interpret the result in (3.106) geometrically.

To obtain the Rényi negativities in the full theory, we simply sum over $d \in \mathbb{N}$ with Poisson weight $p_d(\lambda)$:

$$\begin{aligned}\left\langle \operatorname{Tr}\left(\rho_R^{T_2}\right)^n \right\rangle &= \sum_{d=0}^{\infty} p_d(\lambda) \left\langle \operatorname{Tr}\left(\rho_R^{T_2}\right)^n \right\rangle_{Z=d} \\ &= \sum_{g\in S_n} B_{\chi(g)}(\lambda) k_1^{\chi(g^{-1}X)} k_2^{\chi(g^{-1}X^{-1})}\end{aligned} \tag{3.107}$$

where $B_m(\lambda) = e^{-x}\sum_{k=0}^{\infty}\frac{\lambda^k k^m}{m!}$ are the Bell polynomials, whose asymptotic behavior is $B_m(\lambda) \sim \lambda^m$ as $\lambda \to \infty$. We therefore find

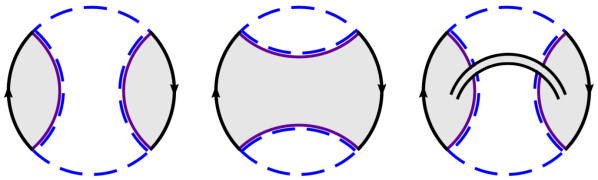
$$\left\langle \operatorname{Tr}\left(\rho_R^{T_2}\right)^n \right\rangle \approx \sum_{g\in S_n} \lambda^{\chi(g)} k_1^{\chi(g^{-1}X)} k_2^{\chi(g^{-1}X^{-1})}, \qquad \lambda \gg 1. \tag{3.108}$$

This is once again equivalent to the microcanonical ensemble in (3.52), with λ now playing the role of $e^{\mathbf{S}}$, and therefore we can obtain concrete results for the negativity spectrum. Since λ is the gravitational partition function of a single connected component of spacetime, we can in fact find a geometric interpretation for the terms in (3.107).

To understand the geometric origins of the terms in (3.107), let us first look at the case $n = 2$, which gives the purity

$$\left\langle \operatorname{Tr}\rho^2 \right\rangle = \lambda^2 k + \lambda k^2 + \lambda k. \tag{3.109}$$

The terms in (3.109) correspond to the following geometries:



(3.110)

The first two diagrams are familiar: they are the disk and wormhole geometries, summed over genus and closed brane boundaries. The last diagram represents two disk geometries joined by an arbitrary number of wormholes; we thus call it a *joining wormhole*.

More generally, the geometries which contribute at leading order in λ in the Rényi negativities (3.107) are in one-to-one correspondence with elements of the permutation group. To be precise, each of these geometries is actually a disjoint union of disks, summed over genus and closed brane boundaries.[12] The subleading contributions in (3.107) can be identified with the same geometries but with arbitrary numbers of joining wormholes between connected components, and thus can not be mapped to elements of the permutation group.

It is clear that $\log \lambda$ plays the same role as $\mathbf{S}$ in the microcanonical JT model, namely it is the Bekenstein-Hawking entropy. In analogy to black hole evaporation, we should assume $\lambda \gg 1$. The joining wormholes are therefore parametrically suppressed, but disks with handles are not since they instead come with factors of e^{-2gS_0} and S_0 is not *a priori* a large parameter (in fact, it may have a small or even negative real part).[13] This is the analogue of the "planar" limit in the topological model. There are thus two distinct classes of higher genus geometries: disks with handles and joining wormholes. The higher

[12]This is in contrast with the JT model where we identify only a single geometry, namely some disjoint union of disks with no handles, with each element of the permutation group. In that case, the sum over genus is highly suppressed by factors of $e^{-\mathbf{S}}$, and they are in fact as suppressed as geometries with joining wormholes. Furthermore, in the JT model closed brane boundaries can not be dynamically generated.

[13]In the JT model both geometries are suppressed in the same parameter $e^{-\mathbf{S}}$.

genus disk geometries can be systematically included in a Schwinger-Dyson equation as in Section 3.4, while the joining wormholes can not.

To study the Page curve, we would like to fix the value of λ and tune k. Since $\lambda \sim e^k$, this involves scaling the prefactor $\frac{e^{2S_0}}{1-e^{-2S_0}}$ with e^{-k}. However, this function has a minimum value for real S_0, which means the black hole can not evaporate completely. To decrease λ beyond the minimum value, we need to go to complex values of S_0, namely $e^{2S_0} \in \frac{1}{2}+i\mathbb{R}$. This is a bit strange because it implies a complex action, but is presumably fine because S_0 is not a physical parameter (it is not the Bekenstein-Hawking entropy here). This is simply a quirk of the model, and can be attributed as a consequence of having a nonvanishing S_∂.

3.7 Discussion

In this chapter, we analyzed the behavior of negativity measures in toy models of evaporating black holes in both JT gravity and a topological theory of gravity, with EOW branes. We found four distinct phases dominated by different saddle-point geometries: the disconnected, cyclically connected, anti-cyclically connected, and pairwise connected. The last of these geometries are new replica wormholes that break the replica symmetry spontaneously.

We also studied the negativity resolvent using a Schwinger-Dyson equation that resums the contributions of different geometries, and used it to extract the negativity spectrum and negativity measures. This analysis is valid not only within each of the four phases, but also near phase transitions. For the topological model or a microcanonical ensemble in JT gravity, we found a cubic equation for the resolvent which can be solved exactly. For a canonical ensemble in JT gravity, we found a quadratic resolvent equation near the disconnected-pairwise transition, and we solved a more complicated resolvent

equation approximately near the cyclic-pairwise transition. Near this last transition, we found enhanced corrections to various negativity measures: the refined Rényi-2 negativity receives an $\mathcal{O}(1/\sqrt{\beta})$ correction, whereas the logarithmic negativity and the partially transposed entropy receive $\mathcal{O}(1/\beta)$ corrections.

These enhanced corrections to negativities are similar to previously found corrections to the von Neumann entropy at the Page transition [61, 131, 132]. For the von Neumann entropy, the enhanced corrections can be explained using a diagonal approximation with respect to a basis of fixed-area states [131, 132, 141]. We argued that the $\mathcal{O}(1/\sqrt{\beta})$ correction to the refined Rényi-2 negativity can be explained in the same way by noting its close connection to von Neumann entropies. It would be interesting to understand further the $\mathcal{O}(1/\beta)$ corrections to the logarithmic negativity and the partially transposed entropy in a similar way. Moreover, it would be useful to study the implications of these $\mathcal{O}(1/\beta)$ corrections for the partially transposed entropy more generally: it was conjectured in [128] that $S^{T_2}(\rho_{R_1R_2})$ is given as a sum of von Neumann entropies $(S_{R_1} + S_{R_2} + S_{R_1R_2})/2$ in general non-fixed-area states by assuming a diagonal approximation, but this would imply an $\mathcal{O}(1/\sqrt{\beta})$ correction and seems to be in tension with the $\mathcal{O}(1/\beta)$ correction that we find here.

We focused our study on two specific toy models of evaporating black holes, but it would be interesting to generalize our analysis to other models, including the examples studied in [60].

Finally, it would be very interesting to use these results on negativity to diagnose the structure of multipartite entanglement in a realistic evaporating black hole and learn more about its quantum state. We hope that this will lead to new insights on understanding the interior of black holes and the dynamics of their evaporation.

Chapter 4

Entanglement Negativity Transitions in Chaotic Eigenstates

4.1 Introduction

The application of ideas from quantum chaos to gravitational settings has been particularly fruitful. Gravitational observables have been shown to be well approximated by observables obeying the eigenstate thermalization hypothesis (ETH) [142–145]. This is due to the fact that a holographic quantum field theory with a semiclassical Einstein gravity dual is expected to be maximally chaotic, i.e. it saturates the bound of [146], up to higher derivative/stringy corrections which take one away from this regime. The power of ETH is that it allows us to approximate observables in the microcanonical ensemble by an observable's long-time quantum expectation value. The resulting microcanonical expectation value should resemble that of the canonical ensemble, up to corrections expected to be suppressed in the system size V by the thermodynamic ensembles' equivalence at large N. This gives a quantitative idea of the process of thermalization in isolated quantum many-body systems.

From the perspective of subsystem ETH [147–151], for a subsystem with volume fraction $f < 1/2$, the corrections to ETH are suppressed in system size. Formally, this means the trace-norm distance between the canonical and microcanonical density matrices vanishes in the large volume/thermodynamic limit, implying that off-diagonal matrix elements of operators vanishes and expectation values are roughly thermal. This line of thinking is expected to apply to Rényi entropies.

When $f = 1/2$ exactly, the usual wisdom would say there's an $\mathcal{O}(1)$ correction to the Rényi entropies. One way of seeing this is that there exists a phase transition in the Rényi entropy at $f = 1/2$, and the correction at this phase transition should be given by the uncertainty in choosing between an $\mathcal{O}(1)$ number of equivalent dominant phases. However, in a model-specific result, a correction to the entanglement entropy of $\mathcal{O}(\sqrt{V})$ was observed in [152], a correction derived in [153] and explicated in [131]. In particular, the von Neumann entropy of a subregion A with volume fraction $f = 1/2$ takes the form

$$S_A = \frac{S(E)}{2} - \sqrt{\frac{C_V}{2\pi}} + \mathcal{O}(a) \tag{4.1}$$

where $S(E)$ is the thermodynamic entropy at energy E and C_V is the heat capacity at constant volume. As C_V is extensive in the system size, the correction is "enhanced" to $\mathcal{O}(\sqrt{V})$. This formula is valid in the large volume limit, where $\sqrt{V} \gg a$, a being the area of the splitting surface.

In a parallel story, the attempt to match results from tensor networks with the gravitational path integral led to the understanding of "fixed area states" [141, 154]. These states, which are eigenstates of the area operator in semiclassical gravity, have a flat entanglement spectrum, up to fluctuations about a fixed saddle point which can naïvely be at most $\mathcal{O}(1)$ in units of G_N, where $G_N \ll 1$. One can think of these fluctuations as the difference in the "canonical" ensemble where one fixes the canonical conjugate to the

area operator, namely the relative boost between the entanglement wedges of the two sides [155], and a "microcanonical" ensemble where the eigenvalue of the area operator is fixed at its most probable value.

It was noted in [131, 132] that the universal enhanced correction to the entanglement entropy also appears in fixed area states near transition, where the "transition" in this context occurs due to a competition between two extremal surfaces. One way of understanding this correction is that near transition we no longer care about fluctuations about a fixed saddle, but instead we care about resumming an infinite number of saddles which appear in the sum over topologies in the replicated geometries. Both of these results match with a more detailed calculation of the same quantity in [61], where in a model of Jackiw-Teitelboim (JT) gravity + end-of-the-world (EOW) branes a particular subsystem entropy $S(\rho_R)$ had the form

$$S(\rho_R) = \log k - \sqrt{\frac{2\pi}{\beta}} + \mathcal{O}(\log \beta). \tag{4.2}$$

This $\sqrt{1/\beta}$ correction is analogous to the $\sqrt{C_V}$ correction in chaotic eigenstates. Here we've set Newton's constant (which is analagous to N) to one, but it can be restored via $\beta \to G_N \beta$.

Recently, it was shown by [67] that similar enhanced corrections exist near transitions in entanglement negativity, a tripartite entanglement measure defined on a bipartite density matrix $\rho_{A_1A_2}$. In particular, the logarithmic negativity was shown to have the following form at transition:

$$\mathcal{E}(\rho_{R_1R_2}) = \log k_2 - \frac{\pi^2}{8\beta} + \mathcal{O}(\log \beta), \tag{4.3}$$

for two subsystems R_1 and R_2 with $R_1 \cup R_2 = R$. Further corrections were derived for

measures descending from a Rényi version of negativity.

There exists a rich phase diagram for entanglement negativity in holographic states, and we show that a similar phase diagram exists for a generic chaotic eigenstate. Our aim is to systematically derive the corrections at transitions in this phase space. There are two possible transitions, but as was explored in [67] we only expect interesting behavior near one of the transitions, for reasons we'll recapitulate in the main text.

The outline of the chapter is as follows. In section 4.2 we review the derivation of (4.1), in particular the resolvent formalism of [131]. In section 4.3 we review the various negativity measures discussed in [67] and the sum over relevant permutations for the phase transition of interest. In section 4.4 we compute corrections to the entanglement measures of interest. We conclude with some discussion and future directions.

4.2 Diagrammatics for Chaotic Subsystems

We first review the formalism of [131, 153], which was used to compute the universal form of corrections to the entanglement entropy of a subsystem at transition. We focus on [131], as their formalism more easily generalizes to our future calculations. Readers familiar with their formalism may skip this section, whose only purpose is to make this work self-contained.

A generic eigenstate $|E\rangle$ of a Hamiltonian H defined on a bipartite system such that $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ can be Schmidt decomposed via

$$|E\rangle = \sum_{iJ} M_{iJ} \, |E_i\rangle_A \otimes |E_J\rangle_B \,, \tag{4.4}$$

where $|E_i\rangle$ and $|E_J\rangle$ denote eigenstates of the subsystem Hamiltonians H_A and H_B, respectively. As is convention, we use lowercase indices for states of A and uppercase

indices for states of B. ETH instructs us to think of M_{iJ} as a Gaussian random variable with zero mean and energy banded with width Δ [156, 157]. In particular, for a system with spatial dimension $d \geq 2$, we have the ansatz

$$M_{iJ} = e^{-S(E_{Ai}+E_{BJ})/2} \left(\frac{e^{-\epsilon^2/2\Delta^2}}{\sqrt{2\pi}\Delta} \right)^{1/2} C_{iJ}, \tag{4.5}$$

where $\epsilon = E_i + E_J - E$ is the deviation from the total microcanonical energy. When averaged over a small energy band in E_A and E_B, the random coefficients C_{iJ} satisfy

$$\overline{C_{iJ}} = 0, \quad \overline{C_{iJ}C_{i'J'}} = \delta_{ii'}\delta_{JJ'}. \tag{4.6}$$

The effects of finite Δ will not affect the current and future computation, so we work in the limit $\Delta \to 0$, where we approximate

$$M_{iJ} \approx e^{-S(E)/2} C_{iJ}. \tag{4.7}$$

This approximation assumes the true density of states in a narrow energy band is well approximated by the thermodynamic entropy in the canonical ensemble. To leading order in the system volume, we can further approximate the density of states of the total system as the product of the density of states of the subsystems A and B, evaluated at the subsystem energy E_A. In other words,

$$S(E) \approx S_A(E_A) + S_B(E - E_A). \tag{4.8}$$

This leads to the following form for a subsystem density matrix ρ_A

$$\rho_A = \frac{1}{\mathcal{N}} \sum_{E_i-2\Delta<E_j<E_i+2\Delta} \sum_{E-E_i-\Delta<E_J<E-E_i+\Delta} C_{iJ}C_{jJ} \left|E_i\right\rangle_A \left\langle E_j\right|_A . \tag{4.9}$$

The double sum takes into account energies in a region of width 2Δ. Averaging over the C_{ij}'s gives the averaged subsystem density matrix

$$\overline{\rho_A} = \frac{1}{\mathcal{N}} \sum_i d_B(E-E_i) \left|E_i\right\rangle_A \left\langle E_i\right|_A , \tag{4.10}$$

where the normalization is given by

$$\mathcal{N} = \sum_i d_A(E_i) d_B(E-E_i). \tag{4.11}$$

Here d_A and d_B are the degeneracies at a given energy. As a shorthand and as a motivation for our future computation, we can instead write

$$d_A(E_i) = e^{S_A(E_i)} \equiv e^{S_A}; \quad d_B(E-E_i) = e^{S_B(E-E_i)} \equiv e^{S_B}. \tag{4.12}$$

As our goal is to compute subsystem von Neumann entropy, we should proceed by generalizing this procedure to compute $\operatorname{Tr} \overline{\rho_A^n}$. Before averaging, from (4.9) we have

$$\rho_A^n = \frac{1}{\mathcal{N}^n} \sum_{E_{i_1}} \sum_{i_2,\cdots,i_{n+1};J_1,\cdots,J_n} \prod_{m=1}^{n} C_{i_m J_m} C_{i_{m+1} J_m} \left|E_{i_1}\right\rangle_A \left\langle E_{i_{n+1}}\right|_A , \tag{4.13}$$

where the second sum is understood to be over a strip of width $2n\Delta$, but we assume Δ vanishes quickly enough at finite n that this isn't a significant effect.

The difference between $\overline{\log \operatorname{Tr} (\rho_A)^n}$ and $\log \operatorname{Tr} \overline{(\rho_A)^n}$ is exponentially suppressed in the system volume [151], so our goal will be to compute $\operatorname{Tr} \overline{(\rho_A)^n}$, as it is a more tractable

calculation. This involves a sum over Wick contractions, as we assume higher point connected correlations of the C_{iJ}'s vanish. The result is

$$\mathrm{Tr}\,\overline{(\rho_A)^n} = \begin{cases} \frac{1}{\mathcal{N}^n} e^{S_A + nS_B}\, {}_2F_1\left(1-n,-n;2;e^{S_A-S_B}\right), S_A < S_B \\ \frac{1}{\mathcal{N}^n} e^{nS_A + S_B}\, {}_2F_1\left(1-n,-n;2;e^{S_B-S_A}\right), S_A > S_B. \end{cases} \tag{4.14}$$

As a sanity check, we recover $S_n(\rho_A) = S/2 + \mathcal{O}(1)$ where $S_A = S_B = S/2$ at $f = 1/2$. The derivation of this expression from the resolvent sum over noncrossing permutations is given in Appendix G.

To study the corrections to this quantity at transition, we upgrade the putative constant density of states to an integral over an energy dependent density of states. In other words, we send

$$S_A \to S_A(E_A), \quad S_B \to S_B(E - E_A) \tag{4.15}$$

and integrate over E_A. The new averaged trace is given by

$$\mathrm{Tr}\,\overline{(\rho_A)^n} = \frac{1}{\mathcal{N}^n} \int dE_A e^{S_A(E_A) + S_B(E-E_A)} G_n(E_A), \tag{4.16}$$

where $G_n(f, E_A)$ encompasses the n-dependent piece of the trace:

$$G_n(E_A) = \begin{cases} e^{(n-1)S_B(E-E_A)}\, {}_2F_1\left(1-n,-n;2;e^{S_A(E_A)-S_B(E-E_A)}\right), S_A(E_A) < S_B(E-E_A) \\ e^{(n-1)S_A(E_A)}\, {}_2F_1\left(1-n,-n;2;e^{S_B(E-E_A)-S_A(E_A)}\right), S_A(E_A) > S_B(E-E_A) \end{cases} \tag{4.17}$$

and the normalization is now

$$\mathcal{N} = \int dE_A e^{S_A(E_A) + S_B(E-E_A)}. \tag{4.18}$$

From this, we can directly calculate the ensemble averaged Rényi entropies $S_n(\rho_A)$:

$$\overline{S_n} = \frac{1}{1-n} \log \left(\frac{1}{\mathcal{N}^n} \int dE_A e^{S_A(E_A)+S_B(E-E_A)} G_n(E_A) \right). \tag{4.19}$$

4.2.1 Saddle Point Analysis

We make the ansatz that the entropy is extensive in the subsystem size, that is

$$S_A(E_A) = fVs\left(\frac{E_A}{fV}\right), \quad S_B(E-E_A) = (1-f)Vs\left(\frac{E-E_A}{(1-f)V}\right), \tag{4.20}$$

where $f \equiv V_A/V$ is the volume fraction, $s(e)$ is the entropy density as a function of the energy density e, and the other factors come from dimensional analysis. We're mainly interested in what happens at the transition $f = 1/2$. The "featureless" or infinite temperature case is when $s(e) = 1$ such that all subsystem entropies are proportional to subsystem volume. We're only interested in the corrections from finite temperature, which can be thought of as the difference between the answer in the canonical ensemble and the microcanonical ensemble. The microcanonical Rényi entropy is the contribution of the global "unaveraged" microcanonical state $\rho = \sum_{E-\Delta<E_i<E+\Delta} |E_i\rangle \langle E_i|$:

$$S_n^{MC} = \frac{1}{1-n} \log \left(\frac{1}{\mathcal{N}^n} \int dE_A e^{S_A(E_A)+nS_B(E-E_A)} \right). \tag{4.21}$$

We are interested in the correction away from the dominant microcanonical saddle, so we are interested in computing the following quantity:

$$\overline{S_n} - S_n^{MC} = \frac{1}{1-n} \ln \left(\frac{\int dE_A \exp(F_1(E_A))}{\int dE_A \exp(F_2(E_A))} \right), \tag{4.22}$$

where $F_1(E_A)$ and $F_2(E_A)$ are functions defined by

$$F_1(E_A) = fVs\left(\frac{E_A}{fV}\right) + (1-f)Vs\left(\frac{E-E_A}{(1-f)V}\right) + \ln G_n(E_A)$$
$$F_2(E_A) = fVs\left(\frac{E_A}{fV}\right) + n(1-f)Vs\left(\frac{E-E_A}{(1-f)V}\right). \tag{4.23}$$

As both functions scale with volume, we can perform a saddle point analysis. The saddle point equations for these functions are

$$s'\left(\frac{E_1}{fV}\right) = s'\left(\frac{E-E_1}{(1-f)V}\right) - \frac{G_n'(f,E_1)}{G_n(f,E_1)}$$
$$s'\left(\frac{E_2}{fV}\right) = ns'\left(\frac{E-E_2}{(1-f)V}\right), \tag{4.24}$$

where E_1 and E_2 are the saddle point energies of $F_1(E_A)$ and $F_2(E_A)$, respectively. The analysis of these saddle point equations was done in totality for $n > 1$ in [131].[1] Here we fill in a small gap and study the case of $n < 1$. This will be useful later when we are computing analytic continuations of Rényi negativities below $n = 1$.

4.2.2 Corrections at Transition for $n < 1$

$s(x)$ is a monotonically increasing function of x with a monotonically decreasing first derivative (take $s(x) = \sqrt{x}$ as a concrete example). For the case $n < 1$, we can therefore write the iequality

$$\frac{E_2}{f} > \frac{E-E_2}{1-f}, \tag{4.25}$$

which immediately implies

$$S_A(E_2) > \frac{f}{1-f}S_B(E-E_2), \tag{4.26}$$

[1] See also [158] for a similar study of relative entropy with the same ansatz.

and therefore $S_A(E_1) > S_B(E - E_1)$ for $f > 1/2$.

The first thing to notice is that there is only one saddle point for both $F_1(E_A)$ and $F_2(E_A)$, as $G_n(f, E_A)$ is now a strictly concave function. The single saddle for $F_1(E_A)$ depends sensitively on the saddle point of $G_n(f, E_A)$, which itself only depends on the crossover point between the two hypergeometrics. As the crossover point is completely determined by the n-independent quantity

$$S_A(E_A) - S_B(E - E_A) \tag{4.27}$$

and the rest of the E_1 saddle point equation is independent of n, the full saddle similarly becomes completely independent of n. This should be contrasted with the obviously n-dependent saddle point of $F_2(E_A)$. This difference will generically cause the two saddles to differ by an $\mathcal{O}(1)$ factor, so for all volume fractions we expect the different in Rényi entropies to be volume law:

$$\overline{S_n} - S_n^{MC} = \mathcal{O}(V), \quad n < 1. \tag{4.28}$$

Note that this applies for all volume fractions, implying that the $n < 1$ Rényi entropies do not obey the principle of canonical typicality.

This clarifies a conceptual point. For $n \to 1^+$, the $\sqrt{V}$ correction lies in between an exponentially suppressed $\mathcal{O}(e^{-cV})$ region ($f < 1/2$) and a strongly enhanced $\mathcal{O}(V)$ region ($f > 1/2$). Why, then, do we not get a similar enhancement for $n \to 1^-$? The answer is that the dominant behavior in $F_1(E_A)$, which previously supplied the emergent "soft mode" for the flat interval between two saddles, becomes independent of the Rényi index. This nonanalyticity might be worrying if one is used to a Rényi entropy analytic in n, but the thermodynamic limit breaks this assumption. The form of these corrections

agrees with the analysis of a gravitational model in Appendix C of [67].

4.3 Entanglement Negativity

In this section we compute similar quantities as [131] for entanglement negativity measures. We begin by reviewing some salient properties of entanglement negativity and its utility as a tripartite measure of entanglement before diving into the calculation.

4.3.1 Review of Negativity

Entanglement negativity refers to an entanglement measure based on properties of the partial transpose operation applied to a bipartite density matrix $\rho_{A_1A_2}$, defined via

$$\left\langle a_1, a_2|\rho_{A_1A_2}^{T_{A_2}}|a_1', a_2'\right\rangle = \langle a_1, a_2'|\rho_{A_1A_2}|a_1', a_2\rangle \tag{4.29}$$

for basis states $\{|a_1\rangle\}$ in A_1 and $\{|a_2\rangle\}$ in A_2 [106, 107, 159]. The partial transpose is a positive but not completely positive map, which means some of the eigenvalues of $\rho_{A_1A_2}^{T_{A_2}}$ (hereafter $\rho_{A_1A_2}^{T_2}$) can be negative. Entanglement negativity quantifies the different between the eigenvalues of the partially transposed density matrix and the original density matrix via

$$\mathcal{N}(\rho_{A_1A_2}) = \sum_i \frac{|\lambda_i| - \lambda_i}{2} = \sum_{i:\lambda_i<0} |\lambda_i|. \tag{4.30}$$

As with the von Neumann entropy, there exist Rényi generalizations of entanglement negativity:

$$\mathcal{N}_n = \mathrm{Tr}\left(\rho_{A_1A_2}^{T_2}\right)^n. \tag{4.31}$$

Due to the absolute value, one needs to define two different analytic continuations for even and odd Rényi index n, so there are in fact two Rényi negativities given by

$$\begin{aligned}\mathcal{N}_{2k}^{(\text{even})} &= \sum_i |\lambda_i|^{2k} \\ \mathcal{N}_{2k-1}^{(\text{odd})} &= \sum_i \text{sgn}\lambda_i |\lambda_i|^{2k-1}\end{aligned} \tag{4.32}$$

for integer k. We define relevant entanglement measures via analytic continuation from these quantities. The most common quantity to talk about is the logarithmic negativity, given via a $k \to 1/2$ analytic continuation of the even Rényi negativity

$$\mathcal{E}(\rho_{A_1A_2}) = \lim_{k\to 1/2} \log \mathcal{N}_{2k}^{(\text{even})}(\rho_{A_1A_2}) = \log \sum_i |\lambda_i|. \tag{4.33}$$

One other quantity of interest is the partially transposed entropy, also known as the odd entropy, which is related to the $k \to 1$ analytic continuation of the odd Rényi negativity and is explicitly given by

$$S^{T_2} \equiv \lim_{k\to 1} \frac{1}{2k-2} \log \mathcal{N}_{2k-1} = -\sum_i \lambda_i \log |\lambda_i|. \tag{4.34}$$

We need to include the Rényi entropy-like singular term out front as $\mathcal{N}_1^{(\text{odd})} = \text{Tr}\, \rho_{A_1A_2}^{T_2} = 1$.

4.3.2 Disorder Averaged Negativity

Now we can discuss the disorder average[2] in the Gaussian approximation described in the previous section. The Schmidt decomposition of the energy eigenstate $|E\rangle$ is now

$$|E\rangle = \sum_{i_1 j_1 J} M_{ijJ} \, |E_i\rangle_{A_1} \otimes |E_j\rangle_{A_2} \otimes |E_J\rangle_B \, . \tag{4.35}$$

Once again we'll consider M_{ijJ} as a Gaussian random variable, in particular with

$$M_{ijJ} \approx e^{-S(E)/2} C_{ijJ} \tag{4.36}$$

$$\overline{C_{ijJ}} = 0, \quad \overline{C_{ijJ} C_{i'j'J'}} = \delta_{ii'} \delta_{jj'} \delta_{JJ'}. \tag{4.37}$$

The partially transposed density matrix is

$$\rho_{A_1 A_2}^{T_2} = \frac{1}{\mathcal{N}} \sum_{E_i E_j E_J} C_{i_1 j_1 J} C_{i_2 j_2 J} \, |E_{i_1}, E_{j_2}\rangle \langle E_{i_2}, E_{j_1}| \, , \tag{4.38}$$

or, by replacing dummy variables

$$\rho_{A_1 A_2}^{T_2} = \frac{1}{\mathcal{N}} \sum_{E_i E_j E_J} C_{i_1 j_2 J} C_{i_2 j_1 J} \, |E_{i_1}, E_{j_1}\rangle \langle E_{i_2}, E_{j_2}| \, , \tag{4.39}$$

where the sum over energies is understood to be in a window of width 3Δ, though again we take this width to vanish. We also have

$$\left(\rho_{A_1 A_2}^{T_2}\right)^n = \frac{1}{\mathcal{N}^n} \sum_{E_{i_1}, E_{j_1}} \sum_{i_1, \cdots, i_n, j_1, \cdots, j_n, J_1, \cdots J_n} \prod_{m=1}^{n} C_{i_m j_{m+1} J_m} C_{i_{m+1} j_m J_m} \, |E_{i_1}, E_{j_1}\rangle \langle E_{i_{n+1}}, E_{j_{n+1}}| \, . \tag{4.40}$$

[2] In the condensed matter literature, disorder averaging has a different meaning and what we're doing should more properly be called "ensemble averaging". Ensemble averaging, however, already has a meaning in the high energy literature, so we keep with the terminology of [131].

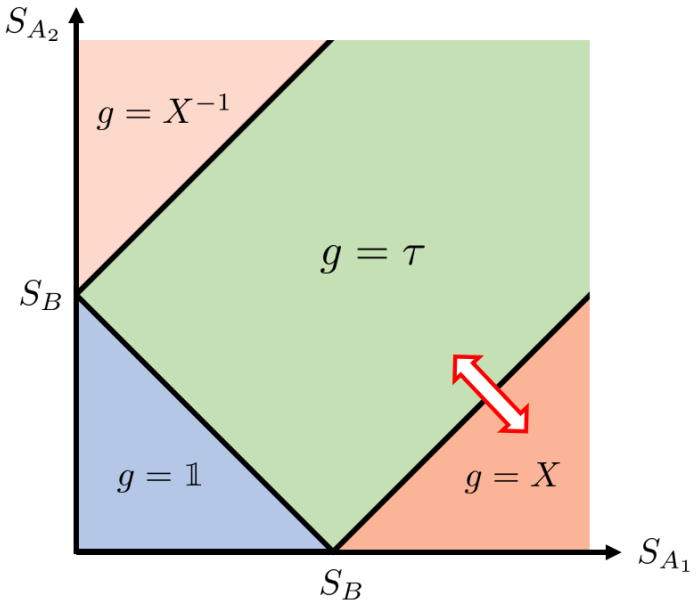


Figure 4.1: Phase diagram of Rényi negativity for various subsystem densities of state. The g's label the dominant permutation which appears in the sum over Wick contractions; their exact forms are given in Appendix F. The resolvent equation (4.41) is valid in the regime $S_{A_2} << S_{A_1} + S_B$; we've indicated the forbidden region $g = X^{-1}$ in a lighter shade. Reproduced with minor alterations from [67]

Note that the partial transpose has made it so the i (A_1) indices are contracted cyclically, while the j (A_2) indices are contracted anti-cyclically. The resolvent equation for these Wick contractions is the same as derived in [67], and a more detailed explanation is given in Appendix G. We quote the result here:

$$\lambda R(\lambda) = e^{S_{A_1}+S_{A_2}} + \frac{e^{S_B}}{e^{S_{A_2}}} \frac{R(\lambda)(1+R(\lambda))}{1-e^{2S_{A_2}}R(\lambda)^2}. \tag{4.41}$$

This resolvent equation furnishes a negativity spectrum described by the phase diagram in Figure 4.1. There are two transitions to consider. The first is when the A and B subsystems are the same size, i.e. $S_{A_1} + S_{A_2} = S_B$, corresponding to the transition from $g = \mathbb{1}$ to $g = X$ in the phase diagram. From the calculation in [67], we don't expect any enhanced corrections at this transition, so we don't study it in any detail, though the

calculation would presumably follow the same steps. The second transition of interest is when the A_1 subsystem is the same size as the combined A_2B subsystem, $S_{A_1} = S_{A_2} + S_B$, corresponding to the transition from $g = \tau$ to $g = X$ in the phase diagram. In this regime the sum over diagrams is known explicitly, and the disorder averaged partially transposed density matrices are, for $S_{A_1} < S_{A_2} + S_B$ and $S_{A_1} > S_{A_2} + S_B$ respectively,

$$\overline{\mathrm{Tr}\,(\rho_{A_1A_2}^{T_2})^{2k}} = \begin{cases} \frac{1}{\mathcal{N}^{2k}} e^{2k(S_{A_2}+S_B)+S_{A_1}} e^{S_{A_2}}\, {}_2F_1\left(1-k,-2k;2;e^{S_{A_1}-S_{A_2}-S_B}\right), \\ \frac{1}{\mathcal{N}^{2k}} e^{2kS_{A_1}+S_{A_2}+S_B} e^{S_{A_2}}\, {}_2F_1\left(1-k,-2k;2;e^{S_{A_2}+S_B-S_{A_1}}\right), \end{cases} \tag{4.42}$$

for even $n = 2k$ and

$$\overline{\mathrm{Tr}\,(\rho_{A_1A_2}^{T_2})^{2k-1}} = \begin{cases} \frac{1}{\mathcal{N}^{2k-1}} e^{(2k-1)(S_{A_2}+S_B)+S_{A_1}}\, {}_2F_1\left(1-2k,1-k;1;e^{S_{A_1}-S_{A_2}-S_B}\right), \\ \frac{1}{\mathcal{N}^{2k-1}} e^{(2k-1)S_{A_1}+S_{A_2}+S_B}\, {}_2F_1\left(1-2k,1-k;1;e^{S_{A_2}+S_B-S_{A_1}}\right), \end{cases} \tag{4.43}$$

for odd $n = 2k - 1$. We give derivations for these formulae in Appendix F; the gist is that we sum over all permutations which lie on a geodesic between two dominant regions in phase space. The permutations on this geodesic can be enumerated, and the previous formulae are functions whose moments reproduce the combinatoric factors for these permutations.

4.4 Negativity Phase Transitions

We can use (4.42) and (4.43) to understand the difference between the microcanonical and canonical Rényi negativities in a chaotic eigenstate, using much the same techniques as were used in [131]. We denote by f_{A_1} the volume fraction of A_1 such that the naïve phase transition happens at $f_{A_1} = 1/2$. We also denote the volume fraction of A_2 by f_{A_2} and use $f_A = f_{A_1} + f_{A_2}$ to denote the total volume fraction of system A.

We impose energy conservation in all three subsystems, such that our ansatz is for subsystem entropies is

$$\begin{aligned} S_{A_1}(E_{A_1}) &= f_{A_1} V s\left(\frac{E_{A_1}}{f_{A_1} V}\right) \\ S_{A_2}(E_{A_2}) &= f_{A_2} V s\left(\frac{E_{A_2}}{f_{A_2} V}\right) \\ S_B(E - E_{A_1} - E_{A_2}) &= (1 - f_A) V s\left(\frac{E - E_{A_1} - E_{A_2}}{(1 - f_A) V}\right). \end{aligned} \tag{4.44}$$

These again follow from ergodicity and imposing that the subsystem entropy is only a function of the subsystem energy density.

4.4.1 A Comment On Our Ensemble Averaging

Unlike the case of Rényi entropy, the disorder average over the partially transposed density matrix is not equivalent to upgrading the resummed traces (4.42) and (4.43) using the ansatz (B.7). The reason for this is that energy conservation between replicated subsystems for negativity forces "nonadjacent" subsystems to interact. The rules for energy conservation are outlined in [134]. For a given permutation, one can "follow the lines" of the three subsystems to see how the replicas interact. For an untransposed system, the constraints conspire such that a resolvent sum over diagrams holds away from the infinite temperature limit.

This is no longer the case upon transposing. For simple permutations like $g = \mathbb{1}$ and $g = X$, imposing the constraints returns integrals similar to those in the Rényi entropy case. Things get more involved for general permutations on $G(\tau, X)$; the simplest permutations are a subset of our τ permutations, labelled τ_{ES} in [134]:

$$\tau_{ES} = (12)(34)\cdots(n-1\ n) \text{ or } (23)(45)\cdots(n\ 1) \tag{4.45}$$

These permutations require an integral over four subsystem energies, and the number of integrals only grows for arbitrary permutations on $G(\tau, X)$. This implies our ansatz (B.7) doesn't describe energy conservation described by disorder averaging over (4.40).

Given this obstruction, we make the choice to connect the replicas in a manner that reduces to the smallest number of integrals over subsystem energies. This has the advantage of allowing us to write a resolvent equation and resum the relevant permutations, while not having a correct interpretation as a trace-preserving energy flow between replicas. This produces some obvious pathologies, such as disagreement between the second even Rényi negativity and the second Rényi entropy, which must agree for a well-defined density matrix. However, as we'll see, even in this model one can reproduce some features of negativity seen in the gravitational model, namely large deviations from the thermodynamic answer for entanglement negativity and an enhanced correction for the partially transposed entropy.

4.4.2 Even Rényi Negativity

We'll start with studying the even Rényi negativities, from which the logarithmic negativity descends. Our expressions for the logarithms of the canonical ensemble and microcanonical ensemble Rényi negativities using our previous ansatzes are as follows:

$$\begin{aligned}\overline{\mathcal{N}_{2k}} &= \frac{1}{\overline{\mathcal{N}^{2k}}}\int dE_{A_1}dE_{A_2}e^{S_{A_1}(E_{A_1})+2S_{A_2}(E_{A_2})+S_B(E-E_{A_1}-E_{A_2})}G_k(f_{A_1},f_{A_2},E_{A_1},E_{A_2})\\ \mathcal{N}_{2k}^{MC} &= \frac{1}{\overline{\mathcal{N}^{2k}}}\int dE_{A_1}dE_{A_2}e^{S_{A_1}(E_{A_1})+S_{A_2}(E_{A_2})+2k(S_{A_2}(E_{A_2})+S_B(E-E_{A_1}-E_{A_2}))},\end{aligned} \tag{4.46}$$

where the function $G_k(f_{A_1}, f_{A_2}, E_{A_1}, E_{A_2})$ is defined as the k-dependent part of (4.42)

$$G_k(f_{A_1}, f_{A_2}, E_{A_1}, E_{A_2}) = \begin{cases} e^{(2k-1)(S_{A_2}+S_B)}{}_2F_1\left(1-k, -2k; 2; e^{S_{A_1}-S_{A_2}-S_B}\right), \\ e^{(2k-1)S_{A_1}}{}_2F_1\left(1-k, -2k; 2; e^{S_{A_2}+S_B-S_{A_1}}\right), \end{cases} \tag{4.47}$$

and $\mathcal{N}$ (with no other sub/superscripts) is an overall normalization given by

$$\mathcal{N} = \int dE_{A_1} dE_{A_2} e^{S_{A_1}(E_{A_1})+S_{A_2}(E_{A_2})+S_B(E-E_{A_1}-E_{A_2})}. \tag{4.48}$$

Whenever unspecified, the subsystem entropies should now be understood to be valued at the subsystem energies, which we only omit for notational clarity. We write the difference between the logarithms of these quantities as

$$\log \overline{\mathcal{N}_{2k}} - \log \mathcal{N}_{2k}^{MC} \equiv \log\left(\frac{\int dE_{A_1} dE_{A_2} \exp(F_1(E_{A_1}, E_{A_2}))}{\int dE_{A_1} dE_{A_2} \exp(F_2(E_{A_1}, E_{A_2}))}\right), \tag{4.49}$$

where the functions $F_1(E_{A_1}, E_{A_2})$ and $F_2(E_{A_1}, E_{A_2})$ are defined via the corresponding integrands in (4.46). The strategy will be to find saddle points for $F_1(E_{A_1}, E_{A_2})$ and $F_2(E_{A_1}, E_{A_2})$ and use the relative behavior of those saddle points to determine the scaling of the correction at transition.

We have two coupled saddle point equations for each both functions, which are given

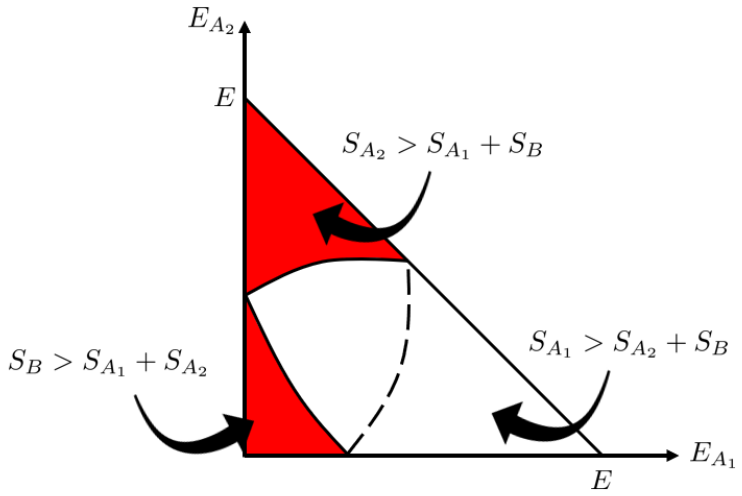


Figure 4.2: A schematic plot of regions (shaded in red) in the $E_{A_1} - E_{A_2}$ plane where our ansatz for the dominant sum over permutations does not hold. The lines separating the regions will depend sensitively on the form of $s(e)$ and the volume fractions of the subsystems.

by

$$\begin{aligned}
s'\left(\frac{E_1^{(1)}}{f_{A_1}V}\right) &= s'\left(\frac{E - E_1^{(1)} - E_1^{(2)}}{(1-f_A)V}\right) - \frac{\partial_{E_{A_1}} G_k(f_{A_1}, f_{A_2}, E_1^{(1)}, E_1^{(2)})}{G_k(f_{A_1}, f_{A_2}, E_1^{(1)}, E_1^{(2)})} \\
2s'\left(\frac{E_1^{(2)}}{f_{A_2}V}\right) &= s'\left(\frac{E - E_1^{(1)} - E_1^{(2)}}{(1-f_A)V}\right) - \frac{\partial_{E_{A_2}} G_k(f_{A_1}, f_{A_2}, E_1^{(1)}, E_1^{(2)})}{G_k(f_{A_1}, f_{A_2}, E_1^{(1)}, E_1^{(2)})} \\
s'\left(\frac{E_2^{(1)}}{f_{A_1}V}\right) &= 2ks'\left(\frac{E - E_2^{(1)} - E_2^{(2)}}{(1-f_A)V}\right) \\
(2k+1)s'\left(\frac{E_2^{(2)}}{f_{A_2}V}\right) &= 2ks'\left(\frac{E - E_2^{(1)} - E_2^{(2)}}{(1-f_A)V}\right),
\end{aligned} \tag{4.50}$$

where the pair $\mathcal{E}_1 = (E_1^{(1)}, E_1^{(2)})$ denotes a saddle point for $F_1(E_{A_1}, E_{A_2})$, while $\mathcal{E}_2 = (E_2^{(1)}, E_2^{(2)})$ denotes the saddle point for $F_2(E_{A_1}, E_{A_2})$. As $F_2(E_{A_1}, E_{A_2})$ is a strictly concave function, there is only one global maximum. $F_1(E_{A_1}, E_{A_2})$ on the other hand can have two maxima, as $G_k(f_{A_1}, f_{A_2}, E_{A_1}, E_{A_2})$ is strictly nonmonotonic.

The first thing we have to be careful about is whether we are still within our regime

of validity for probing the transition of interest. In the case of Rényi entropy, the fact that the dominant contribution comes from noncrossing partitions was assumed to hold for all of parameter space, that is for all values of subsystem entropy. This can be traced back to the fact that the dominant permutations all lie on a single geodesic $G(\mathbb{1}, X)$. In our case, we're trying to probe the transition on one geodesic $G(\tau, X)$ while suppressing diagrams from other geodesics, which imposes some natural constraints on the size of our subsystems.

We are justified in only considering the diagrams from Appendix F only if the saddle point energies satisfy the conditions:

$$\begin{aligned} S_{A_2}(E_{1,2}^{(2)}) &< S_{A_1}(E_{1,2}^{(1)}) + S_B(E - E_{1,2}^{(1)} - E_{1,2}^{(2)}) \\ S_B(E - E_{1,2}^{(1)} - E_{1,2}^{(2)}) &< S_{A_1}(E_{1,2}^{(1)}) + S_{A_2}(E_{1,2}^{(2)}), \end{aligned} \tag{4.51}$$

such that all contributions from subleading permutations remain subleading. We include a rough phase diagram of the allowed region to explore in Figure 4.2. If the saddle point lies outside the allowed region, our answer for the dominant sum over permutations no longer holds, so we shouldn't try to explore those regions of phase space.

This means before attempting to compute corrections at transition for all subsystem volume fractions, we should derive some bounds on the regime of validity of our approximation. We'll make use of the following inequality:

$$S_{A_1}(E_{A_1}) + S_{A_2}(E_{A_2}) + S_B(E - E_{A_1} - E_{A_2}) \leq Vs\left(\frac{E}{V}\right), \tag{4.52}$$

which follows from the fact that our subsystem entropy function $s(e)$ is concave. Plugging

in the saddle points and using the first constraint in (4.51) we can write

$$2S_{A_2}(E_2^{(2)}) < S_{A_1}(E_2^{(1)}) + S_{A_2}(E_2^{(2)}) + S_B(E - E_2^{(1)} - E_2^{(2)}) < Vs\left(\frac{E}{V}\right)$$
$$\Rightarrow S_{A_2}(E_2^{(2)}) < \frac{V}{2}s\left(\frac{E}{V}\right). \tag{4.53}$$

We can use this relation to find

$$S_{A_2}(E_2^{(2)}) = f_{A_2}s\left(\frac{E_2^{(2)}}{f_{A_2}V}\right) > f_{A_2}s\left(\frac{E_2^{(2)}}{V}\right)$$
$$\Rightarrow f_{A_2}s\left(\frac{E_2^{(2)}}{V}\right) < \frac{V}{2}s\left(\frac{E}{V}\right), \tag{4.54}$$

as $E_2^{(2)} < E$, a constraint on f_{A_2} which makes this true for all subsystem entropy densities is

$$f_{A_2} < 1/2. \tag{4.55}$$

Therefore our calculations are only valid when subsystem A_2 is less than half of the total system size. We can find a similar inequality on S_B using the second constraint in (4.51). We have

$$2S_B(E - E_2^{(1)} - E_2^{(2)}) < S_{A_1}(E_2^{(1)}) + S_{A_2}(E_2^{(2)}) + S_B(E - E_2^{(1)} - E_2^{(2)}) \leq Vs\left(\frac{E}{V}\right)$$
$$\Rightarrow S_B(E - E_2^{(1)} - E_2^{(2)}) < \frac{V}{2}s\left(\frac{E}{V}\right). \tag{4.56}$$

We can therefore write

$$S_B(E - E_2^{(1)} - E_2^{(2)}) = (1 - f_A)Vs\left(\frac{E - E_2^{(1)} - E_2^{(2)}}{(1 - f_A)V}\right) > (1 - f_A)Vs\left(\frac{E - E_2^{(1)} - E_2^{(2)}}{V}\right)$$
$$\Rightarrow (1 - f_A)Vs\left(\frac{E - E_2^{(1)} - E_2^{(2)}}{V}\right) < \frac{V}{2}s\left(\frac{E}{V}\right). \tag{4.57}$$

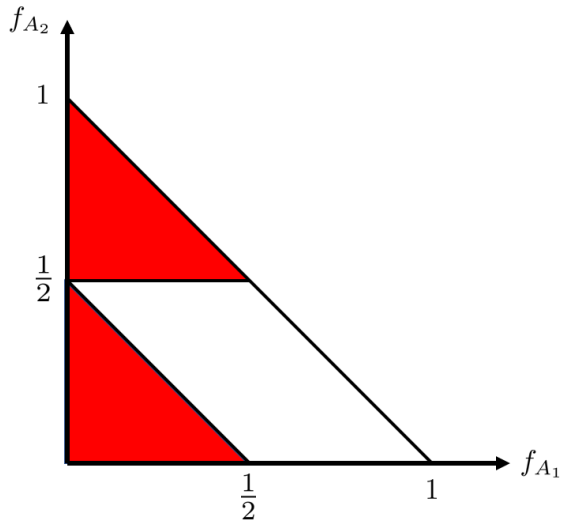


Figure 4.3: Excluded volume fractions from our analysis of the cyclic to pairwise phase transition. Describing the colored "forbidden" regions would require a sum over permutations we assert to be subdominant.

Again, a result that makes this inequality true for all saddle point energies is

$$f_A > 1/2. \tag{4.58}$$

This ties together a nice family of restrictions: both subsystems A_2 and B have to have volume fraction less then half of the system. We illustrate these constraints in Figure 4.3. This makes some sense, as we want to probe transitions dominated by the behavior of A_1 relative to the rest of the system.

Another way of seeing there should be a restricted regime for our procedure is as follows: entanglement negativity is agnostic as to which subsystem A_1 or A_2 one applies the partial transpose to. This would of course result in an averaged density matrix trace symmetric under exchange of S_{A_1} and S_{A_2}, which our expressions (4.42) and (4.43) are not. However, by writing a resolvent equation valid only in a certain parameter regime, we can no longer comfortably integrate over all energies. This is an important point because the deviations from the featureless case can in principle be of order the system

size, and so corrections are not necessarily perturbative as they were assumed to be in [67].

We can, however, be comfortable in the validity of our calculation if the saddle points for $F_1(E_{A_1}, E_{A_2})$ and $F_2(E_{A_1}, E_{A_2})$ obey the conditions above, so restricting to the set of entropy functions which satisfy (4.51), let's first look at the saddle point equations for $\mathcal{E}_2$. Setting the third and fourth equations equal yields

$$s'\left(\frac{E_2^{(1)}}{f_{A_1}V}\right) = (2k+1)s'\left(\frac{E_2^{(2)}}{f_{A_2}V}\right). \tag{4.59}$$

As $s'(e)$ is a monotonically decreasing function, for all $k > 0$ we have the inequality

$$E_2^{(2)} > \frac{f_{A_2}}{f_{A_1}} E_2^{(1)}. \tag{4.60}$$

We can use this inequality to write a simple inequality on $E_2^{(1)}$ by rewriting the $E_2^{(1)}$ saddle point equation as

$$s'\left(\frac{E_2^{(1)}}{f_{A_1}V}\right) > 2ks'\left(\frac{E - \frac{f_A}{f_{A_1}}E_2^{(1)}}{(1-f_A)V}\right). \tag{4.61}$$

Now we have an inequality which depends on k, as we can write

$$E_2^{(1)} < f_{A_1}E, \quad k \geq 1/2. \tag{4.62}$$

Note that this result is also valid for $k = 1/2$, as the relation (4.60) is a strict inequality which is never saturated for positive k. We can use a similar strategy to write an

inequality for $E_2^{(2)}$. Rewriting the $E_2^{(2)}$ equation with (4.60) yields

$$(2k+1)s'\left(\frac{E_2^{(2)}}{f_{A_2}V}\right) < 2ks'\left(\frac{E-\frac{f_A}{f_{A_2}}E_2^{(2)}}{1-f_A}\right). \tag{4.63}$$

The resulting inequality has a slightly different k dependence:

$$E_2^{(2)} > f_{A_2}E, \quad k > 0. \tag{4.64}$$

The last inequalities we can write are those for the saddle point values of S_{A_1} and S_{A_2}:

$$\begin{aligned} S_{A_1}(E_2^{(1)}) &= f_{A_1}Vs\left(\frac{E_2^{(1)}}{f_{A_1}V}\right) < f_{A_1}Vs\left(\frac{E}{V}\right) \\ S_{A_2}(E_2^{(2)}) &= f_{A_2}Vs\left(\frac{E_2^{(2)}}{f_{A_2}V}\right) > f_{A_2}Vs\left(\frac{E}{V}\right). \end{aligned} \tag{4.65}$$

We'd like to find conditions on the hypergeometric being stuck on the first branch, i.e. $S_{A_1} < S_{A_2} + S_B$. This is guaranteed to happen if the weaker inequality $S_{A_1} < S_{A_2}$ is satisfied, which from (4.65) is necessarily true when

$$f_{A_2} > f_{A_1}. \tag{4.66}$$

If we assume $S_{A_1} < S_{A_2}$ for the $\mathcal{E}_1$ saddle point as well, the argument of the hypergeometric is exponentially suppressed and we can approximate it by

$${}_2F_1(1-k,-2k;2;x) \approx 1 + k(k-1)x, \tag{4.67}$$

where the small parameter x is now

$$x \equiv e^{S_{A_1}(E_1^{(1)}) - S_{A_2}(E_1^{(2)}) - S_B(E - E_1^{(1)} - E_1^{(2)})}. \tag{4.68}$$

Under this assumption the saddle point equations for $\mathcal{E}_1$ and $\mathcal{E}_2$ are the same up to exponentially suppressed terms, and therefore the saddle points $\mathcal{E}_1$ and $\mathcal{E}_2$ are exponentially close. This leads to the following form of corrections to ETH:

$$\log \overline{\mathcal{N}_{2k}} - \log \mathcal{N}_{2k}^{MC} \propto \mathcal{O}(e^{-cV}), \quad k \geq 1/2, f_{A_2} > f_{A_1} \tag{4.69}$$

We can write a similar inequality for which $S_{A_1} < S_B$ is always satisfied. We recall the $E_2^{(1)}$ saddle point equation:

$$s'\left(\frac{E_2^{(1)}}{f_{A_1}}\right) = 2ks'\left(\frac{E - E_2^{(1)} - E_2^{(2)}}{(1 - f_A)V}\right). \tag{4.70}$$

At $k = 1/2$ there's clearly an equality between the arguments of the functions on the right and left, so for $k \geq 1/2$ we have the inequality

$$\frac{E_2^{(1)}}{fV} \leq \frac{E - E_2^{(1)} - E_2^{(2)}}{(1 - f_A)V}, \quad k \geq 1/2. \tag{4.71}$$

We'd like to satisfy the inequality $S_{A_1} < S_B$, or

$$f_{A_1} V s\left(\frac{E_2^{(1)}}{fV}\right) < (1 - f_A)Vs\left(\frac{E - E_2^{(1)} - E_2^{(2)}}{(1 - f_A)V}\right). \tag{4.72}$$

This is always satisfied if

$$f_{A_1} < 1 - f_A. \tag{4.73}$$

So far we have two constraints which carve out a corner of the phase space for all $k \geq 1/2$.

Now let's try to find a condition such that $S_{A_1} > S_{A_2} + S_B$. Using our previous ansatz this condition is written as

$$f_{A_1} s\left(\frac{E_2^{(1)}}{f_{A_1} V}\right) > f_{A_2} s\left(\frac{E_2^{(2)}}{f_{A_2} V}\right) + (1-f_A) s\left(\frac{E - E_2^{(1)} - E_2^{(2)}}{(1-f_A)V}\right). \tag{4.74}$$

For all $k > 0$ we can use (4.60) to rewrite this as

$$(f_{A_1} - f_{A_2}) s\left(\frac{E_2^{(2)}}{f_{A_2} V}\right) > (1-f_A) s\left(\frac{E - E_2^{(1)} - E_2^{(2)}}{(1-f_A)V}\right). \tag{4.75}$$

Using the $E_2^{(2)}$ saddle point equation, there exists for $k > 0$:

$$\frac{E_2^{(2)}}{f_{A_2}} > \frac{E - E_2^{(1)} - E_2^{(2)}}{1 - f_A}. \tag{4.76}$$

Therefore, $S_{A_1} > S_{A_2} + S_B$ is always satisfied if

$$f_{A_1} - f_{A_2} > 1 - f_A \Rightarrow f_{A_1} > 1/2, \quad k > 0. \tag{4.77}$$

For these volume fractions the corrections to the Rényi negativity are extensive in the system size, as $\mathcal{E}_1$ and $\mathcal{E}_2$ have no relation:

$$\log \overline{\mathcal{N}_{2k}} - \log \mathcal{N}_{2k}^{MC} \propto \mathcal{O}(V), \quad k > 0, f_{A_1} > 1/2. \tag{4.78}$$

We summarize the results so far in Figure 4.4. In that phase diagram, none of the boundaries should be though of as sharp, that is as (4.60) is never saturated for $k > 0$, neither are any constraints that depend on it. The interpolation between $\mathcal{O}(e^{-cV})$ corrections and $\mathcal{O}(V)$ corrections will happen somewhere in this "unknown region", though the only relevant point is that at $f_{A_1} = 1/2$ we should still be in a region with extensive correc-

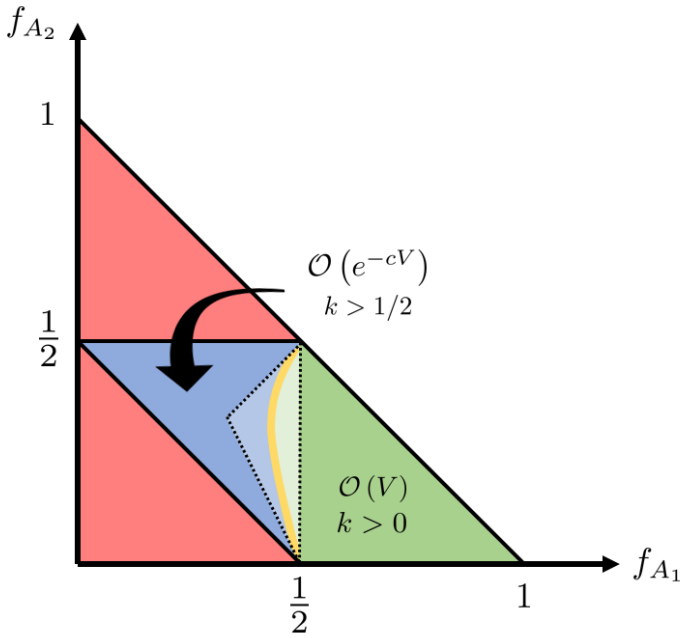


Figure 4.4: Phase diagram for corrections to even Rényi negativities. The concave region with $\mathcal{O}(e^{-cV})$ corrections comes from requiring $S_{A_1} < S_{A_2}$ and/or $S_{A_1} < S_B$. The $\mathcal{O}(V)$ region requires $S_{A_1} > S_{A_2} + S_B$. The interpolation between these regions will lie somewhere with $f_{A_1} < 1/2$ and is outlined by the dashed lines. The yellow curve represents a system specific boundary which will depend on k and potentially on the specifics of $s(e)$.

tions. In particular this implies the logarithmic negativity receives $\mathcal{O}(V)$ corrections, as was noted in [67].[3]

We won't comment on the case $k < 1/2$ for $f_{A_1} < 1/2$, though the expectation is that, like the $n < 1$ Rényi entropy, these measures always receive volume law corrections. It's also entirely possible the interpolating line continues moving towards the point $(0, 1/2)$, meaning there's some set of volume fractions for which arbitrarily small but positive k are well-approximated by ETH.

4.4.3 Odd Rényi Negativity

We can repeat the previous analysis for odd n. We have different expressions for the canonical and microcanonical Rényi negativities:

$$\begin{aligned}\log\overline{\mathcal{N}_{2k-1}} &= \frac{1}{\mathcal{N}_{2k-1}}\int dE_{A_1}dE_{A_2}e^{S_{A_1}(E_{A_1})+S_{A_2}(E_{A_2})+S_B(E-E_{A_1}-E_{A_2})}G_k(f_{A_1},f_{A_2},E_{A_1},E_{A_2})\\ \log\mathcal{N}^{MC}_{2k-1} &= \frac{1}{\mathcal{N}_{2k-1}}\int dE_{A_1}dE_{A_2}e^{S_{A_1}(E_{A_1})+(2k-1)(S_{A_2}(E_{A_2})+S_B(E-E_{A_1}-E_{A_2}))},\end{aligned} \tag{4.79}$$

where $G_k(f_{A_1}, f_{A_2}, E_{A_1}, E_{A_2})$ is now defined by (4.43) as:

$$G_k(f_{A_1},f_{A_2},E_{A_1},E_{A_2}) = \begin{cases} e^{(2k-2)(S_{A_2}+S_B)}{}_2F_1\left(1-k,1-2k;1;e^{S_{A_1}-S_{A_2}-S_B}\right), \\ e^{(2k-2)S_{A_1}}{}_2F_1\left(1-2k,1-k;1;e^{S_{A_2}+S_B-S_{A_1}}\right). \end{cases} \tag{4.80}$$

Again the subsystem entropies should be valued at their respective subsystem energies. Notably $\log\overline{\mathcal{N}_{2k-1}}$ enjoys a symmetry under $S_{A_1} \leftrightarrow S_{A_2} + S_B$. We again write the

[3] At $k = 1$, the even Rényi negativity is equal to the second Rényi entropy $S_2(\rho_A)$, which for $f_{A_1} + f_{A_2} > 1/2$ is expected to always receive volume law corrections, which we don't see for all volume fractions. This could be a consequence of the restriction to a particular phase transition, but more likely is a result of the issues discussed in our ansatz above.

difference between the canonical and microcanonical answers as

$$\log\overline{\mathcal{N}_{2k-1}} - \log\mathcal{N}^{MC}_{2k-1} = \log\left(\frac{\int dE_{A_1}dE_{A_2}\exp(F_1(E_{A_1},E_{A_2}))}{\int dE_{A_1}dE_{A_2}\exp(F_2(E_{A_1},E_{A_2}))}\right) \tag{4.81}$$

and use the same ansatz (B.7) to write the saddle point equations for F_1 and F_2 as

$$\begin{aligned}
s'\left(\frac{E_1^{(1)}}{f_{A_1}V}\right) &= s'\left(\frac{E-E_1^{(1)}-E_1^{(2)}}{(1-f_A)V}\right) - \frac{\partial_{E_{A_1}}G_k(f_{A_1},f_{A_2},E_1^{(1)},E_1^{(2)})}{G_k(f_{A_1},f_{A_2},E_1^{(1)},E_1^{(2)})}\\
s'\left(\frac{E_1^{(2)}}{f_{A_2}V}\right) &= s'\left(\frac{E-E_1^{(1)}-E_1^{(2)}}{(1-f_A)V}\right) - \frac{\partial_{E_{A_2}}G_k(f_{A_1},f_{A_2},E_1^{(1)},E_1^{(2)})}{G_k(f_{A_1},f_{A_2},E_1^{(1)},E_1^{(2)})}\\
s'\left(\frac{E_2^{(1)}}{f_{A_1}V}\right) &= (2k-1)s'\left(\frac{E-E_2^{(1)}-E_2^{(2)}}{(1-f_A)V}\right)\\
s'\left(\frac{E_2^{(2)}}{f_{A_2}V}\right) &= s'\left(\frac{E-E_2^{(1)}-E_2^{(2)}}{(1-f_A)V}\right).
\end{aligned} \tag{4.82}$$

Let's again investigate the saddle point for F_2. We immediately see

$$\frac{E_2^{(2)}}{f_{A_2}} = \frac{E-E_2^{(1)}-E_2^{(2)}}{1-f_A} \tag{4.83}$$

for all k! This is a striking result, as it means we can write the sum of subsystem entropies in A_2 and B as

$$S_{A_2}(E_2^{(2)}) + S_B(E-E_2^{(1)}-E_2^{(2)}) \equiv S_{\overline{A_1}}(E_2^{(2)}) = (1-f_{A_1})s\left(\frac{E_2^{(2)}}{f_{A_2}V}\right) \tag{4.84}$$

This is important as for the odd Rényi negativity, S_{A_2} and S_B always appear summed, so if we're only interested in the leading saddle point approximation we can treat them as one subsystem entropy $S_{\overline{A_1}}$. As such we can rewrite the single saddle point equation

as

$$s'\left(\frac{E_2^{(1)}}{f_{A_1}V}\right) = (2k-1)s'\left(\frac{E_2^{(2)}}{f_{A_2}V}\right). \tag{4.85}$$

We recognize this as similar to the saddle point equation (4.24) for $F_2(E)$ in the Rényi entropy, but we'll go through the discussion nonetheless. At $k = 1$ we can exactly solve for the subsystem energies and they are, unsurprisingly, proportional to the volume fractions of their respective subsystems:

$$\begin{aligned} E_2^{(1)} &= f_{A_1}E, \quad k = 1 \\ E_2^{(2)} &= f_{A_2}E, \quad k = 1. \end{aligned} \tag{4.86}$$

When $k > 1$, we again have

$$E_2^{(2)} > \frac{f_{A_2}}{f_{A_1}}E_2^{(1)}, \tag{4.87}$$

which was true for general k in the even case. Similar inequalities on volume fraction hold in the odd case; we still have

$$\begin{aligned} E_2^{(1)} &< f_{A_1}E, \quad k > 1 \\ E_2^{(2)} &> f_{A_2}E, \quad k > 1. \end{aligned} \tag{4.88}$$

From this the inequality $S_{A_1} < S_{\overline{A_1}}$ is clearly satisfied when

$$f_{A_1} < 1/2, \tag{4.89}$$

and corrections are exponentially suppressed. For $f_{A_1} > 1/2$, this won't be true generically and the corrections are extensive.

We can also say interesting things about $k < 1$. In this case the inequalities are

flipped:

$$\begin{aligned} E_2^{(1)} &> f_{A_1} E, \quad k < 1 \\ E_2^{(2)} &< f_{A_2} E, \quad k < 1 \\ E_2^{(2)} &< \frac{f_{A_2}}{f_{A_1}} E_2^{(1)}. \end{aligned} \tag{4.90}$$

We can check where $S_{A_1} > S_{\overline{A_1}}$. From (4.90) we have

$$\begin{aligned} S_{A_1}(E_2^{(1)}) &= f_{A_1} V s\left(\frac{E_2^{(1)}}{f_{A_1} V}\right) > f_{A_1} V s\left(\frac{E}{V}\right) \\ S_{\overline{A_1}}(E_2^{(2)}) &= (1-f_{A_1}) V s\left(\frac{E_2^{(2)}}{f_{A_2} V}\right) < (1-f_{A_1}) V s\left(\frac{E}{V}\right). \end{aligned} \tag{4.91}$$

We see that $S_{A_1} > S_{\overline{A_1}}$ is guaranteed to be satisfied if $f_{A_1} > 1/2$, and indeed there is no generic behavior for $f_{A_1} < 1/2$. Thus the corrections are extensive for all volume fractions for $k < 1$.

4.4.4 Odd Rényi Negativity at Transition

We would like to study this case in analogy with the entanglement entropy, for reasons that will be clear shortly. Let's follow the same procedure of dividing F_1 into two pieces, F_{dom} and F_Δ, defined as

$$\begin{aligned} F_{\text{dom}} &= S_{A_1}(E_{A_1}) + S_{A_2}(E_{A_2}) + S_B(E - E_{A_1} - E_{A_2}) \\ &\quad + (2k-2)\max\{S_{A_1}(E_{A_1}), S_{A_2}(E_{A_2}) + S_B(E - E_{A_1} - E_{A_2})\} \\ F_\Delta &= \log {}_2F_1\left(1-2k, 1-k; 1; e^{-|S_{A_1}(E_{A_1}) - S_{A_2}(E_{A_2}) - S_B(E-E_{A_1}-E_{A_2})|}\right). \end{aligned} \tag{4.92}$$

That is, we take the dominant contribution and relegate the subleading contributions to a term bounded by $\mathcal{O}(1)$ in volume factors:

$$1 \leq e^{F_\Delta} \leq a_k, \quad a_k \equiv \binom{3k-2}{k-1} = \frac{\Gamma(3k-1)}{\Gamma(k)\Gamma(2k)} = 1 + (k-1) + \mathcal{O}(k-1)^2. \tag{4.93}$$

The averaged Rényi negativity, with a $\frac{1}{2k-2}$ factor which will be important later, can be rewritten as

$$\frac{1}{2k-2} \log \overline{\mathcal{N}_{2k-1}} = \frac{1}{2k-2} \log \left(\frac{1}{\mathcal{N}_{2k-1}} \int dE_{A_1} dE_{A_2} e^{F_{\text{dom}} + F_\Delta} \right), \tag{4.94}$$

and we can bound $\log \overline{\mathcal{N}_{2k-1}}$ via

$$\log \overline{\mathcal{N}_{2k-1}} - \log \mathcal{N}_{2k-1}^{\text{dom}} \leq \frac{1}{2} + \mathcal{O}(k-1). \tag{4.95}$$

As such $\mathcal{N}_{2k-1}^{\text{dom}}$ is enough to look for corrections larger than $\mathcal{O}(1)$.

Unlike the Rényi entropy, at $f = 1/2$ there's no obvious reflection symmetry of the energies in F_{dom}, and indeed we don't find one numerically. There is, however, a symmetry in the saddle points, which we'll argue for as follows. Call the two saddle points for F_{dom} (or F_1, it makes no difference here) $\mathcal{E}_1^{(a)} = (E_1^{(1,a)}, E_1^{(2,a)})$ and $\mathcal{E}_1^{(b)} = (E_1^{(1,b)}, E_1^{(2,b)})$. Under the exchange $S_{A_1} \leftrightarrow S_{\overline{A}_1}$, the saddles are swapped due to the symmetry of the odd Rényi negativity. It's clear then at $f_{A_1} = 1/2$ there exists the equivalence

$$\begin{aligned} \frac{E_1^{(1,a)}}{f_{A_1}} &= \frac{E_1^{(2,b)}}{f_{A_2}} \\ \frac{E_1^{(1,b)}}{f_{A_2}} &= \frac{E_1^{(2,a)}}{f_{A_2}}. \end{aligned} \tag{4.96}$$

This means that the two saddle points contribute with equal magnitude, which contributes an $\mathcal{O}(1)$ factor to the difference between the canonical and microcanonical neg-

ativities:

$$\frac{1}{2k-2}\left(\log \mathcal{N}_{2k-1}^{\text{dom}} - \log \mathcal{N}_{2k-1}^{MC}\right) = \frac{\log 2}{2-2k} \sim \mathcal{O}(1) \tag{4.97}$$

However, as in the case of von Neumann entropy, there is a subtletly related to the fact that the two saddles collide in the limit $k \to 1$, i.e. the partially transposed entropy. As they collide, there is an emergent region between the saddles which contributes to the integral, so we can't treat the presence of multiple equivalent saddles at leading order, we must integrate over the interpolating region. We show a plot of this phenomenon in Figure 4.5. Let's solve the F_2 saddle point equations perturbatively in $\delta \equiv 2k-2$. The $E_2^{(1)}$ saddle point equation (4.85) becomes

$$s'\left(\frac{E_2^{(1)}}{f_{A_1}V}\right) = (1+\delta)s'\left(\frac{E_2^{(2)}}{f_{A_2}V}\right) \approx s'\left(\frac{E_2^{(2)}}{f_{A_2}V} + \delta\frac{s'\left(E/V\right)}{s''\left(E/V\right)}\right), \tag{4.98}$$

where we've again used that $E_2^{(2)} = f_{A_2}E$. Combining this with the unchanged (4.83) and plugging in $f_{A_1} = 1/2$ yields

$$\begin{aligned} E_2^{(1)} &= \frac{E}{2} + \frac{V\delta}{4}\frac{s'(E/V)}{s''(E/V)} \\ E_2^{(2)} &= f_{A_2}E - \frac{f_{A_2}V\delta}{2}\frac{s'(E/V)}{s''(E/V)} \end{aligned} \tag{4.99}$$

From this we can write our subsystem entropies S_{A_1} and $S_{\overline{A_1}}$ in the familiar form

$$\begin{aligned} S_{A_1}(E_2^{(1)}) &= \frac{1}{2}s\left(E + \frac{V\delta}{2}\frac{s'(E/V)}{s''(E/V)}\right) \\ S_{\overline{A_1}}(E_2^{(2)}) &= \frac{1}{2}s\left(E - \frac{V\delta}{2}\frac{s'(E/V)}{s''(E/V)}\right) \end{aligned} \tag{4.100}$$

What happens as $k \to 1$ for the odd Rényi negativity is precisely the same as what happens for the $n \to 1$ von Neumann entropy, namely that the F_Δ term "fills in" the

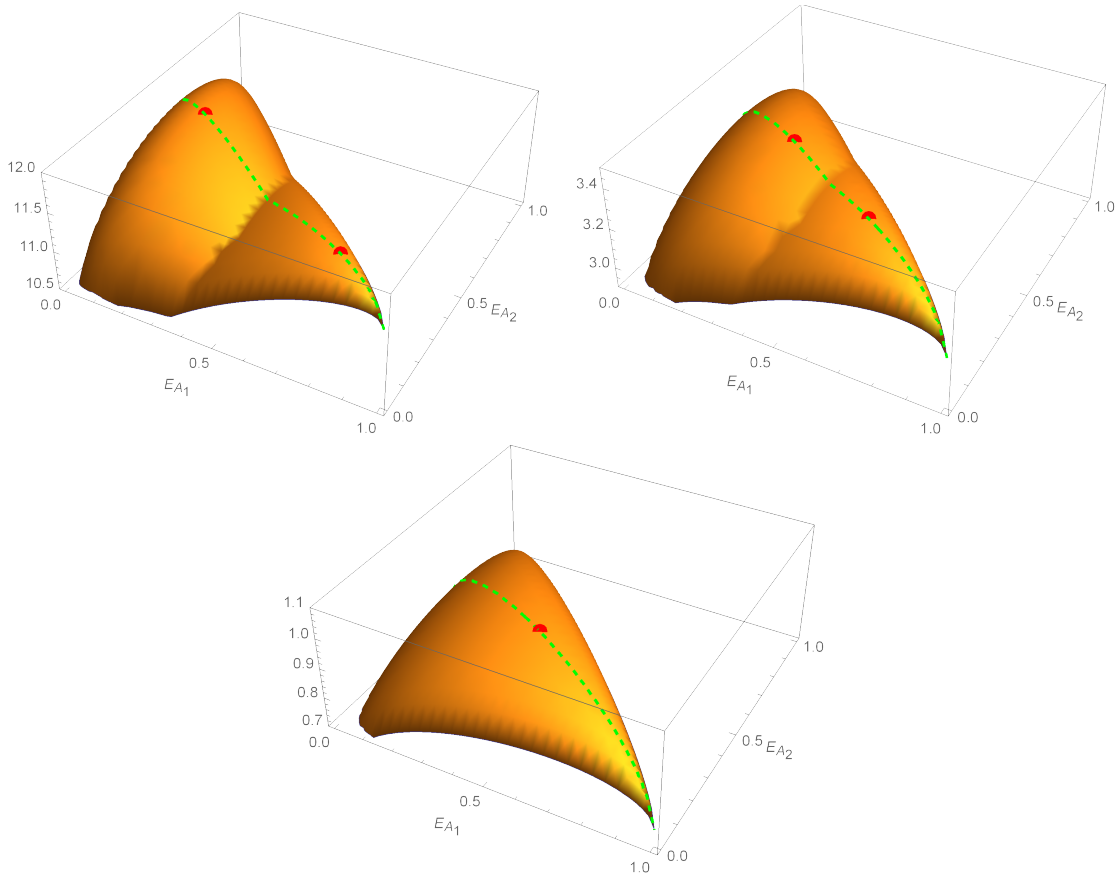


Figure 4.5: Plots of $F_1(E_{A_1}, E_{A_2})$ at phase transition. We've set $E = V = 1, f_{A_1} = 1/2$, and $f_{A_2} = 3/10$. For large k (upper left), the two saddle points are well-separated and can be treated separately. As we decrease k (upper right) the saddle points approach one another and produce an emergent flat region. At exactly $k = 1$ (bottom) the saddle points coincide at $(f_{A_1}E, f_{A_2}E)$. The dotted line connecting the saddle points is given by $E_{A_2} = -2f_{A_2}(E_{A_1} - E)$; all saddles at $f_{A_1} = 1/2$ lie along this line.

space between the two saddles. The only difference is that this flat direction runs between two saddles separated along a line in the $E_{A_1} - E_{A_2}$ plane specified by f_{A_2}. The rest of the calculation is completely unchanged from that of the von Neumann entropy, and there is an enhanced correction exactly of the same form:

$$\overline{S^{T_2}} - S_{MC}^{T_2} = -\sqrt{\frac{C_V}{2\pi}} + \mathcal{O}(\delta) \sim \mathcal{O}(\sqrt{V}) \tag{4.101}$$

In [67], it was noted that a naïve calculation shows the partially transposed entropy receives $\mathcal{O}(\sqrt{V})$ corrections, but a more accurate analysis shows it receives $\mathcal{O}(V)$ corrections. It would be interesting to understand the difference between our calculation and theirs.[4]

4.5 Discussion and Future Work

In this work we've studied a class of tripartite entanglement measures, the Rényi negativities, in a toy model of a chaotic eigenstate. We've resummed the relevant noncrossing permutations obtained via Wick contractions relevant at the transition of interest and studied the corrections to the dominant microcanonical saddle.

The main takeaway is as follows: logarithmic negativity and its Rényi generalizations thereof are not always "good" chaotic observables in the sense that their fluctuations (the difference between the canonical and microcanincal expectation values) are often of the same order as the quantities themselves, implying they are not self-averaging for all volume fractions. We've shown this is the case for the even Rényi negativity at transition, as well as for both even and odd Rényi negativities for $f_{A_1} > 1/2$. In particular we've shown that odd Rényi negativity behaves mostly the same as Rényi entropy at the τ to

[4] A possible resolution is that our calculation was done at fixed f_{A_2}, roughly the same as fixing k_2 in [67]. Only when $k = k_1 k_2$ was fixed, similar to fixing f_A, do they see $\mathcal{O}(V)$ corrections.

X transition, exhibiting a $\mathcal{O}(\sqrt{V})$ enhanced correction at exactly $k = 1$. One surprising outcome is that, for both Rényi negativities, canonical typicality holds in some cases where the partially transposed density matrix is defined on a subsystem A_1A_2 larger than half of the total system.

One interesting question is what bearing these volumetric corrections have on the validity of the cosmic brane prescription. It's expected that the holographic dual of subregion entanglement measures is given by the action of a geometric solution with a massive cosmic brane (or branes) inserted [40, 137]. Away from transition, it's expected that there is a single dominant saddle, or at the very least an $\mathcal{O}(1)$ number of equivalent saddles, all of which have have small enough fluctuations that we can treat the calculation of the brane area perturbatively. What happens if this saddle doesn't exist?[5] For the $n < 1$ Rényi entropy, for example, the dual gravitational description is expected to be a cosmic brane with negative tension [137], so the minimal energy configuration would be a brane that falls towards the boundary. This is roughly the "holographic dual" of the $\mathcal{O}(V)$ corrections to ETH; it represents a failure of a single approximately geometric state to describe the dual system.

We now discuss some extensions to our work. A necessary restriction in our analysis is only summing over a subset of all relevant permutations near a particular phase transition. It would be useful to find a closed form expression for the moments of a block transposed Wishart matrix without these assumptions, which would involve finding a closed form solution to the recursion relation in [140]. This would be especially nice as we could probe the region $f_A < 1/2$, which is where one could expect ETH to hold as the partially transposed density matrix is defined on less than half of the total system.

A technical point in our analysis was the use of 2-Dyck paths and 2-Narayana numbers, as opposed to (1-)Dyck paths which appear in the calculation of entanglement

[5] We thank Pratik Rath for discussions on this point.

entropy. It's possible some further generalization of Narayana numbers (as in e.g. [160]) will be relevant for calculating transitions in higher party entanglement measures in a similar model.

So far, we've only discussed Rényi negativity, but there exists a family of holographically inspired measures termed refined Rényi negativities, which are given by

$$S^{T_2(n)}(\rho_{A_1A_2}) = -n^2\partial_n\left(\frac{1}{n}\log\mathcal{N}_n^{(\text{odd/even})}(\rho_{A_1A_2})\right) \tag{4.102}$$

We have not touched on the structure of transitions in these measures, but they could presumably be treated in the same way we've presented. Of particular interest is the refined Rényi 2-negativity $S^{T_2(2)}$, the $n \to 2$ limit of the even refined Rényi entropy. This quantity is explicitly given by

$$S^{T_2(2)} = -\lim_{m\to 1} m^2\partial_m\left(\frac{1}{m}\log\mathcal{N}_{2m}^{(\text{even})}\right) = -\sum_i \frac{\lambda_i^2}{\sum_j\lambda_j^2}\log\left(\frac{\lambda_i^2}{\sum_j\lambda_j^2}\right) \tag{4.103}$$

which is the von Neumann entropy of the normalized density matrix $\left(\rho_{A_1A_2}^{T_2}\right)^2$. Consequently, the expectation is that the corrections will be $\mathcal{O}(\sqrt{V})$, which is indeed what is seen in the gravitational setting. It would be nice to derive this relation from our formalism.

Additionally, this formalism could be applied to study the reflected entropy [161] and its Rényi generalizations thereof [162–164]. Reflected entropy has been studied in a similar gravitational system [163] and was shown to have $\mathcal{O}(\sqrt{V})$ corrections at transition, as in the case of the von Neumann entropy, derived via a resolvent calculation. Presumably the relevant permutations could be enumerated and the corrections calculated as we've done in this work.

Lastly, it would be interesting to see the corrections to our results from imposing

energy conservation between replicas as described in [134], and to understand if our simplified ansatz produces similar corrections at transition.

Chapter 5

Holographic Tensor Networks with Bulk Gauge Symmetries

5.1 Introduction

The ultimate goal of the AdS/CFT correspondence is to understand, concretely, the relationship between a bulk gravitational theory and its dual boundary conformal field theory. Holographic duality posits that the partition functions of the two theories are equal and that there exists an isomorphism between the Hilbert space of states of a theory of quantum gravity $\mathcal{H}_{\text{bulk}}$ and the Hilbert space of a seemingly unrelated quantum mechanical system $\mathcal{H}_{\text{boundary}}$. If we were to understand the precise relation between these Hilbert spaces, we would have a tractable handle with which to study quantum gravity, in whatever form it may ultimately arise.

In practice, the UV degrees of freedom in the bulk are not well-understood, so one must often be satisfied with studying a subspace of states given by small fluctuations around a fixed semiclassical saddle. These states span a code subspace of the quantum gravity Hilbert space, and are thus embedded in the larger Hilbert space of the dual

boundary theory, in the same way as the logical qubits of a quantum error correcting code (QECC) are embedded in a larger Hilbert space of physical qubits [165].

In the last decade, a useful tool for developing intuition about the bulk-to-boundary map has been tensor networks. Tensor networks, specifically projected entangled pair states (PEPS) and PEPS-inspired tensor networks, originally arose in many-body physics as a generalization of matrix product states, which allowed one to efficiently prepare spin chain states with area law entanglement [166].

As a toy model for holography, tensor networks found their niche due to the fact that they obey the Ryu-Takayanagi (RT) formula [38] and its refinements [40, 43, 45, 167]. In particular, random tensor networks (RTNs) [168] reproduce several desirable properties of a holographic QECC, namely satisfying a quantum-corrected RT formula and the Petz recontruction of local operators [169].

We now give a short overview of holographic RTNs and their entanglement properties, as well as their issues. A rank-k tensor can be represented by its components $T_{\mu_1\cdots\mu_k}$, with $\mu_i = 1, \ldots, D_i$ (the bond dimension). We can associate to each leg a D_i-dimensional Hilbert space $\mathcal{H}_i$ spanned by an orthonormal basis of states $\{|\mu_i\rangle,\ \mu_i = 1, \cdots, D_i\}$. The tensor T can then be thought of as a state on the tensor product Hilbert space $\bigotimes_{i=1}^{k} \mathcal{H}_i$:

$$|T\rangle = \sum_{\mu_1,\cdots,\mu_k} T_{\mu_1\cdots\mu_k} |\mu_1\rangle \otimes \cdots \otimes |\mu_k\rangle. \tag{5.1}$$

To construct a tensor network, we consider a set of vertices and links which form a network. To each vertex x we associate a state $|T_x\rangle$, such that the collection of all tensors defines a product state $\otimes_x |T_x\rangle$. Adjacent tensors are those connected by a link; their corresponding legs are contracted by projecting onto a maximally entangled state. For simplicity, we assume that all contracted legs have the same bond dimension D. Denoting the tensor product Hilbert space on the two legs connecting the tensors at vertices x and

y as $\mathcal{H}_{xy} \otimes \mathcal{H}_{yx}$, this means that we project onto the state $|xy\rangle = D^{-1/2} \sum_{\mu=1}^{D} |\mu_{xy}\rangle \otimes |\mu_{yx}\rangle$. Uncontracted legs are called "dangling" and come in two types: bulk legs (viewed as input) and boundary legs (viewed as output). We write the boundary state in the following way:[1]

$$|\Psi_\partial\rangle = \left(\langle \Phi_b | \otimes \bigotimes_{\langle xy \rangle} \langle xy | \right) \left(\bigotimes_x |T_x\rangle \right), \tag{5.2}$$

where we project the bulk input legs onto a bulk state $|\Phi_b\rangle$. In an RTN, we choose T_x to be independent random tensors and take D to be large. We will not go into details on how one computes Rényi entropy in the RTN here; the important point is that, for a boundary subregion R, one finds the following answer for the Rényi entropy $S_n(R)$:

$$S_n(R) = |\gamma_R| \log D + S_n(\rho_r), \tag{5.3}$$

where $|\gamma_R|$ is the number of links cut by the minimal surface γ_R homologous to R and $S_n(\rho_r)$ is the Rényi entropy of the bulk subregion r bounded by $R \cup \gamma_R$ (we will call r the entanglement wedge). Analytically continuing to $n = 1$ recovers the Faulkner-Lewkowycz-Maldacena (FLM) formula

$$S_{\text{vN}}(R) = \frac{\langle \hat{A} \rangle}{4G_N} + S_{\text{vN}}(\rho_r), \tag{5.4}$$

with $|\gamma_R| \log D$ identified with the expectation value of the area operator $\langle \hat{A} \rangle / 4G_N$.

In a state with vanishing bulk Rényi entropy (such as a product state), the boundary Rényi entropy (5.3) is consequently independent of n. The RTN thus exhibits a flat entanglement spectrum due to the projection of contracted legs onto maximally mixed states.[2] This differs sharply from what we expect from generic situations in AdS/CFT.

[1] Here, we have chosen a pure state as the bulk input, but generalizing to mixed states is straightforward.

[2] The HaPPY code [170] also features a flat Rényi spectrum for similar reasons.

For example, the Rényi entropy for an interval R of length ℓ in the vacuum state of a two-dimensional CFT takes the form

$$S_n(R) = \frac{c}{6}\left(1+\frac{1}{n}\right)\log\left(\frac{\ell}{\epsilon}\right), \tag{5.5}$$

which is manifestly n-dependent. One possible solution is to instead project contracted legs onto a nonmaximally entangled link state [128, 171]. By tuning the entanglement spectrum appropriately, this allows one to reproduce the correct single-interval CFT vacuum Rényi entropy (5.5), but does not work in more general cases such as that of multiple disjoint intervals. To see this, consider two disjoint intervals R_1 and R_2 (see Figure 5.1), and for simplicity consider the case where the mutual information between the intervals is small in the sense that the RT surfaces are always in a disconnected phase. The boundary Rényi entropy can be obtained by inserting appropriate cosmic branes into the bulk [137]. The tension of the cosmic branes is proportional to $1-1/n$. In a fully gravitating system, the two cosmic branes homologous to R_1, R_2 will backreact and affect each other in an n-dependent way. This results in a nonzero Rényi mutual information between the two intervals that cannot be reproduced in RTNs by simply adding nonmaximally entangled links, because they would not allow the minimal surfaces to affect each other.

From the gravity point of view, the RTN prepares a so-called fixed-area state [141, 154], which is an eigenstate of the area operator $\hat{A}$ in (5.4). Such eigenstates form a complete basis for semiclassical states prepared via the gravitational path integral, so in principle any semiclassical state can be represented as a superposition over fixed-area basis states $|\alpha\rangle$, where α labels the eigenvalues of the area operator. As the area operator lives on the RT surface dividing the entanglement wedge r and its complement $\overline{r}$, it naturally belongs to the center of the algebra of bulk operators in r. This view was

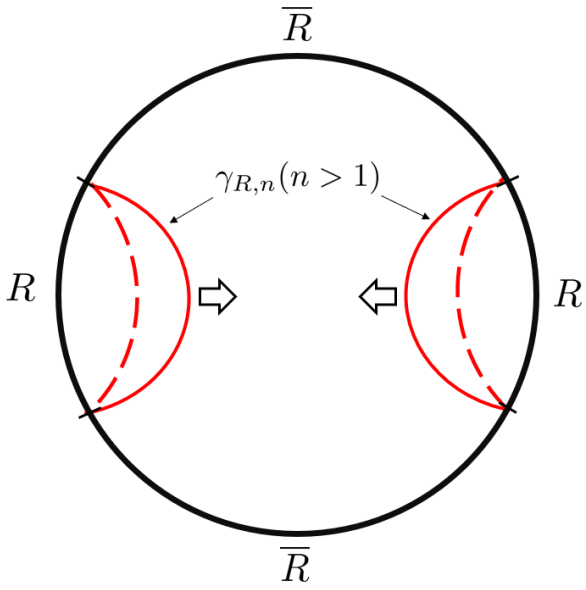


Figure 5.1: The cosmic branes that arise in computing the Rényi entropy for disjoint subregions. These branes have nonzero, n-dependent tension, and so would backreact in a realistic holographic system.

espoused in [172], where it was shown that the FLM formula (5.4) can be derived from a quantum error correcting code with complementary recovery. In that language, the area operator is a specific element of the center of the bulk von Neumann algebra on r. The usual RTN implements a special case of this where the algebra has a trivial center, i.e., the center consists of c-numbers only and is therefore isomorphic to $\mathbb{C}$. In particular, this means that the area operator must be a c-number, which, as previously discussed, is incongruous with what one observes in gravitational holography.

The goal of this chapter is to construct a model where the algebra on r has a nontrivial center and to identify a nontrivial area operator living in the center.[3] An *ad hoc* way of getting a nontrivial center is to "stack" multiple layers of tensor networks by hand to form superpositions of fixed-area states. We will not do this but will instead pursue a more physically motivated approach. In particular, one would like to incorporate something

[3]Having a nontrivial center does not guarantee that the area operator is not a c-number; for example, see [173].

akin to "edge states", degrees of freedom which live on the minimal surface, in order to go beyond fixed-area states and produce a nontrivial area operator.[4] Our goal in this work is to give a model which provides a physical origin for these edge states. Inspired by similar operators found in gauge theory [175], we will add a second layer on top of the standard RTN which imposes gauge invariance. This alters the algebra of operators in the bulk, and as we will show, it introduces a nontrivial contribution to the area operator of the following form:

$$\Delta\widetilde{A} = \bigoplus_{\alpha} \widetilde{P}^{\alpha} \log d_{\alpha}, \tag{5.6}$$

where roughly speaking α denotes a superselection sector in the gauge-invariant Hilbert space, $\widetilde{P}^{\alpha}$ is the projection onto that superselection sector, and d_{α} is the dimension of α viewed as an irreducible representation. The important thing to note at the moment is that this operator is not a c-number and is therefore nontrivial.

The structure of this chapter is as follows. In Section 5.2 we will set up our model – a two-layer gauged random tensor network – and introduce the formalism for gauge theory on a graph. In Section 5.3 we will analyze the Hilbert space of gauge-invariant states and the algebras of gauge-invariant operators for a subregion. In Section 5.4 we will compute entanglement and Rényi entropies in both the pre-gauged and gauge-invariant algebras, which we will use to derive the new area operator for our model. We conclude with some discussion and future directions.

5.2 The Gauged Random Tensor Network

We now construct our model. It has two layers: a top layer consisting of a gauge theory on a graph, and a bottom layer made of a standard random tensor network. We illustrate some examples of this two-layer model in Figure 5.2. The top layer produces a

[4] Initial work in this direction was taken in [174] by generalizing the HaPPY code.

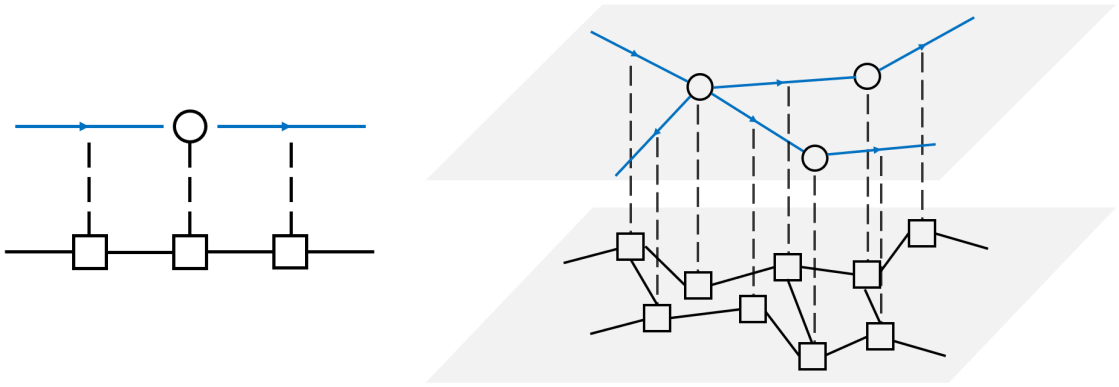


Figure 5.2: Some examples of our two-layer model, with a gauge theory on a directed graph on the top layer and a random tensor network with dangling boundary legs on the bottom. In these examples, we choose each tensor in the bottom layer to have a bulk input leg which is either a vertex or edge on the graph. The light gray planes in the right example are included for visual clarity.

gauge-invariant state which is then fed into the bottom layer as input. The final output of the model is the boundary state produced by the bottom RTN. We can then analyze properties of the boundary state (such as its entropy) using the usual techniques for the random tensor network.

This construction has some nice properties. In particular, one might be worried that if the structure of the RTN is altered, Petz reconstruction of local operators might no longer hold. Here we avoid this potential issue by keeping the tensor network the same, but changing the space of states that can be fed into the network.

Given this construction, we would like to understand what set of gauge-invariant states we will be feeding into the bottom layer. The following is based on a nondynamical version of the standard Kogut-Susskind construction in lattice gauge theory [176].[5] As we do not require our graph to be a lattice, i.e. there is not necessarily a regular tiling, we will refrain from calling our top layer a lattice gauge theory.

Our starting point is an arbitrary directed graph $\Lambda = (V, E)$ consisting of vertices

[5]See also related discussion in [177].

$V = \{v\}$ and edges $E = \{e\}$. We require the graph to be directed so we have a well-defined orientation on each edge, though we emphasize that the choice of orientation is arbitrary. We impose no additional conditions on the graph. In particular, the graph could have loops, adjacent vertices could be connected by multiple edges, and the graph does not need to be planar.

We start with a gauge group, which we choose to be a compact Lie group G. It does not have to be connected, and in particular, we could consider finite groups such as $\mathbb{Z}_2$ if we wish. We assign a (pre-gauged) Hilbert space to each vertex and edge of the graph Λ. The Hilbert space $\mathcal{H}_e$ on each edge e is taken to be $L^2(G)$, the space of square-integrable functions on G. A state $|\psi\rangle_e$ in this $\mathcal{H}_e = L^2(G)$ can be written as an integral[6] over orthonormal basis elements $|g\rangle$ labeled by $g \in G$:

$$|\psi\rangle_e = \int dg \psi(g) |g\rangle_e , \tag{5.7}$$

where dg is the Haar measure[7] on G. For our purposes, it will be useful to work with another orthonormal basis

$$|\alpha i j\rangle_e , \quad i, j = 1, 2, \cdots, d_\alpha \tag{5.8}$$

for the same $\mathcal{H}_e = L^2(G)$, where α labels irreducible representations (irreps) of G and d_α is the dimension of the representation α. This representation basis is orthonormal:

$${}_e\langle \alpha i j | \beta k \ell \rangle_e = \delta_{\alpha\beta}\delta_{ik}\delta_{j\ell}, \tag{5.9}$$

[6] In cases where G is finite, the integral is understood as a sum: $|\psi\rangle_e = \sum_{g\in G} \frac{1}{|G|}\psi(g)|g\rangle_e$, where $|G|$ is the order of G.

[7] The Haar measure on G is invariant under left and right group multiplication ($g \to g'g$ and $g \to gg'$) and is normalized such that $\int dg = 1$.

and can be written in terms of the previously defined group basis $|g\rangle_e$:

$$|\alpha ij\rangle_e = \sqrt{d_\alpha} \int dg\, D^\alpha_{ij}(g)\, |g\rangle_e\,, \tag{5.10}$$

where $D^\alpha_{ij}(g)$ are elements of a unitary matrix $D^\alpha(g)$ representing g in α. This can be viewed as a "Fourier transform" between the representation basis and the group basis.

The group action induces a set of unitaries $L_e(g)$ and $R_e(g)$ which act as left and right group multiplications on the group basis:

$$L_e(g)\,|h\rangle_e = |gh\rangle_e\,, \quad R_e(g^{-1})\,|h\rangle_e = |hg^{-1}\rangle_e\,. \tag{5.11}$$

In the representation basis, the group unitaries instead act as unitary matrix multiplication on one of the two indices i, j:

$$L_e(g)\,|\alpha ij\rangle_e = \sum_k D^{\overline{\alpha}}_{ki}(g)\,|\alpha kj\rangle_e\,, \quad R_e(g^{-1})\,|\alpha ij\rangle_e = \sum_k D^\alpha_{kj}(g)\,|\alpha ik\rangle_e\,, \tag{5.12}$$

where $\overline{\alpha}$ denotes the complex conjugate representation of α defined by $D^{\overline{\alpha}}_{ij}(g) = D^{\alpha *}_{ij}(g)$. Thus there are two copies of G acting on $\mathcal{H}_e$: $L_e(g)$ gives the group action of the first copy under which $|\alpha ij\rangle_e$ transforms in the representation $\overline{\alpha}$, and $R_e(g^{-1})$ gives the action of the second copy under which $|\alpha ij\rangle_e$ transforms in the representation α. Altogether, $|\alpha ij\rangle_e$ transforms in the external tensor product[8] representation $\overline{\alpha} \boxtimes \alpha$ of $G \times G$. Using

[8]For representations α_1, α_2 of G, their external tensor product $\alpha_1 \boxtimes \alpha_2$ is a representation of $G \times G$ with an underlying vector space $\mathcal{H}^{\alpha_1} \otimes \mathcal{H}^{\alpha_2}$, where $\mathcal{H}^{\alpha_1}$ transforms under the first G in the α_1 representation and $\mathcal{H}^{\alpha_2}$ transforms under the second G in the α_2 representation. Note that this is different from the (usual) tensor product $\alpha_1 \otimes \alpha_2$ which is a representation of G (not $G \times G$), with an underlying vector space $\mathcal{H}^{\alpha_1} \otimes \mathcal{H}^{\alpha_2}$ where $\mathcal{H}^{\alpha_1}$ and $\mathcal{H}^{\alpha_2}$ transform under the same G.

this, we decompose $\mathcal{H}_e$ as

$$\mathcal{H}_e \cong \bigoplus_\alpha \mathcal{H}^{\overline{\alpha}} \otimes \mathcal{H}^\alpha \cong \bigoplus_\alpha (\mathcal{H}^\alpha)^{\oplus d_\alpha}\,, \tag{5.13}$$

where the sum runs over all irreducible representations α of G and $\mathcal{H}^\alpha$ is a Hilbert space of dimension d_α transforming in the α representation. It will be convenient to use the representation basis for the remainder of this work.

Now we turn to the (pre-gauged) Hilbert space $\mathcal{H}_v$ on a vertex v. In general, $\mathcal{H}_v$ may be chosen quite arbitrarily (corresponding to specifying any number of matter degrees of freedom including the case of no matter), but it needs to furnish some representation under the group action of G. This representation could be reducible or trivial, but it can always be decomposed into a direct sum of irreducible representations of G. Using this, we may decompose a general $\mathcal{H}_v$ as

$$\mathcal{H}_v = \bigoplus_\alpha (\mathcal{H}_v^\alpha)^{\oplus n_\alpha}\,. \tag{5.14}$$

Here the sum again runs over all distinct irreducible representations α of G and n_α is the multiplicity of the representation α in $\mathcal{H}_v$. Note that n_α could be any nonnegative integer, and in particular, it could be zero (representing the absence of a given representation α in $\mathcal{H}_v$). Thus, the simplest choice of $\mathcal{H}_v$ is a trivial Hilbert space with no matter (corresponding to $n_\alpha = 0$ for all α), but in the discussion below we will consider the general case (5.14) with arbitrary n_α. Furthermore, we will allow $\mathcal{H}_v$ to vary from one vertex to another. An orthonormal basis of states for the Hilbert space $\mathcal{H}_v$ can be written as

$$|\alpha i j\rangle_v\,, \quad i = 1, \cdots, n_\alpha, \quad j = 1, \cdots, d_\alpha, \tag{5.15}$$

where the first index i runs over the multiplicity n_α and the second runs over the dimen-

sion d_α. The group action of G on $\mathcal{H}_v$ is given by unitary operators $U_v(g)$, which act on the $|\alpha ij\rangle_v$ basis as

$$U_v(g)\,|\alpha ij\rangle_v = \sum_k D^\alpha_{kj}(g)\,|\alpha ik\rangle_v\,. \tag{5.16}$$

Note that $U_v(g)$ only acts on the second index j and is analogous to the action of $R_e(g^{-1})$ in (5.11). Thus, we find an important distinction between the vertex Hilbert space $\mathcal{H}_v$ and the edge Hilbert space $\mathcal{H}_e$. To see this, first note that the two Hilbert spaces share some similarities. In particular, $\mathcal{H}_e$ is a direct sum of irreducible representations α with multiplicity d_α as shown on the right-hand side of (5.13), and this is the analogue of (5.14) for $\mathcal{H}_v$. The representation basis (5.8) of $\mathcal{H}_e$ is similar to the basis (5.15) of $\mathcal{H}_v$. However, the difference is that an edge has the additional structure of allowing another group action $L_e(g)$ that acts on the first index i of $|\alpha ij\rangle_e$, whereas at a vertex the first index i of $|\alpha ij\rangle_v$ is a multiplicity index that does not admit a natural group action.

The pre-gauged Hilbert space for the entire graph is then

$$\mathcal{H} = \left(\bigotimes_{v\in V}\mathcal{H}_v\right)\otimes\left(\bigotimes_{e\in E}\mathcal{H}_e\right). \tag{5.17}$$

We refer to the algebra of all bounded operators on $\mathcal{H}$ as $\mathcal{A} = \mathcal{B}(\mathcal{H})$. As $\mathcal{H}$ completely factorizes over the vertices and edges, so too does the algebra of operators

$$\mathcal{A} = \left(\bigotimes_{v\in V}\mathcal{A}_v\right)\otimes\left(\bigotimes_{e\in E}\mathcal{A}_e\right). \tag{5.18}$$

Using the representation basis (5.8) of $\mathcal{H}_e$, $\mathcal{A}_e$ can be written as

$$\mathcal{A}_e = \text{span}\{|\alpha ij\rangle_e\,\langle\beta k\ell|\}, \tag{5.19}$$

where the indices i, j, k, ℓ run over the irrep dimension. Similarly, using (5.15) we write

$$\mathcal{A}_v = \text{span}\{|\alpha ij\rangle_v \langle \beta k\ell|\}, \tag{5.20}$$

where i, k run over the irrep multiplicity and j, ℓ run over the irrep dimension.

For each vertex v, we now define a gauge transformation $A_v(g)$ as the following unitary operator acting on v and all its associated edges:

$$A_v(g) \equiv U_v(g) \prod_{e \in E^-(v)} L_e(g) \prod_{e \in E^+(v)} R_e(g^{-1}), \tag{5.21}$$

where $E^-(v)$ consists of edges associated to v oriented away from the vertex and $E^+(v)$ consists of edges oriented into the vertex. Physical states are defined to be those invariant under gauge transformations $A_v(g)$ for all g and v. The easiest way of generating a gauge-invariant state is to average over all gauge transformations acting on a state in $\mathcal{H}$. The operator that implements this averaging on a vertex v is the following projector:

$$\Pi_v = \int dg A_v(g). \tag{5.22}$$

Π_v obeys the usual properties of a projector such that $\Pi_v^2 = \Pi_v$ and $\Pi_v = \Pi_v^\dagger$. The gauge-invariant projector on the entire graph is simply the product of individual projectors on all vertices:

$$\Pi_{\text{GI}} = \prod_{v \in V} \Pi_v. \tag{5.23}$$

It is easy to verify that $[A_v(g), A_{v'}(g')] = 0$ for all v, v', g, g', and therefore $[\Pi_v, \Pi_{v'}] = 0$.

Throughout the chapter, we will denote fully gauge-invariant spaces, states, and operators with a tilde; for instance, the gauge-invariant states $|\widetilde{\psi}\rangle$ are elements of $\widetilde{\mathcal{H}}$ defined

via

$$\widetilde{\mathcal{H}} \equiv \Pi_{\mathrm{GI}} \mathcal{H}. \tag{5.24}$$

The gauge-invariant algebra $\widetilde{\mathcal{A}}$ is defined as the space of bounded operators on $\widetilde{\mathcal{H}}$. $\widetilde{\mathcal{A}}$ can alternatively be represented by conjugation of the pre-gauged algebra $\mathcal{A}$ with the projector Π_{GI}:

$$\widetilde{\mathcal{A}} = \Pi_{\mathrm{GI}} \mathcal{A} \Pi_{\mathrm{GI}}. \tag{5.25}$$

We should comment on the interpretation of the operators in this gauge-invariant algebra. Every operator $\widetilde{\mathcal{O}} \in \widetilde{\mathcal{A}}$ can be extended to a pre-gauged operator $\mathcal{O} \in \mathcal{A}$ which acts identically on gauge-invariant states. There is generally more than one extension to $\mathcal{A}$, and to choose a unique extension one must specify the action of the pre-gauged operator on the orthogonal complement of $\widetilde{\mathcal{H}}$. We make the natural choice that the extension $\mathcal{O}$ should annihilate the orthogonal complement. Moreover, for notational simplicity, we identify every $\widetilde{\mathcal{O}} \in \widetilde{\mathcal{A}}$ with its natural extension $\mathcal{O} \in \mathcal{A}$ (which annihilates the orthogonal complement), as we have done in (5.25). The reason for this natural extension will become clearer in later sections.

We now feed any gauge-invariant state $|\widetilde{\psi}\rangle$ as the bulk input into the RTN on the bottom layer, in a manner illustrated by Figure 5.2. In particular, the bulk dangling legs of the RTN should match and connect to the edges and vertices of the graph G on the top layer, for $|\widetilde{\psi}\rangle$ lives on these edges and vertices. In other words, each edge or vertex of G is fed into a bulk dangling leg of the RTN.[9]

In order to utilize the full machinery of the original RTN, we would like the Hilbert spaces associated with the tensors on the bottom layer to be finite-dimensional (as is the case for the original RTN). When G is an infinite group, $\mathcal{H}_e = L^2(G)$ is infinite-

[9] In principle, the RTN could also take any pre-gauged state as the bulk input, but we choose to feed only gauge-invariant states because as we will see, this restriction leads to a nontrivial area operator.

dimensional and there are an infinite number of irreducible representations to sum over, so in order to avoid a tensor in the bottom layer having an infinite-dimensional leg, we impose a cutoff on our edge and vertex Hilbert spaces. This can take the form of, e.g., a cutoff in the sums in (5.13) and (5.14). Therefore, we are only feeding in states that live in a finite-dimensional subspace of $\widetilde{\mathcal{H}}$. This does not affect the discussion in the next section of the gauge-invariant algebra; the cutoff is only relevant when we compute entanglement measures in Section 5.4.

5.3 Deriving the Gauge-Invariant Algebra

Now that we have defined our gauge-invariant states, we would like to understand the structure of the algebra of gauge-invariant operators. Our overarching goal is to write down the gauge-invariant subalgebra for a subregion r of the top layer which we will later use to derive an FLM formula for the gauged RTN.

5.3.1 The Structure of the Gauge-Invariant Hilbert Space

We now study the decomposition of $\widetilde{\mathcal{H}}$ when our graph Λ is divided into a subregion and its complement. We define a subregion r of Λ to be an arbitrary subset of vertices and edges (without further restrictions). We call the complement subregion $\overline{r}$.

In order to work out a useful basis for gauge-invariant states, it is convenient to divide the set V of all vertices into three types: those strictly in r (meaning that the vertex and its associated edges are all in r), those strictly in $\overline{r}$, and vertices "on the cut" (meaning that the vertex and its associated edges are partly in r and partly in $\overline{r}$). We call these sets V_r, $V_{\overline{r}}$, and $V_c \equiv V/\left(V_r \cup V_{\overline{r}}\right)$, respectively. Consequently, the gauge-invariant projector

can be decomposed in the following way:

$$\Pi_{\mathrm{GI}} = \Pi_{V_r}\Pi_{V_c}\Pi_{V_{\overline{r}}}, \tag{5.26}$$

where Π_{V_i} is defined as the product of individual projections Π_v over all vertices $v \in V_i$, for $i = r, c, \overline{r}$. First, let us discuss a partial gauging of the pre-gauged Hilbert space. Using the tensor decomposition of $\mathcal{H} = \mathcal{H}_r \otimes \mathcal{H}_{\overline{r}}$, we can write $\widetilde{\mathcal{H}} = \Pi_{\mathrm{GI}}\mathcal{H}$ as

$$\begin{aligned}\widetilde{\mathcal{H}} &= \Pi_{V_r}\Pi_{V_c}\Pi_{V_{\overline{r}}}\left(\mathcal{H}_r \otimes \mathcal{H}_{\overline{r}}\right) \\ &= \Pi_{V_c}\left(\left(\Pi_{V_r}\mathcal{H}_r\right) \otimes \left(\Pi_{V_{\overline{r}}}\mathcal{H}_{\overline{r}}\right)\right).\end{aligned} \tag{5.27}$$

We define the two terms in the parentheses as

$$\hat{\mathcal{H}}_r \equiv \Pi_{V_r}\mathcal{H}_r, \quad \hat{\mathcal{H}}_{\overline{r}} \equiv \Pi_{V_{\overline{r}}}\mathcal{H}_{\overline{r}}. \tag{5.28}$$

These are "partially gauged" Hilbert spaces, in the sense that states in $\hat{\mathcal{H}}_r$ ($\hat{\mathcal{H}}_{\overline{r}}$) are invariant under gauge transformations associated to vertices in V_r ($V_{\overline{r}}$), but not so under gauge transformations on the cut. We denote the partially gauged Hilbert space on the full graph as

$$\hat{\mathcal{H}} = \hat{\mathcal{H}}_r \otimes \hat{\mathcal{H}}_{\overline{r}}. \tag{5.29}$$

As $\hat{\mathcal{H}}$ tensor factorizes, the algebra of operators on $\hat{\mathcal{H}}$ also factorizes as

$$\hat{\mathcal{A}} = \hat{\mathcal{A}}_r \otimes \hat{\mathcal{A}}_{\overline{r}}. \tag{5.30}$$

Now that we have a partially gauged Hilbert space $\hat{\mathcal{H}}$, it remains to impose gauge invariance "on the cut" and obtain the fully gauged Hilbert space $\widetilde{\mathcal{H}} = \Pi_{V_c}\hat{\mathcal{H}}$. The gauge

transformation (5.21) associated to each vertex $v_i \in V_c$ can be decomposed into unitary operators in r and $\overline{r}$:

$$A_{v_i}(g_i) = A_{v_i,r}(g_i)A_{v_i,\overline{r}}(g_i). \tag{5.31}$$

Let $n \equiv |V_c|$ be the number of vertices on the cut. The gauge-invariant projector on the cut Π_{V_c} acts by integrating over the gauge transformations associated to the n vertices in V_c:

$$\begin{aligned}\Pi_{V_c} &= \int dg_1 \cdots dg_n A_{v_1}(g_1) \cdots A_{v_n}(g_n) \\ &= \int dg_1 \cdots dg_n A_{v_1,r}(g_1) \cdots A_{v_n,r}(g_n) A_{v_1,\overline{r}}(g_1) \cdots A_{v_n,\overline{r}}(g_n) \\ &\equiv \int dg A_r(g) A_{\overline{r}}(g),\end{aligned} \tag{5.32}$$

where we have defined $A_r(g) = \prod_{i=1}^n A_{v_i,r}(g_i)$ (and similarly for $A_{\overline{r}}(g)$), $g = (g_1, \cdots, g_n)$ is a element of G^n (the direct product of n copies of G on the cut), and dg is the Haar measure on G^n. Thus $A_r(g)$ is a G^n action on $\hat{\mathcal{H}}_r$, and $\hat{\mathcal{H}}_r$ can be decomposed into irreps of G^n. We decompose $\hat{\mathcal{H}}_r$ into the following way:

$$\hat{\mathcal{H}}_r \cong \bigoplus_{\alpha,i} \hat{\mathcal{H}}_r^{\alpha i}, \tag{5.33}$$

where α as an irreducible representation of G^n can also be thought of as the external tensor product of n irreps of G, i.e., α denotes the external tensor product $\alpha_1 \boxtimes \alpha_2 \boxtimes \cdots \boxtimes \alpha_n$. Thus, we will sometimes write α as a tuple of G irreps $(\alpha_1, \alpha_2, \cdots, \alpha_n)$. The index $i = 1, \cdots, n_\alpha$ denotes the multiplicity of the α irrep. The sum ranges over all G^n irreps but some irreps may appear with zero multiplicity, as in the single vertex Hilbert space (5.14).

From the decomposition (5.33), we write an orthonormal basis for $\hat{\mathcal{H}}_r$ as $\{|\alpha i k\rangle_r\}$,

where again the first index $i = 1, \cdots, n_\alpha$ runs over the irrep multiplicity and the second index $k = 1, \cdots, d_\alpha$ labels an orthonormal basis for each $\hat{\mathcal{H}}_r^{\alpha i}$. Similarly, we write an orthonormal basis for $\hat{\mathcal{H}}_{\overline{r}}$ as $\{|\overline{\beta} j \ell\rangle_{\overline{r}}\}$, where $j = 1, \cdots, \overline{n}_{\overline{\beta}}$, and $\overline{n}_{\overline{\beta}}$ is the multiplicity of the $\overline{\beta}$ irrep on $\overline{r}$. Explicitly, $A_r(g)$ ($A_{\overline{r}}(g)$) acts on the basis states of $\hat{\mathcal{H}}_r$ ($\hat{\mathcal{H}}_{\overline{r}}$) via

$$\begin{aligned} A_r(g) \, |\alpha i k\rangle_r &= \sum_{k'} D^{\alpha}_{k'k}(g) \, |\alpha i k'\rangle_r \, , \\ A_{\overline{r}}(g) \, |\overline{\beta} j \ell\rangle_{\overline{r}} &= \sum_{\ell'} D^{\overline{\beta}}_{\ell'\ell}(g) \, |\overline{\beta} j \ell'\rangle_{\overline{r}} \, . \end{aligned} \tag{5.34}$$

my Combining the basis for $\hat{\mathcal{H}}_r$ and for $\hat{\mathcal{H}}_{\overline{r}}$, we write an orthonormal basis for $\hat{\mathcal{H}}$ as

$$\{|\alpha i k\rangle_r \, |\overline{\beta} j \ell\rangle_{\overline{r}}\}.$$

It is worth noting that the multiplicities $\overline{n}_{\overline{\alpha}}$ on $\overline{r}$ are generally independent from the multiplicities $n_{\overline{\alpha}}$ on r; in particular, $\overline{n}_{\overline{\alpha}}$ could vanish while $n_{\overline{\alpha}}$ is nonzero, and vice versa.

In a sense, we have done as much gauging as we can while keeping the factorization of the Hilbert space between r and $\overline{r}$. $\hat{\mathcal{H}}$ is similar to what is often called the extended Hilbert space [175, 178–180], which is a choice of Hilbert space into which one can embed gauge-invariant states such that the extended Hilbert space factorizes across the cut. Here we arrive at a similar prescription by restricting from a larger Hilbert space $\mathcal{H}$.

Now we will write a basis of states for the fully gauge-invariant Hilbert space $\widetilde{\mathcal{H}}$.

Lemma 1. The fully gauge-invariant Hilbert space $\widetilde{\mathcal{H}} = \Pi_{V_c} \left(\hat{\mathcal{H}}_r \otimes \hat{\mathcal{H}}_{\overline{r}} \right)$ is given by

$$\widetilde{\mathcal{H}} = \left\{ \sum_{\alpha i j k} \widetilde{\psi}_{\alpha i j} \, |\alpha i k\rangle_r \, |\overline{\alpha} j k\rangle_{\overline{r}} : \widetilde{\psi}_{\alpha i j} \in \mathbb{C} \right\}. \tag{5.35}$$

Proof: Since we already have a basis for the partially gauged Hilbert space, it

suffices to demonstrate the action of Π_{V_c} on these basis states, which is given by

$$\Pi_{V_c} |\alpha ik\rangle_r |\overline{\beta} j\ell\rangle_{\overline{r}} = \sum_{k'\ell'} \int dg D^{\alpha}_{k'k}(g) D^{\overline{\beta}}_{\ell'\ell}(g) |\alpha ik'\rangle_r |\overline{\beta} j\ell'\rangle_{\overline{r}}. \tag{5.36}$$

We recall the Schur orthogonality relation for compact groups:

$$\int dg D^{\alpha}_{k'k}(g) D^{\overline{\beta}}_{\ell'\ell}(g) = \frac{\delta_{\alpha\beta}\delta_{k'\ell'}\delta_{k\ell}}{d_\alpha}, \tag{5.37}$$

so that the fully gauge-invariant basis states are

$$\Pi_{V_c} |\alpha ik\rangle_r |\overline{\beta} j\ell\rangle_{\overline{r}} = \frac{1}{d_\alpha}\delta_{\alpha\beta}\delta_{k\ell} \sum_{k'\ell'} \delta_{k'\ell'} |\alpha ik'\rangle_r |\overline{\beta} j\ell'\rangle_{\overline{r}} = \frac{1}{d_\alpha}\delta_{\alpha\beta}\delta_{k\ell} \sum_{k'} |\alpha ik'\rangle_r |\overline{\beta} jk'\rangle_{\overline{r}}. \tag{5.38}$$

Choosing $\alpha = \beta$ and $k = \ell$ gives the desired form (5.35).

Remark 1. (5.35) immediately implies a natural Hilbert space isomorphism

$$\widetilde{\mathcal{H}} \cong \bigoplus_{\alpha} \widetilde{\mathcal{H}}^{\alpha}_r \otimes \widetilde{\mathcal{H}}^{\overline{\alpha}}_{\overline{r}}. \tag{5.39}$$

Here $\widetilde{\mathcal{H}}^{\alpha}_r$ denotes a Hilbert space of dimension n_α with orthonormal basis states $|\alpha i\rangle_r$ transforming in the α representation of G^n, and $\widetilde{\mathcal{H}}^{\overline{\alpha}}_{\overline{r}}$ similarly denotes a Hilbert space of dimension $\overline{n}_{\overline{\alpha}}$ with orthonormal basis states $|\overline{\alpha} j\rangle_{\overline{r}}$ transforming in the $\overline{\alpha}$ representation. Note that although irrep labels such as α appear in the basis states, they are fixed within each Hilbert space $\widetilde{\mathcal{H}}^{\alpha}_r$ or $\widetilde{\mathcal{H}}^{\overline{\alpha}}_{\overline{r}}$.

More explicitly, the natural isomorphism (5.39) maps an arbitrary state of (5.35) in the following way:

$$|\widetilde{\psi}\rangle = \sum_{\alpha ijk} \widetilde{\psi}_{\alpha ij} |\alpha ik\rangle_r |\overline{\alpha} jk\rangle_{\overline{r}} \quad \rightarrow \quad \sum_{\alpha ij} \sqrt{d_\alpha}\widetilde{\psi}_{\alpha ij} |\alpha i\rangle_r |\overline{\alpha} j\rangle_{\overline{r}}. \tag{5.40}$$

The $\sqrt{d_\alpha}$ is a crucial factor which ensures that the isomorphism preserves the inner product.

Given this decomposition, our next goal will be to define an algebra of gauge-invariant operators on r, which we will call $\widetilde{\mathcal{A}}_r$. Given the lack of factorization of $\widetilde{\mathcal{H}}$ as indicated by (5.39), we cannot easily write $\widetilde{\mathcal{A}}_r$ as $\mathcal{B}(\widetilde{\mathcal{H}}_r)$ for some putative Hilbert space $\widetilde{\mathcal{H}}_r$. Rather, we will use the known algebra of operators on $\mathcal{H}_r$ and $\hat{\mathcal{H}}_r$ to define $\widetilde{\mathcal{A}}_r$.

5.3.2 The Gauge-Invariant Subregion Algebra

It is tempting to define the algebra of gauge-invariant operators in a subregion r via restriction of the pre-gauged algebra in that region

$$\widetilde{\mathcal{A}}_r = \Pi_{\text{GI}} \mathcal{A}_r \Pi_{\text{GI}}, \tag{5.41}$$

similar to (5.25). There is a second possible description of the gauge-invariant algebra, which is that $\widetilde{\mathcal{A}_r}$ consists of the set of operators $\{\widetilde{\mathcal{O}}_r = \mathcal{O}_r \Pi_{\text{GI}}\}$ for all operators $\mathcal{O}_r \in \mathcal{A}_r$ which commute with the gauge-invariant projector: $[\mathcal{O}_r, \Pi_{\text{GI}}] = 0$. We will call this algebra $\widetilde{\mathcal{A}}_r^{(1)}$, and the algebra (5.41) defined by conjugation by the gauge-invariant projector $\widetilde{\mathcal{A}}_r^{(2)}$. At first blush it is only obvious that $\widetilde{\mathcal{A}}_r^{(1)}$ is a subset of $\widetilde{\mathcal{A}}_r^{(2)}$, as

$$\mathcal{O}_r \Pi_{\text{GI}} = \mathcal{O}_r \Pi_{\text{GI}}^2 = \Pi_{\text{GI}} \mathcal{O}_r \Pi_{\text{GI}} \Rightarrow \widetilde{\mathcal{A}}_r^{(1)} \subseteq \widetilde{\mathcal{A}}_r^{(2)}, \tag{5.42}$$

but it is not obvious the two definitions are equivalent. Here we aim to show that.

Lemma 2. $\widetilde{\mathcal{A}}_r^{(1)} = \widetilde{\mathcal{A}}_r^{(2)}$.

Proof: We again use the group action on the cut $A(g) = A_r(g) A_{\bar{r}}(g)$ and the

gauge-invariant projector on the cut Π_{V_c} which integrates over the group action:

$$\Pi_{V_c} = \int dg A(g). \tag{5.43}$$

We define an element of $\widetilde{\mathcal{A}}_r^{(2)}$ by acting on an arbitrary pre-gauged operator $\mathcal{O}_r \in \mathcal{A}_r$ via

$$\begin{aligned}\Pi_{\mathrm{GI}}\mathcal{O}_r\Pi_{\mathrm{GI}} &= \Pi_{V_{\bar{r}}}\left(\Pi_{V_c}\Pi_{V_r}\mathcal{O}_r\Pi_{V_r}\Pi_{V_c}\right)\Pi_{V_{\bar{r}}} \\ &= \left(\Pi_{V_c}\Pi_{V_r}\mathcal{O}_r\Pi_{V_r}\Pi_{V_c}\right)\Pi_{V_{\bar{r}}} \\ &\equiv \left(\Pi_{V_c}\hat{\mathcal{O}}_r\Pi_{V_c}\right)\Pi_{V_{\bar{r}}}\end{aligned} \tag{5.44}$$

where $\hat{\mathcal{O}}_r \equiv \Pi_{V_r}\mathcal{O}_r\Pi_{V_r} \in \hat{\mathcal{A}}_r$ is an operator on the partially gauged Hilbert space $\hat{\mathcal{H}}_r$. Conjugation via the gauge-invariant projector on the cut yields

$$\Pi_{V_c}\hat{\mathcal{O}}_r\Pi_{V_c} = \int dg dg' A(g)\hat{\mathcal{O}}_r A(g'). \tag{5.45}$$

Using the right-invariance of the Haar measure, we can shift $g \to g(g')^{-1}$ to obtain

$$\Pi_{V_c}\hat{\mathcal{O}}_r\Pi_{V_c} = \int dg A(g)\int dg' A((g')^{-1})\hat{\mathcal{O}}_r A(g') \tag{5.46}$$

$$= \Pi_{V_c}\int dg' A((g')^{-1})\hat{\mathcal{O}}_r A(g') \equiv \Pi_{V_c}\hat{\mathcal{O}}'_r, \tag{5.47}$$

where $\hat{\mathcal{O}}'_r$ is defined by the integral over g'. We could equivalently send $g' \to g^{-1}g'$ to obtain

$$\begin{aligned}\Pi_{V_c}\hat{\mathcal{O}}_r\Pi_{V_c} &= \int dg A(g)\hat{\mathcal{O}}_r A(g^{-1})\Pi_{V_c} \\ &= \int dg A(g^{-1})\hat{\mathcal{O}}_r A(g)\Pi_{V_c} = \hat{\mathcal{O}}'_r\Pi_{V_c}\end{aligned} \tag{5.48}$$

where we use the fact that the Haar measure is invariant under inversion $dg \to d(g^{-1})$. This shows $\hat{\mathcal{O}}'_r \Pi_{V_c} = \Pi_{V_c} \hat{\mathcal{O}}'_r$, so $\hat{\mathcal{O}}'_r$ commutes with the gauge-invariant projector on the cut. By construction, $\hat{\mathcal{O}}'_r$ also commutes with Π_{V_r} and $\Pi_{V_{\overline{r}}}$, so it commutes with Π_{GI}.

Now we show that $\hat{\mathcal{O}}'_r$ is an element of $\mathcal{A}_r$, which is not obvious as $A(g)$ on the cut acts on both r and $\overline{r}$. However, we can write

$$\begin{aligned}\hat{\mathcal{O}}'_r &= \int dg A(g^{-1}) \hat{\mathcal{O}}_r A(g) = \int dg A_r(g^{-1}) A_{\overline{r}}(g^{-1}) \hat{\mathcal{O}}_r A_{\overline{r}}(g) A_r(g) \\ &= \int dg A_r(g^{-1}) \hat{\mathcal{O}}_r A_r(g), \end{aligned} \tag{5.49}$$

as $\hat{\mathcal{O}}_r$ commutes with operators in $\overline{r}$. Thus $\hat{\mathcal{O}}'_r$ is in $\mathcal{A}_r$.

Combining the above, we can write any element of $\widetilde{\mathcal{A}}_r^{(2)}$ as

$$\Pi_{\mathrm{GI}} \mathcal{O}_r \Pi_{\mathrm{GI}} = \hat{\mathcal{O}}'_r \Pi_{V_c} \Pi_{V_{\overline{r}}} = \hat{\mathcal{O}}'_r \Pi_{V_r} \Pi_{V_c} \Pi_{V_{\overline{r}}} = \hat{\mathcal{O}}'_r \Pi_{\mathrm{GI}}, \tag{5.50}$$

which belong to $\widetilde{\mathcal{A}}_r^{(1)}$ as $\hat{\mathcal{O}}'_r$ is an operator in $\mathcal{A}_r$ that commutes with Π_{GI}. Therefore, $\widetilde{\mathcal{A}}_r^{(2)} \subseteq \widetilde{\mathcal{A}}_r^{(1)}$. Moreover, as we argued earlier, we have $\widetilde{\mathcal{A}}_r^{(1)} \subseteq \widetilde{\mathcal{A}}_r^{(2)}$. Thus, we have shown $\widetilde{\mathcal{A}}_r^{(1)} = \widetilde{\mathcal{A}}_r^{(2)}$.

Remark 2. It will be more convenient to use $\widetilde{\mathcal{A}}_r^{(1)}$ as our definition of $\widetilde{\mathcal{A}}_r$ in later discussions. We now rewrite it by introducing the following notation. For the rest of the chapter, we will denote the subset of operators in an algebra that commute with the gauge-invariant projector with a superscript Π; for example, the algebra $\mathcal{A}^{\Pi}$ is defined by

$$\mathcal{A}^{\Pi} \equiv \{\mathcal{O} \in \mathcal{A} : [\mathcal{O}, \Pi_{\mathrm{GI}}] = 0\}. \tag{5.51}$$

It is clear that this subset is itself a von Neumann algebra, as it contains the identity, which necessarily commutes with any projector, and is closed under addition, multipli-

cation, and involution[10]. Similarly, we define the subalgebra $\mathcal{A}_r^{\Pi}$ as

$$\mathcal{A}_r^{\Pi} = \{\mathcal{O}_r \in \mathcal{A}_r : [\mathcal{O}_r, \Pi_{\text{GI}}] = 0\}, \tag{5.52}$$

and define $\hat{\mathcal{A}}_r^{\Pi}$ as

$$\hat{\mathcal{A}}_r^{\Pi} = \{\hat{\mathcal{O}}_r \in \hat{\mathcal{A}}_r : [\hat{\mathcal{O}}_r, \Pi_{\text{GI}}] = 0\}. \tag{5.53}$$

So far, we have shown

$$\widetilde{\mathcal{A}}_r = \mathcal{A}_r^{\Pi}\Pi_{\text{GI}}. \tag{5.54}$$

Lemma 3. $\mathcal{A}_r^{\Pi}\Pi_{\text{GI}} = \hat{\mathcal{A}}_r^{\Pi}\Pi_{\text{GI}}$. *Proof:* It is clear that $\hat{\mathcal{A}}_r^{\Pi}\Pi_{\text{GI}} \subseteq \mathcal{A}_r^{\Pi}\Pi_{\text{GI}}$, so we only need to show the opposite inclusion. Consider any operator $\mathcal{O}_r \in \mathcal{A}_r^{\Pi}$. As this operator commutes with Π_{GI}, we can use the decomposition of the gauge-invariant projector to write $\mathcal{O}_r\Pi_{\text{GI}}$ as

$$\mathcal{O}_r\Pi_{\text{GI}} = \Pi_{\text{GI}}\mathcal{O}_r\Pi_{\text{GI}} = \Pi_{\text{GI}}\left(\Pi_{V_r}\mathcal{O}_r\Pi_{V_r}\right)\Pi_{\text{GI}} = \left(\Pi_{V_r}\mathcal{O}_r\Pi_{V_r}\right)\Pi_{\text{GI}}. \tag{5.55}$$

Note that $\Pi_{V_r}\mathcal{O}_r\Pi_{V_r}$ is an operator on $\hat{\mathcal{H}}_r$ that commutes with Π_{GI}, so it belongs to $\hat{\mathcal{A}}_r^{\Pi}$. Thus, every element of $\mathcal{A}_r^{\Pi}\Pi_{\text{GI}}$ is an element of $\hat{\mathcal{A}}_r^{\Pi}\Pi_{\text{GI}}$. This shows the inclusion $\mathcal{A}_r^{\Pi}\Pi_{\text{GI}} \subseteq \hat{\mathcal{A}}_r^{\Pi}\Pi_{\text{GI}}$, from which we conclude $\mathcal{A}_r^{\Pi}\Pi_{\text{GI}} = \hat{\mathcal{A}}_r^{\Pi}\Pi_{\text{GI}}$.

Corollary 4. $\widetilde{\mathcal{A}}_r = \hat{\mathcal{A}}_r^{\Pi}\Pi_{\text{GI}}$.

Using the corollary above, we will now construct a generic operator in $\widetilde{\mathcal{A}}_r$.

[10]Closure of $\mathcal{A}^{\Pi}$ under addition and multiplication is obvious, and closure under involution follows from the projector being Hermitian.

Lemma 5. $\widetilde{\mathcal{A}}_r$ can be written in the following two forms:

$$\widetilde{\mathcal{A}}_r = \left\{ \hat{\mathcal{O}}_r \Pi_{\text{GI}} : \hat{\mathcal{O}}_r = \sum_{\alpha ijk} \hat{\mathcal{O}}_{\alpha ij} \left|\alpha ik\right\rangle_r \left\langle \alpha jk\right| \otimes \mathbb{1}_{\overline{r}},\, \hat{\mathcal{O}}_{\alpha ij} \in \mathbb{C} \right\} \tag{5.56}$$

$$= \left\{ \widetilde{\mathcal{O}}_r = \sum_{\alpha i i' j k \ell} \widetilde{\mathcal{O}}_{\alpha ij} \left|\alpha ik\right\rangle_r \left\langle \alpha j\ell\right| \otimes \left|\overline{\alpha} i' k\right\rangle_{\overline{r}} \left\langle \overline{\alpha} i' \ell\right| : \widetilde{\mathcal{O}}_{\alpha ij} \in \mathbb{C} \right\}, \tag{5.57}$$

with $\widetilde{\mathcal{O}}_r$ in (5.57) identified with $\hat{\mathcal{O}}_r \Pi_{\text{GI}}$ in (5.56) under $\widetilde{\mathcal{O}}_{\alpha ij} = \hat{\mathcal{O}}_{\alpha ij}/d_\alpha$.[11]

Proof: We show this by noting $\widetilde{\mathcal{A}}_r = \hat{\mathcal{A}}_r^{\Pi} \Pi_{\text{GI}}$ and constructing a generic operator therein. Recall that $\{\left|\alpha ik\right\rangle_r\}$ is a basis for $\hat{\mathcal{H}}_r$, so an operator $\hat{\mathcal{O}}_r \in \hat{\mathcal{A}}_r$ (not necessarily gauge-invariant) can be written as

$$\hat{\mathcal{O}}_r = \sum_{\alpha\beta ijk\ell} \hat{\mathcal{O}}_{\alpha\beta ijk\ell} \left|\alpha ik\right\rangle_r \left\langle \beta j\ell\right| \otimes \mathbb{1}_{\overline{r}} \tag{5.58}$$

with some $\hat{\mathcal{O}}_{\alpha\beta ijk\ell} \in \mathbb{C}$. Now we require $\hat{\mathcal{O}}_r \in \hat{\mathcal{A}}_r^{\Pi}$, so we will try to impose $\hat{\mathcal{O}}_r \Pi_{\text{GI}} = \Pi_{\text{GI}} \hat{\mathcal{O}}_r$. We find

$$\begin{aligned}
\hat{\mathcal{O}}_r \Pi_{\text{GI}} &= \hat{\mathcal{O}}_r \Pi_{V_r} \Pi_{V_{\overline{r}}} \Pi_{V_c} \\
&= \left(\sum_{\alpha\beta ijk\ell} \hat{\mathcal{O}}_{\alpha\beta ijk\ell} \left|\alpha ik\right\rangle_r \left\langle \beta j\ell\right| \otimes \mathbb{1}_{\overline{r}} \right) \Pi_{V_{\overline{r}}} \Pi_{V_c} \\
&= \left(\sum_{\alpha\beta\gamma ijk\ell i' k'} \hat{\mathcal{O}}_{\alpha\beta ijk\ell} \left|\alpha ik\right\rangle_r \left\langle \beta j\ell\right| \otimes \left|\overline{\gamma} i' k'\right\rangle_{\overline{r}} \left\langle \overline{\gamma} i' k'\right| \right) \Pi_{V_c} \\
&= \sum_{\alpha\beta\gamma ijk\ell i' k' \ell'} \hat{\mathcal{O}}_{\alpha\beta ijk\ell} \frac{1}{d_\beta} \delta_{\beta\gamma} \delta_{\ell k'} \left|\alpha ik\right\rangle_r \left\langle \beta j\ell'\right| \otimes \left|\overline{\gamma} i' k'\right\rangle_{\overline{r}} \left\langle \overline{\gamma} i' \ell'\right| \\
&= \sum_{\alpha\beta ijk\ell i' \ell'} \hat{\mathcal{O}}_{\alpha\beta ijk\ell} \frac{1}{d_\beta} \left|\alpha ik\right\rangle_r \left\langle \beta j\ell'\right| \otimes \left|\overline{\beta} i' \ell\right\rangle_{\overline{r}} \left\langle \overline{\beta} i' \ell'\right|,
\end{aligned} \tag{5.59}$$

[11] In a slight abuse of notation, we have referred to the matrix elements of an operator with the same symbol as the operator itself, but with irrep labels and indices such as α, i, and j. We could have referred to $\hat{\mathcal{O}}_{\alpha ij}$ as $\hat{\mathcal{O}}_{r,\alpha ij}$ in (5.56), but for simplicity, we will use the former.

where we have used the basis of $\hat{\mathcal{H}}_{\overline{r}}$ in going to the third line and used (5.38) in going to the fourth line. We can apply the same procedure to write $\Pi_{\text{GI}}\hat{\mathcal{O}}_r$ as

$$\begin{aligned}\Pi_{\text{GI}}\hat{\mathcal{O}}_r &= \Pi_{V_c}\sum_{\alpha\beta\gamma ijk\ell i'k'}\hat{\mathcal{O}}_{\alpha\beta ijk\ell}\left|\alpha ik\right\rangle_r\left\langle\beta j\ell\right|\otimes\left|\overline{\gamma}i'k'\right\rangle_{\overline{r}}\left\langle\overline{\gamma}i'k'\right| \\ &= \sum_{\alpha\beta\gamma ijk\ell i'k'\ell'}\hat{\mathcal{O}}_{\alpha\beta ijk\ell}\frac{1}{d_\alpha}\delta_{\alpha\gamma}\delta_{kk'}\left|\alpha i\ell'\right\rangle_r\left\langle\beta j\ell\right|\otimes\left|\overline{\gamma}i'\ell'\right\rangle_{\overline{r}}\left\langle\overline{\gamma}i'k'\right| \\ &= \sum_{\alpha\beta ijk\ell i'\ell'}\hat{\mathcal{O}}_{\alpha\beta ijk\ell}\frac{1}{d_\alpha}\left|\alpha i\ell'\right\rangle_r\left\langle\beta j\ell\right|\otimes\left|\overline{\alpha}i'\ell'\right\rangle_{\overline{r}}\left\langle\overline{\alpha}i'k\right|. \end{aligned} \tag{5.60}$$

One way to proceed is to find conditions on $\hat{\mathcal{O}}_{\alpha\beta aijk\ell}$ such that the two expressions (5.59), (5.60) are equal, but doing this explicitly turns out to be slightly complicated (in cases where the multiplicities $\overline{n}_{\overline{\alpha}}$, $\overline{n}_{\overline{\beta}}$ vanish but n_α, n_β do not). Instead, we will use the equality of (5.59) and (5.60) to directly show that $\hat{\mathcal{A}}_r^{\Pi}\Pi_{\text{GI}}$ contains and is contained in the right-hand side of (5.56), which we now define as

$$\widetilde{\mathcal{A}}_r^{(3)} \equiv \left\{\hat{\mathcal{O}}_r\Pi_{\text{GI}} : \hat{\mathcal{O}}_r = \sum_{\alpha ijk}\hat{\mathcal{O}}_{\alpha ij}\left|\alpha ik\right\rangle_r\left\langle\alpha jk\right|\otimes\mathbb{1}_{\overline{r}},\ \hat{\mathcal{O}}_{\alpha ij}\in\mathbb{C}\right\}. \tag{5.61}$$

First, we show that $\widetilde{\mathcal{A}}_r^{(3)}$ defined by (5.61) is equal to (5.57) as claimed. To see this, we simply apply (5.59) to the special case of $\hat{\mathcal{O}}_r = \sum_{\alpha ijk}\hat{\mathcal{O}}_{\alpha ij}\left|\alpha ik\right\rangle_r\left\langle\alpha jk\right|\otimes\mathbb{1}_{\overline{r}}$ and find $\hat{\mathcal{O}}_r\Pi_{V_c}$ to be identical to $\widetilde{\mathcal{O}}_r$ in (5.57) under $\widetilde{\mathcal{O}}_{\alpha ij} = \hat{\mathcal{O}}_{\alpha ij}/d_\alpha$. Moreover, applying (5.60) to this case yields the same operator, so we find that this special $\hat{\mathcal{O}}_r$ commutes with Π_{GI}. Thus, $\widetilde{\mathcal{A}}_r^{(3)}$ is contained in $\hat{\mathcal{A}}_r^{\Pi}\Pi_{\text{GI}}$.

Finally, we will show that $\hat{\mathcal{A}}_r^{\Pi}\Pi_{\text{GI}}$ is contained in $\widetilde{\mathcal{A}}_r^{(3)}$. Any $\hat{\mathcal{O}}_r\Pi_{\text{GI}}\in\hat{\mathcal{A}}_r^{\Pi}\Pi_{\text{GI}}$ can be

written explicitly as

$$\begin{aligned}\hat{\mathcal{O}}_r\Pi_{\rm GI} = \Pi_{\rm GI}\hat{\mathcal{O}}_r\Pi_{\rm GI} &= \Pi_{V_c}\sum_{\alpha\beta ijk\ell i'\ell'}\hat{\mathcal{O}}_{\alpha\beta ijk\ell}\frac{1}{d_\beta}\left|\alpha ik\right\rangle_r\left\langle\beta j\ell'\right|\otimes\left|\overline{\beta}i'\ell\right\rangle_{\overline{r}}\left\langle\overline{\beta}i'\ell'\right| \\ &= \sum_{\alpha\beta ijk\ell i'k'\ell'}\hat{\mathcal{O}}_{\alpha\beta ijk\ell}\frac{1}{d_\alpha d_\beta}\delta_{\alpha\beta}\delta_{k\ell}\left|\alpha ik'\right\rangle_r\left\langle\beta j\ell'\right|\otimes\left|\overline{\beta}i'k'\right\rangle_{\overline{r}}\left\langle\overline{\beta}i'\ell'\right| \\ &= \sum_{\alpha ijki'k'}\hat{\mathcal{O}}_{\alpha\alpha ijk'k'}\frac{1}{d_\alpha^2}\left|\alpha ik\right\rangle_r\left\langle\alpha j\ell\right|\otimes\left|\overline{\alpha}i'k\right\rangle_{\overline{r}}\left\langle\overline{\alpha}i'\ell\right|, \end{aligned} \tag{5.62}$$

which is identical to $\widetilde{\mathcal{O}}_r$ in (5.57) under $\widetilde{\mathcal{O}}_{\alpha ij} = \sum_{k'}\hat{\mathcal{O}}_{\alpha\alpha ijk'k'}/d_\alpha^2$, and thus belongs to $\widetilde{\mathcal{A}}_r^{(3)}$. Combining the above results, we conclude $\widetilde{\mathcal{A}}_r = \widetilde{\mathcal{A}}_r^{(3)}$.

After all of this machinery, it is clear that one is justified in writing the algebra $\widetilde{\mathcal{A}}_r$ of gauge-invariant operators on a subregion r as a restriction via $\Pi_{\rm GI}$ of the pre-gauged algebra $\mathcal{A}_r$ on r. Crucially, however, $\widetilde{\mathcal{A}}_r$ is *not* a subalgebra of $\mathcal{A}_r$, as is obvious from the nontrivial action of (5.57) on $\hat{\mathcal{H}}_{\overline{r}}$. This is manifest from the fact that $\Pi_{\rm GI}$ is an element of $\mathcal{A}$, not of $\mathcal{A}_r$, and so the projection takes one out of the pre-gauged subregion algebra $\mathcal{A}_r$.

5.3.3 The Center of the Algebra

For spatial subregions the following inclusion is obvious:

$$\mathcal{A}_{\overline{r}} \subseteq (\mathcal{A}_r)', \tag{5.63}$$

as causally disconnected operators must commute. Here $(\mathcal{A}_r)'$ denotes the commutant of $\mathcal{A}_r$. Haag duality is the saturation of the above bound:

$$\mathcal{A}_{\overline{r}} = (\mathcal{A}_r)', \tag{5.64}$$

that is, the commutant of the algebra of operators in a subregion is equal to the algebra of operators in the complement region.[12]

In our model, Haag duality certainly holds for the pre-gauged algebras, but does it also hold for the gauge-invariant algebras? We will now show that it does, i.e.,

$$\widetilde{\mathcal{A}}_{\overline{r}} = \left(\widetilde{\mathcal{A}}_r\right)'. \tag{5.65}$$

Proposition 6. The Hilbert space isomorphism (5.39) induces the following isomorphisms between algebras:

$$\widetilde{\mathcal{A}}_r \cong \bigoplus_\alpha \widetilde{\mathcal{A}}_r^\alpha \otimes \mathbb{1}_{\overline{r}}^{\overline{\alpha}}, \qquad \widetilde{\mathcal{A}}_{\overline{r}} \cong \bigoplus_\alpha \mathbb{1}_r^\alpha \otimes \widetilde{\mathcal{A}}_{\overline{r}}^{\overline{\alpha}}, \tag{5.66}$$

where $\widetilde{\mathcal{A}}_r^\alpha \equiv \mathcal{B}(\widetilde{\mathcal{H}}_r^\alpha)$, the algebra of bounded operators on $\widetilde{\mathcal{H}}_r^\alpha$, and similarly we define $\widetilde{\mathcal{A}}_{\overline{r}}^{\overline{\alpha}} \equiv \mathcal{B}(\widetilde{\mathcal{H}}_{\overline{r}}^{\overline{\alpha}})$. Moreover, $\mathbb{1}_r^\alpha$, $\mathbb{1}_{\overline{r}}^{\overline{\alpha}}$ denote the identity operators on $\widetilde{\mathcal{H}}_r^\alpha$, $\widetilde{\mathcal{H}}_{\overline{r}}^{\overline{\alpha}}$, respectively.

Proof: Recall from (5.39) that $\widetilde{\mathcal{H}}$ is isomorphic to a direct sum of factorizing Hilbert spaces:

$$\widetilde{\mathcal{H}} \cong \bigoplus_\alpha \widetilde{\mathcal{H}}_r^\alpha \otimes \widetilde{\mathcal{H}}_{\overline{r}}^{\overline{\alpha}}; \tag{5.67}$$

where the two sides are identified under the natural isomorphism (5.40), which we reproduce here:

$$|\widetilde{\psi}\rangle = \sum_{\alpha ijk} \widetilde{\psi}_{\alpha ij} \, |\alpha i k\rangle_r \, |\overline{\alpha} j k\rangle_{\overline{r}} \quad \to \quad \sum_{\alpha ij} \sqrt{d_\alpha} \widetilde{\psi}_{\alpha ij} \, |\alpha i\rangle_r \, |\overline{\alpha} j\rangle_{\overline{r}}. \tag{5.68}$$

We now apply this isomorphism to our algebra $\widetilde{\mathcal{A}}_r$. Consider a general element of $\widetilde{\mathcal{A}}_r$

[12] There are counterexamples to Haag duality in quantum field theories with global or gauge symmetries; see for example [181].

defined via (5.57). Under (5.68), this element becomes

$$\sum_{\alpha ii' jk\ell} \widetilde{\mathcal{O}}_{\alpha ij} \left|\alpha ik\right\rangle_r \left\langle \alpha j\ell\right| \otimes \left|\overline{\alpha} i' k\right\rangle_{\overline{r}} \left\langle \overline{\alpha} i'\ell\right| \to \sum_{\alpha ii' j} d_\alpha \widetilde{\mathcal{O}}_{\alpha ij} \left|\alpha i\right\rangle_r \left\langle \alpha j\right| \otimes \left|\overline{\alpha} i'\right\rangle_{\overline{r}} \left\langle \overline{\alpha} i'\right| = \sum_{\alpha ij} d_\alpha \widetilde{\mathcal{O}}_{\alpha ij} \left|\alpha i\right\rangle_r \left\langle \alpha j\right| \otimes \mathbb{1}_{\overline{r}}^{\overline{\alpha}}, \tag{5.69}$$

which is an element of $\widetilde{\mathcal{A}}_r^\alpha \otimes \mathbb{1}_{\overline{r}}^{\overline{\alpha}}$. Thus, we have demonstrated the isomorphism for $\widetilde{\mathcal{A}}_r$ in (5.66). The isomorphism for $\widetilde{\mathcal{A}}_{\overline{r}}$ follows from a similar argument.

Corollary 7. $\widetilde{\mathcal{A}}_r$ obeys Haag duality, such that $\left(\widetilde{\mathcal{A}}_r\right)' = \widetilde{\mathcal{A}}_{\overline{r}}$, where the commutant is defined with respect to the full gauge-invariant algebra $\widetilde{\mathcal{A}}$. *Proof:* This immediately follows from the algebra isomorphisms (5.66) and

$$\left(\bigoplus_\alpha \widetilde{\mathcal{A}}_r^\alpha \otimes \mathbb{1}_{\overline{r}}^{\overline{\alpha}}\right)' = \bigoplus_\alpha \left(\widetilde{\mathcal{A}}_r^\alpha \otimes \mathbb{1}_{\overline{r}}^{\overline{\alpha}}\right)' = \bigoplus_\alpha \mathbb{1}_r^\alpha \otimes \widetilde{\mathcal{A}}_{\overline{r}}^{\overline{\alpha}}. \tag{5.70}$$

The center of an algebra is defined to be the intersection of the algebra with its commutant. As our gauge-invariant subalgebra $\widetilde{\mathcal{A}}_r$ obeys Haag duality, the center is

$$\widetilde{\mathcal{Z}}_r = \widetilde{\mathcal{A}}_r \cap \widetilde{\mathcal{A}}_r' = \widetilde{\mathcal{A}}_r \cap \widetilde{\mathcal{A}}_{\overline{r}}. \tag{5.71}$$

Lemma 8. The center $\widetilde{\mathcal{Z}}_r$ is

$$\widetilde{\mathcal{Z}}_r = \left\{ z_\alpha \widetilde{P}^\alpha : z_\alpha \in \mathbb{C} \right\}, \tag{5.72}$$

where $\widetilde{P}^\alpha$ are mutually orthogonal projections defined via

$$\widetilde{P}^\alpha = \frac{1}{d_\alpha} \sum_{ijk\ell} \left|\alpha ik\right\rangle_r \left\langle \alpha i\ell\right| \otimes \left|\overline{\alpha} jk\right\rangle_{\overline{r}} \left\langle \overline{\alpha} j\ell\right|. \tag{5.73}$$

Proof: Under the algebra isomorphisms (5.66) for $\widetilde{\mathcal{A}}_r$ and $\widetilde{\mathcal{A}}_{\overline{r}}$, we can immediately identify the center as

$$\widetilde{\mathcal{Z}}_r \cong \bigoplus_\alpha \mathbb{C}\left(\mathbb{1}_r^\alpha \otimes \mathbb{1}_{\overline{r}}^{\overline{\alpha}}\right). \tag{5.74}$$

That is, the center $\widetilde{\mathcal{Z}}_r$ is the direct sum of complex multiples of the identity within each superselection sector α. We can write the identity in a superselection sector as

$$\mathbb{1}_r^\alpha \otimes \mathbb{1}_{\overline{r}}^{\overline{\alpha}} = \sum_{ij} |\alpha i\rangle_r \langle \alpha i| \otimes |\overline{\alpha} j\rangle_{\overline{r}} \langle \overline{\alpha} j|, \tag{5.75}$$

and examine the pullback of these operators under the natural isomorphism (5.68) to find the corresponding operators in $\widetilde{\mathcal{A}}$. We obtain

$$\mathbb{1}_r^\alpha \otimes \mathbb{1}_{\overline{r}}^{\overline{\alpha}} \quad \Rightarrow \quad \frac{1}{d_\alpha} \sum_{ijk\ell} |\alpha ik\rangle_r \langle \alpha i\ell| \otimes |\overline{\alpha} jk\rangle_{\overline{r}} \langle \overline{\alpha} j\ell| = \widetilde{P}^\alpha. \tag{5.76}$$

We identify these operators $\widetilde{P}^\alpha$ as the (properly normalized) projections onto the α superselection sector, where we remind the reader that α is an irreducible representation of G^n. These operators can alternatively be written as

$$\widetilde{P}^\alpha = \left(\hat{P}_r^\alpha \otimes \mathbb{1}_{\overline{r}}\right) \Pi_{\text{GI}} = \left(\mathbb{1}_r \otimes \hat{P}_{\overline{r}}^{\overline{\alpha}}\right) \Pi_{\text{GI}}, \tag{5.77}$$

where $\hat{P}_r^\alpha$ and $\hat{P}_{\overline{r}}^{\overline{\alpha}}$ are orthogonal projections in $\hat{\mathcal{H}}_r$ and $\hat{\mathcal{H}}_{\overline{r}}$, respectively:

$$\hat{P}_r^\alpha = \sum_{ik} |\alpha ik\rangle_r \langle \alpha ik|, \quad \hat{P}_{\overline{r}}^{\overline{\alpha}} = \sum_{ik} |\overline{\alpha} ik\rangle_{\overline{r}} \langle \overline{\alpha} ik|. \tag{5.78}$$

One can show the $\widetilde{P}^\alpha$ are orthogonal and idempotent such that $\widetilde{P}^\alpha \widetilde{P}^\beta = \delta_{\alpha\beta} \widetilde{P}^\alpha$.

5.3.4 Traces in $\mathcal{A}_r$ and $\widetilde{\mathcal{A}}_r$

We now define traces in our von Neumann algebras. When an algebra is $\mathcal{B}(\mathcal{H})$ for some Hilbert space $\mathcal{H}$, we can simply identify the minimal projections as projections onto a pure state in $\mathcal{H}$, and the trace is the usual trace of a square matrix. Our algebras are not always of this form; an example is $\widetilde{\mathcal{A}}_r$. Therefore, we will first identify the minimal projections, which are then used to define a normalized trace on the algebra. In particular, for $\widetilde{\mathcal{A}}_r$ our task is to find the minimal projections $\widetilde{P}_r$ in $\widetilde{\mathcal{A}}_r$ and use them to define a "rescaled" trace $\widetilde{\mathrm{Tr}}_r$ which satisfies

$$\widetilde{\mathrm{Tr}}_r \widetilde{P}_r = 1. \tag{5.79}$$

Let us first discuss the case that we understand well: that of $\mathcal{A}_r$, and by extension $\hat{\mathcal{A}}_r$. As $\mathcal{A}_r = \mathcal{B}(\mathcal{H}_r)$, the minimal projections are projections onto a pure state in $\mathcal{H}_r$, and we define the trace Tr_r in $\mathcal{A}_r$ such that the minimal projections have trace 1. As $\hat{\mathcal{A}}_r = \mathcal{B}(\hat{\mathcal{H}}_r)$, we proceed similarly. Recall that the basis states of $\hat{\mathcal{H}}_r$ are $\{|\alpha i k\rangle_r\}$, and so we define the trace $\hat{\mathrm{Tr}}_r$ in $\hat{\mathcal{A}}_r$ via

$$\hat{\mathrm{Tr}}_r \, |\alpha i k\rangle_r \langle \alpha i k| = 1. \tag{5.80}$$

As the minimal projections in $\hat{\mathcal{A}}_r$ are also minimal projections in $\mathcal{A}_r$, the two traces agree (on $\hat{\mathcal{A}}_r$):

$$\mathrm{Tr}_r = \hat{\mathrm{Tr}}_r, \tag{5.81}$$

so we will use only Tr_r (not $\hat{\mathrm{Tr}}_r$) moving forward.

Now consider $\widetilde{\mathcal{A}}_r$. Although $\widetilde{\mathcal{A}}_r$ is not the algebra of all bounded operators on a Hilbert space, the algebra isomorphism (5.66) shows that we can write it as a direct sum of

algebras for which we can easily identify minimal projections. In particular, the pullback of minimal projections onto pure states $|\alpha i\rangle_r \in \widetilde{\mathcal{H}}_r^\alpha$ under the natural isomorphism (5.68) gives minimal projections in $\widetilde{\mathcal{A}}_r$. Thus, we write these minimal projections $\widetilde{P}_r^{\alpha i} \in \widetilde{\mathcal{A}}_r$ as

$$\widetilde{P}_r^{\alpha i} = \frac{1}{d_\alpha} \sum_{jk\ell} |\alpha i k\rangle_r \langle \alpha i \ell| \otimes |\overline{\alpha} j k\rangle_{\overline{r}} \langle \overline{\alpha} j \ell| \tag{5.82}$$

for all nonempty sectors α, defined as those with nonzero n_α, $\overline{n}_{\overline{\alpha}}$. If n_α vanishes, the index i above has an empty range, and if $\overline{n}_{\overline{\alpha}}$ vanishes, $\widetilde{P}_r^{\alpha i}$ vanishes due to the empty sum over j in (5.82).

We can alternatively write $\widetilde{P}_r^{\alpha i}$ as

$$\widetilde{P}_r^{\alpha i} = \hat{P}_r^{\alpha i} \Pi_{\mathrm{GI}}, \tag{5.83}$$

where the projections $\hat{P}_r^{\alpha i}$ are defined similarly to (5.78):

$$\hat{P}_r^{\alpha i} \equiv \sum_k |\alpha i k\rangle \langle \alpha i k|_r \otimes \mathbb{1}_{\overline{r}}. \tag{5.84}$$

Although we already argued that $\widetilde{P}_r^{\alpha i}$ are minimal projections using the natural isomorphism, we now show it more directly.

Lemma 9. The projections $\widetilde{P}_r^{\alpha i}$ (for nonempty sectors α) are minimal projections in $\widetilde{\mathcal{A}}_r$.

Proof: We recall that minimal projections are nonzero and have the property that any subprojection $\widetilde{Q}_r$ of $\widetilde{P}_r^{\alpha i}$ is either zero or $\widetilde{P}_r^{\alpha i}$. As an element of $\widetilde{\mathcal{A}}_r$, $\widetilde{Q}_r$ must be of the form

$$\widetilde{Q}_r = \sum_{\beta j j' k} \hat{Q}_{\beta j j'} \left(|\beta j k\rangle_r \langle \beta j' k| \otimes \mathbb{1}_{\overline{r}} \right) \Pi_{\mathrm{GI}} \tag{5.85}$$

with complex coefficients $\hat{Q}_{\beta j j'}$. The subprojection $\widetilde{Q}_r$ is left fixed under conjugation via

$\widetilde{P}_r^{\alpha i}$, so we have

$$\widetilde{Q}_r = \widetilde{P}_r^{\alpha i}\widetilde{Q}_r\widetilde{P}_r^{\alpha i} = \sum_k \hat{Q}_{\alpha ii}\left(|\alpha ik\rangle_r \langle \alpha ik| \otimes \mathbb{1}_{\bar{r}}\right)\Pi_{\mathrm{GI}} = \hat{Q}_{\alpha ii}\widetilde{P}_r^{\alpha i}. \tag{5.86}$$

Additionally imposing $\widetilde{Q}_r^2 = \widetilde{Q}_r$, we find

$$\widetilde{Q}_r^2 = \hat{Q}_{\alpha ii}^2\widetilde{P}_r^{\alpha i} = \hat{Q}_{\alpha ii}\widetilde{P}_r^{\alpha i}. \tag{5.87}$$

Unless $\widetilde{Q}_r$ is zero, we obtain $\hat{Q}_{\alpha ii} = 1$ and thus $\widetilde{Q}_r = \widetilde{P}_r^{\alpha i}$. So $\widetilde{P}_r^{\alpha i}$ (for a nonempty sector α) is indeed a minimal projection.

Therefore, we define the trace $\widetilde{\mathrm{Tr}}_r$ in $\widetilde{\mathcal{A}}_r$ by imposing

$$\widetilde{\mathrm{Tr}}_r\widetilde{P}_r^{\alpha i} = 1 \tag{5.88}$$

for every nonempty sector α and every $i = 1, \cdots, n_\alpha$.

How do we understand this trace acting on a general operator in $\widetilde{\mathcal{A}}_r$? Such an operator can always be written in the form (5.56):

$$\widetilde{\mathcal{O}}_r = \hat{\mathcal{O}}_r\Pi_{\mathrm{GI}}, \quad \hat{\mathcal{O}}_r = \sum_{\alpha ijk}\hat{\mathcal{O}}_{\alpha ij}\,|\alpha ik\rangle_r \langle \alpha jk| \otimes \mathbb{1}_{\bar{r}}. \tag{5.89}$$

Taking the trace using $\widetilde{\mathrm{Tr}}_r$, we find

$$\widetilde{\mathrm{Tr}}_r\widetilde{\mathcal{O}}_r = \widetilde{\mathrm{Tr}}_r\sum_{\alpha i}\hat{\mathcal{O}}_{\alpha ii}\widetilde{P}_r^{\alpha i} = \sum_{\alpha i}\hat{\mathcal{O}}_{\alpha ii}. \tag{5.90}$$

If we were to take the trace Tr_r of the corresponding $\hat{\mathcal{O}}_r$, we would instead find

$$\mathrm{Tr}_r\, \hat{\mathcal{O}}_r = \mathrm{Tr} \sum_{\alpha i} \hat{\mathcal{O}}_{\alpha i i} \hat{P}_r^{\alpha i} = \sum_{\alpha i} d_\alpha \hat{\mathcal{O}}_{\alpha i i}. \tag{5.91}$$

Thus, it is tempting to relate the trace $\widetilde{\mathrm{Tr}}_r$ to Tr_r using an appropriate rescaling by $1/d_\alpha$ in each sector. A more precise version of this statement is the following: for any operator $\widetilde{\mathcal{O}}_r^\alpha \in \widetilde{\mathcal{A}}_r$ that acts only in the α sector such that it can be written as

$$\widetilde{\mathcal{O}}_r^\alpha = \hat{\mathcal{O}}_r^\alpha \Pi_{\mathrm{GI}}, \quad \hat{\mathcal{O}}_r^\alpha = \sum_{ijk} \hat{\mathcal{O}}_{\alpha ij} \left|\alpha i k\right\rangle_r \left\langle \alpha j k\right| \otimes \mathbb{1}_{\bar{r}}, \tag{5.92}$$

i.e., with no sum over α, the two traces are related by

$$\widetilde{\mathrm{Tr}}_r \widetilde{\mathcal{O}}_r^\alpha = \frac{1}{d_\alpha} \mathrm{Tr}_r\, \hat{\mathcal{O}}_r^\alpha. \tag{5.93}$$

Summing both sides over α recovers (5.90) and (5.91).

5.3.5 Reduced States

Our ultimate goal is to relate the von Neumann entropies for the same gauge-invariant state $\widetilde{\rho}$ in $\widetilde{\mathcal{H}}$ on two different subalgebras: $\mathcal{A}_r$ and $\widetilde{\mathcal{A}}_r$.

The first thing to note is that, when we consider the full graph (instead of restricting to a subregion r), we have $\mathcal{A} = \mathcal{B}(\mathcal{H})$ and $\widetilde{\mathcal{A}} = \mathcal{B}(\widetilde{\mathcal{H}})$ where $\widetilde{\mathcal{H}}$ is a subspace of $\mathcal{H}$, so minimal projections in $\widetilde{\mathcal{A}}$ are also minimal projections in $\mathcal{A}$, and the trace $\widetilde{\mathrm{Tr}}$ in $\widetilde{\mathcal{A}}$ therefore agrees with the trace Tr in $\mathcal{A}$ when acting on gauge-invariant states. Hence, a gauge-invariant state $\widetilde{\rho}$ on the full graph that is properly normalized under the $\widetilde{\mathrm{Tr}}$ trace is also properly normalized under the Tr trace, and can therefore be viewed as a properly normalized state $\rho = \widetilde{\rho}$ in $\mathcal{H}$ (albeit a special one). Thus, we will use only ρ (not $\widetilde{\rho}$) for

notational simplicity in the following discussions. We should still remember that ρ is a special state that belongs to $\widetilde{\mathcal{A}}$.

The above statements do not hold for reduced states on subregions. In particular, we need to distinguish a properly normalized state ρ_r in $\mathcal{A}_r$ from a properly normalized $\widetilde{\rho}_r$ in $\widetilde{\mathcal{A}}_r$. Now we derive the relation between these two states.

Recall that to find $S(\rho, \mathcal{A}_r)$ for a general subalgebra $\mathcal{A}_r \subset \mathcal{A}$, we need to find a reduced state $\rho_r \in \mathcal{A}_r$ satisfying

$$\mathrm{Tr}_r(\rho_r \mathcal{O}_r) = \mathrm{Tr}(\rho \mathcal{O}_r) \tag{5.94}$$

for all $\mathcal{O}_r \in \mathcal{A}_r$. For our particular $\mathcal{A}_r$ (the pre-gauged algebra on r), the answer is, of course, $\rho_r = \mathrm{Tr}_{\bar{r}}\, \rho$.

Now we work out the reduced state in the subalgebra $\widetilde{\mathcal{A}}_r$.

Lemma 10. The reduced state $\widetilde{\rho}_r \in \widetilde{\mathcal{A}}_r$ satisfying

$$\widetilde{\mathrm{Tr}}_r(\widetilde{\rho}_r \widetilde{\mathcal{O}}_r) = \mathrm{Tr}\Big(\rho \widetilde{\mathcal{O}}_r\Big) \tag{5.95}$$

for all $\widetilde{\mathcal{O}}_r \in \widetilde{\mathcal{A}}_r$ is of the form

$$\widetilde{\rho}_r = \hat{\rho}_r \Pi_{\mathrm{GI}}, \quad \hat{\rho}_r = \sum_{\alpha ijk} \hat{\rho}_{\alpha ij} \left|\alpha ik\right\rangle_r \left\langle \alpha jk\right| \otimes \mathbb{1}_{\bar{r}}, \tag{5.96}$$

with $\hat{\rho}_{\alpha ij} = d_\alpha \rho_{\alpha ij}$, where $\rho_{\alpha ij}$ is defined by

$$\rho_r = \sum_{\alpha ijk} \rho_{\alpha ij} \left|\alpha ik\right\rangle_r \left\langle \alpha jk\right|. \tag{5.97}$$

Proof:

A general gauge-invariant state $\rho \in \widetilde{\mathcal{A}}$ can be written as

$$\rho = \sum_{\alpha\beta ijki'j'k'} \rho_{\alpha\beta ii'jj'} \left|\alpha ik\right\rangle_r \left|\overline{\alpha} jk\right\rangle_{\overline{r}} \left\langle \beta i'k'\right|_r \left\langle \overline{\beta} j'k'\right|_{\overline{r}} \tag{5.98}$$

using the basis states for $\widetilde{\mathcal{H}}$. Tracing over $\overline{r}$, we find

$$\begin{aligned}\rho_r = \mathrm{Tr}_{\overline{r}}\, \rho &= \sum_{\alpha\beta ijki'j'k'} \rho_{\alpha\beta ii'jj'} \left\langle \overline{\beta} j'k' | \overline{\alpha} jk\right\rangle_{\overline{r}} \left|\alpha i'k'\right\rangle_r \left\langle \beta ik\right| \\ &= \sum_{\alpha ii'jk} \rho_{\alpha\alpha ii'jj} \left|\alpha ik\right\rangle_r \left\langle \alpha i'k\right| . \end{aligned} \tag{5.99}$$

This verifies (5.97) and determines ρ_{aij}.

Now recall that as an element of $\widetilde{\mathcal{A}}_r$, $\widetilde{\rho}_r$ must be of the form (5.96) with some complex coefficients $\hat{\rho}_{\alpha ij}$. It remains to determine what they are from (5.95). In order to impose it, we define the following basis for $\widetilde{\mathcal{A}}_r$:

$$\widetilde{\mathcal{O}}_r^{\alpha ij} = \hat{\mathcal{O}}_r^{\alpha ij} \Pi_{\mathrm{GI}}, \quad \hat{\mathcal{O}}_r^{\alpha ij} = \sum_k \left|\alpha ik\right\rangle_r \left\langle \alpha jk\right| \otimes \mathbb{1}_{\overline{r}}, \tag{5.100}$$

such that we can rewrite the reduced gauge-invariant density matrix as

$$\widetilde{\rho}_r = \sum_{\alpha ij} \hat{\rho}_{\alpha ij} \widetilde{\mathcal{O}}_{\alpha ij}. \tag{5.101}$$

Note that the basis elements $\widetilde{\mathcal{O}}_r^{\alpha ij}$ and their corresponding basis elements $\hat{\mathcal{O}}_r^{\alpha ij} \in \hat{\mathcal{A}}_r$ obey the following relations:

$$\widetilde{\mathcal{O}}_r^{\alpha ij} \widetilde{\mathcal{O}}_r^{\beta i'j'} = \delta_{\alpha\beta}\delta_{i'j} \widetilde{\mathcal{O}}_r^{\alpha ij'}, \quad \widetilde{\mathrm{Tr}}_r \widetilde{\mathcal{O}}_r^{\alpha ij} = \delta_{ij} \tag{5.102}$$

$$\hat{\mathcal{O}}_r^{\alpha ij} \hat{\mathcal{O}}_r^{\beta i'j'} = \delta_{\alpha\beta}\delta_{i'j} \hat{\mathcal{O}}_r^{\alpha ij'}, \quad \mathrm{Tr}_r\, \hat{\mathcal{O}}_r^{\alpha ij} = d_\alpha \delta_{ij}. \tag{5.103}$$

From these relations we can check both sides of (5.95) for $\widetilde{\mathcal{O}}_r$ set to one of the basis elements $\widetilde{\mathcal{O}}_r^{\alpha ij}$. The trace in the gauge-invariant algebra becomes

$$\widetilde{\mathrm{Tr}}_r\left(\widetilde{\rho}_r\widetilde{\mathcal{O}}_r^{\alpha ij}\right) = \widetilde{\mathrm{Tr}}_r\left(\sum_{\beta i'j'}\hat{\rho}_{\beta i'j'}\widetilde{\mathcal{O}}_r^{\beta i'j'}\widetilde{\mathcal{O}}^{\alpha ij}\right) = \sum_{\beta i'j'}\hat{\rho}_{\beta i'j'}\delta_{\alpha\beta}\delta_{i'j}\delta_{ij'} = \hat{\rho}_{\alpha ji}. \tag{5.104}$$

We need to equate this with the trace in the pre-gauged algebra, which we begin to evaluate by simplifying to the trace in $\mathcal{A}_r$. We have

$$\mathrm{Tr}\left(\rho\widetilde{\mathcal{O}}_r^{\alpha ij}\right) = \mathrm{Tr}\left(\rho\hat{\mathcal{O}}_r^{\alpha ij}\Pi_{\mathrm{GI}}\right) = \mathrm{Tr}\left(\Pi_{\mathrm{GI}}\rho\hat{\mathcal{O}}_r^{\alpha ij}\right) = \mathrm{Tr}\left(\rho\hat{\mathcal{O}}_r^{\alpha ij}\right) = \mathrm{Tr}_r(\rho_r\hat{\mathcal{O}}_r^{\alpha ij}), \tag{5.105}$$

where we have used the cyclicity of the trace, the gauge invariance of ρ, and the fact that $\hat{\mathcal{O}}_r^{\alpha ij} \in \mathcal{A}_r$. We further simplify this and obtain

$$\mathrm{Tr}_r(\rho_r\hat{\mathcal{O}}_r^{\alpha ij}) = \mathrm{Tr}_r\left(\sum_{\beta i'j'}\rho_{\beta i'j'}\hat{\mathcal{O}}_r^{\beta i'j'}\hat{\mathcal{O}}_r^{\alpha ij}\right) = \sum_{\beta i'j'}d_\alpha\rho_{\beta i'j'}\delta_{\alpha\beta}\delta_{i'j}\delta_{ij'} = d_\alpha\rho_{\alpha ji}. \tag{5.106}$$

Thus we identify the reduced density matrix $\widetilde{\rho}_r \in \widetilde{\mathcal{A}}_r$ as a density matrix of the form (5.96) with

$$\hat{\rho}_{\alpha ij} = d_\alpha\rho_{\alpha ij}. \tag{5.107}$$

5.4 Entropies in the Gauged Random Tensor Network

Having written down the reduced states in $\mathcal{A}_r$ and $\widetilde{\mathcal{A}}_r$, we are now ready to compute the von Neumann entropies with respect to the two algebras. As we will see, the difference between the two entropies in the gauged random tensor network is precisely accounted for by an additional contribution to the area operator in the nontrivial center $\widetilde{\mathcal{Z}}_r$.

5.4.1 Entanglement Entropy

From (5.96) and (5.97), we proceed by defining the reduced states projected onto a superselection sector α:

$$\rho_r^\alpha = \sum_{ijk} \rho_{\alpha ij} |\alpha ik\rangle_r \langle \alpha jk|, \qquad \hat{\rho}_r^\alpha = \sum_{ijk} \hat{\rho}_{\alpha ij} |\alpha ik\rangle_r \langle \alpha jk| = d_\alpha \rho_r^\alpha. \tag{5.108}$$

Note that these density matrices are not properly normalized with respect to their appropriate traces. The reduced states (5.96) and (5.97) can be written as a direct sum over representations:

$$\rho_r = \bigoplus_\alpha \rho_r^\alpha, \qquad \widetilde{\rho}_r = \bigoplus_\alpha \hat{\rho}_r^\alpha \Pi_{\text{GI}}. \tag{5.109}$$

Furthermore, functions of the reduced states are superselected in the same way. In particular,

$$\rho_r \log \rho_r = \bigoplus_\alpha \rho_r^\alpha \log \rho_r^\alpha, \qquad \widetilde{\rho}_r \log \widetilde{\rho}_r = \bigoplus_\alpha (\hat{\rho}_r^\alpha \log \hat{\rho}_r^\alpha) \Pi_{\text{GI}}, \tag{5.110}$$

where we used the fact that $[\hat{\rho}_r^\alpha, \Pi_{\text{GI}}] = 0$.

We are now ready to compute the subregion entropies (in the bulk). The von Neumann entropy of ρ with respect to $\mathcal{A}_r$ is simply given by

$$S(\rho, \mathcal{A}_r) = -\operatorname{Tr}_r \rho_r \log \rho_r = -\sum_\alpha \operatorname{Tr}_r \rho_r^\alpha \log \rho_r^\alpha. \tag{5.111}$$

On the other hand, using the relation between the traces (5.93), we can write the von Neumann entropy with respect to $\widetilde{\mathcal{A}}_r$ as

$$S(\rho, \widetilde{\mathcal{A}}_r) = -\widetilde{\operatorname{Tr}}_r \widetilde{\rho}_r \log \widetilde{\rho}_r = -\sum_\alpha d_\alpha^{-1} \operatorname{Tr}_r \hat{\rho}_r^\alpha \log \hat{\rho}_r^\alpha. \tag{5.112}$$

Using $\hat{\rho}_r^\alpha = d_\alpha \rho_r^\alpha$ and $\mathrm{Tr}_r\, \rho_r^\alpha = \widetilde{\mathrm{Tr}}_r \widetilde{\rho}_r^\alpha$, we can rewrite each term in the sum as

$$\begin{aligned} d_\alpha^{-1}\, \mathrm{Tr}_r\, \hat{\rho}_r^\alpha \log \hat{\rho}_r^\alpha &= \mathrm{Tr}_r\, \rho_r^\alpha \log \rho_r^\alpha + \mathrm{Tr}_r\, \rho_r^\alpha \log d_\alpha \\ &= \mathrm{Tr}_r\, \rho_r^\alpha \log \rho_r^\alpha + \widetilde{\mathrm{Tr}}_r \widetilde{\rho}_r^\alpha \log d_\alpha. \end{aligned} \tag{5.113}$$

The von Neumann entropy with respect to $\widetilde{\mathcal{A}}_r$ can thus be written as

$$\begin{aligned} S(\rho, \widetilde{\mathcal{A}}_r) &= -\sum_\alpha \left(\mathrm{Tr}_r\, \rho_r^\alpha \log \rho_r^\alpha + \widetilde{\mathrm{Tr}}_r \widetilde{\rho}_r^\alpha \log d_\alpha \right) \\ &= S(\rho, \mathcal{A}_r) - \widetilde{\mathrm{Tr}}_r \left(\widetilde{\rho}_r \Delta \widetilde{A} \right), \end{aligned} \tag{5.114}$$

where we have defined a new "extra area operator" via

$$\Delta \widetilde{A} \equiv \bigoplus_\alpha \widetilde{P}^\alpha \log d_\alpha. \tag{5.115}$$

The projections $\widetilde{P}^\alpha$ are precisely the projections (5.73) which generate the center $\widetilde{\mathcal{Z}}_r$, so $\Delta \widetilde{A}$ is manifestly an operator in the center.

We have now arrived at our final relation between the entropies with respect to $\mathcal{A}_r$ and $\widetilde{\mathcal{A}}_r$,

$$S(\rho, \mathcal{A}_r) = S(\rho, \widetilde{\mathcal{A}}_r) + \widetilde{\mathrm{Tr}}_r \left(\widetilde{\rho}_r \Delta \widetilde{A} \right), \tag{5.116}$$

which we now use in our two-layer gauged RTN defined in Section 5.2. In particular, we would like to derive an FLM formula relating the boundary entropy with the gauged bulk entropy $S(\rho, \widetilde{\mathcal{A}}_r)$. Recall that when we feed any bulk state ρ in the pre-gauged algebra $\mathcal{A}$ into the RTN, the entropy $S(R)$ of the resulting boundary state on a boundary subregion R satisfies an FLM formula:

$$S(R) = |\gamma_R| \log D + S(\rho, \mathcal{A}_r), \tag{5.117}$$

where the bulk subregion r is chosen to be the entanglement wedge between R and its minimal surface γ_R. Now specializing to a gauge-invariant bulk state $\rho \in \widetilde{\mathcal{A}}$ and using (5.116), we find that the boundary entropy can now be written as a new FLM formula:

$$S(R) = \widetilde{\mathrm{Tr}}_r\left(\widetilde{\rho}_r \widetilde{A}\right) + S\left(\rho, \widetilde{\mathcal{A}}_r\right), \tag{5.118}$$

where the full area operator $\widetilde{A}$ is

$$\widetilde{A} = |\gamma_R| \log D + \bigoplus_{\alpha} \widetilde{P}^{\alpha} \log d_{\alpha} = |\gamma_R| \log D + \bigoplus_{\alpha_1,\cdots,\alpha_n} \widetilde{P}^{(\alpha_1,\cdots,\alpha_n)} \sum_{i=1}^{n} \log d_{\alpha_i}. \tag{5.119}$$

Again, we sum over all irreps $\alpha = (\alpha_1, \cdots, \alpha_n)$ of G^n acting on the cut, although some α sectors may be emtpy (i.e., n_α or $\overline{n}_{\overline{\alpha}}$ is zero) in which case $\widetilde{P}^{\alpha}$ vanishes.

This is our main result. We note that this area operator looks like what arises in a superposition of a stack of standard RTNs with probabilities determined by the projections $\widetilde{P}^{\alpha}$ and with bond dimensions augmented by d_{α_i}.

5.4.2 Rényi Entropy and Rényi Mutual Information

As discussed in Section 5.2, one can modify the entanglement structure of the links in the standard RTN to obtain a nonflat Rényi spectrum for boundary states. However, this is not enough to reproduce the properties of holographic Rényi entropies on general boundary subregions. In particular, it fails to account for the lack of backreaction, displayed in the tensor network as a lack of (Rényi) correlation between disconnected boundary subregions when the RT surface is in a disconnected phase. This problem becomes clear when one calculates the Rényi mutual information between two such boundary

subregions R_1 and R_2, defined as[13]

$$I_n(R_1 : R_2) \equiv S_n(R_1) + S_n(R_2) - S_n(R_1 \cup R_2). \tag{5.120}$$

As the area operator in the original RTN is a c-number, using (5.3) we find that the area operator contribution cancels out in $I_n(R_1 : R_2)$ for all n (as long as the minimal surface γ_R is in a disconnected phase), leaving the boundary mutual information equal to the bulk mutual information:

$$I_n(R_1 : R_2) = I_n(r_1 : r_2, \mathcal{A}_{r_1 r_2}). \tag{5.121}$$

This implies that, if one wants a contribution to the Rényi mutual information of the same order as the area, that is $\mathcal{O}(\log D)$, one must input by hand a highly entangled bulk state. Doing this is unsatisfying and quite arbitrary.

We will now see that our gauged RTN solves this problem in a natural way, due to our nontrivial area operator. In general, the presence of a nontrivial area operator will lead to a nontrivial, n-dependent boundary Rényi mutual information, even for states with vanishing bulk Rényi mutual information.

To see how this is realized in the gauged RTN, we will study a simple example shown in Figure 5.3, where the top layer is disconnected but the bottom layer is connected.[14] We allow the bond dimensions in the bottom layer to be different for different links, and in fact design them so that the minimal surfaces associated with R_1, R_2, and their union $R_1 \cup R_2$ are fixed as we vary the Rényi index n at $\mathcal{O}(1)$ values. We will feed in a

[13]The Rényi index n should not be confused with the number of vertices on the cut $n = |V_c|$.

[14]This connection is unnecessary to prove our point, as the internal leg connecting r_1 and r_2 never contributes to the area term, but it is more intuitively satisfying to discuss a connected spatial slice for the purposes of demonstrating backreaction.

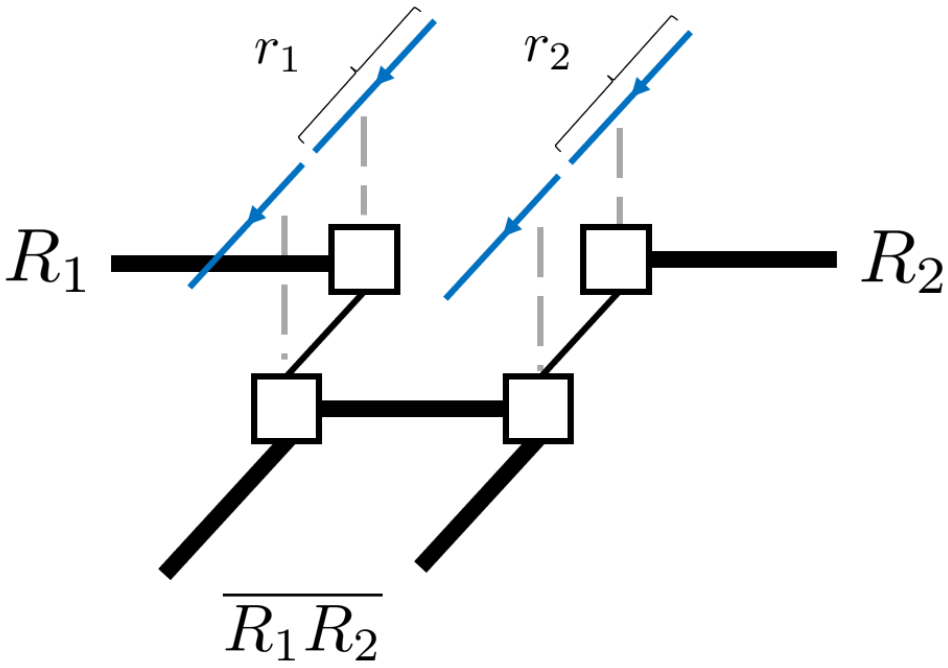


Figure 5.3: A simple gauged RTN in which we compute the Rényi mutual information between R_1 and R_2. The input from the top layer lives on four edges of a disconnected graph G, as we choose to have no matter on any of the vertices. In the bottom layer, the thick legs have a bond dimension much larger than that of the thin legs, such that the minimal surfaces for the three boundary regions R_1, R_2, and $R_1 \cup R_2$ only involve the light internal legs. Consequently, the associated bulk regions will be r_1, r_2, and $r_1 \cup r_2$.

gauge-invariant bulk state ρ with the following reduced state on $r_1 \cup r_2$:

$$\rho_{r_1 r_2} = \sum_{\alpha\beta} (d_\alpha d_\beta)^{-1} P(\alpha, \beta) \sum_{k\ell} |\alpha i k\rangle_{r_1} |\beta j \ell\rangle_{r_2} \langle \alpha i k|_{r_1} \langle \beta j \ell|_{r_2}, \tag{5.122}$$

for some particular choice of i, j. This state has classical correlations between r_1 and r_2 as described by a probability distribution $P(\alpha, \beta)$, but has no quantum correlations. For simplicity, we consider the following distribution $P(\alpha, \beta)$ that has support on only two superselection sectors α_1, α_2 on r_1 and only two sectors β_1, β_2 on r_2:

$$P(\alpha_1, \beta_1) = p, \quad P(\alpha_2, \beta_1) = P(\alpha_1, \beta_2) = p', \quad P(\alpha_2, \beta_2) = p'', \tag{5.123}$$

subject to the constraint $p + 2p' + p'' = 1$.

The Rényi entropy of ρ in the pre-gauged algebra $\mathcal{A}_{r_1 r_2}$ is defined as

$$S_n(\rho, \mathcal{A}_{r_1 r_2}) \equiv \frac{1}{1-n} \log \left(\mathrm{Tr}_{r_1 r_2} \rho_{r_1 r_2}^n \right). \tag{5.124}$$

Using our $\rho_{r_1 r_2}$, we find

$$\begin{aligned} S_n(\rho, \mathcal{A}_{r_1 r_2}) = \frac{1}{1-n} \log \Bigg(& d_{\alpha_1} d_{\beta_1} \left(\frac{p}{d_{\alpha_1} d_{\beta_1}} \right)^n + d_{\alpha_2} d_{\beta_1} \left(\frac{p'}{d_{\alpha_2} d_{\beta_1}} \right)^n \\ & + d_{\alpha_1} d_{\beta_2} \left(\frac{p'}{d_{\alpha_1} d_{\beta_2}} \right)^n + d_{\alpha_2} d_{\beta_2} \left(\frac{p''}{d_{\alpha_2} d_{\beta_2}} \right)^n \Bigg). \end{aligned} \tag{5.125}$$

We can also compute the reduced density matrices on r_1 and r_2, as well as their corresponding Rényi entropies in the pre-gauged algebra. We find the reduced density matrices

to be

$$\rho_{r_1} = \sum_{k=1}^{d_{\alpha_1}} d_{\alpha_1}^{-1}(p+p') \left|\alpha_1 ik\right\rangle_{r_1} \left\langle\alpha_1 ik\right| + \sum_{k'=1}^{d_{\alpha_2}} d_{\alpha_2}^{-1}(p'+p'') \left|\alpha_2 ik'\right\rangle_{r_1} \left\langle\alpha_2 ik'\right|,$$
$$\rho_{r_2} = \sum_{k=1}^{d_{\beta_1}} d_{\beta_1}^{-1}(p+p') \left|\beta_1 jk\right\rangle_{r_2} \left\langle\beta_1 jk\right| + \sum_{k'=1}^{d_{\beta_2}} d_{\beta_2}^{-1}(p'+p'') \left|\beta_2 jk'\right\rangle_{r_2} \left\langle\beta_2 jk'\right|, \tag{5.126}$$

and the bulk Rényi entropies are

$$S_n(\rho, \mathcal{A}_{r_1}) = \frac{1}{1-n} \log\left(d_{\alpha_1} \left(\frac{p+p'}{d_{\alpha_1}}\right)^n + d_{\alpha_2} \left(\frac{p'+p''}{d_{\alpha_2}}\right)^n \right)$$
$$S_n(\rho, \mathcal{A}_{r_2}) = \frac{1}{1-n} \log\left(d_{\beta_1} \left(\frac{p+p'}{d_{\beta_1}}\right)^n + d_{\beta_2} \left(\frac{p'+p''}{d_{\beta_2}}\right)^n \right). \tag{5.127}$$

In the gauge-invariant algebra, the dependence on irrep dimensions drops out and the Rényi entropies become purely Shannon terms:

$$S_n(\rho, \widetilde{\mathcal{A}}_{r_1}) = S_n(\rho, \widetilde{\mathcal{A}}_{r_2}) = \frac{1}{1-n} \log\left((p+p')^n + (p'+p'')^n \right)$$
$$S_n(\rho, \widetilde{\mathcal{A}}_{r_1 r_2}) = \frac{1}{1-n} \log\left(p^n + 2(p')^n + (p'')^n \right), \tag{5.128}$$

which we choose to be parametrically suppressed relative to the Rényi entropies in the pre-gauged algebra.

When the sum inside the logarithm is dominated by one term, we can approximate it using

$$\log\left(\textstyle\sum_i x_i\right) \approx \log\left(\max_i \{x_i\}\right). \tag{5.129}$$

To simplify our calculation, we will enter a parameter regime where all three (pre-gauged) Rényi entropies satisfy the approximation above and have phase transitions.

First consider $S_n(\rho, \mathcal{A}_{r_1})$. We take $d_{\alpha_1} > d_{\alpha_2}$. The two terms in the sum are equal at

some critical n_*, given by

$$\left(\frac{p+p'}{p'+p''}\right)^{n_*} = \left(\frac{d_{\alpha_1}}{d_{\alpha_2}}\right)^{n_*-1} \quad \Rightarrow \quad \frac{n_*}{n_*-1} = \frac{\log\left(\frac{d_{\alpha_1}}{d_{\alpha_2}}\right)}{\log\left(\frac{p+p'}{p'+p''}\right)}. \tag{5.130}$$

Thus, in order to have a phase transition at $n_* > 1$ we require

$$\log\left(\frac{d_{\alpha_1}}{d_{\alpha_2}}\right) > \log\left(\frac{p+p'}{p'+p''}\right). \tag{5.131}$$

The width of this transition is controlled by the corrections to (5.129). This depends on the curvature of $S_n(\rho, \mathcal{A}_{r_1})$ at n_*; explicitly we can diagnose this with the following quantity:

$$\frac{d^2}{dn^2}(1-n)S_n(\rho, \mathcal{A}_{r_1})\Big|_{n=n_*} = \frac{1}{4}\left(\log\frac{d_{\alpha_1}(p'+p'')}{d_{\alpha_2}(p+p')}\right)^2. \tag{5.132}$$

For fixed n_*, this quantity increases with increasing $d_{\alpha_1}/d_{\alpha_2}$, so we should make this ratio large for a sharp transition. A simple way to ensure the previous conditions is the following:

$$\frac{d_{\alpha_1}}{d_{\alpha_2}} \equiv q \gg 1, \quad p \gg p', \quad p' \gg p''. \tag{5.133}$$

Furthermore, we impose

$$\frac{d_{\alpha_1}}{d_{\alpha_2}} = \frac{d_{\beta_1}}{d_{\beta_2}} = q, \tag{5.134}$$

which forces the phase transitions in $S_n(\rho, \mathcal{A}_{r_1})$ and $S_n(\rho, \mathcal{A}_{r_2})$ to occur at the same critical n_*.

Now let us examine the phase transition in $S_n(\rho, \mathcal{A}_{r_1 r_2})$. In the limit of sharp transitions we have

$$S_n(\rho, \mathcal{A}_{r_1 r_2}) \approx \frac{1}{1-n}\log\left(\max\left\{\frac{p^n}{(d_{\alpha_1}d_{\beta_1})^{n-1}}, \frac{(p')^n}{(d_{\alpha_2}d_{\beta_1})^{n-1}}, \frac{(p')^n}{(d_{\alpha_1}d_{\beta_2})^{n-1}}, \frac{(p'')^n}{(d_{\alpha_2}d_{\beta_2})^{n-1}}\right\}\right). \tag{5.135}$$

For simplicity, we will choose

$$\frac{p}{p'} > \frac{p'}{p''} \gg 1. \tag{5.136}$$

In this case, we find that $S_n(\rho, \mathcal{A}_{r_1 r_2})$ has a phase transition occurring at a critical n_c determined by

$$\frac{n_c}{n_c - 1} = \frac{\log(q^2)}{\log\left(\frac{p}{p''}\right)} = \frac{\log(q^2)}{\log\left(\frac{p}{p'}\frac{p'}{p''}\right)} \tag{5.137}$$

which satisfies $1 < n_c < n_*$.

We now combine the above results to find the (pre-gauged) Rényi mutual information

$$I_n(r_1 : r_2, \mathcal{A}_{r_1 r_2}) \equiv S_n(\rho, \mathcal{A}_{r_1}) + S_n(\rho, \mathcal{A}_{r_2}) - S_n(\rho, \mathcal{A}_{r_1 r_2}). \tag{5.138}$$

We find the following phases:

$$I_n(r_1 : r_2, \mathcal{A}_{r_1 r_2}) \approx \begin{cases} 0 & n < n_c, \\ \log(q^2) + \frac{n}{1-n} \log\left(\frac{(p+p')^2}{p''}\right) & n_c < n < n_*, \\ \frac{n}{1-n} \log\left(\frac{(p'+p'')^2}{p''}\right) & n_* < n. \end{cases} \tag{5.139}$$

Now we rewrite the boundary Rényi mutual information (5.121) as

$$S_n(R_1 : R_2) = \underbrace{I_n(r_1 : r_2, \mathcal{A}_{r_1 r_2}) - I_n(r_1 : r_2, \widetilde{\mathcal{A}}_{r_1 r_2})}_{\text{area contribution}} + \underbrace{I_n(r_1 : r_2, \widetilde{\mathcal{A}}_{r_1 r_2})}_{\text{bulk matter contribution}}, \tag{5.140}$$

where the contribution of the nontrivial area operator to the boundary Rényi mutual information is identified with the difference of the bulk Rényi mutual information in the two algebras. As stated previously, $I_n(r_1 : r_2, \widetilde{\mathcal{A}}_{r_1 r_2})$ is suppressed relative to $I_n(r_1 : r_2, \mathcal{A}_{r_1 r_2})$, so this model implements phase transitions in the boundary Rényi mutual information without a large bulk matter contribution (in the gauge-invariant algebra).

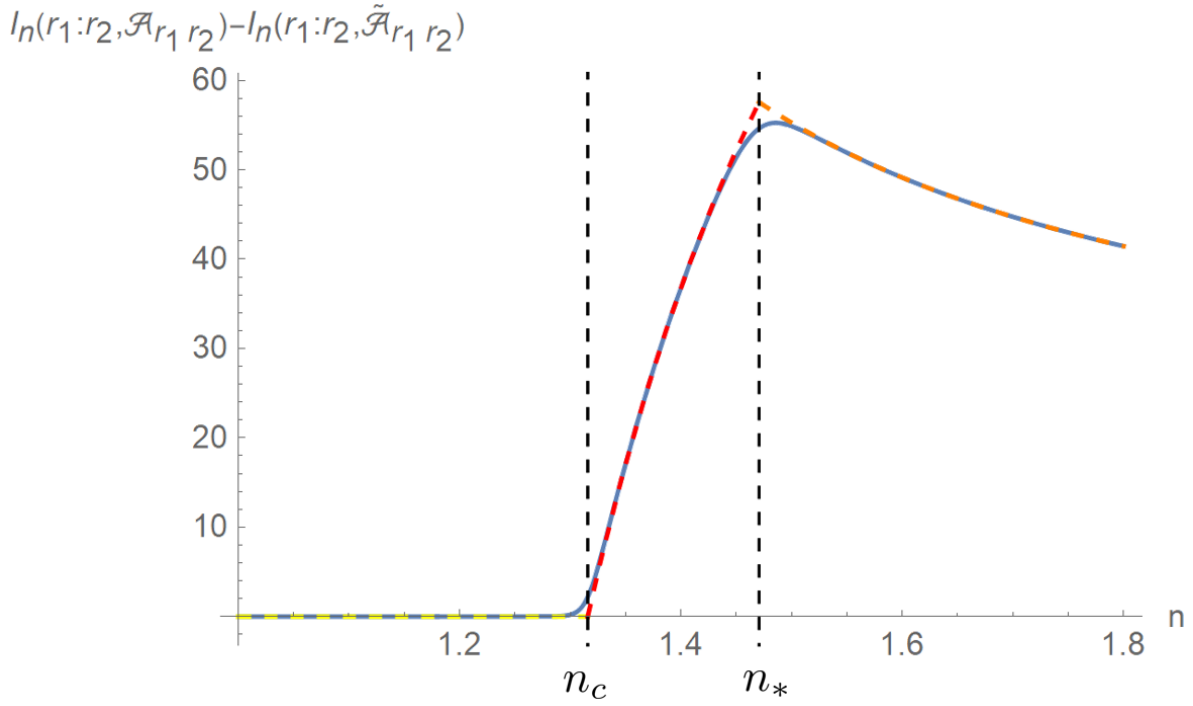


Figure 5.4: Phase transitions in the Rényi mutual information. Here we set $q = 10^{50}$, $p' = 10^{-16}$, and $p'' = 10^{-24}$. We plot the dominant contribution to the Rényi mutual information in the three phases (dashed) as well as the fully analytic interpolating function (solid).

We plot these two phase transitions for an example in Figure 5.4.

This is a proof of concept showing that adding bulk gauge symmetries to the RTN in this manner allows the boundary Rényi mutual information to be nontrivial and n-dependent, even for states with small bulk Rényi mutual information (in the gauge-invariant algebra). In our simple example here, the minimal surface does not shift—i.e. it is the same for all n—but there is no obstruction to writing a more complicated example in which the location of the minimal surface changes with n due to the nontrivial area operator.

5.5 Discussion and Outlook

In this work, we have presented a modification of the random tensor network which allows us to reproduce known features of semiclassical holographic states. We discuss some open questions and possible future directions below.

We have presented a toy model which, for simple choices of bulk input state, exhibits sharp phase transitions in the Rényi entropy and Rényi mutual information. With a sufficiently tuned set of probabilities and irrep dimensions, one could engineer a smooth varying Rényi entropy that matches with, for example, the correct one-interval CFT_2 Rényi entropy (5.5). It would be an even more complicated task to reproduce the correct Rényi entropy for multiple intervals in the CFT [182, 183].

The bulk algebras that we encountered in our model are type I von Neumann algebras. This is in contrast to the type II von Neumann algebras for gravity constructed using the crossed product [49, 51, 184]. A "type I approximation" to the crossed product was recently studied in [185]. It is thus tempting to incorporate the crossed product and the resultant birth of a type II algebra into the tensor network toy models of holography.

Our gauge-invariant subregion algebras generally have nontrivial centers. On the other hand, a prescription was given in [186] to construct gauge-invariant subregion algebras with trivial centers in lattice gauge theory. This prescription involves adding operators to the algebra that we do not include, so it does not contradict our results in any way.

Here we have implemented a graph version of the lattice gauge theory construction along the lines of Kogut and Susskind, but crucially without dynamics, due to the lack of a Hamiltonian. Because of this, our construction does not have anything more to say about time evolution in tensor networks than previous models. It would be interesting to understand how to incorporate a Hamiltonian and the associated time evolution into

tensor networks. It would also be interesting to study the commutators of intersecting area operators in our gauged RTN, which in standard AdS/CFT do not commute [187].

Appendix A

Geodesic Length in Kruskal-like Coordinates

We start with AdS_3 in Rindler coordinates. The metric in the embedding space $\mathbb{R}^{2,2}$ is

$$ds^2 = -dT_1^2 - dT_2^2 + dX_1^2 + dX_2^2. \tag{A.1}$$

The usual Rindler patch has a horizon at $r = r_+$, but to describe a rotating black hole solution we requires two horizons r_+ and r_-. The embedding functions for this geometry are given by

$$\begin{aligned}
T_1 &= \ell\sqrt{\alpha}\cosh\left(\frac{r_+\phi}{\ell} - \frac{r_-t}{\ell^2}\right) \\
X_1 &= \ell\sqrt{\alpha}\sinh\left(\frac{r_+\phi}{\ell} - \frac{r_-t}{\ell^2}\right) \\
T_2 &= \ell\sqrt{\alpha-1}\sinh\left(\frac{r_+t}{\ell^2} - \frac{r_-\phi}{\ell}\right) \\
X_2 &= \ell\sqrt{\alpha-1}\cosh\left(\frac{r_+t}{\ell^2} - \frac{r_-\phi}{\ell}\right)
\end{aligned} \tag{A.2}$$

for $\alpha \equiv \frac{r^2-r_-^2}{r_+^2-r_-^2}$. For this problem, we would like a coordinate reparametrization that makes the null directions manifest, so we use the coordinates (U, V, ϕ), where U and V are defined to be

$$\begin{aligned} U &= \exp\left[\kappa(t + r_*)\right] \\ V &= \exp\left[-\kappa(t - r_*)\right] \end{aligned} \tag{A.3}$$

where $\kappa = \frac{r_+^2-r_-^2}{\ell^2 r_+}$ and $r_* = \frac{\ell^2}{2\kappa}\left(\frac{\sqrt{r_+^2-r_-^2}-\sqrt{r^2-r_-^2}}{\sqrt{r_+^2-r_-^2}+\sqrt{r^2-r_-^2}}\right)$. In our new Kruskal-like coordinates, and using the corotating coordinates $\phi \to \phi - \frac{r_-}{\ell r_+}t$, the AdS_3 embedding functions become

$$\begin{aligned} T_1 &= \ell\left(\frac{U+V}{1+UV}\cosh\frac{r_-\phi}{\ell} - \frac{U-V}{1+UV}\sinh\frac{r_-\phi}{\ell}\right) \\ X_1 &= \ell\left(\frac{U-V}{1+UV}\cosh\frac{r_-\phi}{\ell} - \frac{U+V}{1+UV}\sinh\frac{r_-\phi}{\ell}\right) \\ T_2 &= \ell\frac{1-UV}{1+UV}\cosh\frac{r_+\phi}{\ell} \\ X_2 &= \ell\frac{1-UV}{1+UV}\sinh\frac{r_+\phi}{\ell}. \end{aligned} \tag{A.4}$$

The metric that results from these embedding functions is

$$ds^2 = \frac{1}{(1+UV)^2}\left(-4\ell^2 dU dV + 4\ell r_-(U dV - V dU) + (r_+^2(1-UV)^2 + 4UVr_-^2)d\phi^2\right). \tag{A.5}$$

Under an identification $\phi \sim \phi + 2\pi$, this metric becomes the BTZ metric with $r_+^2 + r_-^2 = \ell^2 M$.

With our embedding functions (A.4), one can also derive an expression for the geodesic distance s in our AdS_3 Kruskal-like coordinates using the geodesic distance in the em-

bedding spacetime. The result is [188]

$$\begin{aligned} s(x,x') &= \ell \cos^{-1}\sigma(x,x') && s \text{ timelike} \\ s(x,x') &= \ell \cosh^{-1}\sigma(x,x') && s \text{ spacelike} \end{aligned} \tag{A.6}$$

The geodesic we will be concerned with for our calculations is the spacelike case. This $\sigma(x,x')$ is the same quantity that appears in the calculation of the scalar Green's function and is given by

$$\sigma(x,x') = \frac{1}{\ell^2}\left[T_1(x)T_1(x') + T_2(x)T_2(x') - X_1(x)X_1(x') + X_2(x)X_2(x')\right]. \tag{A.7}$$

Substituting our Kruskal-like embedding functions, we arrive at an expression for the geodesic distance in AdS_3

$$\begin{aligned} s = \ell \cosh^{-1}\Bigg(\frac{1}{(UV+1)(U'V'+1)}\Bigg[&(UV-1)(U'V'-1)\cosh\left(\frac{r_+(\phi-\phi')}{\ell}\right) + \\ &2(UV'+VU')\cosh\left(\frac{r_-(\phi-\phi')}{\ell}\right) + 2(VU'-UV')\sinh\left(\frac{r_-(\phi-\phi')}{\ell}\right)\Bigg]\Bigg). \end{aligned} \tag{A.8}$$

As a consistency check, the norm of this geodesic, $n_\mu \equiv \partial_\mu s$, satisfies $n^2 = 1$, as it should for a spacelike geodesic, $\nabla_\mu n_\nu \equiv n^\mu \nabla_\mu n_\nu = 0$, and the constraints on the bitensor of parallel transport $g_\nu^{\nu'}$ in Appendix C of [102].

Appendix B

Spinor Propagator in AdS_d

We follow the derivation given in [101] for the parallel propagator in $\mathbb{H}_d$, with the necessary few changes required for analytic continuation to AdS_d pointed out in footnotes. The calculation of $\langle T_{\mu\nu} \rangle$ for timelike geodesics in $d = 4$ can be found in [189].

We start with defining the norms of the AdS_d geodesic, which are given by

$$\begin{aligned} n_\mu &= \partial_\mu s(x, x') \\ n_{\mu'} &= \partial_{\mu'} s(x, x'). \end{aligned} \tag{B.1}$$

We also define the bitensor of parallel transport $g_\mu^{\nu'}$, which takes vectors between the two tangent spaces defined at x and x'. In particular, $n_\mu = -g_\mu^{\nu'} n_{\nu'}$, as the geodesic norm at a point is in the opposite direction of the geodesic length.

We also define two functions A and C relating the derivatives of n_μ and $n_{\mu'}$ to

similarly-indexed quantities[1]

$$\begin{aligned}\nabla_\mu n_\nu &= A(g_{\mu\nu} - n_\mu n_\nu)\\ \nabla_\mu n_{\nu'} &= C(g_{\mu\nu'} + n_\mu n_{\nu'}).\end{aligned} \tag{B.2}$$

For spacetimes with negative curvature, $A = \frac{1}{\ell}\coth\left(\frac{s}{\ell}\right)$ and $C = -\frac{1}{\ell}\operatorname{csch}\left(\frac{s}{\ell}\right)$[2]. By inverting the second equation and substituting the first, we find

$$\nabla_\mu g_{\nu\lambda'} = -(A+C)(g_{\mu\nu}n_{\lambda'} + g_{\mu\lambda'}n_\nu). \tag{B.3}$$

In order to properly treat spinors at two disconnected points, we also need to write a spinor parallel propagator $\Lambda^{\alpha'}_{\beta}$, which transports a spinor from x to x' as

$$\Psi'(x')^{\alpha'} = \Lambda(x',x)^{\alpha'}_{\beta}\Psi(x)^{\beta}. \tag{B.4}$$

The covariant derivatives of $\Lambda^{\alpha'}_{\beta}$ with respect to primed and unprimed coordinates are fixed by the parallel transport of the gamma matrices and (B.2) to be

$$\begin{aligned}D_\mu\Lambda(x,x') &= \frac{1}{2}(A+C)(\gamma_\mu\gamma^\nu n_\nu - n_\mu)\Lambda(x,x')\\ D_{\mu'}\Lambda(x,x') &= -\frac{1}{2}(A+C)\Lambda(x,x')(\gamma_{\mu'}\gamma^{\nu'}n_{\nu'} - n_{\mu'}).\end{aligned} \tag{B.5}$$

In principle, we could define a vierbein, find the exact form of the gamma matrices and the covariant derivative, and solve a system of PDEs for the spinor parallel propagator.

[1] In the case of timelike geodesics, these differ from [102] by the transforms

$$g_{\mu\nu} \to -g_{\mu\nu}, \quad g_{\mu\nu'} \to -g_{\mu\nu'}$$

[2] In [101], these functions in $\mathbb{H}_d$ are instead given by $A = \frac{1}{\ell}\cot\left(\frac{s}{\ell}\right)$ and $C = -\frac{1}{\ell}\csc\left(\frac{s}{\ell}\right)$

However, as for the scalar case, making use of the maximal symmetry of AdS turns out to greatly simplify the derivation [101].

The spinor propagator $S^{\alpha}_{\beta'}(x,x') = \langle \psi^{\alpha}(x)\bar{\psi}_{\beta'}(x')\rangle$ is a solution of the spacelike Dirac equation

$$\left[(\not{D} - m)S(x,x')\right]^{\alpha}_{\beta'} = \frac{\delta(x-x')}{\sqrt{-g}}\delta^{\alpha}_{\beta'} \tag{B.6}$$

with appropriate short-distance singularities. In the AdS vacuum it will share the maximal symmetry of the spacetime. It must therefore be of the form

$$S(x,x') = [\alpha(s) + \not{n}\beta(s)]\Lambda(x,x'), \tag{B.7}$$

for $\not{n} = n_{\mu}\gamma^{\mu}$ and some functions of the geodesic distance $\alpha(s)$ and $\beta(s)$. Inserting (B.7) into (B.6) yields

$$\left[\left(\alpha' + \frac{1}{2}(d-1)(A+C)\alpha - m\beta\right)\not{n} + \left(\beta' + \frac{1}{2}(d-1)(A-C)\beta - m\alpha\right)\right]\Lambda(x,x') = \frac{\delta(x-x')}{\sqrt{-g}}\delta^{\alpha}_{\beta'} \tag{B.8}$$

where a prime indicates a derivative with respect to s. Taking the trace of the above equation gives a set of two coupled differential equations

$$\begin{aligned} \beta' + \frac{1}{2}(d-1)(A-C)\beta - m\alpha &= \frac{\delta(x-x')}{\sqrt{-g}}, \\ \alpha' + \frac{}{2}(d-1)(A+C)\alpha - m\beta &= 0, \end{aligned} \tag{B.9}$$

Substituting the second equation into the first and using identities involving A and C

from table 1 of [102], one finds a second order equation for $\alpha(s)$:

$$\alpha'' + (d-1)A\alpha' - \left[m^2 + \frac{1}{2}(d-1)(C^2+AC) - \frac{(d-1)^2}{4\ell^2}\right]\alpha = m\frac{\delta(s)}{\sqrt{-g}}. \tag{B.10}$$

B.0.1 Solution in Minkowski Space

In order to find the proper normalization of the solutions in AdS$_d$, we need to find the short distance behavior of $\alpha(s)$, which is just the solution of (B.10) in Mink$_d$, or equivalently, in the limit $\ell \to \infty$. (B.10) becomes

$$\alpha'' + \frac{d-1}{s}\alpha' - m^2\alpha = m\delta(s). \tag{B.11}$$

The solution to this equation, properly normalized such that the left and right sides agree in the coincident limit, is

$$\alpha(s) = -\left(\frac{m}{2\pi}\right)^{d/2} s^{1-d/2} K_{d/2-1}(ms) \tag{B.12}$$

where $K_n(z)$ is the modified Bessel function of the second kind. The series expansion of this solution around $s = 0$ is given by

$$\alpha(s) \approx -\frac{m}{4}\pi^{-d/2} s^{2-d}\Gamma\left(\frac{d}{2}-1\right). \tag{B.13}$$

B.0.2 Solution in AdS$_d$

To solve (B.10) in AdS$_d$, we make the substitutions $z \equiv \cosh^2\left(\frac{s}{2\ell}\right)$ and

$$\gamma(z) \equiv \frac{1}{\sqrt{z}}\alpha(z) \tag{B.14}$$

to obtain[3]

$$z(1-z)\gamma''(z)+\left(\frac{d}{2}+1-(d+1)z\right)\gamma'(z)+\left(m^2\ell^2-\frac{d^2}{4}\right)\gamma(z)=-m\frac{\delta(s)}{\sqrt{-g}}, \tag{B.15}$$

where here primes denote derivatives with respect to z. This is a hypergeometric equation in $\gamma(z)$, with differential operator

$$H(a,b,c;z)=z(1-z)\frac{d^2}{dz^2}+(c-(a+b-1)z)\,\frac{d}{dz}-ab, \tag{B.16}$$

with $a=\frac{d}{2}-m\ell$, $b=\frac{d}{2}+m\ell$, $c=\frac{d}{2}+1$. This is the same as the $\mathbb{H}^d$ solution in [101], so we can proceed as follows. We want solutions that decay as a power of z as $z\to\infty$. There are two independent solutions, which up to overall normalizations $\lambda_\pm$ are

$$\gamma_\pm(z)=\lambda_\pm z^{-\left(\frac{d}{2}\pm m\ell\right)}{}_2F_1\left(\frac{d}{2}\pm m\ell,\pm m\ell,1\pm 2m\ell;\frac{1}{z}\right). \tag{B.17}$$

Both are allowed for sufficiently small $|m\ell|$, though large $|m\ell|$ requires the + sign. The expansion of $\alpha_\pm(s)$ around $s=0$ is

$$\alpha_\pm(s)\approx\lambda_\pm\left(\frac{s}{2\ell}\right)^{2-d}\frac{\Gamma(1\pm 2m\ell)\Gamma\left(\frac{d}{2}-1\right)}{\Gamma\left(\frac{d}{2}\pm m\ell\right)\Gamma(\pm m\ell)}. \tag{B.18}$$

The above expression must match with the Minkowski solution in the limit of vanishing s, so the coefficients $\lambda_\pm$ are given by

$$\lambda_\pm=\mp 2^{-(d\pm 2m\ell)}\ell^{1-d}\frac{\Gamma\left(\frac{d}{2}\pm m\ell\right)}{\pi^{(d-1)/2}\Gamma\left(\frac{1}{2}\pm m\ell\right)}. \tag{B.19}$$

[3]For timelike geodesics, we would instead make the same substitutions as in [101], with $z\equiv\cos^2\left(\frac{s}{2\ell}\right)$.

The choice of sign in these quantities corresponds to a choice of boundary conditions analogous to that referenced in (2.8). For simplicity, we will choose the (+) boundary condition for all p, as this is allowed for any effective 3D fermion mass. We do so for all computations in the main text.

B.0.3 $\langle T_{kk} \rangle$ in KKZBO

We still don't have a closed form expression for $S(x, x')$, as we don't know anything about $\Lambda(x, x')$, but we can still calculate observables such as $\langle \psi(x)\psi(x') \rangle$, $\langle J_\mu \rangle$, and $\langle T_{\mu\nu} \rangle$. We'll only carry out the calculation for our quotient spacetime[4]. The expectation value of the Belinfante stress-energy tensor in terms of the spinor propagator is [99]

$$\langle T_{\mu\nu} \rangle = \frac{i}{2} \lim_{x \to x'} \mathrm{Tr} \left[\left(\gamma_{(\mu} D_{\nu)} S(x, x') - \overline{D_{\nu'}} S(x, x') \gamma_{(\mu} g^{\nu'}_{\nu)} \right) \Lambda(x', x) \right] \tag{B.20}$$

where a primed index on the covariant derivative denotes a derivative on the second coordinate and action from the left. The parallel propagators become trivial at coincident points, namely $g^{\nu'}_{\nu} = \delta^{\nu'}_{\nu}$ and $\Lambda(x, x') = \mathbb{I}_2$ at $x = x'$. As shown in section 2.3.2, the null-null component of the stress-energy tensor in the quotient KKZBO spacetime is

$$\langle T_{kk} \rangle \equiv k^\mu k^\nu \langle T_{\mu\nu} \rangle = \frac{i}{2} \sum_{A,B} \mathrm{Tr} \left[\not{k} D_k S(x, J_3 x') \widetilde{j} \right] + \mathrm{Tr} \left[\widetilde{j} \overline{D_{k'}} S(J_3 x', x) \not{k} \right], \tag{B.21}$$

where we've suppressed the sum over Kaluza-Klein modes p but kept the sum over three-dimensional fermion representations A and B. In order for this stress-energy tensor to be real, we expect that the second trace is related to the first by complex conjugation. Using the fact that

$$\gamma^0 S(x, x')^\dagger \gamma^0 = S(x', x) \tag{B.22}$$

[4]For a similar calculation in AdS$_4$ for timelike geodesics, see [189]

the trace of the Hermitian conjugate of the first term becomes

$$\begin{aligned}\mathrm{Tr}\left[\left(\not{k}D_k S(x,J_3x')\widetilde{j}\right)^\dagger\right] &= \mathrm{Tr}\left[\widetilde{j}^\dagger D_{k'}^\dagger S(x,J_3x')^\dagger \not{k}^\dagger\right] \\ &= \mathrm{Tr}\left[\widetilde{j}^\dagger D_{k'}^\dagger \gamma^0 S(J_3x',x)\gamma^0\not{k}^\dagger\right] \\ &= -\mathrm{Tr}\left[\widetilde{j}\overleftarrow{D_{k'}} S(J_3x',x)\not{k}\right] \end{aligned} \tag{B.23}$$

where we've also used $\gamma^0\not{k}\gamma^0 = \not{k}^\dagger$ and $\gamma^0\widetilde{j}\gamma^0 = -\widetilde{j} = -\widetilde{j}^\dagger$. As the trace of the Hermitian conjugate is the complex conjugate of the original trace, we have

$$\mathrm{Tr}\left[\not{k}D_k S(x,J_3x')\widetilde{j}\right] = -\mathrm{Tr}\left[\widetilde{j}\overleftarrow{D_{k'}} S(J_3x',x)\not{k}\right]^* \tag{B.24}$$

as expected. The sum of the two traces can then be represented as the imaginary part of the first, and the stress-energy tensor becomes

$$\langle T_{kk}\rangle = -\sum_{A,B}\mathrm{Im}\Big\{\mathrm{Tr}\left[\widetilde{j}\not{k}D_k S(x,J_3x')\right]\Big\}. \tag{B.25}$$

Plugging in our ansatz (B.7) and substituting (B.5), we obtain

$$\begin{aligned}\langle T_{kk}\rangle &= -\sum_{A,B}\mathrm{Im}\left\{\mathrm{Tr}\left[\widetilde{j}\not{k}k^\mu D_\mu\left((\alpha(s)+\beta(s)\not{n})\Lambda(x,J_3x)\right)\right]\right\} \\ &= -(k^\mu n_\mu)\sum_{A,B}\mathrm{Im}\Big\{\mathrm{Tr}\left[\widetilde{j}\not{k}\left(\left(\alpha'+\frac{\alpha}{2}(A+C)\right)+\left(\beta'+\frac{\beta}{2}(C-A)\right)\not{n}\right)\Lambda(x,J_3x)\right]\Big\}.\end{aligned} \tag{B.26}$$

The A and B representations are distinguished by gamma matrices with opposite sign: $\gamma_A^\mu = -\gamma_B^\mu$. The sum over representations will cancel terms containing $\alpha(s)$, as they involve an odd number of gamma matrices, and introduce a factor of 2 to the $\beta(s)$ terms,

as they contain an even number of gamma matrices. The stress-energy tensor is therefore

$$\begin{aligned}\langle T_{kk}\rangle &= -2\,(k^\mu n_\mu)\left(\beta' + \frac{\beta}{2}(C-A)\right)\mathrm{Im}\Big\{\mathrm{Tr}\left[\widetilde{j}\not{k}\not{n}\Lambda(x,J_3x)\right]\Big\}\\ &= -2\,(k^\mu n_\mu)\left(\frac{\partial}{\partial s} - \frac{1}{2\ell}\coth\frac{s}{2\ell}\right)\beta(s)\,\mathrm{Im}\Big\{\mathrm{Tr}\left[\widetilde{j}\not{k}\not{n}\Lambda(x,J_3x)\right]\Big\}.\end{aligned} \tag{B.27}$$

In general, calculating $\Lambda(x,x')$ involves a path ordered integral along the geodesic in question, so it's easier to choose coordinates such that the parallel propagator becomes trivial even for noncoincident points. As our calculation ultimately takes place in the covering space AdS_3, we can choose global AdS_3 coordinates (t,ρ,φ) with metric

$$ds^2 = -\left(1+\rho^2/\ell^2\right)dt^2 + \frac{d\rho^2}{1+\rho^2/\ell^2} + \rho^2 d\varphi^2. \tag{B.28}$$

For a timelike slice of AdS_3, we define the origin by the intersection of the timelike axis defined by the isometry and the geodesic itself as in Figure 2.1. We therefore have geodesic norm $n_\mu = \frac{1}{\sqrt{1+\rho^2/\ell^2}}\partial_\rho$ and a perpendicular unit vector $\varphi_\mu = \rho\partial_\varphi$. In these coordinates, the spinor parallel propagator for our spacelike geodesic is trivial, as can be seen from the covariance of $\Lambda(x,x')$ along the geodesic:

$$\begin{aligned} n^\mu D_\mu \Lambda(x,x') &= n^\rho \partial_\rho \Lambda(x,x') = 0\\ \Rightarrow \Lambda(x,x') &= \mathbb{I}_2.\end{aligned} \tag{B.29}$$

Additionally, as the vielbein for this coordinate system is diagonal, we can rewrite the isometry $\widetilde{j} = i\gamma^1\gamma^2$ as

$$\widetilde{j} = i\not{n}\not{\varphi}. \tag{B.30}$$

The trace inside of the stress-energy tensor therefore becomes

$$\begin{aligned}\mathrm{Tr}\left[\widetilde{j}\not{k}\not{n}\right] &= \mathrm{Tr}\left[i\not{k}\not{\varphi}\right] \\ &= 2ik^\mu\varphi_\mu. \end{aligned} \tag{B.31}$$

This expression is manifestly vielbein independent. The final expression for the null-null component of the stress-energy tensor is therefore

$$\langle T_{kk}\rangle = -4\left(k^\mu n_\mu\right)\left(k^\mu\varphi_\mu\right)\left(\frac{\partial}{\partial s} - \frac{1}{2\ell}\coth\frac{s}{2\ell}\right)\beta(s). \tag{B.32}$$

The full calculation involves a sum over Kaluza-Klein modes p and BTZ images n. Explicitly, in terms of our Kruskal-like coordinates (U, V, ϕ) and radial coordinate

$$\rho = \ell\sqrt{\sinh^2\left(\frac{r_+\phi}{\ell}\right) + U^2\exp\left(-\frac{2r_-\phi}{\ell}\right)}, \tag{B.33}$$

we have

$$\begin{aligned} k^\mu n_\mu = n_U &= \frac{e^{-2r_-\phi/\ell}\ell^3 U}{\rho\sqrt{\rho^2+\ell^2}} \\ k^\mu\varphi_\mu = \varphi_U &= \frac{e^{-r_-\phi/\ell}\ell^2\sinh\left(\frac{r_+\phi}{\ell}\right)}{\rho}. \end{aligned} \tag{B.34}$$

Appendix C

Derivation of Dominant Saddles for Negativity

In this appendix, we derive the set of saddle-point geometries that give dominant contributions to the Rényi negativity in various regimes of the parameter space. This includes each of the distinct phases and near phase transitions.

Our derivation uses facts about geodesics on the permutation group, which we review first. Let S_n be the symmetric group of order n, which is the set of permutations on n elements. For any permutation $g \in S_n$, we define $\ell(g)$ as the minimum number of swaps from the identity $\mathbb{1} = (1)(2)\cdots(n)$ to g and $\chi(g)$ as the number of disjoint cycles in g, including 1-cycles. These quantities satisfy the relations

$$\ell(g) + \chi(g) = n, \tag{C.1}$$

$$\chi(g) = \chi(g^{-1}). \tag{C.2}$$

As an example, the permutation[1] $g = (12)(345) \in S_5$ has $\ell(g) = 3$ and $\chi(g) = 2$.

[1] This g is the permutation $12345 \to 21453$ written in cycle notation. Each digit in a given cycle is replaced by the following digit, except for the last digit which is replaced by the first.

We can define the distance between two permutations g and h by

$$d(g,h) \equiv \ell(g^{-1}h) \tag{C.3}$$

which satisfies the usual properties of a distance measure. In particular, given any sequence of permutations $(g_1, \cdots, g_m)$, the distance satisfies the triangle inequality

$$d(g_1,g_2) + \cdots + d(g_{m-1},g_m) \geq d(g_1,g_m). \tag{C.4}$$

A sequence of permutations that saturates (C.4) is said to be on a *geodesic*. We denote a geodesic between two permutations g and h by $G(g,h)$. We say that a permutation g' is on $G(g,h)$, or equivalently $g' \in G(g,h)$, if the sequence (g,g',h) saturates the triangle inequality (C.4).

Our goal is to identify the permutations g that dominate the sum in the Rényi negativity (3.25), in different regimes of the parameter space labeled by e^{S_0}, k_1, and k_2. We repeat the sum in (3.25) here:

$$\sum_{g\in S_n} \left(e^{S_0}\right)^{\chi(g)} k_1^{\chi(g^{-1}X)} k_2^{\chi(g^{-1}X^{-1})}. \tag{C.5}$$

For reasons that will become clear shortly, it is useful to first identify the permutations on one or more of the following geodesics: $G(\mathbb{1},X)$, $G(\mathbb{1},X^{-1})$, and $G(X,X^{-1})$. Here $X = (12\cdots n)$ is the cyclic permutation of n elements, and $X^{-1} = (n\cdots 21)$ is the anti-cyclic permutation.

For a given permutation g, let us use m, p, q to denote the three exponents in the sum (C.5):

$$m = \chi(g), \qquad p = \chi(g^{-1}X), \qquad q = \chi(g^{-1}X^{-1}). \tag{C.6}$$

They satisfy three triangle inequalities, which can be obtained from (C.1), (C.3), and (C.4):

$$d(\mathbb{1}, g) + d(g, X) \geq d(\mathbb{1}, X) \quad \Rightarrow \quad m + p \leq n + 1, \tag{C.7}$$

$$d(\mathbb{1}, g) + d(g, X^{-1}) \geq d(\mathbb{1}, X^{-1}) \quad \Rightarrow \quad m + q \leq n + 1, \tag{C.8}$$

$$d(X, g) + d(g, X^{-1}) \geq d(X, X^{-1}) \quad \Rightarrow \quad p + q \leq n + f(n), \tag{C.9}$$

where $f(n)$ is a useful function defined as

$$f(n) \equiv \begin{cases} 1, & n \text{ odd}, \\ 2, & n \text{ even}, \end{cases} \tag{C.10}$$

and we have used $\chi(X) = \chi(X^{-1}) = 1$, $\chi(X^2) = f(n)$.

We now identify the permutations g on one or more of the three geodesics.

Permutations on $G(\mathbb{1}, X)$: These are known to be in one-to-one correspondence with noncrossing partitions, so we say that the corresponding geometries are planar. We can write such a element as a product of m noncrossing cycles (including 1-cycles):

$$g = \prod_{i=1}^{m} c_i. \tag{C.11}$$

It is clear that such an element exists for every $m \in [1, n]$. Since it saturates (C.7), we immediately find

$$p = n - m + 1. \tag{C.12}$$

Moreover, it is straightforward to derive

$$q = \sum_{i=1}^{m} f(|c_i|) - m + 1, \tag{C.13}$$

where $|c_i|$ is the length of the i-th cycle c_i.

Permutations on $G(\mathbb{1}, X^{-1})$: These can be obtained by simply taking the inverse of the permutations on $G(\mathbb{1}, X)$, sending c_i in (C.11) to c_i^{-1}. We say that these correspond to "anti-planar" geometries.

Permutations on $G(\mathbb{1}, X)$ and $G(\mathbb{1}, X^{-1})$: Their cycles c_i must be their own inverses, so the length of each cycle is at most 2. Therefore, these permutations are precisely those noncrossing partitions that consist of only 1-cycles and 2-cycles. Such permutations exist for every $m \geq \lceil \frac{n}{2} \rceil$, with the lower bound saturated by noncrossing pairings consisting of $\lceil \frac{n}{2} \rceil$ pairs and at most one 1-cycle.

Permutations on $G(\mathbb{1}, X)$ and $G(X, X^{-1})$: As they saturate (C.9), it is straightforward to use (C.12) and (C.13) to show that these permutations are precisely those noncrossing partitions with at most one odd cycle. Here we define an odd cycle as one of odd length and an even cycle as one of even length. For even n, these permutations consist of only even cycles, whereas for odd n, they have exactly one odd cycle. Such permutations exist for every $m \leq \lceil \frac{n}{2} \rceil$, with the upper bound saturated by noncrossing pairings.

Permutations on $G(\mathbb{1}, X^{-1})$ and $G(X, X^{-1})$: These are obtained by taking the inverse of the permutations on $G(\mathbb{1}, X)$ and $G(X, X^{-1})$.

Permutations on $G(\mathbb{1}, X)$, $G(\mathbb{1}, X^{-1})$, and $G(X, X^{-1})$: It is clear by combining the previous cases that these permutations are those noncrossing partitions that consist of only 2-cycles and at most one 1-cycle. Therefore, they are in one-to-one correspondence with noncrossing pairings [128]. These all have $m = \lceil \frac{n}{2} \rceil$ and $p = q = \lfloor \frac{n}{2} \rfloor + 1$. We denote

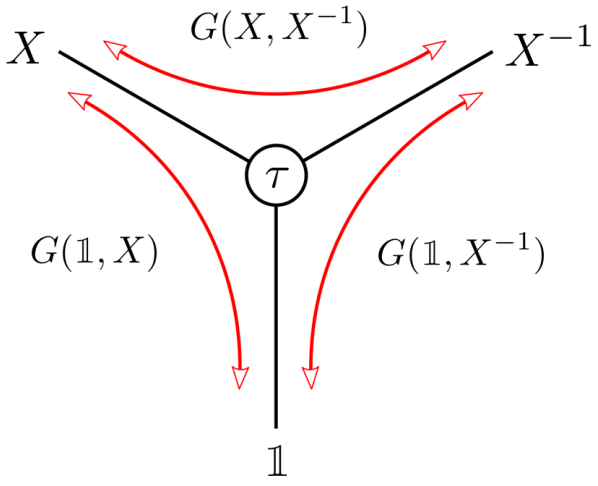


Figure C.1: A schematic Cayley graph for the relevant permutations and the geodesics connecting them.

these noncrossing pairings by τ, and say that they correspond to pairwise geometries. A simple example is $\tau = (12)(34)\cdots(n-1, n)$ for even n and $\tau = (12)(34)\cdots(n-2, n-1)(n)$ for odd n.

These results are illustrated schematically in Figure C.1. We now state and prove the main points of this appendix.

Lemma 11. In the regime $k_1/k_2 \gg e^{-S_0}$, planar geometries dominate the sum (C.5). In other words, for any $g \notin G(\mathbb{1}, X)$, there exists $g' \in G(\mathbb{1}, X)$ such that g' dominates over g.

Proof: As $g \notin G(\mathbb{1}, X)$, the difference

$$d(\mathbb{1}, g) + d(g, X) - d(\mathbb{1}, X) \tag{C.14}$$

is positive. However, this difference must be even, regardless of whether g is an even or odd permutation. Therefore, we denote this difference by $2r$ with a positive integer r,

and use it to rewrite the triangle inequality (C.7) as

$$m + p = n + 1 - 2r. \tag{C.15}$$

Our goal is to choose a more dominant g'. For g', we define m', p', q' similarly as

$$m' = \chi(g'), \qquad p' = \chi(g'^{-1}X), \qquad q' = \chi(g'^{-1}X^{-1}). \tag{C.16}$$

Let us discuss $m + r \geq \lceil \frac{n}{2} \rceil$ and $m + r < \lceil \frac{n}{2} \rceil$ separately. For $m + r \geq \lceil \frac{n}{2} \rceil$, we choose g' to be on $G(\mathbb{1}, X)$ and $G(\mathbb{1}, X^{-1})$, with $m' = m + r$. As we discussed earlier, such a permutation exists; in particular, (C.15) guarantees $m + r < n$. As g' saturates (C.7) and (C.8) (after primes are added), we find

$$p' = n - m - r + 1 = p + r, \qquad q' = n - m - r + 1 \geq q - r, \tag{C.17}$$

where the second equality for p' comes from (C.15) and the inequality comes from (C.8) for g. We thus find that g' gives a more dominant contribution to the sum (C.5) than g, as

$$\frac{e^{m'S_0} k_1^{p'} k_2^{q'}}{e^{mS_0} k_1^{p} k_2^{q}} \geq \left(e^{S_0} \frac{k_1}{k_2} \right)^r \gg 1. \tag{C.18}$$

In the other case with $m + r < \lceil \frac{n}{2} \rceil$, we choose g' to be on $G(\mathbb{1}, X)$ and $G(X, X^{-1})$, with $m' = m + r$. Again, such a permutation exists. As g' saturates (C.7) and (C.9) (after primes are added), we find

$$p' = n - m - r + 1 = p + r, \qquad q' = n + f(n) - p - r \geq q - r, \tag{C.19}$$

where the inequality comes from (C.9) for g. We again find that (C.18) holds and therefore g' dominates over g.

It is clear that Lemma 11 is tight in the sense that every planar geometry could give a dominant contribution to the sum (C.5) at some point in the regime $k_1/k_2 \gg e^{-S_0}$. In particular, at the point where $k_1 = e^{S_0}$ and $k_2 = 1$, they give an equal contribution $e^{(n+1)S_0}$.

From Lemma 11, we immediately obtain the following corollary by taking the inverse of all permutations and switching $k_1 \leftrightarrow k_2$.

Corollary 12. In the regime $k_2/k_1 \gg e^{-S_0}$, anti-planar geometries dominate the sum (C.5).

Combining Lemma 11 and Corollary 12, and recalling that the permutations on $G(\mathbb{1}, X)$ and $G(\mathbb{1}, X^{-1})$ are precisely those that consist of only 1-cycles and 2-cycles, we immediately obtain the following corollary (which is useful for studying the disconnected-pairwise transition in Section 3.5.2).

Corollary 13. In the regime $e^{-S_0} \ll k_1/k_2 \ll e^{S_0}$, the permutations consisting of only 1-cycles and 2-cycles dominate the sum (C.5).

It is again clear that Corollary 13 is tight in the sense that every permutation on $G(\mathbb{1}, X)$ and $G(\mathbb{1}, X^{-1})$ could give a dominant contribution to the sum (C.5) at some point in the regime $e^{-S_0} \ll k_1/k_2 \ll e^{S_0}$. In particular, at the point where $k_1 = k_2 = e^{S_0/2}$, they all give an equal contribution $e^{(n+1)S_0}$.

Lemma 14. In the regime $k_1 k_2 \ll e^{S_0}$, the disconnected geometry dominates the sum (C.5).

Proof: The disconnected geometry is represented by the identity $\mathbb{1}$ and contributes $e^{nS_0} k_1 k_2$. For any other permutation g, we have $m \leq n-1$. From (C.7) and (C.8), we obtain $p, q \leq n - m + 1$. We thus find that $\mathbb{1}$ gives a more dominant contribution to the sum (C.5) than g, as

$$\frac{e^{nS_0} k_1 k_2}{e^{mS_0} k_1^p k_2^q} \geq \left(\frac{e^{S_0}}{k_1 k_2}\right)^{n-m} \gg 1. \tag{C.20}$$

Lemma 15. In the regime $k_1/k_2 \gg e^{S_0}$, the cyclic geometry dominates the sum (C.5).

Proof: The cyclic geometry is represented by X and contributes $e^{S_0} k_1^n k_2^{f(n)}$. For any other permutation g, we have $p \leq n-1$. From (C.7) and (C.9), we obtain $m \leq n-p+1$ and $q \leq n + f(n) - p$. We thus find that X gives a more dominant contribution to the sum (C.5) than g, as

$$\frac{e^{S_0} k_1^n k_2^{f(n)}}{e^{mS_0} k_1^p k_2^q} \geq \left(\frac{k_1}{k_2} e^{-S_0}\right)^{n-p} \gg 1. \tag{C.21}$$

From Lemma 15, we immediately obtain the following corollary by taking the inverse of all permutations and switching $k_1 \leftrightarrow k_2$.

Corollary 16. In the regime $k_2/k_1 \gg e^{S_0}$, the anti-cyclic geometry dominates the sum (C.5).

We now show that pairwise geometries dominate a fourth phase. We first derive the following lemma as a useful intermediate step.

Lemma 17. In the regime $k_1 k_2 \gg e^{S_0}$, the permutations on $G(X, X^{-1})$ dominate the sum (C.5). In other words, for any $g \notin G(X, X^{-1})$, there exists $g' \in G(X, X^{-1})$ such that g' dominates over g.

Proof: As $g \notin G(X, X^{-1})$, the triangle inequality (C.9) must fail to saturate by a positive but even integer, which is at least 2:

$$p + q \leq n + f(n) - 2 = 2\left\lfloor \frac{n}{2} \right\rfloor. \tag{C.22}$$

Therefore, one of p, q must be no greater than $\lfloor \frac{n}{2} \rfloor$. Without loss of generality, we consider the case of $p \leq \lfloor \frac{n}{2} \rfloor$. We then choose g' to be on $G(\mathbb{1}, X)$ and $G(X, X^{-1})$, with $p' = p + 1$. As g' saturates (C.7) and (C.9) (after primes are added), we find

$$m' = n - p \geq m - 1, \qquad q' = n + f(n) - p - 1 \geq q + 1, \tag{C.23}$$

where the two inequalities comes from (C.7) and (C.22), respectively. As we discussed earlier, such a permutation g' exists, as $m' = n - p \geq \lceil \frac{n}{2} \rceil$. From this, we find that g' gives a more dominant contribution to the sum (C.5) than g, as

$$\frac{e^{m'S_0} k_1^{p'} k_2^{q'}}{e^{mS_0} k_1^{p} k_2^{q}} \geq \frac{k_1 k_2}{e^{S_0}} \gg 1. \tag{C.24}$$

Combining Lemmas 11, 17 and Corollary 12, and recalling that the permutations on all three geodesics $G(\mathbb{1}, X)$, $G(\mathbb{1}, X^{-1})$, and $G(X, X^{-1})$ are precisely noncrossing pairings that lead to pairwise geometries, we immediately obtain the following corollary.

Corollary 18. In the regime satisfying both $k_1 k_2 \gg e^{S_0}$ and $e^{-S_0} \ll k_1/k_2 \ll e^{S_0}$, the pairwise geometries dominate the sum (C.5).

It is clear that Corollary 18 is tight in the sense that all pairwise geometries give an equal, dominant contribution $(e^{S_0})^{\lceil \frac{n}{2} \rceil} (k_1 k_2)^{\lfloor \frac{n}{2} \rfloor + 1}$ to the sum (C.5), as they all have the same $m = \lceil \frac{n}{2} \rceil$ and $p = q = \lfloor \frac{n}{2} \rfloor + 1$.

Appendix D

Details of the Cyclic-Pairwise Transition in the Canonical Ensemble

In Section 4.4, we used some of the techniques developed in [61] to derive an approximation for the eigenvalue spectrum of the partially transposed density matrix near transition:

$$D(\lambda) = e^{S_0} \int_0^{s_k} ds\rho(s) \left[\frac{k_2(k_2+1)}{2} \delta\left(\lambda - \lambda_0 - \frac{w(s)}{k_2}\right) + \frac{k_2(k_2-1)}{2} \delta\left(\lambda - \lambda_0 + \frac{w(s)}{k_2}\right) \right]. \tag{D.1}$$

This approximation was derived under a set of assumptions which we repeat here:

1. $w(s_t)R \ll kk_2$
2. $k_2^2 e^{S_0} \int_0^{s_t} ds\rho(s) \frac{w(s)R(k+w(s)R)}{k^2k_2^2 - w(s)^2R^2} \ll k$
3. $s_t = s_k - \kappa$, where κ is $\mathcal{O}(1)$ but large

Our analysis in this section will be based on checking the consistency of the iterative procedure we applied to the resolvent equation (3.67), namely the zeroth order approximation

$$R_0 = \frac{k}{\lambda - \lambda_0}. \tag{D.2}$$

We start with Assumption (2). We want to rigorously show the following inequality on the second term in (3.67):

$$\left| k_2^2 e^{S_0} \int_0^{s_t} \rho(s) \frac{w(s)R(k + w(s)R)}{k^2 k_2^2 - w(s)^2 R^2} \right| \equiv \left| e^{S_0} \int_0^{s_t} \rho(s) f(s) \right| \ll k. \tag{D.3}$$

This function has a pole located at $s = s_*$, which is captured by the integral under the assumption $|w(s_t)R| \ll kk_2$, as $w(s)$ is a monotonically decreasing function of s. We can therefore rewrite the integral with an $i\epsilon$ prescription as

$$\begin{aligned} e^{S_0} \int ds \rho(s) f(s) = PV \left(e^{S_0} \int ds \rho(s) f(s) \right) \\ \pm i\pi k_2^2 e^{S_0} \int ds \rho(s) \frac{w(s)R(k + w(s)R)}{\partial_s \left(k^2 k_2^2 - w(s)^2 R^2 \right)} \delta(s - s_*), \end{aligned} \tag{D.4}$$

where PV denotes the Cauchy principal value. We choose the sign of $i\epsilon$ arbitrarily, as we are only looking to bound the absolute value of this integral.

Let us treat the first term. The principle value is dominated by the $(s - s_*)^0$ term in the Laurent expansion of $\rho(s)f(s)$. We can perform a Laurent expansion around $s = s_*$ using the semiclassical approximation

$$w(s) \approx e^{-\beta s^2/2 - S_0 - 2\pi^2/\beta}. \tag{D.5}$$

We find

$$\rho(s)f(s) = \rho(s_*)\left(\frac{k_2(k_2+1)}{2(\beta s_*)(s-s_*)} - \frac{k_2(1+k_2(1+2\beta s_*^2))}{4\beta s_*^2} + \mathcal{O}(s-s_*)\right). \tag{D.6}$$

Ignoring scaling in β, the $\mathcal{O}(s-s_*)^0$ term goes like $k_2^2 \sim k/e^{S_0} \ll k$, so we can safely ignore this term.

Let us now look at the second term in (D.4). We have

$$\begin{aligned} k_2^2 e^{S_0}\rho(s)\frac{w(s)R(k+w(s)R)}{\partial_s\left(k^2k_2^2 - w(s)^2R^2\right)}\bigg|_{s=s_*} &\approx k_2^2 e^{S_0}\rho(s)\frac{(w(s)R)^2}{(2\beta s)w(s)^2R^2}\bigg|_{s=s_*} \\ &\approx \frac{k_2^2 e^{S_0}\rho(s_*)}{2\beta s_*}, \end{aligned} \tag{D.7}$$

where we have used $|w(s_*)R| = kk_2 \gg k$ to simplify the numerator. We can rewrite this in terms of s_k using (3.61) and the asymptotic form of $\rho(s)$:

$$\begin{aligned} \frac{k_2^2 e^{S_0}\rho(s_*)}{2\beta s_*} &\approx \frac{k_2^2}{2\beta}e^{2\pi s_*} \\ &\approx \frac{k_2^2}{2\beta s_k}e^{S_0}\rho(s_k)e^{2\pi(s_*-s_k)} \\ &= \frac{k}{2\beta s_k}e^{2\pi(s_*-s_k)} \ll k. \end{aligned} \tag{D.8}$$

As stated previously, $s_k \sim \mathcal{O}(1/\beta)$, so for this to be much smaller than k we require $s_k - s_*$ being at least $\mathcal{O}(1)$ but large, and by proxy $\kappa \equiv s_k - s_t$ being at least $\mathcal{O}(1)$ but large, as stated in assumption (3).

What is stopping κ from being much larger, say $\mathcal{O}(1/\beta)$? Now we check the validity of assumption (1), that is $|w(s_t)R| \ll kk_2$. Under our approximation (D.2), this assumption translates into the condition

$$|\lambda - \lambda_0| \gg \frac{w(s_t)}{k_2}. \tag{D.9}$$

However, at the boundary of our spectrum located at $s = s_t$, we should also have $|\lambda - \lambda_0| = w(s_t)/k_2$. The way to reconcile these two assumptions is to state that our approximation only holds for some range of s between 0 and some control parameter s_c where

$$|\lambda - \lambda_0| = \frac{w(s_c)}{k_2}. \tag{D.10}$$

What is the value of s_c? We have

$$w(s_c) \gg w(s_t) \Rightarrow e^{-\beta(s_t^2 - s_c^2)/2} \ll 1. \tag{D.11}$$

We define $\kappa' \equiv s_t - s_c$. Now (D.11) becomes

$$2s_t\kappa' + \kappa'^2 \gg 1/\beta. \tag{D.12}$$

Under our previous assumption that κ is $\mathcal{O}(1)$ in β such that $s_t \sim 1/\beta$, κ' too must be $\mathcal{O}(1)$ but large. This answers our previous question about the size of κ. There is something of an inverse relationship between s_t and κ': our goal is to design an approximation in which s_c is as close as possible to s_t, as well as one in which s_t is as close as possible to s_k. Under the assumption that s_t is as close as possible to s_k, that is κ is $\mathcal{O}(1)$ but large, we also have κ' $\mathcal{O}(1)$ but large. If s_t was far from s_k so that, say, $s_t \sim 1/\sqrt{\beta}$, we would also require $\kappa' \sim \mathcal{O}(1)/\sqrt{\beta}$ with a large $\mathcal{O}(1)$ constant, thus missing a large part of the spectrum in our approximation.

Our conclusion is that there exists a region of size $\mathcal{O}(1)$ in s where assumption (1) does not hold. As the spectrum is over a region of size $\frac{w(0)-w(s_k)}{k_2}$, the size of a region $\mathcal{O}(1)$ in s is exponentially suppressed in $1/\beta$, and we conclude that very few eigenvalues are in the uncontrolled region.

Appendix E

Rényi Entropies near the Page Transition

We draw an analogy between the $\mathcal{O}(1/\beta)$ corrections to the logarithmic negativity and the partially transposed entropy and the $\mathcal{O}(1/\beta)$ corrections to the Rényi entropy S_n with $n < 1$. Here we show this result explicitly in the model of [61]. We recall many of their results, which can equivalently be obtained from ours by sending k_1 to k and k_2 to 1.

We consider the model of Section 3.3, but without partitioning the radiation system. The approximation for the density of states (3.76) is now

$$D(\lambda) = e^{S_0} \int_0^{s_k} ds \rho(s) \delta(\lambda - \lambda_0 - w(s)), \tag{E.1}$$

where $w(s)$ and λ_0 are defined as in the main text. Here s_k is defined as

$$k = e^{S_0} \int_0^{s_k} ds \rho(s) \Rightarrow s_k \approx \frac{1}{2\pi} \left(\log k - S_0 \right). \tag{E.2}$$

The transition at the Page time can be thought of as the transition from the fully disconnected phase to the cyclic phase along the x axis of our phase diagram (Figure 3.4). Using our results from Table 3.1 and the semiclassical approximation $\mu \gg \frac{1}{\beta} \gg 1$, we have

$$\begin{aligned} \operatorname{Tr} \rho_R^n &= \left(\frac{Z_1}{k}\right)^{n-1} = \frac{Z_n}{Z_1^n} \\ \Rightarrow S_n &= S_0 + \left(1 + \frac{1}{n}\right) \frac{2\pi^2}{\beta}. \end{aligned} \tag{E.3}$$

Again we can solve for $\log k$ at transition in the semiclassical regime to obtain

$$\log k = \log \left(\frac{Z_1^n}{Z_n}\right)^{\frac{1}{n-1}} = S_0 + \left(1 + \frac{1}{n}\right) \frac{2\pi^2}{\beta}. \tag{E.4}$$

From this we find $s_k^{(n)}$ at transition to be

$$s_k^{(n)} = \frac{\pi}{\beta} \left(1 + \frac{1}{n}\right). \tag{E.5}$$

As $s^{(n)} = \frac{2\pi}{n\beta}$, for $n < 1$ we have $s_k^{(n)} < s^{(n)}$. Part of our derivation relied on s_k scaling like $1/\beta$, which remains true for n of $\mathcal{O}(1)$.

The Rényi entropy is given by

$$\begin{aligned} S_n &= \frac{1}{1-n} \log \int_{-\infty}^{\infty} d\lambda D(\lambda) \lambda^n \\ &= \frac{1}{1-n} \log \left(e^{S_0} \int_0^{s_k} ds \rho(s) \left(\lambda_0 + w(s)\right)^n \right). \end{aligned} \tag{E.6}$$

As $s_k > s^{(1)}$, λ_0 will be exponentially suppressed in $1/\beta$, so this integral is dominated by

the $w(s)$ term, and as $s_k^{(n)} < s^{(n)}$ we approximate

$$\begin{aligned} S_n &\approx \frac{1}{1-n} \log \left(e^{S_0} \int_0^{s_k} ds \rho(s) w(s)^n \right) \\ &\approx \frac{1}{1-n} \log \left(e^{S_0} \rho(s_k) w(s_k)^n \right) \\ &\approx \frac{1}{1-n} \left(\log k - n S_0 - \frac{n \beta s_k^2}{2} - \frac{2\pi^2 n}{\beta} \right). \end{aligned} \tag{E.7}$$

Using our expressions (E.4) and $s_k^{(n)}$, we find

$$S_n = S_0 + \left(\frac{3+5n}{2n} \right) \frac{\pi^2}{\beta}. \tag{E.8}$$

Comparing this to our previous answer (E.3), we find a correction ΔS_n at transition of the form

$$\Delta S_n = \frac{\pi^2}{2\beta} \left(1 - \frac{1}{n} \right). \tag{E.9}$$

We conclude that there are enhanced corrections of the form $\mathcal{O}(1/\beta)$ to the Rényi entropy for $n < 1$.

Appendix F

Deriving the Relevant Sum Over Permutations

Let's recall a few facts about the permutation group. For an element $g \in S_n$, we denote the number of swaps from the identity permutation $\mathbb{1} = (1)(2)\cdots(n)$ to g by $\ell(g)$ and the number of distinct cycles in g by $\chi(g)$. These quantities satisfy the relation

$$\ell(g) + \chi(g) = n \tag{F.1}$$

The number of swaps between two permutations $\ell(g^{-1}h) \equiv d(g,h)$ introduces a natural distance measure between two permutations. In particular, there exists the traingle inequality

$$d(g,g_1) + d(g_1,h) \geq d(g,h) \tag{F.2}$$

A *geodesic* between two permutations $G(g,h)$ is the set of $g_1's$ which saturate this inequality.

The sum over permutations we're interested in takes the form [67, 127, 134]

$$\texttt{SUM} = \sum_{g \in S_n} (e^{S_B})^m (e^{S_{A_1}})^p (e^{S_{A_2}})^q \tag{F.3}$$

where we've made the substitutions

$$m = \chi(g), \quad p = \chi(g^{-1}X), \quad q = \chi(g^{-1}X^{-1}) \tag{F.4}$$

Here X is the cyclic permutation $(12\cdots n)$ and X^{-1} is the anti-cyclic permutation $(n\ n-1\cdots 1)$. This is the sum relevant for calculating the moments of a block transposed Wishart matrix [140], i.e. the weighting of Wick contractions when averaging over a random density matrix with Gaussian correlations. In our work, we're interested in the permutations which live on the geodesic $G(\mathbb{1}, X)$ and the geodesic $G(X, X^{-1})$ but not necessarily on the geodesic $G(\mathbb{1}, X^{-1})$. These permutations satisfy the following three equations:

$$\begin{aligned} m + p &= n + 1 \\ m + q &\leq n + 1 \\ p + q &= n + f(n) \end{aligned} \tag{F.5}$$

where the function $f(n) = 1$ if n is odd and 2 if n is even. When the second inequality is saturated, we're talking about the set of noncrossing pairings τ. There are C_n of these permutations, where C_n are the Catalan numbers

$$C_n = \frac{1}{n+1}\binom{2n}{n} \tag{F.6}$$

An example of a τ permutation on an even number of elements is $(12)(34)\cdots(n-1\ n)$.

A noncrossing pairing on an odd number of elements will have a single cycle of length 1 and all other cycles of length 2. Permutations which live on $G(\mathbb{1}, X)$ and $G(X, X^{-1})$ but not $G(\mathbb{1}, X^{-1})$ are precisely those which live on the single geodesic $G(\tau, X)$, which is the phase transition we're interested in. How do we enumerate these permutations? From the two equalities, we have:

$$p = n - 1 - mq = m + f(n) - 1 \tag{F.7}$$

From this, we can derive an upper bound on m:

$$m \leq \frac{n + 2 - f(n)}{2} \tag{F.8}$$

So we've reduced the sum over all permutations to a sum over a single parameter $m = \chi(g)$. We now have

$$\texttt{SUM} = \sum_{m=1}^{\frac{n-f(n)+2}{2}} T'(n,m) \left(e^{S_{A_2}}\right)^{f(n)-1} \left(e^{S_{A_1}}\right)^{n+1} \left(\frac{e^{S_B} e^{S_{A_2}}}{e^{S_{A_1}}}\right)^m \tag{F.9}$$

for some counting function $T'(n,m)$ which denotes the multiplicity at every $\chi(g)$. What is this function? Let's consider it for both even and odd n. For even $n = 2k$, the sum is

$$\texttt{EVEN SUM} = \sum_{m=1}^{k} T_e(k,m) e^{S_{A_2}} \left(e^{S_{A_1}}\right)^{2k+1} \left(\frac{e^{S_B} e^{S_{A_2}}}{e^{S_{A_1}}}\right)^m \tag{F.10}$$

This is a sum over permutations starting with the cyclic permutation X at $m = 1$ and ending with the pairwise connected permutations τ at $m = k$. Each m corresponds to a permutation with m cycles of even length. In this case, the numbers $T_e(k,m)$ are

equivalent to the number of 2-Dyck paths of order k with m peaks and are given by

$$T_e(k,m) = \frac{1}{k}\binom{k}{m}\binom{2k}{m-1} \tag{F.11}$$

The $T_e(k,m)$ that appear here are analagous to the Narayana numbers which appear in the sum over noncrossing permutations. They are sometimes referred to as 2-Narayana numbers and appeared in various contexts elsewhere [160, 190, 191]. We therefore have

$$\begin{aligned}\texttt{EVEN SUM} &= e^{S_{A_2}}\left(e^{S_{A_1}}\right)^{2k+1}\sum_{m=1}^{k} T_e(k,m)\left(\frac{e^{S_B}e^{S_{A_2}}}{e^{S_{A_1}}}\right)^m \\ &= e^{2kS_{A_1}}e^{2S_{A_2}}e^{S_B}{}_2F_1\left(1-k,-2k,2;\frac{e^{S_B}e^{S_{A_2}}}{e^{S_{A_1}}}\right)\end{aligned} \tag{F.12}$$

Now let's look at the odd case. When $n = 2k-1$, the sum over permutations is

$$\texttt{ODD SUM} = \sum_{m=1}^{k} T_o(k,m)\left(e^{S_{A_1}}\right)^{2k}\left(\frac{e^{S_B}e^{S_{A_2}}}{e^{S_{A_1}}}\right)^m \tag{F.13}$$

Now the counting function is slightly different. We can derive it as follows: consider a permutation allowed in the even sum (F.10) with m cycles. The second binomial factor in (F.11) can morally be thought of as choosing $m-1$ distinct elements to belong to different cycles, while the rest is a symmetry factor that controls the number of noncrossing permutations modulo that choice. Therefore, one can think of each noncrossing permutation as living in a "labelled" superselection sector of size $\binom{2k}{m-1}$. By ignoring this choice and dividing by this factor, we can find a degenerate set of "unlabelled" noncrossing permutations. From this set we can remove an element from each cycle, so one unlabelled permutation in the even sum generates m distinct unlabelled permutations in the odd sum, which then have to be relabelled to give the correct counting. This strategy

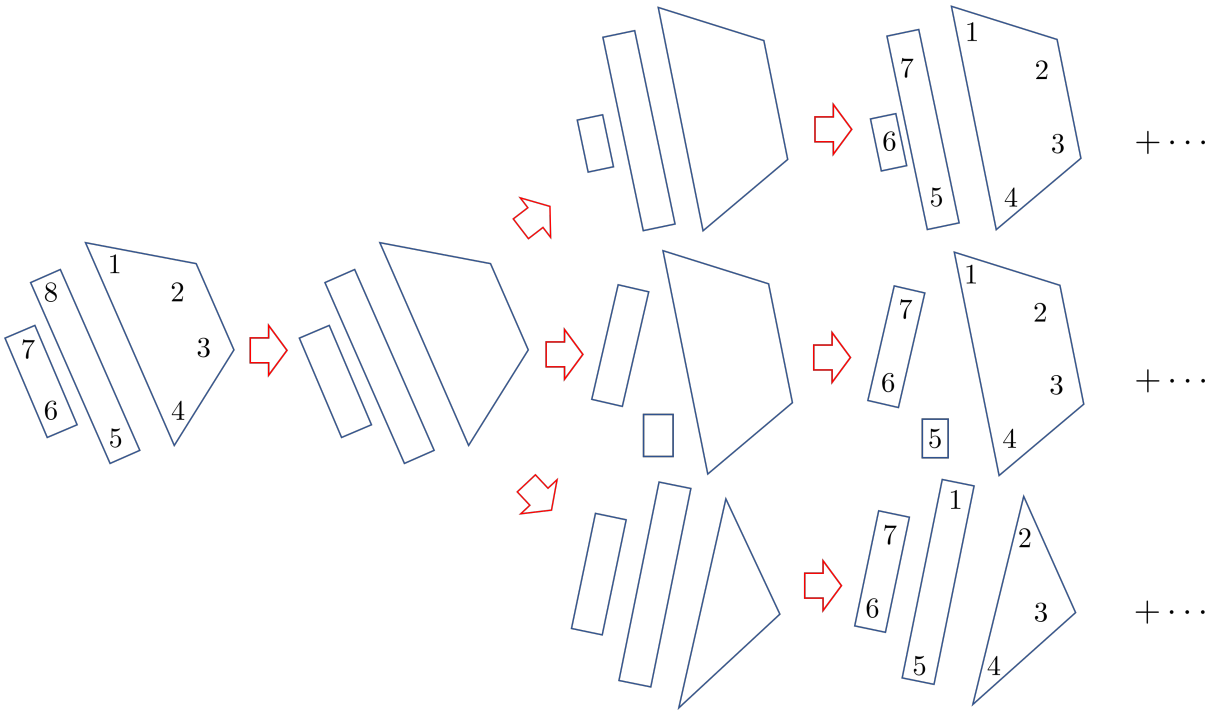


Figure F.1: The procedure of generating permutations on $G(\tau, X)$ for odd n from even n. We identify even permutations on $G(\tau, X)$ which are the same up to the choice of $m-1$ elements. Each of these pieces produces m unlabelled odd pieces by removing an element, and identifying $m-1$ elements again gives us all odd permutations on $G(\tau, X)$.

of unlabelling, removing an element, and relabelling gives us

$$T_o(k,m) = m\frac{\binom{2k-1}{m-1}}{\binom{2k}{m-1}}T_e(k,m) = \binom{2k-1}{m-1}\binom{k-1}{m-1} \tag{F.14}$$

If that was a bit too abstract, we illustrate this procedure in Figure F.1. The sum over permutations for odd n is now

$$\begin{aligned}\texttt{ODD SUM} &= \left(e^{S_{A_1}}\right)^{2k}\sum_{m=1}^{k} T_e(k,m)\left(\frac{e^{S_B}e^{S_{A_2}}}{e^{S_{A_1}}}\right)^m \\ &= \left(e^{S_{A_1}}\right)^{2k-1} e^{S_{A_2}}e^{S_B}{}_2F_1\left(1-2k, 1-k, 1; \frac{e^{S_B}e^{S_{A_2}}}{e^{S_{A_1}}}\right)\end{aligned} \tag{F.15}$$

As a sanity check for our counting functions $T_e(k,m)$ and $T_o(k,m)$, both $T_e(k,k)$ and

$T_o(k,k)$ are equal to the symmetry factors for the pairwise connected geometries:

$$T_e(k,k) = C_k, \quad T_o(k,k) = (2k-1)C_{k-1} \tag{F.16}$$

and $T_e(k,1) = T_o(k,1) = 1$.

Appendix G

Resolvent for Disorder Averaging

The resolvent matrix $R_{ij}(\lambda)$ encodes the eigenvalue spectrum given by

$$R(\lambda)_{ij} = \frac{1}{\lambda}\delta_{ij} + \sum_{n=1}^{\infty} \frac{1}{\lambda^{n+1}} \left(\rho_A^n\right)_{ij} \tag{G.1}$$

or in traced version

$$R(\lambda) = \frac{e^{S_A}}{\lambda} + \sum_{n=1}^{\infty} \frac{1}{\lambda^{n+1}} \operatorname{Tr}\left(\rho_A^n\right) \tag{G.2}$$

There is a resolvent equation for the Wick contractions

$$R(\lambda)_{ij} = \frac{\delta_{ij}}{\lambda} + \frac{e^{S_B}}{\lambda} \sum_{n=1}^{\infty} R(\lambda)^n R(\lambda)_{ij} \tag{G.3}$$

We can take the trace and resum this equation:

$$\lambda R(\lambda) = e^{S_A} + \frac{e^{S_B} R(\lambda)}{1 - R(\lambda)} \tag{G.4}$$

This is the resolvent equation for the eigenvalues of a Wishart matrix. This equation can be solved exactly for $R(\lambda)$. It admits an expansion

$$R(\lambda) = \frac{e^{S_A}}{\lambda} + \frac{e^{S_B}}{\lambda} \sum_{n=1}^{\infty} \sum_{k=1}^{n} N(n,k) \left(\frac{e^{S_A}}{\lambda}\right)^n e^{(k-1)(S_B - S_A)} \tag{G.5}$$

where $N(n,k)$ are the Narayana numbers. From this we can read off $\mathrm{Tr}\,\overline{(\rho_A)^n}$:

$$\mathrm{Tr}\,\overline{(\rho_A)^n} = e^{nS_A + S_B} \sum_{k=1}^{n} N(n,k) e^{(k-1)(S_B - S_A)} \tag{G.6}$$

This sum has a nice closed form expression

$$\mathrm{Tr}\,\overline{(\rho_A)^n} = \begin{cases} e^{S_A + nS_B} {}_2F_1\left(1-n, -n; 2; e^{S_A - S_B}\right), S_A < S_B \\ e^{nS_A + S_B} {}_2F_1\left(1-n, -n; 2; e^{S_B - S_A}\right), S_A > S_B \end{cases} \tag{G.7}$$

where there are two branches such that the final argument of the hypergeometric always always lies within the unit circle on the complex plane. A similar resolvent exists for the disorder averaging over Wick contractions for the partially transposed density matrix $\rho_{A_1A_2}^{T_2}$. We work in the regime where $S_{A_2} \ll S_{A_1} + S_B$. The resolvent equation is [67]

$$\lambda R(\lambda)^{i_1 i_2}_{j_1 j_2} = \delta^{i_1 i_2}_{j_1 j_2} + e^{S_B} \left(\sum_{m=1}^{\infty} \frac{R(\lambda)^{2m-2}}{e^{(2m-2)S_{A_2}}} R(\lambda)^{i_1 i_2}_{j_1 j_2} + \sum_{m=1}^{\infty} \frac{R(\lambda)^{2m-1}}{e^{(2m-2)S_{A_2}}} R(\lambda)^{i_1 i_2}_{j_1 j_2} \right) \tag{G.8}$$

Taking the trace:

$$\lambda R(\lambda) = e^{S_{A_1} + S_{A_2}} + e^{S_B} \left(\sum_{m=1}^{\infty} \frac{R(\lambda)^{2m-1}(1 + R(\lambda))}{e^{(2m-2)S_{A_2}}} \right) \tag{G.9}$$

and resumming gives the final resolvent equation

$$\lambda R(\lambda) = e^{S_{A_1}+S_{A_2}} + \frac{e^{S_B}}{e^{S_{A_2}}} \frac{R(\lambda)(1+R(\lambda))}{1-e^{2S_{A_2}}R(\lambda)^2} \tag{G.10}$$

We recognize this as the resolvent equation for the moments of a block transposed Wishart matrix. In the case $S_{A_2} = 0$ this reduces to the resolvent equation for the untransposed density matrix. This is a cubic equation and can be solved exactly, but the solution is not enlightening.

Previously, we saw that the moments of a Wishart matrix are given in closed form by a sum of Narayana numbers. An equivalent statement is that the inverse Stieltjes transform of the the resolvent defined by (G.4), a generating function for the Narayana numbers, gives the eigenvalue spectrum of a Wishart matrix. The inverse Stieltjes transform of the solution to (G.10), a generating function for the moments of a block transposed Wishart matrix ("block transposed Narayana numbers") will produce the eigenvalue spectrum of a block transposed Wishart matrix (the "negativity spectrum"). The block transposed Narayana numbers are not known in closed form; see [140] for a recursive definition.

Bibliography

[1] K. G. Wilson, "Renormalization group and critical phenomena. 1. Renormalization group and the Kadanoff scaling picture," *Phys. Rev. B* **4** (1971) 3174–3183.

[2] K. G. Wilson, "Renormalization group and critical phenomena. 2. Phase space cell analysis of critical behavior," *Phys. Rev. B* **4** (1971) 3184–3205.

[3] K. G. Wilson and M. E. Fisher, "Critical exponents in 3.99 dimensions," *Phys. Rev. Lett.* **28** (1972) 240–243.

[4] J. Polchinski, "Renormalization and Effective Lagrangians," *Nucl. Phys. B* **231** (1984) 269–295.

[5] X. Fan, T. G. Myers, B. A. D. Sukra, and G. Gabrielse, "Measurement of the Electron Magnetic Moment," *Phys. Rev. Lett.* **130** no. 7, (2023) 071801, arXiv:2209.13084 [physics.atom-ph].

[6] T. Aoyama, T. Kinoshita, and M. Nio, "Revised and Improved Value of the QED Tenth-Order Electron Anomalous Magnetic Moment," *Phys. Rev. D* **97** no. 3, (2018) 036001, arXiv:1712.06060 [hep-ph].

[7] D. Rauch *et al.*, "Cosmic Bell Test Using Random Measurement Settings from High-Redshift Quasars," *Phys. Rev. Lett.* **121** no. 8, (2018) 080403, arXiv:1808.05966 [quant-ph].

[8] S. Storz *et al.*, "Loophole-free Bell inequality violation with superconducting circuits," *Nature* **617** no. 7960, (2023) 265–270.

[9] **MICROSCOPE** Collaboration, P. Touboul *et al.*, "MICROSCOPE Mission: Final Results of the Test of the Equivalence Principle," *Phys. Rev. Lett.* **129** no. 12, (2022) 121102, arXiv:2209.15487 [gr-qc].

[10] **LISA** Collaboration, K. G. Arun *et al.*, "New horizons for fundamental physics with LISA," *Living Rev. Rel.* **25** no. 1, (2022) 4, arXiv:2205.01597 [gr-qc].

[11] T. Hartman, "Lectures on Quantum Gravity and Black Holes," 2015. http://www.hartmanhep.net/topics2015/gravity-lectures.pdf.

[12] **BICEP, Keck** Collaboration, P. A. R. Ade *et al.*, "Improved Constraints on Primordial Gravitational Waves using Planck, WMAP, and BICEP/Keck Observations through the 2018 Observing Season," *Phys. Rev. Lett.* **127** no. 15, (2021) 151301, arXiv:2110.00483 [astro-ph.CO].

[13] J. Oppenheim, "A Postquantum Theory of Classical Gravity?," *Phys. Rev. X* **13** no. 4, (2023) 041040, arXiv:1811.03116 [hep-th].

[14] J. Polchinski, *String theory. Vol. 1: An introduction to the bosonic string*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 12, 2007.

[15] J. Polchinski, *String theory. Vol. 2: Superstring theory and beyond*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 2007.

[16] L. Eberhardt, M. R. Gaberdiel, and R. Gopakumar, "The Worldsheet Dual of the Symmetric Product CFT," *JHEP* **04** (2019) 103, arXiv:1812.01007 [hep-th].

[17] L. Eberhardt, M. R. Gaberdiel, and R. Gopakumar, "Deriving the AdS_3/CFT_2 correspondence," *JHEP* **02** (2020) 136, arXiv:1911.00378 [hep-th].

[18] S. Caron-Huot, Z. Komargodski, A. Sever, and A. Zhiboedov, "Strings from Massive Higher Spins: The Asymptotic Uniqueness of the Veneziano Amplitude," *JHEP* **10** (2017) 026, arXiv:1607.04253 [hep-th].

[19] N. Arkani-Hamed, C. Cheung, C. Figueiredo, and G. N. Remmen, "Multiparticle Factorization and the Rigidity of String Theory," *Phys. Rev. Lett.* **132** no. 9, (2024) 091601, arXiv:2312.07652 [hep-th].

[20] J. D. Bekenstein, "Black holes and entropy," *Phys.Rev.* **D7** (1973) 2333–2346.

[21] S. Hawking, "Particle Creation by Black Holes," *Commun.Math.Phys.* **43** (1975) 199–220.

[22] D. N. Page, "Information in black hole radiation," *Phys. Rev. Lett.* **71** (1993) 3743–3746, arXiv:hep-th/9306083.

[23] D. N. Page, "Average entropy of a subsystem," *Phys. Rev. Lett.* **71** (1993) 1291–1294, arXiv:gr-qc/9305007.

[24] D. N. Page, "Time Dependence of Hawking Radiation Entropy," *JCAP* **09** (2013) 028, arXiv:1301.4995 [hep-th].

[25] L. Susskind, L. Thorlacius, and J. Uglum, "The Stretched horizon and black hole complementarity," *Phys. Rev. D* **48** (1993) 3743–3761, arXiv:hep-th/9306069.

[26] A. Almheiri, D. Marolf, J. Polchinski, and J. Sully, "Black Holes: Complementarity or Firewalls?," *JHEP* **02** (2013) 062, arXiv:1207.3123 [hep-th].

[27] D. Marolf and J. Polchinski, "Gauge/Gravity Duality and the Black Hole Interior," *Phys. Rev. Lett.* **111** (2013) 171301, arXiv:1307.4706 [hep-th].

[28] L. Susskind, "The Typical-State Paradox: Diagnosing Horizons with Complexity," *Fortsch. Phys.* **64** (2016) 84–91, arXiv:1507.02287 [hep-th].

[29] D. Stanford and Z. Yang, "Firewalls from wormholes," arXiv:2208.01625 [hep-th].

[30] L. V. Iliesiu, A. Levine, H. W. Lin, H. Maxfield, and M. Mezei, "On the non-perturbative bulk Hilbert space of JT gravity," arXiv:2403.08696 [hep-th].

[31] A. Blommaert, C.-H. Chen, and Y. Nomura, "Firewalls at exponentially late times," arXiv:2403.07049 [hep-th].

[32] J. M. Maldacena, "The Large N limit of superconformal field theories and supergravity," *Adv.Theor.Math.Phys.* **2** (1998) 231–252, arXiv:hep-th/9711200 [hep-th].

[33] S. Gubser, I. R. Klebanov, and A. M. Polyakov, "Gauge theory correlators from noncritical string theory," *Phys.Lett.* **B428** (1998) 105–114, arXiv:hep-th/9802109 [hep-th].

[34] E. Witten, "Anti-de Sitter space and holography," *Adv.Theor.Math.Phys.* **2** (1998) 253–291, arXiv:hep-th/9802150 [hep-th].

[35] G. 't Hooft, "Dimensional reduction in quantum gravity," *Conf. Proc. C* **930308** (1993) 284–296, arXiv:gr-qc/9310026.

[36] L. Susskind, "The World as a hologram," *J. Math. Phys.* **36** (1995) 6377–6396, arXiv:hep-th/9409089 [hep-th].

[37] O. Aharony, S. S. Gubser, J. M. Maldacena, H. Ooguri, and Y. Oz, "Large N field theories, string theory and gravity," *Phys. Rept.* **323** (2000) 183–386, arXiv:hep-th/9905111 [hep-th].

[38] S. Ryu and T. Takayanagi, "Holographic derivation of entanglement entropy from AdS/CFT," *Phys.Rev.Lett.* **96** (2006) 181602, arXiv:hep-th/0603001 [hep-th].

[39] S. Ryu and T. Takayanagi, "Aspects of Holographic Entanglement Entropy," *JHEP* **08** (2006) 045, arXiv:hep-th/0605073 [hep-th].

[40] A. Lewkowycz and J. Maldacena, "Generalized gravitational entropy," *JHEP* **08** (2013) 090, arXiv:1304.4926 [hep-th].

[41] V. E. Hubeny, M. Rangamani, and T. Takayanagi, "A Covariant holographic entanglement entropy proposal," *JHEP* **07** (2007) 062, arXiv:0705.0016 [hep-th].

[42] X. Dong, A. Lewkowycz, and M. Rangamani, "Deriving covariant holographic entanglement," *JHEP* **11** (2016) 028, arXiv:1607.07506 [hep-th].

[43] T. Faulkner, A. Lewkowycz, and J. Maldacena, "Quantum corrections to holographic entanglement entropy," *JHEP* **11** (2013) 074, arXiv:1307.2892 [hep-th].

[44] N. Engelhardt and A. C. Wall, "Quantum Extremal Surfaces: Holographic Entanglement Entropy beyond the Classical Regime," *JHEP* **01** (2015) 073, arXiv:1408.3203 [hep-th].

[45] X. Dong and A. Lewkowycz, "Entropy, Extremality, Euclidean Variations, and the Equations of Motion," *JHEP* **01** (2018) 081, arXiv:1705.08453 [hep-th].

[46] J. D. Bekenstein, "Generalized second law of thermodynamics in black hole physics," *Phys. Rev. D* **9** (1974) 3292–3300.

[47] S. Leutheusser and H. Liu, "Causal connectability between quantum systems and the black hole interior in holographic duality," *Phys. Rev. D* **108** no. 8, (2023) 086019, arXiv:2110.05497 [hep-th].

[48] S. A. W. Leutheusser, "Emergent Times in Holographic Duality," *Phys. Rev. D* **108** no. 8, (2023) 086020, arXiv:2112.12156 [hep-th].

[49] E. Witten, "Gravity and the crossed product," *JHEP* **10** (2022) 008, arXiv:2112.12828 [hep-th].

[50] V. Chandrasekaran, R. Longo, G. Penington, and E. Witten, "An algebra of observables for de Sitter space," *JHEP* **02** (2023) 082, arXiv:2206.10780 [hep-th].

[51] V. Chandrasekaran, G. Penington, and E. Witten, "Large N algebras and generalized entropy," arXiv:2209.10454 [hep-th].

[52] A. Rényi, "On Measures of Entropy and Information," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, pp. 547–561. University of California Press, 1961. http://projecteuclid.org/euclid.bsmsp/1200512181.

[53] G. Penington, "Entanglement Wedge Reconstruction and the Information Paradox," arXiv:1905.08255 [hep-th].

[54] 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** (2019) 063, arXiv:1905.08762 [hep-th].

[55] A. Almheiri, R. Mahajan, J. Maldacena, and Y. Zhao, "The Page curve of Hawking radiation from semiclassical geometry," *JHEP* **03** (2020) 149, arXiv:1908.10996 [hep-th].

[56] C. Akers and G. Penington, "Leading order corrections to the quantum extremal surface prescription," arXiv:2008.03319 [hep-th].

[57] C. Akers, A. Levine, G. Penington, and E. Wildenhain, "One-shot holography," arXiv:2307.13032 [hep-th].

[58] X. Dong, J. Kudler-Flam, and P. Rath, "A Modified Cosmic Brane Proposal for Holographic Renyi Entropy," arXiv:2312.04625 [hep-th].

[59] C. Akers, N. Engelhardt, and D. Harlow, "Simple holographic models of black hole evaporation," *JHEP* **08** (2020) 032, arXiv:1910.00972 [hep-th].

[60] A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian, and A. Tajdini, "Replica Wormholes and the Entropy of Hawking Radiation," *JHEP* **05** (2020) 013, arXiv:1911.12333 [hep-th].

[61] G. Penington, S. H. Shenker, D. Stanford, and Z. Yang, "Replica wormholes and the black hole interior," arXiv:1911.11977 [hep-th].

[62] D. Marolf and H. Maxfield, "Transcending the ensemble: baby universes, spacetime wormholes, and the order and disorder of black hole information," *JHEP* **08** (2020) 044, arXiv:2002.08950 [hep-th].

[63] D. Marolf and H. Maxfield, "Observations of Hawking radiation: the Page curve and baby universes," arXiv:2010.06602 [hep-th].

[64] J. McNamara and C. Vafa, "Baby Universes, Holography, and the Swampland," arXiv:2004.06738 [hep-th].

[65] D. Marolf and J. E. Santos, "AdS Euclidean wormholes," *Class. Quant. Grav.* **38** no. 22, (2021) 224002, arXiv:2101.08875 [hep-th].

[66] D. Marolf and S. McBride, "Simple Perturbatively Traversable Wormholes from Bulk Fermions," *JHEP* **11** (2019) 037, arXiv:1908.03998 [hep-th].

[67] X. Dong, S. McBride, and W. W. Weng, "Replica Wormholes and Holographic Entanglement Negativity," arXiv:2110.11947 [hep-th].

[68] S. McBride and F. Iniguez, "Entanglement Negativity Transitions in Chaotic Eigenstates," arXiv:2303.00018 [hep-th].

[69] X. Dong, S. McBride, and W. W. Weng, "Holographic tensor networks with bulk gauge symmetries," *JHEP* **02** (2024) 222, arXiv:2309.06436 [hep-th].

[70] A. Einstein and N. Rosen, "The Particle Problem in the General Theory of Relativity," *Phys. Rev.* **48** (Jul, 1935) 73–77.

[71] R. W. Fuller and J. A. Wheeler, "Causality and Multiply Connected Space-Time," *Phys. Rev.* **128** (Oct, 1962) 919–929.

[72] H. G. Ellis, "Ether flow through a drainhole: A particle model in general relativity," *Journal of Mathematical Physics* **14** no. 1, (1973) 104–118.

[73] M. S. Morris and K. S. Thorne, "Wormholes in spacetime and their use for interstellar travel: A tool for teaching general relativity," *American Journal of Physics* **56** no. 5, (1988) 395–412.

[74] M. S. Morris, K. S. Thorne, and U. Yurtsever, "Wormholes, Time Machines, and the Weak Energy Condition," *Phys. Rev. Lett.* **61** (Sep, 1988) 1446–1449.

[75] O. James, E. von Tunzelmann, P. Franklin, and K. S. Thorne, "Visualizing Interstellar's Wormhole," *Am. J. Phys.* **83** (2015) 486, arXiv:1502.03809 [gr-qc].

[76] J. L. Friedman, K. Schleich, and D. M. Witt, "Topological censorship," *Phys. Rev. Lett.* **71** (1993) 1486–1489, arXiv:gr-qc/9305017 [gr-qc].

[77] G. J. Galloway, K. Schleich, D. M. Witt, and E. Woolgar, "Topological censorship and higher genus black holes," *Phys. Rev.* **D60** (1999) 104039, arXiv:gr-qc/9902061 [gr-qc].

[78] E. Ayon-Beato, F. Canfora, and J. Zanelli, "Analytic self-gravitating Skyrmions, cosmological bounces and AdS wormholes," *Phys. Lett.* **B752** (2016) 201–205, arXiv:1509.02659 [gr-qc].

[79] F. Canfora, N. Dimakis, and A. Paliathanasis, "Topologically nontrivial configurations in the 4d Einstein-nonlinear σ-model system," *Phys. Rev. D* **96** (Jul, 2017) 025021. https://link.aps.org/doi/10.1103/PhysRevD.96.025021.

[80] N. Graham and K. D. Olum, "Achronal averaged null energy condition," *Phys. Rev.* **D76** (2007) 064001, arXiv:0705.3193 [gr-qc].

[81] A. C. Wall, "Proving the Achronal Averaged Null Energy Condition from the Generalized Second Law," *Phys. Rev.* **D81** (2010) 024038, arXiv:0910.5751 [gr-qc].

[82] J. M. Maldacena, "The Large N Limit of Superconformal Field Theories and Supergravity," http://arxiv.org/abs/hep-th/9711200v3.

[83] P. Gao, D. L. Jafferis, and A. Wall, "Traversable Wormholes via a Double Trace Deformation," *JHEP* **12** (2017) 151, arXiv:1608.05687 [hep-th].

[84] J. Maldacena, D. Stanford, and Z. Yang, "Diving into traversable wormholes," *Fortsch. Phys.* **65** no. 5, (2017) 1700034, arXiv:1704.05333 [hep-th].

[85] T. G. Mertens, G. J. Turiaci, and H. L. Verlinde, "Solving the Schwarzian via the Conformal Bootstrap," *JHEP* **08** (2017) 136, arXiv:1705.08408 [hep-th].

[86] J. Maldacena and X.-L. Qi, "Eternal traversable wormhole," arXiv:1804.00491 [hep-th].

[87] E. Caceres, A. S. Misobuchi, and M.-L. Xiao, "Rotating traversable wormholes in AdS," http://arxiv.org/abs/1807.07239v2.

[88] J. Maldacena and L. Susskind, "Cool horizons for entangled black holes," *Fortsch. Phys.* **61** (2013) 781–811, arXiv:1306.0533 [hep-th].

[89] L. Susskind and Y. Zhao, "Teleportation through the wormhole," *Phys. Rev.* **D98** no. 4, (2018) 046016, arXiv:1707.04354 [hep-th].

[90] L. Susskind, "Dear Qubitzers, GR=QM," arXiv:1708.03040 [hep-th].

[91] J. Maldacena, A. Milekhin, and F. Popov, "Traversable wormholes in four dimensions," arXiv:1807.04726 [hep-th].

[92] Z. Fu, B. Grado-White, and D. Marolf, "A perturbative perspective on self-supporting wormholes," arXiv:1807.07917 [hep-th].

[93] Z. Fu, B. Grado-White, and D. Marolf, "Traversable Asymptotically Flat Wormholes with Short Transit Times," arXiv:1908.03273 [hep-th].

[94] M. Banados, C. Teitelboim, and J. Zanelli, "The Black hole in three-dimensional space-time," *Phys. Rev. Lett.* **69** (1992) 1849–1851, arXiv:hep-th/9204099 [hep-th].

[95] M. Banados, M. Henneaux, C. Teitelboim, and J. Zanelli, "Geometry of the (2+1) black hole," *Phys. Rev.* **D48** (1993) 1506–1525, arXiv:gr-qc/9302012 [gr-qc]. [Erratum: Phys. Rev.D88,069902(2013)].

[96] J. Louko and D. Marolf, "Single exterior black holes and the AdS / CFT conjecture," *Phys. Rev.* **D59** (1999) 066002, arXiv:hep-th/9808081 [hep-th].

[97] I. Ichinose and Y. Satoh, "Entropies of scalar fields on three-dimensional black holes," *Nucl. Phys.* **B447** (1995) 340–372, arXiv:hep-th/9412144 [hep-th].

[98] M. de Jesus Anguiano Galicia and A. Bashir, "Fermions in odd space-time dimensions: Back to basics," *Few Body Syst.* **37** (2005) 71–78, arXiv:hep-ph/0502089 [hep-ph].

[99] D. Freedman and A. Van Proeyen, *Supergravity*. Cambridge University Press, 2012.

[100] P. Di Francesco, P. Mathieu, and D. Senechal, *Conformal Field Theory*. Graduate Texts in Contemporary Physics. Springer-Verlag, New York, 1997.

[101] W. Mück, "Spinor parallel propagator and Green's function in maximally symmetric spaces," *J. Phys.* **A33** (2000) 3021–3026, arXiv:hep-th/9912059 [hep-th].

[102] B. Allen and T. Jacobson, "Vector two-point functions in maximally symmetric spaces," *Comm. Math. Phys.* **103** no. 4, (1986) 669–692.

[103] S. Hirano, Y. Lei, and S. van Leuven, "Information Transfer and Black Hole Evaporation via Traversable BTZ Wormholes," arXiv:1906.10715 [hep-th].

[104] B. Freivogel, D. A. Galante, D. Nikolakopoulou, and A. Rotundo, "Traversable wormholes in AdS and bounds on information transfer," arXiv:1907.13140 [hep-th].

[105] G. Vidal and R. Werner, "Computable measure of entanglement," *Phys. Rev. A* **65** (2002) 032314, arXiv:quant-ph/0102117.

[106] M. Plenio, "Logarithmic Negativity: A Full Entanglement Monotone That is not Convex," *Phys. Rev. Lett.* **95** no. 9, (2005) 090503, arXiv:quant-ph/0505071.

[107] K. Audenaert, M. Plenio, and J. Eisert, "Entanglement Cost under Positive-Partial-Transpose-Preserving Operations," *Phys. Rev. Lett.* **90** no. 2, (2003) 027901.

[108] P. Calabrese, J. Cardy, and E. Tonni, "Entanglement negativity in quantum field theory," *Phys. Rev. Lett.* **109** (2012) 130502, arXiv:1206.3092 [cond-mat.stat-mech].

[109] P. Calabrese, J. Cardy, and E. Tonni, "Entanglement negativity in extended systems: A field theoretical approach," *J. Stat. Mech.* **1302** (2013) P02008, arXiv:1210.5359 [cond-mat.stat-mech].

[110] M. Rangamani and M. Rota, "Comments on Entanglement Negativity in Holographic Field Theories," *JHEP* **10** (2014) 060, arXiv:1406.6989 [hep-th].

[111] P. Calabrese, J. Cardy, and E. Tonni, "Finite temperature entanglement negativity in conformal field theory," *J. Phys. A* **48** no. 1, (2015) 015006, arXiv:1408.3043 [cond-mat.stat-mech].

[112] P. Chaturvedi, V. Malvimat, and G. Sengupta, "Entanglement negativity, Holography and Black holes," *Eur. Phys. J. C* **78** no. 6, (2018) 499, arXiv:1602.01147 [hep-th].

[113] P. Chaturvedi, V. Malvimat, and G. Sengupta, "Holographic Quantum Entanglement Negativity," *JHEP* **05** (2018) 172, arXiv:1609.06609 [hep-th].

[114] P. Jain, V. Malvimat, S. Mondal, and G. Sengupta, "Holographic entanglement negativity conjecture for adjacent intervals in AdS_3/CFT_2," *Phys. Lett. B* **793** (2019) 104–109, arXiv:1707.08293 [hep-th].

[115] P. Jain, V. Malvimat, S. Mondal, and G. Sengupta, "Holographic entanglement negativity for adjacent subsystems in $\mathrm{AdS}_{d+1}/\mathrm{CFT}_d$," *Eur. Phys. J. Plus* **133** no. 8, (2018) 300, arXiv:1708.00612 [hep-th].

[116] V. Malvimat and G. Sengupta, "Entanglement negativity at large central charge," arXiv:1712.02288 [hep-th].

[117] J. Kudler-Flam and S. Ryu, "Entanglement negativity and minimal entanglement wedge cross sections in holographic theories," *Phys. Rev. D* **99** no. 10, (2019) 106014, arXiv:1808.00446 [hep-th].

[118] K. Tamaoka, "Entanglement Wedge Cross Section from the Dual Density Matrix," *Phys. Rev. Lett.* **122** no. 14, (2019) 141601, arXiv:1809.09109 [hep-th].

[119] V. Malvimat, S. Mondal, B. Paul, and G. Sengupta, "Holographic entanglement negativity for disjoint intervals in AdS_3/CFT_2," *Eur. Phys. J. C* **79** no. 3, (2019) 191, arXiv:1810.08015 [hep-th].

[120] J. Kudler-Flam, M. Nozaki, S. Ryu, and M. T. Tan, "Quantum vs. classical information: operator negativity as a probe of scrambling," *JHEP* **01** (2020) 031, arXiv:1906.07639 [hep-th].

[121] Y. Kusuki, J. Kudler-Flam, and S. Ryu, "Derivation of Holographic Negativity in $\mathrm{AdS}_3/\mathrm{CFT}_2$," *Phys. Rev. Lett.* **123** no. 13, (2019) 131603, arXiv:1907.07824 [hep-th].

[122] J. Kumar Basak, H. Parihar, B. Paul, and G. Sengupta, "Holographic entanglement negativity for disjoint subsystems in $\mathrm{AdS}_{\mathrm{d+1}}/\mathrm{CFT}_{\mathrm{d}}$," arXiv:2001.10534 [hep-th].

[123] J. Kumar Basak, V. Malvimat, H. Parihar, B. Paul, and G. Sengupta, "On minimal entanglement wedge cross section for holographic entanglement negativity," arXiv:2002.10272 [hep-th].

[124] J. Kudler-Flam, Y. Kusuki, and S. Ryu, "The quasi-particle picture and its breakdown after local quenches: mutual information, negativity, and reflected entropy," arXiv:2008.11266 [hep-th].

[125] T.-C. Lu and T. Grover, "Entanglement transitions as a probe of quasiparticles and quantum thermalization," *Phys. Rev. B* **102** no. 23, (2020) 235110, arXiv:2008.11727 [cond-mat.stat-mech].

[126] J. Kumar Basak, D. Basu, V. Malvimat, H. Parihar, and G. Sengupta, "Islands for Entanglement Negativity," arXiv:2012.03983 [hep-th].

[127] H. Shapourian, S. Liu, J. Kudler-Flam, and A. Vishwanath, "Entanglement Negativity Spectrum of Random Mixed States: A Diagrammatic Approach," *PRX Quantum* **2** no. 3, (2021) 030347, arXiv:2011.01277 [cond-mat.str-el].

[128] X. Dong, X.-L. Qi, and M. Walter, "Holographic entanglement negativity and replica symmetry breaking," arXiv:2101.11029 [hep-th].

[129] J. Kumar Basak, D. Basu, V. Malvimat, H. Parihar, and G. Sengupta, "Page Curve for Entanglement Negativity through Geometric Evaporation," arXiv:2106.12593 [hep-th].

[130] S. Vardhan, J. Kudler-Flam, H. Shapourian, and H. Liu, "Bound entanglement in thermalized states and black hole radiation," arXiv:2110.02959 [hep-th].

[131] X. Dong and H. Wang, "Enhanced corrections near holographic entanglement transitions: a chaotic case study," *JHEP* **11** (2020) 007, arXiv:2006.10051 [hep-th].

[132] D. Marolf, S. Wang, and Z. Wang, "Probing phase transitions of holographic entanglement entropy with fixed area states," *JHEP* **12** (2020) 084, arXiv:2006.10089 [hep-th].

[133] J. Kudler-Flam, V. Narovlansky, and S. Ryu, "Negativity Spectra in Random Tensor Networks and Holography," arXiv:2109.02649 [hep-th].

[134] S. Vardhan, J. Kudler-Flam, H. Shapourian, and H. Liu, "Mixed-state entanglement and information recovery in thermalized states and evaporating black holes," arXiv:2112.00020 [hep-th].

[135] A. Peres, "Separability criterion for density matrices," *Phys. Rev. Lett.* **77** (1996) 1413–1415, arXiv:quant-ph/9604005.

[136] M. Horodecki, P. Horodecki, and R. Horodecki, "On the necessary and sufficient conditions for separability of mixed quantum states," *Phys. Lett. A* **223** (1996) 1, arXiv:quant-ph/9605038.

[137] X. Dong, "The Gravity Dual of Renyi Entropy," *Nature Commun.* **7** (2016) 12472, arXiv:1601.06788 [hep-th].

[138] I. Kourkoulou and J. Maldacena, "Pure states in the SYK model and nearly-AdS_2 gravity," arXiv:1707.02325 [hep-th].

[139] A. Almheiri, "Holographic Quantum Error Correction and the Projected Black Hole Interior," arXiv:1810.02055 [hep-th].

[140] T. Banica and I. Nechita, "Asymptotic eigenvalue distributions of block-transposed Wishart matrices," *Journal of Theoretical Probability* **26** no. 3, (2013) 855–869.

[141] C. Akers and P. Rath, "Holographic Renyi Entropy from Quantum Error Correction," *JHEP* **05** (2019) 052, arXiv:1811.05171 [hep-th].

[142] J. M. Deutsch, "Quantum statistical mechanics in a closed system," *Phys. Rev. A* **43** (Feb, 1991) 2046–2049. https://link.aps.org/doi/10.1103/PhysRevA.43.2046.

[143] M. Srednicki, "Chaos and quantum thermalization," *Phys. Rev. E* **50** (Aug, 1994) 888–901. https://link.aps.org/doi/10.1103/PhysRevE.50.888.

[144] M. Srednicki, "Thermal fluctuations in quantized chaotic systems," *J. Phys. A* **29** (1996) L75–L79, arXiv:chao-dyn/9511001.

[145] M. Srednicki, "The approach to thermal equilibrium in quantized chaotic systems," *Journal of Physics A: Mathematical and General* **32** no. 7, (Feb, 1999) 1163. https://dx.doi.org/10.1088/0305-4470/32/7/007.

[146] J. Maldacena, S. H. Shenker, and D. Stanford, "A bound on chaos," *JHEP* **08** (2016) 106, arXiv:1503.01409 [hep-th].

[147] J. R. Garrison and T. Grover, "Does a single eigenstate encode the full Hamiltonian?," *Phys. Rev. X* **8** no. 2, (2018) 021026, arXiv:1503.00729 [cond-mat.str-el].

[148] A. Dymarsky, N. Lashkari, and H. Liu, "Subsystem ETH," *Phys. Rev. E* **97** (2018) 012140, arXiv:1611.08764 [cond-mat.stat-mech].

[149] A. Dymarsky and H. Liu, "New characteristic of quantum many-body chaotic systems," *Phys. Rev. E* **99** no. 1, (2019) 010102, arXiv:1702.07722 [cond-mat.stat-mech].

[150] N. Lashkari, A. Dymarsky, and H. Liu, "Universality of Quantum Information in Chaotic CFTs," *JHEP* **03** (2018) 070, arXiv:1710.10458 [hep-th].

[151] T.-C. Lu and T. Grover, "Renyi Entropy of Chaotic Eigenstates," arXiv:1709.08784 [cond-mat.stat-mech].

[152] L. Vidmar and M. Rigol, "Entanglement Entropy of Eigenstates of Quantum Chaotic Hamiltonians," *Physical Review Letters* **119** no. 22, (Nov, 2017) . https://doi.org/10.1103%2Fphysrevlett.119.220603.

[153] C. Murthy and M. Srednicki, "Structure of chaotic eigenstates and their entanglement entropy," *Phys. Rev. E* **100** no. 2, (2019) 022131, arXiv:1906.04295 [cond-mat.stat-mech].

[154] X. Dong, D. Harlow, and D. Marolf, "Flat entanglement spectra in fixed-area states of quantum gravity," *JHEP* **10** (2019) 240, arXiv:1811.05382 [hep-th].

[155] X. Dong, D. Marolf, P. Rath, A. Tajdini, and Z. Wang, "The spacetime geometry of fixed-area states in gravitational systems," *JHEP* **08** (2022) 158, arXiv:2203.04973 [hep-th].

[156] J. M. Deutsch, "Thermodynamic entropy of a many-body energy eigenstate," *New Journal of Physics* **12** no. 7, (Jul, 2010) 075021. https://dx.doi.org/10.1088/1367-2630/12/7/075021.

[157] T.-C. Lu and T. Grover, "Renyi entropy of chaotic eigenstates," *Physical Review E* **99** no. 3, (Mar, 2019) . https://doi.org/10.1103%2Fphysreve.99.032111.

[158] J. Kudler-Flam, V. Narovlansky, and S. Ryu, "Distinguishing Random and Black Hole Microstates," *PRX Quantum* **2** no. 4, (2021) 040340, arXiv:2108.00011 [hep-th].

[159] G. Vidal and R. F. Werner, "Computable measure of entanglement," *Phys. Rev. A* **65** (Feb, 2002) 032314. https://link.aps.org/doi/10.1103/PhysRevA.65.032314.

[160] J.-C. Novelli and J.-Y. Thibon, "Hopf Algebras of m-permutations, (m + 1)-ary trees, and m-parking functions," arXiv:1403.5962 [math.CO].

[161] S. Dutta and T. Faulkner, "A canonical purification for the entanglement wedge cross-section," arXiv:1905.00577 [hep-th].

[162] C. Akers, T. Faulkner, S. Lin, and P. Rath, "Reflected entropy in random tensor networks," *JHEP* **05** (2022) 162, arXiv:2112.09122 [hep-th].

[163] C. Akers, T. Faulkner, S. Lin, and P. Rath, "The Page curve for reflected entropy," *JHEP* **06** (2022) 089, arXiv:2201.11730 [hep-th].

[164] C. Akers, T. Faulkner, S. Lin, and P. Rath, "Reflected entropy in random tensor networks II: a topological index from the canonical purification," arXiv:2210.15006 [hep-th].

[165] A. Almheiri, X. Dong, and D. Harlow, "Bulk Locality and Quantum Error Correction in AdS/CFT," *JHEP* **04** (2015) 163, arXiv:1411.7041 [hep-th].

[166] F. Verstraete and J. I. Cirac, "Renormalization algorithms for quantum-many body systems in two and higher dimensions," arXiv:cond-mat/0407066.

[167] B. Swingle, "Entanglement Renormalization and Holography," *Phys. Rev.* **D86** (2012) 065007, arXiv:0905.1317 [cond-mat.str-el].

[168] P. Hayden, S. Nezami, X.-L. Qi, N. Thomas, M. Walter, and Z. Yang, "Holographic duality from random tensor networks," *JHEP* **11** (2016) 009, arXiv:1601.01694 [hep-th].

[169] H. F. Jia and M. Rangamani, "Petz reconstruction in random tensor networks," *JHEP* **10** (2020) 050, arXiv:2006.12601 [hep-th].

[170] F. Pastawski, B. Yoshida, D. Harlow, and J. Preskill, "Holographic quantum error-correcting codes: Toy models for the bulk/boundary correspondence," *JHEP* **06** (2015) 149, arXiv:1503.06237 [hep-th].

[171] N. Cheng, C. Lancien, G. Penington, M. Walter, and F. Witteveen, "Random tensor networks with nontrivial links," arXiv:2206.10482 [quant-ph].

[172] D. Harlow, "The Ryu–Takayanagi Formula from Quantum Error Correction," *Commun. Math. Phys.* **354** no. 3, (2017) 865–912, arXiv:1607.03901 [hep-th].

[173] C. Cao, "Stabilizer Codes Have Trivial Area Operators," arXiv:2306.14996 [hep-th].

[174] W. Donnelly, B. Michel, D. Marolf, and J. Wien, "Living on the Edge: A Toy Model for Holographic Reconstruction of Algebras with Centers," *JHEP* **04** (2017) 093, arXiv:1611.05841 [hep-th].

[175] W. Donnelly, "Decomposition of entanglement entropy in lattice gauge theory," *Phys. Rev. D* **85** (2012) 085004, arXiv:1109.0036 [hep-th].

[176] J. B. Kogut and L. Susskind, "Hamiltonian Formulation of Wilson's Lattice Gauge Theories," *Phys. Rev. D* **11** (1975) 395–408.

[177] K. Dolev, V. Calvera, S. S. Cree, and D. J. Williamson, "Gauging the bulk: generalized gauging maps and holographic codes," *JHEP* **05** (2022) 158, arXiv:2108.11402 [quant-ph].

[178] W. Donnelly, "Entanglement entropy and nonabelian gauge symmetry," *Class. Quant. Grav.* **31** no. 21, (2014) 214003, arXiv:1406.7304 [hep-th].

[179] S. Aoki, T. Iritani, M. Nozaki, T. Numasawa, N. Shiba, and H. Tasaki, "On the definition of entanglement entropy in lattice gauge theories," *JHEP* **06** (2015) 187, arXiv:1502.04267 [hep-th].

[180] S. Ghosh, R. M. Soni, and S. P. Trivedi, "On The Entanglement Entropy For Gauge Theories," *JHEP* **09** (2015) 069, arXiv:1501.02593 [hep-th].

[181] H. Casini, M. Huerta, J. M. Magan, and D. Pontello, "Entropic order parameters for the phases of QFT," *JHEP* **04** (2021) 277, arXiv:2008.11748 [hep-th].

[182] T. Hartman, "Entanglement Entropy at Large Central Charge," arXiv:1303.6955 [hep-th].

[183] T. Faulkner, "The Entanglement Renyi Entropies of Disjoint Intervals in AdS/CFT," arXiv:1303.7221 [hep-th].

[184] K. Jensen, J. Sorce, and A. Speranza, "Generalized entropy for general subregions in quantum gravity," arXiv:2306.01837 [hep-th].

[185] R. M. Soni, "A Type I Approximation of the Crossed Product," arXiv:2307.12481 [hep-th].

[186] H. Casini, M. Huerta, and J. A. Rosabal, "Remarks on entanglement entropy for gauge fields," *Phys. Rev. D* **89** no. 8, (2014) 085012, arXiv:1312.1183 [hep-th].

[187] M. Kaplan and D. Marolf, "The action of HRT-areas as operators in semiclassical gravity," *JHEP* **08** (2022) 102, arXiv:2203.04270 [hep-th].

[188] H. Nastase, *Introduction to the ADS/CFT Correspondence*. Cambridge University Press, 2015.

[189] V. E. Ambruş, *Dirac fermions on rotating space-times*. PhD thesis, University of Sheffield, 2014.

[190] E. Brezin, C. Itzykson, G. Parisi, and J. B. Zuber, "Planar Diagrams," *Commun. Math. Phys.* **59** (1978) 35.

[191] F. Cachazo and B. G. Umbert, "Connecting Scalar Amplitudes using The Positive Tropical Grassmannian," arXiv:2205.02722 [hep-th].