Thermodynamics of dynamical black holes

Zhiwei Wang

Doctor of Philosophy

University of York
Computer Science

September 2019

To the people who illuminate my life

Abstract

The thermal nature of black hole horizons has engendered a revolution in physics since its discovery in the 1970s. The first result of this thesis is to extend the analysis of black hole thermodynamics from horizons to arbitrary ordinary surfaces for static spacetimes. It is proved that ordinary surfaces do not obey the same first law as black hole horizons, even for static spacetimes. This result undermines the thermodynamic assumptions of the emergent gravity program.

The second result of this thesis is to rigorously generalize black hole thermodynamics to dynamical spacetimes. To achieve this, we firstly transform the physical ADM mass into a suitable covariant form for applying Stokes' theorem. Then we analyze the energy changes under metric perturbations for dynamical non-rotating spacetimes. It is proved that dynamical horizons still have a perfect analogue to the first law and a Hawking temperature, hence still behave thermodynamically. Analytic calculations for binary interacting black holes show that the temperature along horizons of dynamical black holes is generally non-uniform, with equilibrium replaced by local-equilibrium behavior. The local nature of black hole thermodynamics supports long-held intuitive claims on how information is encoded on a black hole's surface.

Finally, the generally non-uniform temperature of interacting black holes induces a non-equilibrium entropic force between interacting black holes, which cannot be encoded in general relativity. Extrapolating this force to coalescing black holes provides a proposal for the first direct test for the thermodynamic nature of black hole entropy from future precision measurements of gravitational waveforms.

Contents

List of Figures

Acknowledgements

My supervisor Professor Samuel L. Braunstein is the greatest physicist I can imagine and I am so proud to be his student. He deserves my sincere appreciation and respect for his patience and insightful guidance, especially for teaching me to be crazy. I also appreciate Professor Netta Cohen for her delicious dinners, helpful suggestions and good advice.

I am grateful to my family, my sister in particular, for taking care of our parents. I am sorry for missing all the recent five Chinese New Years.

Jilin University and the China Scholarships Council supported my entire Ph.D., and I am grateful to them for this support. I also appreciate all the help I have obtained from the Tang Aoqing Honors Program in Science at Jilin University.

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Some parts of this thesis have been published in conference proceedings and journals; where items were published jointly with collaborators, the author of this thesis is responsible for the material presented here. For each published item the primary author is the first listed author.

Chapter 2 includes:

Zhi-Wei Wang & Samuel L. Braunstein. Surfaces away from horizons are not thermodynamic. In *Nature communications*, pages 2977, July 2018.
For static asymptotically-flat spacetimes, we prove that the first law of black hole mechanics remains an excellent approximation for stretched horizons. For surfaces away from horizons in the emergent gravity program the first law fails (except for spherically-symmetric scenarios) thus undermining the key thermodynamic assumption of this program.

Chapter 3 includes:

Zhi-Wei Wang & Samuel L. Braunstein. Black hole mechanics for dynamical spacetimes. Submitted.
For asymptotically-flat dynamical spacetimes, we prove that the physical energy reduces to a simple covariant form and use this energy definition to generalize the first law of black hole thermodynamics to dynamical settings. We find that generally the temperature along the horizons of dynamical black holes is non-uniform,

with equilibrium replaced by local equilibrium behavior. These generalizations and discoveries will have impact on both black hole physics and quantum gravity.

Chapter 4 includes:

Zhi-Wei Wang & Samuel L. Braunstein. Non-equilibrium entropic forces: From molecular chains to black holes. In preparation.
We find a non-equilibrium entropic force may exist among interacting black holes. We illustrate such non-equilibrium entropic forces for two scenarios: a microscopic, molecular system, and a macroscopic, black hole system. In both cases, the entropic force vanishes at equilibrium. In the latter scenario, extrapolating the force to coalescing black holes provides the first direct test for the thermodynamic nature of black hole entropy from future precision measurements of gravitational waveforms. Since this entropic force cannot be encoded in the classical Einstein field equations, its observation would constitute the first unequivocal signature of quantum gravity. By contrast, its absence would compel us to reject the statistical nature of black hole entropy initiating a major reassessment of the theoretical foundations of physics.

List of symbols

	Greek indices run from 0 to 3
	lower-case Latin indices run from 1 to 3
	upper-case Latin indices run from 2 to 3
x^μ	coordinate of the entire spacetimes
y^a	coordinate on each hypersurface Σ_f
$_{,\mu}$	ordinary derivative ($= \frac{\partial}{\partial x^\mu}$)
$g_{\mu\nu}$	metric
$R_{\mu\nu\alpha\beta}$	Riemann curvature tensor
$R_{\mu\nu}$	Ricci curvature tensor ($= R^\alpha{}_{\mu\alpha\nu}$)
R	Ricci scalar ($= R^\alpha{}_\alpha$)
$T^{\mu\nu}$	energy-momentum tensor
$\mathcal{M}$	4-dimensional spacetime in this thesis
Σ	3-dimensional hypersurface in this thesis
$\Sigma_{\mathbf{EG}}$	3-dimensional hypersurface in Ref. [59]
$\partial\Sigma$	boundary of Σ
$\partial\Sigma_\infty$	outer boundary of Σ
$\partial\Sigma_{\mathbf{inner}}$	inner boundaries of Σ
$\partial\Sigma_{\mathbf{BH}}$	inner boundary of Σ (black hole horizon)
$\partial\Sigma_{\mathbf{HS}}$	inner boundary of Σ (holographic screen)
γ_{ij}	induced metric on the hypersurface Σ
$\gamma^{(\Sigma)}$	determinant of the induced metric on Σ

$\gamma^{(\partial\Sigma)}$ determinant of the induced metric on $\partial\Sigma$

dA area element on $\partial\Sigma$ ($= \sqrt{\gamma^{(\partial\Sigma)}}\, d^2z$)

$\hat{T}^\mu$ timelike unit 4-vector normal to Σ

$\hat{N}^\mu$ spacelike unit 4-vector normal to $\partial\Sigma$

$\gamma^{\mu\nu}$ projector onto Σ, $P^{\mu\nu} = g^{\mu\nu} + \hat{T}^\mu \hat{T}^\nu$

$P^{\mu\nu}$ projector onto $\partial\Sigma$, $P^{\mu\nu} = g^{\mu\nu} + \hat{T}^\mu \hat{T}^\nu - \hat{N}^\mu \hat{N}^\nu$

$\mathcal{N}$ lapse function ($= 1/\hat{T}^0$)

$M^{\mathbf{Komar}}$ net Komar mass–energy in a spacelike hypersurface

$M^{\mathbf{ADM}}$ net ADM mass–energy in a spacelike hypersurface

$p_\mu^{\mathbf{ADM}}$ the ADM 4-momentum vector

$E(\xi), E$ net mass-energy at spatial infinity ($= M^{\mathrm{ADM}}$)

$\kappa, \kappa(\xi)$ surface gravity, see Eqs. (2.7) (3.43)

ξ^μ the vector field used to define $E(\xi)$

K^μ the Killing vector field used to define Komar mass-energy

l^μ outgoing null normal vector

n^μ ingoing null normal vector

$\theta^{(l)}$ expansion of the outgoing null normal vector l^μ

$\sigma_+^{(l)}$, $\sigma_\times^{(l)}$ shears of the outgoing null normal vector l^μ

$\mathfrak{L}_\xi(X)$ Lie derivative of X along ξ^μ

$h_{\mu\nu}$ diffeomorphic variation of the metric ($= \delta g_{\mu\nu}$)

Chapter 1

Introduction

1.1 Introduction of the thesis

The major theme of this thesis is to rigorously generalize black hole thermodynamics to dynamical non-rotating spacetimes and give a framework for testing this theory. This thesis is based on three research papers and all of them are joint work with Professor Samuel L. Braunstein. By dynamical spacetime, we follow the conventional definition, i.e., the absence of a timelike Killing vector in the exterior of any horizons [36].

In 1973, Bardeen, Carter and Hawking summarized some of the interesting discoveries of stationary black holes found in the early 1970s as "The four laws of black hole mechanics" which are in analogy with the four laws of classical thermodynamics [11]. Despite the perfect analogy between these two mechanics, many physicists were skeptical about the thermal nature of black hole horizons. In brief, in 1975, through studying the evolution of quantum fields in a spacetime with a black hole, Hawking proved that black holes would radiate particles and hence have a temperature [31]. This discovery caused physicists to widely accept the claim that stationary black holes are real thermodynamic systems and obey the four laws of thermodynamics.

Since the 1970s, the thermodynamic properties proved for stationary black hole horizons have been taken and used in several more general scenarios without rigorous proof. For example, in 1995 Jacobson came up with a beautiful argument to derive the equations of general relativity [37]. He assumed that the horizon associated with an arbitrary accelerated observer (the Rindler horizon) has well-defined local thermodynamic behavior at each point of its surface [37]. He then expressed the boosted flux of matter across a small patch of the horizon in terms of the energy-momentum tensor and used the Raychaudhuri equation to express the change of the local horizon area (i.e., entropy) in terms of the curvature tensor of spacetime. In the end, he used the first law of thermodynamics to quantitatively link the change of the local horizon entropy and the energy flux, finally yielding the Einstein field equations [37].

If Jacobson's thermodynamic assumptions for Rindler horizons is a small step, then Verlinde must have taken a big step by assuming even ordinary surfaces also behave thermodynamically. In 2011, Verlinde heuristically derived Einstein's equations by assuming that ordinary surfaces with equal Newtonian potential have the same thermodynamic behavior as black hole horizons [59]. He called this the emergent gravity program. He used a list of famous results in his arguments, e.g., the Unruh temperature [55]. Although these results are individually proven and accepted, whether they can be connected to each other by his heuristic arguments is unclear. Instead of discussing whether every step in his argument is valid, I will directly tackle the questions of whether the thermodynamic assumptions in his program are consistent with general relativity.

Based on general relativity, the first result of this thesis is to generalize the analysis of black hole thermodynamics from horizons to ordinary surfaces [64]. To do this, we closely followed in Bardeen, Carter and Hawking's footsteps and extended their original analysis from black hole horizons to ordinary surfaces. We prove that, for static spacetimes, ordinary surfaces generally do not obey the same first law as the black hole horizon. This means general relativity is not consistent with the thermodynamic assumptions in Verlinde's emergent gravity program and hence undermines the foundations of this program [64]. These results are summarized in chapter 2 of this thesis.

Despite the disruptive effects of black hole thermodynamics on fundamental physics, Bardeen, Carter, and Hawking's analysis relied on the assumption that both the initial and perturbed spacetimes were stationary. In 1994, Iyer and Wald rigorously generalized this result to allow for arbitrary, though infinitesimal, perturbations from initially stationary spacetimes [36]. Since the initial spacetime in these two analyses is stationary, important quantities such as surface gravity are only determined for stationary black holes. As the surface gravity plays the role of temperature, this means that these rigorous analyses are limited to scenarios at equilibrium and cannot exhibit out-of-equilibrium behavior that might be expected for truly dynamical black holes.

Distinct from Bardeen et al. and Iyer and Wald, Ashtekar proposed the so-called isolated horizon based on a geometric analysis to extend black hole thermodynamics to more general scenarios. The key goal of an isolated horizon is to model the state of a post-collapse black hole ‘after’ ring down [4–8]. In particular, such a horizon is specifically defined to be at equilibrium with its exterior and hence has a well-defined and provably uniform surface gravity (temperature) [4–8]. Although the uniform temperature makes the isolated horizon naturally support the zeroth law of black hole thermodynamics, it also limits its applicability. For example, the Brill-Lindquist initial conditions describe interacting black holes, however, it has been found that the horizons for these initial conditions do not correspond to isolated horizons, except as an approximation for large inter-black hole distance [39]. Therefore, the concept of an isolated horizon is insufficient to describe the thermodynamic behaviors of truly dynamical or interacting black holes.

After completing my first paper, I aimed at trying to generalize black hole thermodynamics to truly dynamical spacetimes [66]. Again following in Bardeen et al.’s footsteps and the study in chapter 2, I firstly transform the physical ADM energy into a covariant form suitable for applying Stokes’ theorem. By applying Stokes’ theorem, the resulting ADM quantities then provide a candidate expression for the surface gravity (corresponding to the temperature) in dynamical spacetimes – a quantity for which there is currently no consensus [42]. I then study how this energy changes under perturbations to the metric and matter. My analysis shows that the surface of vanishing expansion of outgoing null-normal congruences obeys a first-law analogue for dynamical non-rotating spacetimes and hence might be the relevant thermodynamic surface [66].

For this analogue to become an actual law of thermodynamics, I must first confirm that the surface gravity – the analogue to temperature – actually corresponds to the Hawking temperature for the dynamical black holes and horizons under study. This was Hawking’s key realization for stationary black holes upon the discovery of his eponymous radiation [31]. Repeating Hawking’s original analysis remains daunting, so instead I compute the tunneling temperature [66] in this dynamical non-rotating setting and find that it agrees with the temperature conjectured from my candidate surface gravity. This confirms that horizons of non-rotating dynamical black holes do behave thermodynamically. I then analytically calculate the temperature of a pair of interacting binary black holes and find their temperatures are generally non-uniform. All these results are summarized in chapter 3.

In chapter 4, I further analyze the interacting binary black hole system and find that the non-uniform temperatures of interacting black holes should induce a novel entropic

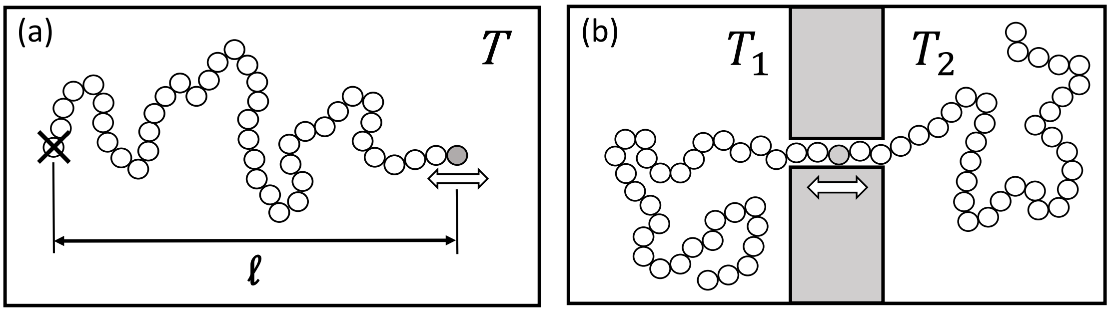

Figure 1.1: Examples of a system experiencing: (a) a conventional equilibrium entropic force: A flaccid negligible-weight long-chain molecule with end-to-end separation ℓ sits in a bath at temperature T. One end of the molecule is fixed so that the force is experienced at the free end; (b) a non-equilibrium entropic force: Here our long-chain molecule is threaded across a nanopore between two chambers held at respective temperatures T_i.

force [65]. This new kind of entropic effect may be illustrated by considering a molecular chain model. When one places such a molecule in a heat bath and fixes one end of the chain, we feel an entropic force at the free end [69], see Fig. 1.1 (a). Since the molecular chain is in an equilibrium thermal bath, we call such a force an equilibrium entropic force. Now consider two thermal baths of different temperatures connected by a nanoscale channel with a long-chain molecule passing through the channel, see Fig. 1.1 (b). As long as both of the lengths of the chain into each chamber are not too short, the different temperatures of the two chambers will induce an entropic force dragging the molecular chain toward the higher temperature chamber, this effect disappearing when the two chambers have the same temperature. Since this new entropic effect is induced by the non-uniform temperature of the thermal system, we call it a non-equilibrium entropic force. We note that the concept of an entropic force which vanishes at equilibrium and which only exists for out-of-equilibrium scenarios is novel and does not appear in the literature. We then repeat this analysis for the interacting black holes we studied in chapter 3. We find that the non-uniform temperature of these black holes can also induce such a non-equilibrium entropic force among interacting black holes [65]. Applying this non-equilibrium entropic force to coalescing binary black holes would change the waveform of the gravitational waves generated [65]. Although this effect is very weak, it should still be observable in the future precision measurements. More details of this effect are given in chapter 4.

After having introduced the structure of my thesis, I would like to talk about some of the tools and theorems that will be used in my analysis.

1.2 Some tools and theorems

1.2.1 Curvature tensor

Since most of our analysis is based on Einstein's general relativity, I would first like to introduce some basic concepts. In general relativity, the distribution of matter determines the curvature of spacetime and the curvature of the spacetime conversely instructs the motion of matter. So how one may describe curvature is a very fundamental thing for general relativity and I will now give a short introduction.

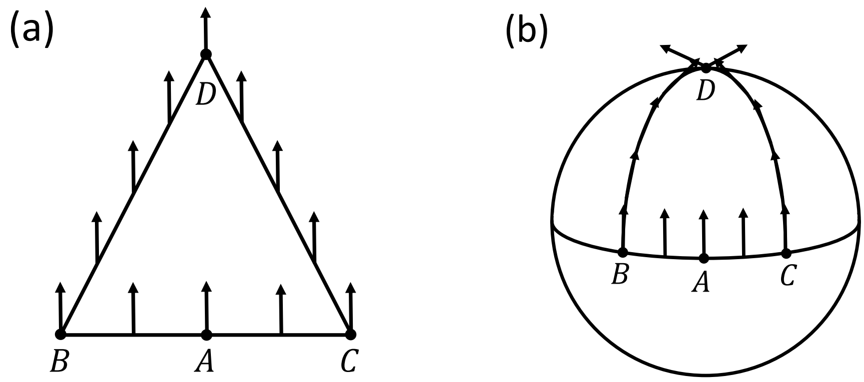

Figure 1.2: Parallel transporting an arbitrary vector field from point A to point D along two different routes ABD and ACD: (a) In a two-dimensional flat plane. (b) In a two-dimensional spherical surface. The vectors arriving at D along the two different routes are identical in (a) but pointing in different directions in (b).

In Euclidean space (flat space), a vector is just a magnitude along a direction and it is not tied to a point. So if we parallel transport a vector from point A to point D along two different routes ABD and ACD in the Euclidean plane, as in Fig. 1.2 (a), it is obvious that the vectors arriving at D along these two different routes are identical. However, if we parallel transport a vector from point A to point D along two different routes ABD and ACD on a spherical surface, as in Fig. 1.2 (b), generally the vectors arriving D are different. This effect may be used to describe the curvature of a manifold and is characterized by the following equation

$$A^{\mu}{}_{;\alpha\beta} - A^{\mu}{}_{;\beta\alpha} = -R^{\mu}{}_{\nu\alpha\beta}A^{\nu} \ , \tag{1.1}$$

where A^{μ} is an arbitrary vector field and $R^{\mu}{}_{\nu\alpha\beta}$ is the Riemann curvature tensor which is commonly used to describe the curvature of spacetime in general relativity. In $d+1$ dimensions, Greek indices go from 0 to d, lower-case Latin indices go from 1 to d (unless otherwise stated). I will be working throughout in $3+1$ spacetime dimensions.

From Eq. (1.1), we can see that the Riemann tensor is anti-symmetric about the second pair of indices. In fact, the Riemann tensor obeys further symmetry conditions [21]:

$$R_{\mu\nu\alpha\beta} = R_{\alpha\beta\mu\nu} = -R_{\nu\mu\alpha\beta} = -R_{\mu\nu\beta\alpha}, \tag{1.2}$$

$$R_{\mu\alpha\nu\beta} + R_{\mu\nu\beta\alpha} + R_{\mu\beta\alpha\nu} = 0. \tag{1.3}$$

Contracting the first and third (or second and forth) indices of the Riemann tensor yields the Ricci curvature tensor $R_{\nu\beta} \equiv R^{\mu}{}_{\nu\mu\beta}$, and further contracting the indices of the Ricci tensor yields the Ricci scalar $R \equiv R^{\mu}{}_{\mu}$. Einstein's field equations in fact only involve the Ricci tensor and Ricci scalar:

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = 8\pi T_{\mu\nu}, \tag{1.4}$$

where I have taken the cosmological constant to be zero, $T_{\mu\nu}$ is the energy-momentum tensor, and $g_{\mu\nu}$ is the metric tensor of the spacetime which describes the physical length of intervals ($ds^2 = g_{\mu\nu} dx^{\mu} dx^{\nu}$). I will typically use natural units where $\hbar = c = k_B = G = 1$ throughout the thesis unless otherwise stated.

Further contracting the indices μ and α of Eq. (1.1) yields

$$A^{\mu}{}_{;\mu\beta} - A^{\mu}{}_{;\beta\mu} = -R_{\nu\beta} A^{\nu} \ . \tag{1.5}$$

1.2.2 Hypersurfaces and the 3+1 split formalism

Einstein's general relativity is described on a four-dimensional manifold combining space and time, but, when we come to an analysis, we usually want to know what is happening at any particular time. So it is naturally to foliate the four-dimensional spacetime $\mathcal{M}$ into a series of three-dimensional (spacelike) surfaces of fixed time which are called hypersurfaces Σ [30, 70]. Without loss of generality, we may label every hypersurface by a function $t = x^0$ (the time coordinate) and introduce a coordinate system $(x^i) = (x^1, x^2, x^3)$ on each hypersurface. The coordinate functions $x^i(\mathcal{P})$ are chosen as smooth functions on the manifold, $\mathcal{P} \in \mathcal{M}$, within their respective chart. With such a construction, we have [30,70]

$$\hat{T}_{\mu} = -\mathcal{N} \partial_{\mu} x^0 = -\mathcal{N} \bigtriangledown_{\mu} x^0 \ , \tag{1.6}$$

with $\hat{T}_\mu \hat{T}^\mu = -1$ and $\mathcal{N}$ is the lapse function [70]. Here $\hat{T}^\mu$ is the unit timelike vector normal to Σ. [1] Although we now label each hypersurface Σ by the coordinate x^0, this does not require the trajectory at fixed spatial coordinates x^i to be normal to Σ. Therefore, generally the tangent vector $(\partial_0)^\mu$ along any curve of constant x^i is not parallel to the unit normal vector $\hat{T}^\mu$ of Σ. To ensure that $\hat{T}_\mu \hat{T}^\mu = -1$ and $(\bigtriangledown_\mu x^0)(\partial_0)^\mu = 1$ hold both, we must have

$$\hat{T}_\mu = -\mathcal{N} \bigtriangledown_\mu x^0 \ , \quad (\partial_0)^\mu = \mathcal{N} \hat{T}^\mu + \beta^\mu, \tag{1.7}$$

where β^μ is orthogonal to $\hat{T}_\mu$ (and by convention $\beta^0 = 0$). Since β^μ characterizes how the curves of constant x^i shift between neighbouring hypersurfaces, it is called the shift vector and it is tangent to the hypersurface [70].

From Eq. (1.7), we have [30, 70]

$$\hat{T}_\mu = (-\mathcal{N}, 0, 0, 0) \ , \quad \hat{T}^\mu = \Big(\frac{1}{\mathcal{N}}, \frac{-\beta^1}{\mathcal{N}}, \frac{-\beta^2}{\mathcal{N}}, \frac{-\beta^3}{\mathcal{N}} \Big) \ . \tag{1.8}$$

The projector onto Σ from $\mathcal{M}$ may be expressed as

$$\gamma_{\mu\nu} = g_{\mu\nu} + \hat{T}_\mu \hat{T}_\nu \ , \quad \gamma^{\mu\nu} = g^{\mu\nu} + \hat{T}^\mu \hat{T}^\nu \ . \tag{1.9}$$

In such a manner, the corresponding induced metric of Σ embedded in $\mathcal{M}$ may be written as $\gamma_{ab} \equiv \frac{\partial x^\mu}{\partial y^a} \frac{\partial x^\nu}{\partial y^b} g_{\mu\nu}$ and the determinant of the metric $g_{\mu\nu}$ obeys the relation

$$\sqrt{-g} = \mathcal{N} \sqrt{\gamma} \ , \tag{1.10}$$

where $\gamma \equiv \det(\gamma_{ab})$. For clarity sometimes the determinant of the induced metric γ_{ab} on hypersurface Σ is written $\gamma^{(\Sigma)}$.

1.2.3 Penrose diagram

Since the Penrose diagram is a popular tool to describe some infinite universes in a finite two dimensional diagram, here I give a short introduction to it and use it to illustrate some physical scenarios in this thesis.

For a spacetime with a black hole, we may consider the Schwarzschild metric

$$ds^2 = -\Big(1 - \frac{2m}{r}\Big) dt^2 + \frac{1}{1 - \frac{2m}{r}} dr^2 + r^2 d\Omega^2, \tag{1.11}$$

[1] Note for a general family of hypersurfaces defined by $f(x) = C$, $\partial_\mu f(x) dx^\mu = 0$ for tangent vectors dx^μ within a given hypersurface, hence $\partial_\mu f$ is normal to the hypersurface. The hypersurfaces, Σ, we study consist of the family given by $f(x) = x^0 = C$.

where $d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2$.

When we consider coordinate transformations, there are two different situations $0 < r < 2m$ and $r > 2m$. For $0 < r < 2m$, the time coordinate is r; while for $r > 2m$, the time coordinate is t.

For both these cases, the metric may be written

$$\begin{aligned} ds^2 &= -\Big(1-\frac{2m}{r}\Big)dt^2 + \frac{1}{1-\frac{2m}{r}}dr^2 + r^2 d\Omega^2 \\ &= -\Big(1-\frac{2m}{r}\Big)\Big(dt^2 - \frac{dr^2}{(1-\frac{2m}{r})^2}\Big) + r^2 d\Omega^2. \end{aligned} \tag{1.12}$$

Then we make the coordinate transformation $dr' = \frac{dr}{1-\frac{2m}{r}}$, the integral of this transformation yields $r' = r + 2m\ln(|r-2m|)$. When $0 < r < 2m$, r' changes from some finite value to negative infinity; and when $r > 2m$, r' changes from negative infinity to positive infinity. The exact range relations between r' and r are (see Fig. 1.3)

$$\begin{cases} r' \in (2m\ln(2m), -\infty), & \text{when } r \in (0, 2m) \\ r' \in (-\infty, \infty), & \text{when } r \in (2m, \infty) \end{cases}. \tag{1.13}$$

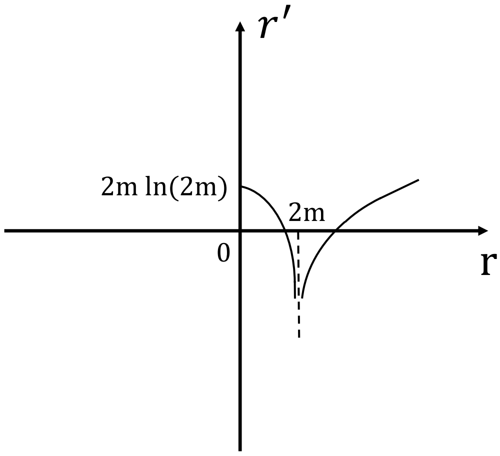

Figure 1.3: The range of r' with respect to the original Schwarzschild coordinate r.

With this coordinate transformation, Eq. (1.12) becomes

$$ds^2 = -(1-\frac{2m}{r})(dt^2 - dr'^2) + r^2 d\Omega^2. \tag{1.14}$$

We then make another coordinate transformation to radial light-cone coordinates $u = t + r'$, $v = t - r'$ and obtain

$$ds^2 = -(1-\frac{2m}{r})du\,dv + r^2 d\Omega^2, \tag{1.15}$$

where r is a function of u, v, and $u, v \in (-\infty, \infty)$.

Next we make a further coordinate transformation $u = \tan\psi, v = \tan\chi$ in order to fit the light-cone coordinates into a finite range, yielding

$$\begin{cases} du = \frac{d\psi}{\cos^2\psi} \\ dv = \frac{d\chi}{\cos^2\chi} \end{cases}. \tag{1.16}$$

With this coordinate transformation, Eq. (1.15) becomes

$$ds^2 = -\Big(1 - \frac{2m}{r}\Big)\frac{d\psi d\chi}{\cos^2\psi\cos^2\chi} + r^2 d\Omega^2, \tag{1.17}$$

where r is a function of ψ, χ. Here $\psi, \chi \in (\frac{2i-1}{2}\pi, \frac{2i+1}{2}\pi)$ and $i \in \mathbb{Z}$ is the set of integers. Without loss of generality, we choose $i = 0$.

We then replace ψ and χ by $\psi = \frac{1}{2}(T + R)$ and $\chi = \frac{1}{2}(T - R)$. Hence Eq. (1.17) becomes

$$ds^2 = -\Big(1 - \frac{2m}{r}\Big)\frac{dT^2 - dR^2}{4\cos^2\big(\frac{T+R}{2}\big)\cos^2\big(\frac{T-R}{2}\big)} + r^2 d\Omega^2, \tag{1.18}$$

where r is now expressed as a function of T, R.

If we ignore the closed spherically-symmetric part of the metric ($r^2 d\Omega^2$), the spacetime outside the black hole may now be represented by a finite two-dimensional diagram with the following boundaries (see Fig. 1.4)

$$\begin{cases} T + R &= \pi \\ T + R &= -\pi \\ T - R &= \pi \\ T - R &= -\pi \end{cases}. \tag{1.19}$$

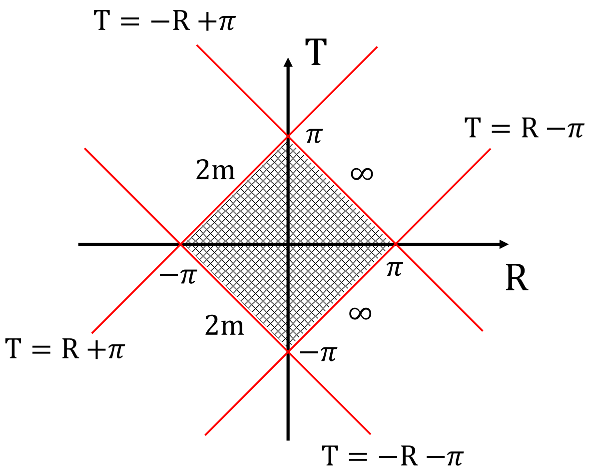

Figure 1.4: The whole spacetime *outside* a black hole is now represented by the shadow, and every point in the shadow is a two-dimensional spherical surface. The timelike coordinate is T in this figure.

For the spacetime inside the horizon, $0 < r < 2m$, all the above steps follow except that we have one more constraint that $r > 0$. This constraint limits the scope of T and R, and the spacetime illustrated by T and R finally looks like Fig. 1.5

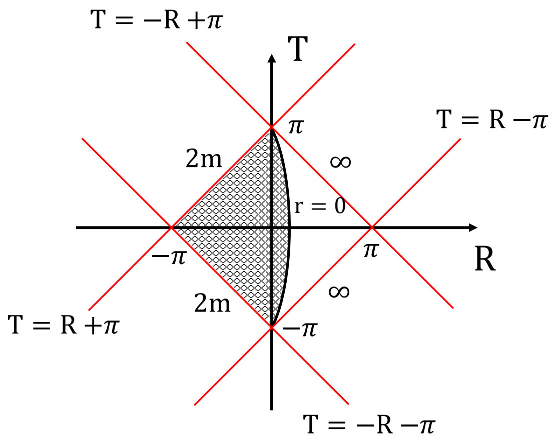

Figure 1.5: The whole spacetime *inside* a black hole is now represented by the shadow, and every point in the shadow is again a two-dimensional spherical surface. Note that the timelike coordinate is R in this figure.

If we reverse the time in both Fig. 1.4 and Fig. 1.5, we will obtain two new figures with time reflection symmetry to the old figures. Merging all these four figures together yields the familiar Penrose diagram, see Fig 1.6. Since two the external parts are obtained in this time-reverse procedure, in fact, only one half of the diagram in Fig 1.6 is sufficient to describe the spacetime of the Schwarzschild metric.

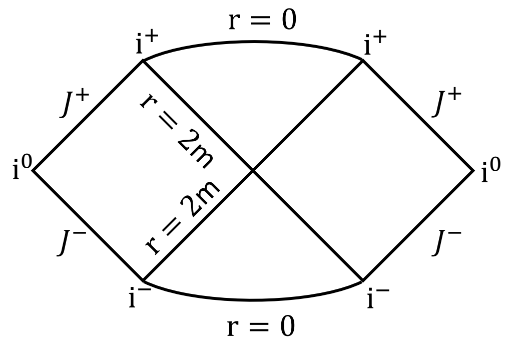

Figure 1.6: The whole spacetime with its time reversed parts is now represented by a finite region, and every point in the figure is a two-dimensional surface. Here i^+ is future timelike infinity, i^- is past timelike infinity, i^0 is spatial infinity, J^+ is future null infinity and J^- is past null infinity. The time orientation is upward, and lines having a forty five degree angle with the time orientation are null trajectories.

Let us now consider the Penrose diagram that describes a shell of matter collapsing to form a black hole. Before the appearance of any black hole, the diagram looks like the Penrose diagram for Minkowski spacetime. After the formation of a black hole, the spacetime connects with the spacetime inside a black hole, see Fig. 1.7.

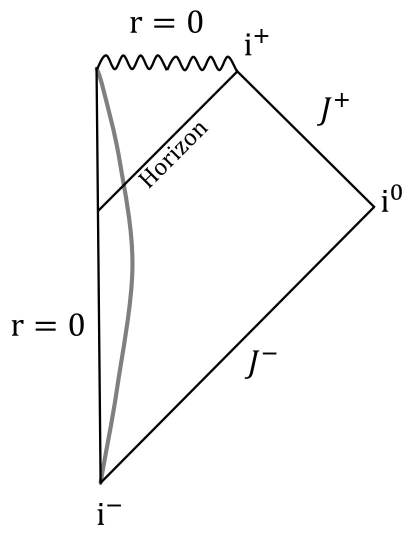

Figure 1.7: The gray line represents a spherical shell of matter that finally collapses into a black hole, and every point in the diagram is a two-dimensional spherical surface. The time orientation is upward. (Only the right-hand side of the diagram is shown.)

1.2.4 Different hypersurfaces appearing in this thesis

I will now use the Penrose diagram to illustrate different hypersurfaces that appear in this thesis. It is also worth noting that three-dimensional hypersurfaces in spacetime are divided into three different classes by the characteristics of their normal vectors. A hypersurface is said to be timelike if its normal vector is spacelike, and conversely, a hypersurface is spacelike if its normal vector is timelike. When the normal vector of a hypersurface is null, the hypersurface is called a null hypersurface.

Since the stretched horizon is a 'boundary' between a real physical horizon and ordinary surfaces, I first give a short introduction to the stretched horizon. Stretched horizons were proposed by Thorne and colleagues as a surrogate for the real black hole horizon to facilitate their research. In the 3+1 formalism, a stretched horizon is a 2-dimensional membrane that resides in 3-dimensional space, and it may be chosen to be arbitrarily close to the real physical horizon depending on the requirements of the calculations at hand [49, 54].

If we allow a stretched horizon to evolve in a spacetime, then it forms a timelike hypersurface, which is also a claim of found in Ref. [54]. Suppose a stretched horizon is n Planck length outside the black hole horizon, then, for a Schwarzschild black hole, the physical distance between black hole horizon and the stretched horizon may be calculated by

$$n l_{\mathrm{P}} = \int_{2m}^{2m+\varepsilon} \frac{dr}{\sqrt{1-\frac{2m}{r}}} \tag{1.20}$$

where l_{P} is the Planck length, $n \in \mathbb{R}^{+}$ is the number of l_{P}, and m is the mass of the black hole. Replacing r with $2m+\varepsilon'$, so that $dr = d\varepsilon'$, Eq. (1.20) may be written as

$$\begin{aligned} n l_{\mathrm{P}} &= \int_{0}^{\varepsilon} \sqrt{\frac{2m+\varepsilon'}{\varepsilon'}} d\varepsilon' \\ &= \sqrt{2m} \int_{0}^{\varepsilon} \sqrt{\frac{1}{\varepsilon'}} \sqrt{1+\frac{\varepsilon'}{2m}} d\varepsilon' \end{aligned} \tag{1.21}$$

Then we expand $\sqrt{1+\frac{\varepsilon'}{2m}}$ and find that $\varepsilon \approx \frac{(n l_{\mathrm{P}})^2}{8m}$. Therefore, the trajectory of a point on the stretched horizon in the Penrose diagram is described by $x^{\mu} = (t, 2m + \frac{(n l_{\mathrm{P}})^2}{8m}, \theta, \phi)$. After going through all the coordinate transformations, $r = 2m + \frac{(n l_{\mathrm{P}})^2}{8m}$ in the original coordinate finally yields

$$\sin R = 2m + \frac{(n l_{\mathrm{P}})^2}{8m} + 2m(\cos T + \cos R) \ln \frac{(n l_{\mathrm{P}})^2}{8m} + O\Big(\frac{(n l_{\mathrm{p}})^4}{m^2}\Big), \tag{1.22}$$

which is represented by the dotted line in Fig. 1.8. From the diagram, it is also easy to see that the stretched horizon is timelike.

In contrast, ordinary surfaces need not have any relation to black hole horizons. For the spacetime we discussed above, ordinary surfaces may, for example, be seen as a stretched horizon for ε finite and large.

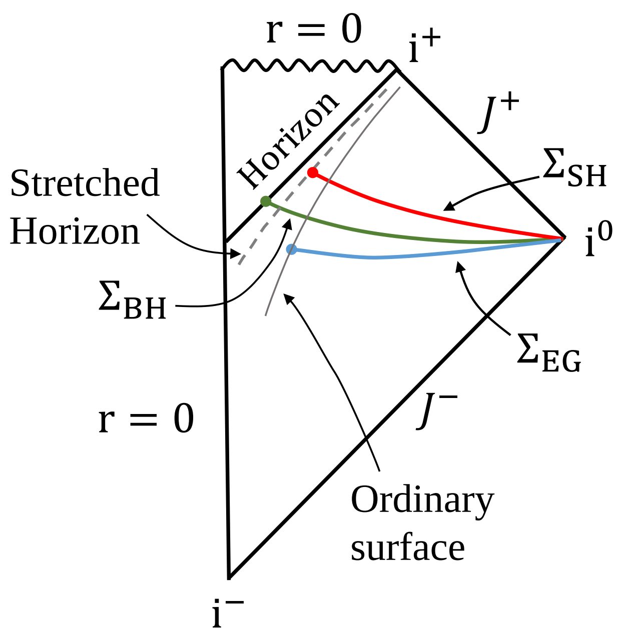

Figure 1.8: The dotted grey line represents the stretched horizon and the solid grey line represents an ordinary surface. Here Σ_{BH}(green) represents the hypersurface which stops at the black hole horizon, Σ_{SH}(red) represents the hypersurface which stops at the stretched horizon, and Σ_{EG}(blue) represents the hypersurface in the emergent gravity program that stops at an ordinary two-dimensional inner boundary which generally does not have any relation with the black hole horizon. Every point in the diagram is a two-dimensional spherical surface. The time orientation is upward.

Finally, we may use the Penrose diagram to describe some different scenarios we are interested in. When Bardeen, Carter and Hawking originally proved black hole thermodynamics, they considered a 3-dimensional spacelike hypersurface that stops at the black hole horizon [11] (we ignore rotation here since this thesis does not consider the rotating case), this hypersurface could now be described by a line (green) that stops at the horizon, see Fig. 1.8. Similarly, a spacelike hypersurface that stops at the stretched horizon is represented by the red line in this diagram. Verlinde's emergent gravity program conjectures that an ordinary inner surface external to the black hole will also behave thermodynamically [59], and relevant spacelike hypersurface stops at such an ordinary surface here represented by the blue line in Fig. 1.8. This Penrose diagram, Fig. 1.8, illustrates examples of the different types of hypersurfaces that will appear in this thesis.

1.2.5 Stokes' theorem

My work in this thesis primarily focuses on extending the first law of black hole thermodynamics originally derived by Bardeen, Carter and Hawking [11]. Originally, they analyzed a hypersurface external to a stationary black hole hence avoiding the physics of the black hole interior. In this manner, any hypersurface in their analysis has two boundaries: the black hole horizon and spatial infinity. The behavior of spatial infinity of a spacetime is related to the global physical energy which is one of the most important quantities in thermodynamics. By using Stokes' theorem, Bardeen et al. transformed the integral at spatial infinity into an integral on the black hole horizon plus an integral over the hypersurface between these two boundaries. Let us now briefly summarize Stokes's theorem.

In a four-dimensional spacetime $\mathcal{M}$ with boundary $\partial\mathcal{M}$, Stokes' theorem may be written

$$\int_{\mathcal{M}} A^{\mu}{}_{;\mu}\sqrt{-g}\,dx^4 = \int_{\partial\mathcal{M}} A^{\mu}\varepsilon\hat{n}_{\mu}\sqrt{|\gamma|}\,dy^3, \tag{1.23}$$

where the unit tangent vector $\hat{n}^{\mu}$ is outward pointing from $\mathcal{M}$ and $\varepsilon \equiv \hat{n}^{\mu}\hat{n}_{\mu}$. Since the boundary of a boundary is the empty set, applying Stokes' theorem on a boundary yields zero.

Similarly, for a three-dimensional spacelike hypersurface Σ in $\mathcal{M}$ with boundary $\partial\Sigma$, Stokes' theorem for an anti-symmetric tensor $F^{\mu\nu}$ may be written as [21]

$$\int_{\Sigma} \hat{T}_{\mu}F^{\mu\nu}{}_{;\nu}\sqrt{\gamma^{(\Sigma)}}d^3y = \int_{\partial\Sigma} \hat{T}_{\mu}F^{\mu\nu}\hat{N}_{\nu}\sqrt{\gamma^{(\partial\Sigma)}}d^2z, \tag{1.24}$$

where $\hat{T}^{\mu}$ is the future directed timelike unit vector normal to Σ and $\hat{N}^{\mu}$ is the outgoing spacelike unit vector normal to $\partial\Sigma$ and tangent to Σ itself.

1.3 The definition of mass (energy)

Since thermodynamics is a theory about the transformation of energy and there are multiple definitions of energy in general relativity, I now give a short discussion about what is the suitable definition of energy. Since some definitions of energy are based on Killing vectors (and indeed the original analysis of Bardeen et al. is also based on spacetimes with Killing vectors), we first give a short introduction to the Killing vector.

1.3.1 Killing vector

A Killing vector describes a global symmetry of the spacetime. If A^μ is a Killing vector, then the Lie derivative of the metric along this vector vanishes everywhere, $\mathfrak{L}_A(g_{\mu\nu}) = 0$, which means that the metric is independent of the direction tangent to the vector A^μ. Using the definition of the Lie derivative [47], the Killing condition $\mathfrak{L}_A(g_{\mu\nu}) = 0$ may be written

$$\mathfrak{L}_A(g_{\mu\nu}) = g_{\mu\nu;\lambda}A^\lambda + A^\lambda{}_{;\mu}g_{\lambda\nu} + A^\lambda{}_{;\nu}g_{\mu\lambda} = A_{\nu;\mu} + A_{\mu;\nu} = 2A_{(\mu;\nu)} = 0. \tag{1.25}$$

Therefore, the covariant derivative of a Killing vector field is naturally anti-symmetric $A_{\mu;\nu} = A_{[\mu;\nu]}$.

In fact, Bardeen et al.'s original analysis of black hole thermodynamics for stationary spacetimes relies on the existence of Killing vectors. With the help of a Killing vector K^μ, Eq. (1.5) reduces to a much simpler form:

$$K^\mu{}_{;\beta\mu} = R_{\nu\beta}K^\nu \ . \tag{1.26}$$

Further, when Eq. (1.26) appears within an integral over a hypersurface, the anti-symmetry of the covariant derivative of the Killing vector in Eq. (1.26) allows one easily apply Stokes' theorem Eq. (1.24) to this integral. This trick was used to good effect by Bardeen, Carter and Hawking for their stationary spacetime analysis [11].

However, generally, dynamical spacetimes do not have any global symmetry and hence no Killing vectors. So generalizing black hole thermodynamics to dynamical spacetimes means losing all the convent tricks discussed above. This is in fact one of the main challenges we have to overcome in Chapter 3 when we generalize black hole thermodynamics to dynamical spacetimes.

1.3.2 The Komar and ADM masses

There are two primary definitions of energy on asymptotically-flat spacetimes: the Komar mass (energy) M^{Komar} and the ADM mass (energy) M^{ADM} (though other definitions exist). The ADM mass is based on the Hamiltonian formulation of general relativity and is widely accepted as the physical energy on a hypersurface for both stationary and dynamical spacetimes. In asymptotically rectilinear coordinates, it may be written [2,3]

$$M^{\text{ADM}} = \frac{1}{16\pi}\int_{\partial\Sigma_\infty}\left(g_{ij}{}^{,j}\hat{N}^i - g_{jj}{}^{,i}\hat{N}_i\right)dA, \tag{1.27}$$

where $dA = \sqrt{|\gamma^{(\partial\Sigma_\infty)}|}\, d^2z$.

In fact, Bardeen, Carter and Hawking used a different definition of energy in their analysis, the Komar mass [11, 38]:

$$M^{\text{Komar}} = \frac{1}{4\pi}\int_{\partial\Sigma_\infty} K_{\mu;\nu}\hat{T}^\nu\hat{N}^\mu dA, \tag{1.28}$$

where K^μ is a timelike Killing vector which satisfies the asymptotic normalization condition $K^\mu K_\mu = -1$ at spatial infinity. As discussed in Section 1.3.1, since $K_{\mu;\nu}$ is antisymmetric, it is easy to apply Stokes' theorem to Eq. (1.28) and hence relate this global energy to the sum of an integral on the black hole horizon and an integral over the hypersurface. Interestingly, the proof of the equivalence between the Komar mass and the physical ADM mass was achieved by Beig only in 1978 [12] five years after Bardeen et al.'s original analysis of black hole thermodynamics.

Since the ADM mass is not in a covariant form, it is not straightforward to apply Stokes' theorem to it in order to relate an integral on the boundary of a hypersurface to an integral over the remaining hypersurface. The lack of a suitable covariant form for the physical energy for dynamical spacetimes is a further obstruction we will need to overcome when we try to generalize black hole thermodynamics to dynamical spacetimes. In Chapter 3, we will show for dynamical asymptotically-flat spacetimes, that the ADM mass may be transformed into a covariant form similar to the expression for the Komar mass. This covariant expression does not require the presence of a Killing vector and allows us to apply Stokes' theorem to our analysis without undue difficulty.

Chapter 2

Horizons are hot, ordinary surfaces are not

Since the 1970's it has been known that black hole (and other) horizons are truly thermodynamic. More generally, surfaces which are not horizons have also been conjectured to behave thermodynamically. Initially, for surfaces microscopically expanded from a horizon to so-called stretched horizons, and more recently, for more general ordinary surfaces in Verlinde's emergent gravity program. To test these conjectures we ask whether such surfaces satisfy an analogue to the first law of thermodynamics as do horizons. For static asymptotically-flat spacetimes we find that such a first law holds on horizons. We prove that this law remains an excellent approximation for stretched horizons. Note that this result explicitly illustrates the insufficiency of the laws of black hole mechanics alone from implying truly thermodynamic behavior. For surfaces away from horizons in Verlinde's emergent gravity program the first law fails (except for spherically-symmetric scenarios) thus undermining the key thermodynamic assumption of this program. [1]

[1] Zhi-Wei Wang & Samuel L. Braunstein. These results have been published in *Nature Communications*

2.1 Introduction

In 1973, Bardeen et al. [11] derived the laws of black-hole mechanics which are in direct analogy with the laws of thermodynamics. Together with the discovery of Hawking radiation [31], the truly thermodynamic behavior of black-hole horizons became well established. Indeed such thermodynamic behavior is now well accepted for all spacetime horizons, including those due to accelerated observers [37, 55] and cosmological horizons [29].

Later, other surfaces were also attributed with thermodynamic properties. Firstly, stretched horizons were claimed to be thermodynamic, effectively acting as radiating black bodies [54] with a temperature $T = \kappa/(2\pi)$ determined by their local surface gravity [2] κ and an entropy (a 'state variable') associated with a presumed statistical mechanical interpretation of black hole entropy [54, 72]. An explicit re-derivation of the laws of black hole mechanics has not been previously carried out for stretched horizons. More recently, a class of ordinary surfaces has been conjectured to behave thermodynamically, forming the key assumption in Verlinde's emergent gravity program [59]. This thermodynamic attribution was justified in part by using it in a heuristic derivation of the full Einstein field equations in static asymptotically-flat spacetime [59].

In this chapter, we ask whether canonical General Relativity is consistent with the assumption that such ordinary surfaces can be rigorously seen to behave thermodynamically. We attack this question by focusing on the analogue to the first law of thermodynamics. Originally this law was derived in an analysis that was specialized to the behavior of horizons [11]. We remove this specialization to reveal the behavior of ordinary surfaces in an analysis of the first law. In this chapter, we report that the first law holds to an excellent approximation for stretched horizons. Finally, with the exception of fully spherically-symmetric scenarios, we find that the first law fails to hold for the ordinary surfaces in Verlinde's emergent gravity program.

In order to attempt to derive a first law for ordinary surfaces we closely follow in the footsteps of Bardeen, Carter and Hawking's 1973 classic paper [11]. The first step is to obtain an integral equation for the net energy in a static system, where instead of an inner boundary located at a black hole horizon, this boundary is a generic ordinary surface. Next, we consider small 'changes' in the net energy corresponding to shifting to a

[2] In stationary spacetimes the surface gravity corresponds to the force per unit mass needed to hold an object 'stationary' at the horizon as measured by an observer at spatial infinity.

parametrically nearby solution of the Einstein field equations. This 'differential' version is determined by studying the behavior of the net energy under spacetime diffeomorphisms of the initial metric [11]. As in Bardeen et al., "gauge" freedom in the choice of coordinates is used to ensure that the hypersurfaces before and after the diffeomorphism are covered by identical sets of coordinates.

Our analysis is limited to static (and hence non-rotating and with zero shift vector $\beta^\mu = 0$ in coordinates in which the metric components are time-independent) spacetimes. For simplicity, we assume that there is no matter exterior to the holographic screen ($T^{\mu\nu} = 0$).

2.2 Energy and possible temperature

Since the Komar mass (energy) equals the ADM mass in stationary spacetimes [12], in this chapter we will use the Komar energy as the global physical energy of the hypersurfaces we are studying.

2.2.1 Integral expression for net energy

Komar form Theorem: For a static asymptotically-flat spacetime with timelike Killing vector $K^\mu = (\partial_0)^\mu$ one may derive the total gravitating Komar energy E as an integral over a spacelike hypersurface Σ that is truncated (or bounded) internally by an ordinary 2-surface $\partial\Sigma_{\text{inner}}$ (see Fig. 2.1) [38] by

$$\boxed{E = \frac{1}{4\pi}\int_\Sigma R_{\mu\nu} K^\mu \hat{T}^\nu \sqrt{|\gamma^{(\Sigma)}|}\, d^3x + \frac{1}{4\pi}\int_{\partial\Sigma_{\text{inner}}} \kappa \, dA,} \tag{2.1}$$

where $\kappa \equiv K_{\mu;\nu}\hat{T}^\nu \hat{N}^\mu$ is the generalized surface gravity on the inner boundary provided $K^\mu K_\mu = -1$ at spatial infinity. Note that this result is generally true without requiring $T_{\mu\nu} = 0$ on Σ.

Proof:

Consider a static spacetime with a Killing vector $K^\mu = (\partial_0)^\mu = (1, 0, 0, 0)$, with $K^\mu K_\mu = -1$ at spatial infinity. As discussed in Section 1.3, the Killing condition implies that

$$K_{\mu;\nu} = K_{[\mu;\nu]} \equiv \frac{1}{2}(K_{\mu;\nu} - K_{\nu;\mu}) \ . \tag{2.2}$$

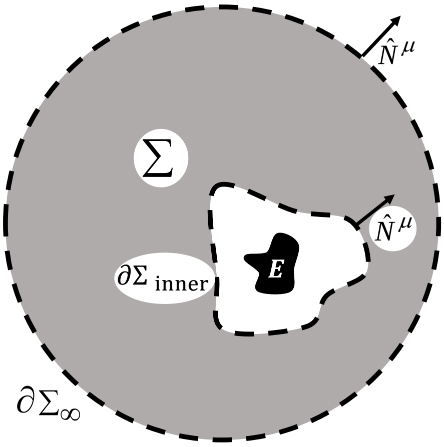

Figure 2.1: Schematic of the spacelike three-dimensional hypersurface of interest, Σ, with an inner boundary $\partial\Sigma_{\rm inner}$ and a boundary at infinity $\partial\Sigma_\infty$. Here $\hat{N}^\mu$ is the spacelike 4-vector normal to the boundaries of Σ (note the direction convention on the inner boundary). We assume a general mass distribution within the inner boundary and no matter outside it. Here E is the Komar energy which equals the ADM energy in stationary spacetimes [12].

Now recall that permuting the order of a pair of covariant derivatives acting on a 4-vector A^μ provides a defining feature of the Riemann curvature tensor as [21]

$$A^\mu{}_{;\alpha\beta} - A^\mu{}_{;\beta\alpha} = -R^\mu{}_{\nu\alpha\beta}A^\nu \ .$$

Contracting the indices μ and α reduces this to an expression in terms of the Ricci tensor

$$A^\mu{}_{;\mu\beta} - A^\mu{}_{;\beta\mu} = -R_{\nu\beta}A^\nu \ . \tag{2.3}$$

Since the Killing vector is anti-symmetric, we must have $K^\mu{}_{;\mu} = 0$ and we immediately find that

$$K^\mu{}_{;\beta\mu} = R_{\nu\beta}K^\nu. \tag{2.4}$$

Integrating this over a spacelike hypersurface Σ, yields

$$\int_\Sigma K^\mu{}_{;\beta\mu}\,\hat{T}^\beta\,\sqrt{|\gamma^{(\Sigma)}|}\,d^3x = \int_\Sigma R_{\nu\beta}K^\nu\,\hat{T}^\beta\,\sqrt{|\gamma^{(\Sigma)}|}\,d^3x \tag{2.5}$$

here $\hat{T}^\mu$ is the timelike unit 4-vector normal to Σ with $\hat{T}^\mu\hat{T}_\mu = -1$.

The hypersurface is assumed to have an outer boundary at spatial infinity $\partial\Sigma_\infty$, and an inner boundary $\partial\Sigma_{\rm inner}$ (see Fig. 2.1). In the original work of Bardeen et al. [11], this inner boundary corresponded to the black hole's horizon $\partial\Sigma_{\rm BH}$. Here we generalize this by taking it to be an arbitrary closed 2-surface $\partial\Sigma_{\text{inner}}$. The boundary of the hypersurface is assumed to be oriented, with unit normal $\hat{N}^\mu$ (see Fig. 2.1), so $\hat{N}_\mu \hat{N}^\mu = 1$ and $\hat{N}^\mu \hat{T}_\mu = 0$.

Then applying Stokes's theorem for an anti-symmetric tensor Eq. (1.24) to the left-hand-side of Eq. (2.5) we find

$$\int_{\partial\Sigma_\infty} K^\mu{}_{;\beta}\,\hat{T}^\beta \hat{N}_\mu \sqrt{|\gamma^{(\partial\Sigma_\infty)}|}\; d^2z - \int_{\partial\Sigma_{\text{inner}}} K^\mu{}_{;\beta}\,\hat{T}^\beta \hat{N}_\mu \sqrt{|\gamma^{(\partial\Sigma_{\text{inner}})}|}\; d^2z$$
$$= \int_\Sigma R_{\nu\beta} K^\nu\,\hat{T}^\beta \sqrt{|\gamma^{(\Sigma)}|}\, d^3x. \tag{2.6}$$

At this stage, we wish to generalize the concept of surface gravity as a quantity defined anywhere. Assuming that the surface $\partial\Sigma$ is non-rotating (corresponding to zero angular velocity of the spacetime itself) , we may interpret the integrand of the integral on the boundary in Eq. (2.6) to be the surface gravity, so

$$\boxed{\kappa \equiv K_{\mu;\nu} \hat{T}^\nu \hat{N}^\mu.} \tag{2.7}$$

It is worth noting that $\kappa/(2\pi)$ is precisely the formula Verlinde gives (his Eq. (5.3) of [59]) for what he calls the local temperature of the holographic screen (what he calls ordinary surfaces of constant Newtonian potential $\phi = \frac{1}{2}\ln(-K^\mu K_\mu) = \ln \mathcal{N}$ [59]) as measured with respect to a reference point at spatial infinity.

This definition of surface gravity allows us to naturally extend the original 1973 analysis away from black hole horizons. In particular, the left-hand-side of Eq. (2.6) reduces to

$$\int_{\partial\Sigma_\infty} \kappa \sqrt{|\gamma^{(\partial\Sigma_\infty)}|}\; d^2z - \int_{\partial\Sigma_{\text{inner}}} \kappa \sqrt{|\gamma^{(\partial\Sigma_{\text{inner}})}|}\; d^2z. \tag{2.8}$$

According to the definition of Komar mass Eq. (1.28), the integral over $\partial\Sigma_\infty$ equals $4\pi E$ leading to

$$E = \frac{1}{4\pi}\int_\Sigma R_{\mu\nu} K^\mu\,\hat{T}^\nu \sqrt{|\gamma^{(\Sigma)}|}\; d^3x + \frac{1}{4\pi}\int_{\partial\Sigma_{\text{inner}}} \kappa\; dA\;, \tag{2.9}$$

This completes the proof of the Komar form Theorem.

Were we to consider the spherically symmetric case, Eq. (2.1) would reduce to

$$E = \frac{1}{4\pi}\int_\Sigma R_{\mu\nu} K^\mu\,\hat{T}^\nu \sqrt{|\gamma^{(\Sigma)}|}\, d^3x + \frac{\kappa}{4\pi} A. \tag{2.10}$$

Just to emphasize what this represents, here the hypersurface Σ extends from an arbitrary inner boundary, $\partial\Sigma_{\text{inner}}$, out to spatial infinity. Thus, the generalized surface gravity, κ, and the area, A, are those associated with the inner boundary itself (rather than any horizon).

Eq. (2.1) has exactly the same form as the conventional formula for the total mass of the system [11] but extended to an arbitrary 2-dimensional surface (the inner boundary instead of a horizon). Finally, note that the matter inside the inner boundary need *not* be associated with a black hole, it may be ordinary matter, with no horizon present at all. Thus, were inner boundaries found to have thermodynamic properties (i.e., a well-defined entropy and temperature), it would *not* be because such properties were inherited from a real horizon behind the screen.

2.2.2 Possible local Temperature

Following the results for horizons [11], it is tempting to seek to interpret $\kappa/(2\pi)$ from Eq. (2.1) as the local temperature at any point along an arbitrary 2-surface $\partial\Sigma_{\text{inner}}$. However, this would be unsatisfactory if true for arbitrary surfaces, since this local temperature would not be in thermal equilibrium with an actual physical screen held fixed at the same location; the temperature now coming from the Unruh effect [55] and the local proper acceleration required to keep each portion of the screen stationary. Only for surfaces of constant Newtonian gravitational potential ϕ, where the proper acceleration of a stationary observer and the local normal to the surface are parallel, is such thermal equilibrium possible (see section 2.4.2). Thus the temptation of such a thermodynamic interpretation should be restricted to the family of ordinary surfaces satisfying $\phi = \text{constant}$.

Indeed, this restricted temptation appears to have been satisfied in the emergent gravity program, where for static asymptotically-flat spacetimes, ordinary surfaces of constant ϕ are dubbed holographic screens and are claimed to have a local temperature [59] given by $T = \kappa/(2\pi)$ and even claimed to possess a 'state variable' quantifying the number of 'bits' on the screen. These putative thermodynamic properties are then used in a heuristic derivation of the full Einstein field equations [59]. If correct, such a claim would mean that Verlinde's emergent gravity program would already subsume many decades of results associated with full General Relativity.

2.3 Differential "first law" of thermodynamics for stationary spacetimes

2.3.1 The first law we expect

As discussed above, the straightforward generalization, especially in the spherically symmetric case, for net energy on a hypersurface might appear to suggest that a temperature and entropy can actually be defined for any surface by

$$T = \frac{\kappa}{2\pi}, \qquad S = \frac{A}{4}. \tag{2.11}$$

However, such quantities need to behave thermodynamically. In particular, for our static system, the net energy E, should admit changes which behave as

$$\delta E = T\delta S, \tag{2.12}$$

(ignoring work terms) so that the temperature would be acting as an integrating factor relating changes in the (state function) entropy to changes in the energy. In other words, we must show that such changes lead to the expected form of the first-law of thermodynamics. In the simplest case, where the hypersurface Σ is empty of matter, this law should read

$$\delta E = \frac{1}{8\pi} \int_{\partial\Sigma_{\text{inner}}} \kappa \, \delta(dA). \tag{2.13}$$

2.3.2 Diffeomorphic conditions

Again here we follow in the footsteps of the original analysis of Bardeen, Carter, and Hawking and consider changes corresponding to parametric differences between diffeomorphicly nearby solutions. In particular, we will consider two nearby configurations corresponding to the metrics

$$g_{\mu\nu}, \qquad g'_{\mu\nu} = g_{\mu\nu} + h_{\mu\nu}, \tag{2.14}$$

where $h_{\mu\nu} \equiv \delta g_{\mu\nu} = -g_{\mu\sigma} g_{\nu\tau} \delta g^{\sigma\tau}$, i.e., $\delta g^{\sigma\tau} = -h^{\sigma\tau}$.

Without loss of generality and as with the original analysis of Bardeen, Carter, and Hawking, we may assume that for the two diffeomorphicly related configurations, the hypersurfaces Σ and Σ' are described by identical sets of coordinates; this is always possible due to "gauge" freedom in the choice of coordinate systems [11]. Henceforth we label both by Σ. Similarly, for their boundaries $\partial\Sigma$. Further, as in Ref. [11] we likewise assume that both configurations have the same Killing vector, so

$$\delta K^{\mu} = 0, \qquad \delta K_{\mu} = h_{\mu\nu} K^{\nu}. \tag{2.15}$$

Since the sets of coordinates of Σ and $\partial\Sigma$ are unchanged by the diffeomorphism, without loss of generality we may take [11]

$$\delta(dx^\mu) \in \Sigma. \tag{2.16}$$

Because $K_\mu dx^\mu = \mathcal{N}\hat{T}_\mu dx^\mu = 0$ for all dx^μ in Σ, we have $\delta K_\mu \parallel K_\mu$, so

$$\delta K_\mu = k_0 K_\mu, \tag{2.17}$$

for some function k_0. Comparing Eq. (2.17) with Eq. (2.15), one finds

$$h_{\mu\nu}K^\nu = k_0 K_\mu \tag{2.18}$$

everywhere. Then contracting $\hat{T}^\mu$ on both sides of this equation yields

$$k_0 = -h_{\mu\nu}\hat{T}^\mu\hat{T}^\nu. \tag{2.19}$$

(In other words, $k_0 = -h_{\hat{T}\hat{T}}$ in the tetrad basis.)

Similarly, since $\hat{T}_\mu dx^\mu = 0$ for all dx^μ in Σ, we have $\delta\hat{T}_\mu \parallel \hat{T}_\mu$, so

$$\delta\hat{T}_\mu = \frac{k_1}{2}\hat{T}_\mu, \tag{2.20}$$

for some function k_1 (the factor of $\frac{1}{2}$ is for later convenience). To get an expression for k_1, we calculate the variation of $g^{\mu\nu}\hat{T}_\mu\hat{T}_\nu = -1$.

$$\begin{aligned} & (\delta g^{\mu\nu})\hat{T}_\mu\hat{T}_\nu + 2g^{\mu\nu}\hat{T}_\mu(\delta\hat{T}_\nu) &=& \;0 \\ \Rightarrow\quad & (-h^{\mu\nu})\hat{T}_\mu\hat{T}_\nu + 2g^{\mu\nu}\hat{T}_\mu(\frac{k_1}{2}\hat{T}_\nu) &=& \;0, \end{aligned} \tag{2.21}$$

hence

$$k_1 = -h_{\mu\nu}\hat{T}^\mu\hat{T}^\nu = k_0. \tag{2.22}$$

Recall $K^\mu = \partial_t$ and $\hat{T}^\mu$ is normal to Σ, so $K^\mu = \mathcal{N}\hat{T}^\mu + \beta^\mu$ where β^μ is the shift vector and $\mathcal{N} = 1/\hat{T}^t$ is the lapse function [30]. Since we only consider non-rotating spacetimes in this chapter, without loss of generality, we assume $\beta^\mu = 0$ from now on, then $\hat{T}^\mu = \hat{T}^t K^\mu$

and $\hat{T}_\mu = \hat{T}^t K_\mu$. Then for $\hat{T}^\mu$, we find

$$\begin{aligned}
\delta\hat{T}^\mu &= \delta(g^{\mu\nu}\hat{T}_\nu) \\
&= -h^{\mu\nu}\hat{T}_\nu + \frac{1}{2}k_1\hat{T}^\mu \\
&= -g^{\mu\lambda}h_{\lambda\nu}\hat{T}^\nu + \frac{1}{2}k_1\hat{T}^\mu \\
&= -g^{\mu\lambda}h_{\lambda\nu}K^\nu\hat{T}^t + \frac{1}{2}k_1\hat{T}^\mu \\
&= -g^{\mu\lambda}k_1 K_\lambda\hat{T}^t + \frac{1}{2}k_1 T^\mu \\
&= -k_1\hat{T}^\mu + \frac{1}{2}k_1\hat{T}^\mu \\
&= -\frac{1}{2}k_1\hat{T}^\mu, \qquad (2.23)
\end{aligned}$$

where we have used Eq. (2.18) in the fourth line.

Again since, $dx^\mu\hat{N}_\mu = 0$ for all dx^μ in $\partial\Sigma$, combined with $\hat{T}^\mu\hat{N}_\mu = 0$, we find $\delta\hat{N}_\mu \parallel \hat{N}_\mu$, and so we may write

$$\delta\hat{N}_\mu = \frac{1}{2}k_2\hat{N}_\mu, \qquad (2.24)$$

for some function k_2 (the factor $\frac{1}{2}$ is introduced for later convenience). In the same way as for Eq. (2.19) we find

$$k_2 = h_{\mu\nu}\hat{N}^\mu\hat{N}^\nu, \qquad h_{\lambda\nu}\hat{T}^\lambda\hat{N}^\nu = 0. \qquad (2.25)$$

We now extend our use of the "gauge" freedom to extend this equality of coordinates for our nearby solutions slightly away from the inner boundary $\partial\Sigma_{\text{inner}}$, though still within the hypersurface Σ. Indeed, Bardeen et al. [11] used such freedom on the (future) null horizon for stationary black holes. Here, we make an analogous construction for the outgoing spacelike tangent vectors along $\hat{N}^\mu$ from each point on the boundary $\partial\Sigma_{\text{inner}}$. In particular, for an infinitesimal 'distance' along the tangent vectors $dx^\mu \propto \hat{N}^\mu$ from the inner boundary, we use gauge freedom to ensure that these vectors satisfy $\delta(dx^\mu) = 0$. In other words, gauge freedom allows us to choose the covariant vectors $\hat{N}^\mu$ normal to the inner boundary $\partial\Sigma_{\text{inner}}$ to remain parallel to themselves under the diffeomorphism. Consequently

$$\delta\hat{N}^\mu = -\frac{1}{2}k_2\hat{N}^\mu, \qquad (2.26)$$

a condition similar to those given by Bardeen et al. [11]. Combining Eqs.(2.24) and (2.26) we find

$$\delta\hat{N}_\mu = -g_{\mu\nu}\,\delta\hat{N}^\nu, \qquad (2.27)$$

at the inner boundary $\partial\Sigma_{\text{inner}}$.

Let us now introduce the whole tetrad basis $\{\hat{T}^\mu, \hat{N}^\mu, \hat{U}^\mu, \hat{V}^\mu\}$; recall $\hat{T}^\mu$ is normal to Σ, $\hat{N}^\mu$ is in Σ but normal to $\partial\Sigma$, and $\hat{U}^\mu, \hat{V}^\mu$ lie in $\partial\Sigma$. The projector onto the tangent space on $\partial\Sigma$ is defined

$$P^{\mu\nu} \equiv (\hat{U} \otimes \hat{U} + \hat{V} \otimes \hat{V})^{\mu\nu} = \hat{U}^\mu \hat{U}^\nu + \hat{V}^\mu \hat{V}^\nu. \tag{2.28}$$

Similarly

$$g^{\mu\nu} = -\hat{T}^\mu \hat{T}^\nu + \hat{N}^\mu \hat{N}^\nu + P^{\mu\nu}. \tag{2.29}$$

Now tangent vectors in $\partial\Sigma$ are contained in the span$\{\hat{U}^\mu, \hat{V}^\mu\}$ and since the coordinates of $\partial\Sigma$ are preserved under the diffeomorphism, δU^μ and δV^μ must also be contained in span$\{\hat{U}^\mu, \hat{V}^\mu\}$. By the same reasoning, $\delta P^{\mu\nu} \in$ span$\{\hat{U} \otimes \hat{U}, \hat{U} \otimes \hat{V}, \hat{V} \otimes \hat{U}, \hat{V} \otimes \hat{V}\}$.

Further, since $\delta U^\mu, \delta V^\mu \in$ span$\{\hat{U}^\mu, \hat{V}^\mu\}$, we may explicitly write them as

$$\delta\hat{U}^\mu = -\frac{1}{2}k_3\hat{U}^\mu - \frac{1}{2}k_4\hat{V}^\mu, \quad \delta\hat{V}^\mu = -\frac{1}{2}k_5\hat{U}^\mu - \frac{1}{2}k_6\hat{V}^\mu, \tag{2.30}$$

for some functions k_3, k_4, k_5 and k_6. By considering $\delta(g_{\mu\nu}\hat{U}^\mu\hat{U}^\nu) = 0$, $\delta(g_{\mu\nu}\hat{U}^\mu\hat{V}^\nu) = 0$ and $\delta(g_{\mu\nu}\hat{V}^\mu\hat{V}^\nu) = 0$, it is easy to show that

$$\begin{aligned} k_3 &= h_{\mu\nu}\hat{U}^\mu\hat{U}^\nu \\ k_6 &= h_{\mu\nu}\hat{V}^\mu\hat{V}^\nu \\ k_4 + k_5 &= 2h_{\mu\nu}\hat{U}^\mu\hat{V}^\nu. \end{aligned} \tag{2.31}$$

Then by considering $\delta(\hat{U}_\mu\hat{U}^\mu) = 0$, $\delta(\hat{U}_\mu\hat{V}^\mu) = 0$, $\delta(\hat{V}_\mu\hat{V}^\mu) = 0$ and $\delta(\hat{V}_\mu\hat{U}^\mu) = 0$, one finds

$$\delta\hat{U}_\mu = \frac{1}{2}k_3\hat{U}_\mu + \frac{1}{2}k_5\hat{V}_\mu, \quad \delta\hat{V}_\mu = \frac{1}{2}k_4\hat{U}_\mu + \frac{1}{2}k_6\hat{V}_\mu. \tag{2.32}$$

Hence, $\delta P^{\mu\nu}$ may be explicitly computed to be

$$\begin{aligned} \delta P^{\mu\nu} &= -k_3\hat{U}^\mu\hat{U}^\nu - k_6\hat{V}^\mu\hat{V}^\nu - \frac{1}{2}(k_4 + k_5)(\hat{U}^\mu\hat{V}^\nu + \hat{U}^\nu\hat{V}^\mu) \\ &= -\frac{1}{2}(k_3 + k_6)(\hat{U}^\mu\hat{U}^\nu + \hat{V}^\mu\hat{V}^\nu) - \frac{1}{2}(k_3 - k_6)(\hat{U}^\mu\hat{U}^\nu - \hat{V}^\mu\hat{V}^\nu) \\ &\quad -\frac{1}{2}(k_4 + k_5)(\hat{U}^\mu\hat{V}^\nu + \hat{U}^\nu\hat{V}^\mu) \\ &= -\frac{1}{2}(k_3 + k_6)P^{\mu\nu} - \frac{1}{2}(k_3 - k_6)(\hat{U}^\mu\hat{U}^\nu - \hat{V}^\mu\hat{V}^\nu) - \frac{1}{2}(k_4 + k_5)(\hat{U}^\mu\hat{V}^\nu + \hat{U}^\nu\hat{V}^\mu). \end{aligned} \tag{2.33}$$

Similarly,

$$\begin{aligned}
\delta P_{\mu\nu} &= k_3\hat{U}_\mu\hat{U}_\nu + k_6\hat{V}_\mu\hat{V}_\nu + \frac{1}{2}(k_4+k_5)(\hat{U}_\mu\hat{V}_\nu + \hat{U}_\nu\hat{V}_\mu) \\
&= \frac{1}{2}(k_3+k_6)P_{\mu\nu} + \frac{1}{2}(k_3-k_6)(\hat{U}_\mu\hat{U}_\nu - \hat{V}_\mu\hat{V}_\nu) \\
&\quad + \frac{1}{2}(k_4+k_5)(\hat{U}_\mu\hat{V}_\nu + \hat{U}_\nu\hat{V}_\mu).
\end{aligned} \tag{2.34}$$

So the key diffeomorphic conditions may be summarized as

$$\begin{aligned}
&\delta K^\mu = 0, && \delta K_\mu = h_{\mu\nu}K^\nu = k_1 K_\mu \\
&\delta\hat{T}^\mu = -\frac{1}{2}k_1\hat{T}^\mu, && \delta\hat{T}_\mu = \frac{1}{2}k_1\hat{T}_\mu \\
&\delta\hat{N}^\mu = -\frac{1}{2}k_2\hat{N}^\mu, && \delta\hat{N}_\mu = \frac{1}{2}k_2\hat{N}_\mu \\
&\delta\hat{T}^t = -\frac{1}{2}k_1\hat{T}^t, && \hat{T}^\mu h_{\mu\nu}\hat{N}^\nu = 0 \\
&\delta P^{\mu\nu} = \text{Eq.}(2.33), && \delta P_{\mu\nu} = \text{Eq.}(2.34).
\end{aligned} \tag{2.35}$$

where $k_1 = -h_{\mu\nu}\hat{T}^\mu\hat{T}^\nu$ and $k_2 = h_{\mu\nu}\hat{N}^\mu\hat{N}^\nu$.

From Eqs. (2.19), (2.25) and (2.31), we know that in the tetrad basis

$$h_{\hat{\mu}\hat{\nu}} = \begin{pmatrix} -k_1 & 0 & 0 & 0 \\ 0 & k_2 & 0 & 0 \\ 0 & 0 & k_3 & \frac{k_4+k_5}{2} \\ 0 & 0 & \frac{k_4+k_5}{2} & k_6 \end{pmatrix}, \tag{2.36}$$

i.e., $h_{\hat{T}\hat{T}} = -k_1, h_{\hat{N}\hat{N}} = k_2$ etc., and where generally $k_2, k_3, \ldots, k_6$ are independent functions from each other.

2.3.3 Diffeomorphic variation of energy

Now we begin to analyse how energy transforms under metric perturbations. We start by following Bardeen et al.'s original analysis [11], generalizing it where necessary to deal with a boundary $\partial\Sigma_{\text{inner}}$ which is an arbitrary ordinary surface instead of a horizon. It will be sufficient for our purposes to consider only the case where there is no matter on Σ itself, so $T^{\mu\nu} = 0$ there. Geometrically, this corresponds to all the matter lying behind or within the inner boundary $\partial\Sigma_{\text{inner}}$ (see Fig. 2.1).

In order to consider diffeomorphisms which need not respect spherical symmetry, we return to Eq. (2.1). Using the Einstein field equations we start by rewriting this integral formula as

$$E = \int_\Sigma (2\,T_{\mu\nu} + \frac{1}{8\pi}R\,g_{\mu\nu})K^\mu\,\hat{T}^\nu\,\sqrt{|\gamma^{(\Sigma)}|}\,d^3x + \frac{1}{4\pi}\int_{\partial\Sigma_{\text{inner}}}\kappa\,dA\,, \tag{2.37}$$

recall that $R = -8\pi T$ where $T = T_{\mu\nu} g^{\mu\nu}$.

Since $\mathcal{N}\sqrt{|\gamma^{(\Sigma)}|} = \sqrt{-g}$ on the hypersurface [30], the variation of the Ricci scalar term may be computed as

$$\begin{aligned}
& \frac{1}{8\pi}\int_\Sigma \delta(R\sqrt{|\gamma^{(\Sigma)}|}\, K^\beta\, \hat{T}_\beta)\, d^3x \\
= & \frac{1}{8\pi}\int_\Sigma \delta(R\mathcal{N}\sqrt{|\gamma^{(\Sigma)}|}\, \hat{T}^\beta\, \hat{T}_\beta)\, d^3x \\
= & \frac{1}{8\pi}\int_\Sigma \delta(R\sqrt{-g})\, \hat{T}^\beta\, \hat{T}_\beta\, d^3x \\
= & -\frac{1}{8\pi}\int_\Sigma \Big((R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R)h^{\mu\nu} - (g^{\mu\nu}\delta\Gamma^\alpha{}_{\mu\nu} - g^{\mu\alpha}\delta\Gamma^\lambda{}_{\lambda\mu})_{;\alpha}\Big)\, K^\beta\, \hat{T}_\beta\, \sqrt{|\gamma^{(\Sigma)}|}\, d^3x \,.
\end{aligned} \tag{2.38}$$

where in the last step we have used the well-known result that [16]

$$\delta(R\sqrt{-g}) = -\Big((R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R)h^{\mu\nu} - (g^{\mu\nu}\delta\Gamma^\alpha{}_{\mu\nu} - g^{\mu\alpha}\delta\Gamma^\lambda{}_{\lambda\mu})_{;\alpha}\Big)\sqrt{-g}. \tag{2.39}$$

Lemma 2.1: $-(g^{\mu\nu}\delta\Gamma^\alpha{}_{\mu\nu} - g^{\mu\alpha}\delta\Gamma^\lambda{}_{\lambda\mu})_{;\alpha} = 2\, h^\mu{}_{[\mu;\nu]}{}^{;\nu}$, a result quoted in Ref. [11], there without proof.

Proof:

Since [44]

$$\delta\Gamma^\alpha_{\mu\nu} = \frac{1}{2}g^{\alpha\rho}(h_{\mu\rho;\nu} + h_{\nu\rho;\mu} - h_{\mu\nu;\rho})\ , \tag{2.40}$$

we have

$$\begin{aligned}
& -(g^{\mu\nu}\delta\Gamma^\alpha{}_{\mu\nu} - g^{\mu\alpha}\delta\Gamma^\lambda{}_{\lambda\mu})_{;\alpha} \\
= & \Big(g^{\mu\alpha}\frac{1}{2}g^{\lambda\rho}(h_{\mu\rho;\lambda} + h_{\lambda\rho;\mu} - h_{\mu\lambda;\rho}) - g^{\mu\nu}\frac{1}{2}g^{\alpha\rho}(h_{\mu\rho;\nu} + h_{\nu\rho;\mu} - h_{\mu\nu;\rho})\Big)_{;\alpha} \\
= & \frac{1}{2}\Big(g^{\mu\alpha}h^\rho{}_{\rho;\mu} - g^{\mu\nu}(h^\alpha{}_{\mu;\nu} + h^\alpha{}_{\nu;\mu} - h_{\mu\nu}{}^{;\alpha})\Big)_{;\alpha} \\
= & \frac{1}{2}\Big(h_\rho{}^{\rho;\alpha} - (h^{\alpha\nu}{}_{;\nu} + h^{\alpha\mu}{}_{;\mu} - h_\mu{}^{\mu;\alpha})\Big)_{;\alpha} \\
= & \frac{1}{2}(2\, h_\rho{}^{\rho;\alpha} - 2\, h^{\alpha\mu}{}_{;\mu})_{;\alpha} \\
= & 2\, h^\mu{}_{[\mu;\nu]}{}^{;\nu}.
\end{aligned} \tag{2.41}$$

This completes the proof of Lemma 2.1.

Using Lemma 2.1, the variation of the Ricci scalar term Eq. (2.38) becomes

$$\begin{aligned}
& -\frac{1}{8\pi}\int_\Sigma \Big((R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R)h^{\mu\nu} + 2h^\mu{}_{[\mu;\nu]}{}^{;\nu}\Big)\, K^\beta\, \hat{T}_\beta\, \sqrt{|\gamma^{(\Sigma)}|}\, d^3x \\
= & -\frac{1}{4\pi}\int_\Sigma h^\mu{}_{[\mu;\nu]}{}^{;\nu} K^\beta\, \hat{T}_\beta\, \sqrt{|\gamma^{(\Sigma)}|}\, d^3x
\end{aligned} \tag{2.42}$$

since the first term is zero if we assume $T^{\mu\nu} = 0$ on Σ outside the holographic screen.

Lemma 2.2: $h^{\mu}{}_{[\mu;\nu]}{}^{;\nu}\xi^{\beta} = (\xi^{\beta}h_{\mu}{}^{[\mu;\nu]} - \xi^{\nu}h_{\mu}{}^{[\mu;\beta]})_{;\nu} + \mathfrak{L}_{\xi}(h_{\mu}{}^{[\mu;\beta]}) + h_{\mu}{}^{[\mu;\beta]}\xi^{\nu}{}_{;\nu}$.

Proof: Expanding out the right-hand-side (rhs) of the claim in Lemma 2.2, we get

$$\begin{aligned}
\text{rhs} &= h_{\mu}{}^{[\mu;\nu]}{}_{;\nu}\xi^{\beta} + h_{\mu}{}^{[\mu;\nu]}\xi^{\beta}{}_{;\nu} - h_{\mu}{}^{[\mu;\beta]}\xi^{\nu}{}_{;\nu} - h_{\mu}{}^{[\mu;\beta]}{}_{;\nu}\xi^{\nu} + \mathfrak{L}_{\xi}(h_{\mu}{}^{[\mu;\beta]}) + h_{\mu}{}^{[\mu;\beta]}\xi^{\nu}{}_{;\nu} \\
&= h^{\mu}{}_{[\mu;\nu]}{}^{;\nu}\xi^{\beta} + h_{\mu}{}^{[\mu;\nu]}\xi^{\beta}{}_{;\nu} - h_{\mu}{}^{[\mu;\beta]}{}_{;\nu}\xi^{\nu} + \mathfrak{L}_{\xi}(h_{\mu}{}^{[\mu;\beta]}) \\
&= h^{\mu}{}_{[\mu;\nu]}{}^{;\nu}\xi^{\beta} - \mathfrak{L}_{\xi}(h_{\mu}{}^{[\mu;\beta]}) + \mathfrak{L}_{\xi}(h_{\mu}{}^{[\mu;\beta]}) \\
&= \text{lhs}\,, \qquad (2.43)
\end{aligned}$$

where we used $h_{\mu}{}^{[\mu;\nu]}{}_{;\nu}\xi^{\beta} = h^{\mu}{}_{[\mu;\nu]}{}^{;\nu}\xi^{\beta}$ in the first line and $\mathfrak{L}_{\xi}(h_{\mu}{}^{[\mu;\beta]}) = h_{\mu}{}^{[\mu;\beta]}{}_{;\nu}\xi^{\nu} - h_{\mu}{}^{[\mu;\nu]}\xi^{\beta}{}_{;\nu}$ in the second line.

This completes the proof of Lemma 2.2.

When ξ^{μ} is the Killing vector K^{μ}, Lemma 2.2 reduces to a result quoted in Ref. [11], there without proof:

$$h^{\mu}{}_{[\mu;\nu]}{}^{;\nu}K^{\beta} = (K^{\beta}h_{\mu}{}^{[\mu;\nu]} - K^{\nu}h_{\mu}{}^{[\mu;\beta]})_{;\nu}. \qquad (2.44)$$

Applying Eq. (2.44) to Eq. (2.42), the variation of the term involving the Ricci scalar reduces to

$$-\frac{1}{4\pi}\int_{\Sigma}(K^{\beta}h_{\mu}{}^{[\mu;\nu]} - K^{\nu}h_{\mu}{}^{[\mu;\beta]})_{;\nu}\,\hat{T}_{\beta}\,\sqrt{|\gamma^{(\Sigma)}|}\,d^3x. \qquad (2.45)$$

Thus, the variation in the total mass may be written

$$\begin{aligned}
\delta E &= -\frac{1}{4\pi}\int_{\Sigma}(K^{\beta}h_{\mu}{}^{[\mu;\nu]} - K^{\nu}h_{\mu}{}^{[\mu;\beta]})_{;\nu}\,\hat{T}_{\beta}\,\sqrt{|\gamma^{(\Sigma)}|}\,d^3x + \frac{1}{4\pi}\int_{\partial\Sigma_{\text{inner}}}\delta(\kappa\,dA) \\
&= -\frac{1}{4\pi}\int_{\Sigma}(K^{\beta}h_{\mu}{}^{[\mu;\nu]} - K^{\nu}h_{\mu}{}^{[\mu;\beta]})_{;\nu}\,\hat{T}_{\beta}\,\sqrt{|\gamma^{(\Sigma)}|}\,d^3x \\
&\quad + \frac{1}{4\pi}\int_{\partial\Sigma_{\text{inner}}}\delta\kappa\,dA + \frac{1}{4\pi}\int_{\partial\Sigma_{\text{inner}}}\kappa\,\delta(dA). \qquad (2.46)
\end{aligned}$$

Since the term inside the bracket is an anti-symmetric tensor, we may use Stokes's theorem, Eq. (1.24), to obtain

$$\begin{aligned}
\delta E &= -\frac{1}{4\pi}\int_{\partial\Sigma_{\infty}}(K^{\beta}h_{\mu}{}^{[\mu;\nu]} - K^{\nu}h_{\mu}{}^{[\mu;\beta]})\hat{N}_{\nu}\,\hat{T}_{\beta}\,\sqrt{|\gamma^{(\partial\Sigma_{\infty})}|}\,d^2z \\
&\quad + \frac{1}{4\pi}\int_{\partial\Sigma_{\text{inner}}}(K^{\beta}h_{\mu}{}^{[\mu;\nu]} - K^{\nu}h_{\mu}{}^{[\mu;\beta]})\hat{N}_{\nu}\,\hat{T}_{\beta}\,\sqrt{|\gamma^{(\partial\Sigma_{\text{inner}})}|}\,d^2z \\
&\quad + \frac{1}{4\pi}\int_{\partial\Sigma_{\text{inner}}}\delta\kappa\,dA + \frac{1}{4\pi}\int_{\partial\Sigma_{\text{inner}}}\kappa\,\delta(dA), \qquad (2.47)
\end{aligned}$$

where the boundary has been split into the inner boundary $\partial\Sigma_{\text{inner}}$ and the boundary at infinity $\partial\Sigma_\infty$. The contribution for the term at infinity may be evaluated using the notation of tensorial volume elements [60] as

$$\begin{aligned}
&-\frac{1}{4\pi}\int_{\partial\Sigma_\infty}(K^\beta h_\mu{}^{[\mu;\nu]} - K^\nu h_\mu{}^{[\mu;\beta]})\hat{N}_\nu\,\hat{T}_\beta\,\sqrt{|\gamma^{(\partial\Sigma_\infty)}|}\,d^2z \\
= \;& -\frac{1}{4\pi}\int_{\partial\Sigma_\infty} K^\beta h_\mu{}^{[\mu;\nu]}(\hat{T}_\beta\,\hat{N}_\nu\, - \hat{T}_\nu\,\hat{N}_\beta)\sqrt{|\gamma^{(\partial\Sigma_\infty)}|}\,d^2z \\
= \;& -\frac{1}{4\pi}\int_{\partial\Sigma_\infty} K^\beta h_\mu{}^{[\mu;\nu]}\,\varepsilon_{\beta\nu}\,\varepsilon_{\alpha\mu} \\
= \;& \frac{1}{8\pi}\int_{\partial\Sigma_\infty}(h_\mu{}^{\mu;\nu} - h_\mu{}^{\nu;\mu})K^\beta\,\varepsilon_{\beta\nu\alpha\mu} \\
= \;& -\delta E\,,
\end{aligned} \tag{2.48}$$

where the orientation of $\varepsilon_{\beta\nu\alpha\mu}$ is chosen so that $\varepsilon_{\beta\nu\alpha\mu} = -6\,\varepsilon_{[\beta\nu}\varepsilon_{\alpha\mu]}$ and $\varepsilon_{\alpha\mu}$ is the 'volume' element of the boundary at infinity, and we have applied the result $\frac{1}{8\pi}\int_{\partial\Sigma_\infty}(h_\mu{}^{\mu;\nu} - h_\mu{}^{\nu;\mu})K^\beta\,\varepsilon_{\beta\nu\alpha\mu} = -\delta E$ in the final step [60].

Eq. (2.48) allows us to transform Eq. (2.47) into

$$\begin{aligned}
\delta E \;=\; & -\delta E + \frac{1}{4\pi}\int_{\partial\Sigma_{\text{inner}}}(K^\beta h_\mu{}^{[\mu;\nu]} - K^\nu h_\mu{}^{[\mu;\beta]})\hat{N}_\nu\,\hat{T}_\beta\,\sqrt{|\gamma^{(\partial\Sigma_{\text{inner}})}|}\,d^2z \\
& + \frac{1}{4\pi}\int_{\partial\Sigma_{\text{inner}}}\delta\kappa\,dA + \frac{1}{4\pi}\int_{\partial\Sigma_{\text{inner}}}\kappa\,\delta(dA).
\end{aligned} \tag{2.49}$$

Or equivalently,

$$\delta E = \frac{1}{8\pi}\int_{\partial\Sigma_{\text{inner}}}\frac{1}{2}(h_\mu{}^{\mu;\nu} - h_\mu{}^{\nu;\mu})\hat{N}_\nu\,\hat{T}_\beta K^\beta\,dA + \frac{1}{8\pi}\int_{\partial\Sigma_{\text{inner}}}\delta\kappa\,dA + \frac{1}{8\pi}\int_{\partial\Sigma_{\text{inner}}}\kappa\,\delta(dA)\,, \tag{2.50}$$

where we have used $K^\nu\hat{N}_\nu = \mathcal{N}\hat{T}^\nu\hat{N}_\nu = 0$, which follows since $\hat{T}^\mu$ is normal to Σ and $\hat{N}^\mu$ lies in Σ. For the first law to be true, we need the first two boundary integrals of Eq. (2.50) to exactly cancel each other.

Since $K^\mu\hat{N}_\mu = 0$ and $(K^\mu\hat{N}_\mu)_{;\nu} = 0$, then $K^\mu{}_{;\nu}\hat{N}_\mu = -K^\mu\hat{N}_{\mu;\nu}$. Therefore, the surface gravity may be written as

$$\kappa = K^\mu{}_{;\nu}\hat{T}^\nu\hat{N}_\mu = -K^\mu\hat{N}_{\mu;\nu}\hat{T}^\nu. \tag{2.51}$$

We next consider the expansion of null normal congruences on the inner boundary which may be written as

$$\begin{aligned}
\theta^{(l)} &= P^{\mu\nu} l_{\mu;\nu} \\
&= P^{\mu\nu}(\hat{T}_\mu + \hat{N}_\mu)_{;\nu} \\
&= P^{\mu\nu}(\hat{T}^t K_\mu)_{;\nu} + P^{\mu\nu}\hat{N}_{\mu;\nu} \\
&= \hat{T}^t P^{\mu\nu} K_{\mu;\nu} + P^{\mu\nu} K_\mu (\hat{T}^t)_{;\nu} + P^{\mu\nu}\hat{N}_{\mu;\nu} \\
&= P^{\mu\nu}\hat{N}_{\mu;\nu} \\
&= (g^{\mu\nu} + \hat{T}^\mu\hat{T}^\nu - \hat{N}^\mu\hat{N}^\nu)\hat{N}_{\mu;\nu} \\
&= \hat{N}^\mu{}_{;\mu} - K_{\mu;\nu}\hat{T}^\nu\hat{N}^\mu\hat{T}^t \\
&= \hat{N}^\mu{}_{;\mu} - \kappa\hat{T}^t \\
&= \frac{1}{\sqrt{-g}}(\sqrt{-g}\hat{N}^\mu)_{,\mu} - \kappa\hat{T}^t \qquad (2.52)
\end{aligned}$$

where $l_\mu = \hat{T}_\mu + \hat{N}_\mu$ is the outgoing null normal vector of the inner boundary, and we have used Eq. (2.29) in the fifth line and Eq. (2.51) in the sixth line. Using this relation we may express the variation of $\kappa\hat{T}^t$ as

$$\delta\hat{T}^t\kappa + \hat{T}^t\delta\kappa = \delta(\frac{1}{\sqrt{-g}})(\sqrt{-g}\hat{N}^\mu)_{,\mu} + \frac{1}{\sqrt{-g}}(\delta\sqrt{-g}\hat{N}^\mu)_{,\mu} + \frac{1}{\sqrt{-g}}(\sqrt{-g}\delta\hat{N}^\mu)_{,\mu} - \delta\theta^{(l)}. \qquad (2.53)$$

Since the left hand of Eq. (2.53) equals $-\frac{1}{2}k_1\hat{T}^t\kappa + \hat{T}^t\delta\kappa$, we further have

$$\begin{aligned}
&-\frac{1}{2}k_1\hat{T}^t\kappa + \hat{T}^t\delta\kappa \\
=& -\frac{1}{2}g^{\tau\nu}\delta g_{\tau\nu}\frac{1}{\sqrt{-g}}(\sqrt{-g}\hat{N}^\mu)_{,\mu} + \frac{1}{\sqrt{-g}}(\frac{1}{2}\sqrt{-g}g^{\tau\nu}\delta g_{\tau\nu}\hat{N}^\mu)_{,\mu} + (\delta\hat{N}^\mu)_{;\mu} - \delta\theta^{(l)} \\
=& -\frac{1}{2}g^{\tau\nu}\delta g_{\tau\nu}\frac{1}{\sqrt{-g}}(\sqrt{-g}\hat{N}^\mu)_{,\mu} + \frac{1}{2}g^{\tau\nu}\delta g_{\tau\nu}\frac{1}{\sqrt{-g}}(\sqrt{-g}\hat{N}^\mu)_{,\mu} \\
&+\frac{1}{2}(g^{\tau\nu}\delta g_{\tau\nu})_{,\mu}\hat{N}^\mu + (\delta\hat{N}^\mu)_{;\mu} - \delta\theta^{(l)} \\
=& \frac{1}{2}(h_\nu{}^\nu)_{,\mu}\hat{N}^\mu + (\delta\hat{N}^\mu)_{;\mu} - \delta\theta^{(l)}\ , \qquad (2.54)
\end{aligned}$$

where $\delta\sqrt{-g} = \frac{1}{2}\sqrt{-g}g^{\mu\nu}\delta g_{\mu\nu}$. Hence, the first term in Eq. (2.50) may be written as

$$\frac{1}{2}(h_\nu{}^\nu)_{;\mu}\hat{N}^\mu = -\frac{1}{2}k_1\hat{T}^t\kappa + \hat{T}^t\delta\kappa - (\delta\hat{N}^\mu)_{;\mu} + \delta\theta^{(l)} \qquad (2.55)$$

The second term ${h_\mu}^{\nu;\mu}\hat{N}_\nu$ in Eq. (2.50) may then be expressed as

$$\begin{aligned}
\frac{1}{2}{h_\mu}^{\nu;\mu}\hat{N}_\nu &= \frac{1}{2}({h_\mu}^\nu\hat{N}_\nu)^{;\mu} - \frac{1}{2}{h_\mu}^\nu{\hat{N}_\nu}^{;\mu} = \frac{1}{2}(\delta g_{\mu\nu}\hat{N}^\nu)^{;\mu} + \frac{1}{2}\delta g^{\mu\nu}\hat{N}_{\mu;\nu} \\
&= \frac{1}{2}\big(\delta(g_{\mu\nu}\hat{N}^\nu) - g_{\mu\nu}\delta\hat{N}^\nu\big)^{;\mu} + \frac{1}{2}\delta(-\hat{T}^\mu\hat{T}^\nu + \hat{N}^\mu\hat{N}^\nu + P^{\mu\nu})\hat{N}_{\mu;\nu} \\
&= \frac{1}{2}(\delta\hat{N}_\mu - \delta\hat{N}^\nu g_{\mu\nu})^{;\mu} + \frac{1}{2}(k_1\hat{T}^\mu\hat{T}^\nu - k_2\hat{N}^\mu\hat{N}^\nu + \delta P^{\mu\nu})\hat{N}_{\mu;\nu} \\
&= \frac{1}{2}(\frac{1}{2}k_2\hat{N}_\mu)^{;\mu} - \frac{1}{2}(\delta\hat{N}^\mu)_{;\mu} + \frac{1}{2}k_1\hat{T}^\mu\hat{T}^\nu\hat{N}_{\mu;\nu} + \frac{1}{2}\delta P^{\mu\nu}\hat{N}_{\mu;\nu} \\
&= \frac{1}{2}(\frac{1}{2}k_2\hat{N}^\mu)_{;\mu} - \frac{1}{2}(\delta\hat{N}^\mu)_{;\mu} - \frac{1}{2}k_1\hat{T}^t\hat{T}^\nu\hat{N}_\mu{K^\mu}_{;\nu} + \frac{1}{2}\delta P^{\mu\nu}\hat{N}_{\mu;\nu} \\
&= -(\delta\hat{N}^\mu)_{;\mu} - \frac{1}{2}k_1\hat{T}^t\kappa + \frac{1}{2}\delta P^{\mu\nu}\hat{N}_{\mu;\nu}, \qquad (2.56)
\end{aligned}$$

where we have used Eq. (2.51) in the fifth line, and Eqs. (2.7) and (2.35) in the last step.

Next, consider the final term $\frac{1}{2}\delta P^{\mu\nu}\hat{N}_{\mu;\nu}$ in Eq. (2.56), using Eq. (2.33) we have

$$\begin{aligned}
\frac{1}{2}\delta P^{\mu\nu}\hat{N}_{\mu;\nu} &= -\frac{1}{4}\Big((k_3+k_6)P^{\mu\nu} + (k_3-k_6)(\hat{U}^\mu\hat{U}^\nu - \hat{V}^\mu\hat{V}^\nu) \\
&\quad +(k_4+k_5)(\hat{U}^\mu\hat{V}^\nu + \hat{U}^\nu\hat{V}^\mu)\Big)\hat{N}_{\mu;\nu} \\
&= -\frac{1}{4}\Big((k_3+k_6)P^{\mu\nu} + (k_3-k_6)(\hat{U}^\mu\hat{U}^\nu - \hat{V}^\mu\hat{V}^\nu) \\
&\quad +(k_4+k_5)(\hat{U}^\mu\hat{V}^\nu + \hat{U}^\nu\hat{V}^\mu)\Big)(l_\mu - \hat{T}_\mu)_{;\nu} \\
&= -\frac{1}{4}\Big((k_3+k_6)P^{\mu\nu} + (k_3-k_6)(\hat{U}^\mu\hat{U}^\nu - \hat{V}^\mu\hat{V}^\nu) \\
&\quad +(k_4+k_5)(\hat{U}^\mu\hat{V}^\nu + \hat{U}^\nu\hat{V}^\mu)\Big)(l_\mu - \hat{T}^tK_\mu)_{;\nu} \\
&= -\frac{1}{4}(k_3+k_6)\theta^{(l)} - \frac{1}{4}(k_3-k_6)\sigma_+^{(l)} - \frac{1}{4}(k_4+k_5)\sigma_\times^{(l)}, \qquad (2.57)
\end{aligned}$$

where $\sigma_+^{(l)}, \sigma_\times^{(l)}$ are the shears of l^μ defined by [33]

$$\sigma_+^{(l)} = (\hat{U}^\mu\hat{U}^\nu - \hat{V}^\mu\hat{V}^\nu)l_{\mu;\nu} \qquad \sigma_\times^{(l)} = (\hat{U}^\mu\hat{V}^\nu + \hat{U}^\nu\hat{V}^\mu)l_{\mu;\nu}\ . \qquad (2.58)$$

Therefore,

$$\frac{1}{2}{h_\mu}^{\nu;\mu}\hat{N}_\nu = -(\delta\hat{N}^\mu)_{;\mu} - \frac{1}{2}k_1\hat{T}^t\kappa - \frac{1}{4}(k_3+k_6)\theta^{(l)} - \frac{1}{4}(k_3-k_6)\sigma_+^{(l)} - \frac{1}{4}(k_4+k_5)\sigma_\times^{(l)}. \quad (2.59)$$

Finally, substituting Eq. (2.55) and Eq. (2.59) into Eq. (2.50), we find

$$\begin{aligned}
\delta E &= -\frac{1}{8\pi}\int_{\partial\Sigma_{\rm inner}}\Big(\delta\kappa + \frac{1}{\hat{T}^t}\Big(\delta\theta^{(l)} + \frac{1}{4}(k_3+k_6)\theta^{(l)} + \frac{1}{4}(k_3-k_6)\sigma_+^{(l)} + \frac{1}{4}(k_4+k_5)\sigma_\times^{(l)}\Big)\Big)\,dA \\
&\quad + \frac{1}{8\pi}\int_{\partial\Sigma_{\rm inner}}\delta\kappa\,dA + \frac{1}{8\pi}\int_{\partial\Sigma_{\rm inner}}\kappa\,\delta(dA)\ . \qquad (2.60)
\end{aligned}$$

Or in summary,

$$\boxed{\begin{aligned}\delta E &= -\frac{1}{8\pi}\int_{\partial\Sigma_{\rm inner}}\left(\delta\theta^{(l)}+\frac{1}{4}(k_3+k_6)\theta^{(l)}+\frac{1}{4}(k_3-k_6)\sigma_+^{(l)}+\frac{1}{4}(k_4+k_5)\sigma_\times^{(l)}\right)\mathcal{N}\,dA\\ &\quad+\frac{1}{8\pi}\int_{\partial\Sigma_{\rm inner}}\kappa\,\delta(dA).\end{aligned}} \tag{2.61}$$

It is worth noting that

$$\boxed{\delta\theta^{(l)}=-\frac{k_2}{2}\theta^{(l)}+\frac{1}{2}(k_3+k_6)_{;\rho}\hat{N}^\rho} \tag{2.62}$$

which separately depends only on k_2, k_3, k_6 which we now prove.

Since $P^{\mu\nu}\hat{T}_{\mu;\nu}=P^{\mu\nu}(K_\mu\hat{T}^t)_{;\nu}=P^{\mu\nu}K_{\mu;\nu}\hat{T}^t+P^{\mu\nu}K_\mu(\hat{T}^t)_{;\nu}=0$, $\theta^{(l)}$ can be simplified as

$$\theta^{(l)}=P^{\mu\nu}l_{\mu;\nu}=P^{\mu\nu}(\hat{T}_{\mu;\nu}+\hat{N}_{\mu;\nu})=P^{\mu\nu}\hat{N}_{\mu;\nu}. \tag{2.63}$$

Thus the variation of $\theta^{(l)}$ is

$$\begin{aligned}\delta\theta^{(l)} &= \delta P^{\mu\nu}\hat{N}_{\mu;\nu}+P^{\mu\nu}\delta(\hat{N}_{\mu;\nu})\\ &= \delta P^{\mu\nu}\hat{N}_{\mu;\nu}+P^{\mu\nu}(\delta\hat{N}_{\mu,\nu}-\delta\Gamma^\lambda_{\mu\nu}\hat{N}_\lambda-\Gamma^\lambda_{\mu\nu}\delta\hat{N}_\lambda)\\ &= \delta P^{\mu\nu}\hat{N}_{\mu;\nu}+P^{\mu\nu}\Big((\delta\hat{N}_\mu)_{;\nu}-\delta\Gamma^\lambda_{\mu\nu}\hat{N}_\lambda\Big)\\ &= \delta P^{\mu\nu}\hat{N}_{\mu;\nu}+P^{\mu\nu}\Big((\frac{k_2}{2}\hat{N}_\mu)_{;\nu}-\frac{1}{2}g^{\lambda\rho}(h_{\mu\rho;\nu}+h_{\nu\rho;\mu}-h_{\mu\nu;\rho})\hat{N}_\lambda\Big)\\ &= \delta P^{\mu\nu}\hat{N}_{\mu;\nu}+P^{\mu\nu}(\frac{k_2}{2}\hat{N}_\mu)_{;\nu}-P^{\mu\nu}\frac{1}{2}(h_{\mu\rho;\nu}+h_{\nu\rho;\mu}-h_{\mu\nu;\rho})\hat{N}^\rho\\ &= \delta P^{\mu\nu}\hat{N}_{\mu;\nu}+P^{\mu\nu}(\frac{k_2}{2}\hat{N}_\mu)_{;\nu}-P^{\mu\nu}h_{\mu\rho;\nu}\hat{N}^\rho+\frac{1}{2}P^{\mu\nu}h_{\mu\nu;\rho}\hat{N}^\rho,\end{aligned} \tag{2.64}$$

where we have used Eq. (2.40) in the fourth line. Further, using $P^{\mu\nu}{}_{;\rho}\hat{N}^\rho h_{\mu\nu}=0$, we can now simplify Eq. (2.64) as

$$\begin{aligned}\delta\theta^{(l)} &= \delta P^{\mu\nu}\hat{N}_{\mu;\nu}+P^{\mu\nu}(\frac{k_2}{2}\hat{N}_\mu)_{;\nu}-(\hat{N}^\rho h_{\mu\rho})_{;\nu}P^{\mu\nu}+\hat{N}^\rho{}_{;\nu}P^{\mu\nu}h_{\mu\rho}+\frac{1}{2}(P^{\mu\nu}h_{\mu\nu})_{;\rho}\hat{N}^\rho\\ &\quad-\frac{1}{2}P^{\mu\nu}{}_{;\rho}\hat{N}^\rho h_{\mu\nu}\\ &= \delta P^{\mu\nu}\hat{N}_{\mu;\nu}+P^{\mu\nu}(\frac{k_2}{2}\hat{N}_\mu)_{;\nu}-(k_2\hat{N}_\mu)_{;\nu}P^{\mu\nu}+\hat{N}^\rho{}_{;\nu}(\delta(P^{\mu\nu}g_{\mu\rho})-\delta P^{\mu\nu}g_{\mu\rho})\\ &\quad+\frac{1}{2}(k_3+k_6)_{;\rho}\hat{N}^\rho\\ &= \delta P^{\mu\nu}\hat{N}_{\mu;\nu}-P^{\mu\nu}(\frac{k_2}{2}\hat{N}_\mu)_{;\nu}-\delta P^{\mu\nu}\hat{N}_{\mu;\nu}+\frac{1}{2}(k_3+k_6)_{;\rho}\hat{N}^\rho\\ &= -\frac{k_2}{2}P^{\mu\nu}\hat{N}_{\mu;\nu}+\frac{1}{2}(k_3+k_6)_{;\rho}\hat{N}^\rho\\ &= -\frac{k_2}{2}\theta^{(l)}+\frac{1}{2}(k_3+k_6)_{;\rho}\hat{N}^\rho.\end{aligned} \tag{2.65}$$

This completes the proof of Eq. (2.62).

As the expansion and shear vanish identically in the special case of the event horizon of a black hole [33], we see that Eq. (2.61) trivially reduces to the first law on the horizon, Eq. (2.13), thus reproducing the famous 1973 result [11]. Similarly, it follows straightforwardly that for surfaces sufficiently close to the horizon (so-called stretched horizons), the corrections to the first law can be made negligible.

2.4 Surfaces away from horizons.

2.4.1 Emergent gravity program

Since 2010, Verlinde's emergent gravity program has attracted huge attention because it claims that gravity may not be a fundamental effect but instead has a thermodynamic origin [59]. With some plucky assumptions this program even gives a heuristic derivation of the laws of Newton and the Einstein field equations.

One of the key assumptions in this argument is that ordinary surfaces away from horizons also behave thermodynamically. It is claimed that ordinary surfaces (which are called holographic screens) also have an entropy proportional to their surface area and a temperature proportional to their surface gravity. Since the Unruh effect shows that an accelerating observer would record a temperature proportional to its acceleration, to make their thermal assumption consistent with the Unruh effect, the holographic screens used in their analysis are taken to have a constant Newtonian potential ϕ. In this way, an observer that is stationary at a holographic screen will see the same temperature as is assigned to the holographic screen.

Then, Verlinde provides a heuristic derivation of the Einstein field equations using these thermodynamic assumptions. The derivation is heuristic in that it starts with an assumption of thermal behavior of holographic screens and proceeds with a sequence of true statements ending with the Einstein field equations. The problem lies in the reasoning connecting each statement in the sequence. One could test the rigour of the argument by focusing on the reasoning behind each step. However this would be ultimately unsatisfying since it would not rule out a different logically correct route. To rule out the existence of any such route we chose to test whether the thermodynamic properties of holographic screens were actually consistent with general relativity.

2.4.2 Local temperature in emergent gravity

We focus here on the temperature defined in the original paper on emergent gravity [59] which is used there in its heuristic derivation of the Einstein field equations. In Fig. 2.2 we show a schematic of the hypersurface considered there. $\partial\Sigma_{\rm HS}$ denotes the holographic screen (ordinary surfaces of constant Newtonian potential ϕ) which now is the *outer* boundary of the spacelike hypersurface $\Sigma_{\rm EG}$ under study, and $\hat{N}^\mu$ is the unit normal vector to the holographic screen.

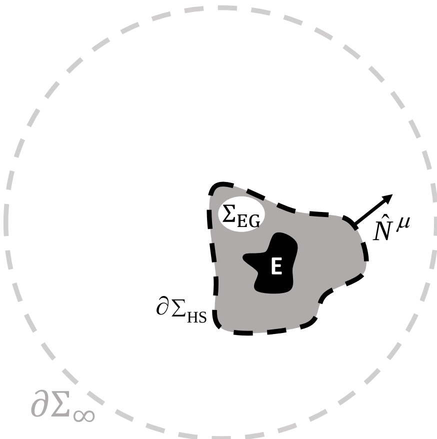

Figure 2.2: Schematic of the spacelike three-dimensional hypersurface $\Sigma_{\rm EG}$ used in Ref. [59] which has the mass under study embedded within it. As can be seen, the 2-surface corresponding to the holographic screen $\partial\Sigma_{\rm HS}$ is now the *outer* boundary to $\Sigma_{\rm EG}$ (compare to Fig. 2.1); and Ref. [59] defines $\partial\Sigma_{\rm HS}$ as ordinary surfaces of constant Newtonian potential ϕ. (For context, we show spatial infinity as $\partial\Sigma_\infty$ in grey, though it plays no role in this section.)

The 'local' temperature of the holographic screen (as measured at spatial infinity) used in Ref. [59] is defined as

$$T \equiv \frac{1}{2\pi} e^{\phi}\, \phi_{;\mu} \hat{N}^\mu, \tag{2.66}$$

where ϕ is the generalized Newtonian potential, given by $\phi = \frac{1}{2}\ln(-K^\mu K_\mu) = \ln \mathcal{N}$, recalling that $K^\mu K_\mu = -\mathcal{N}^2$. It is now an easy matter to check that

$$\begin{aligned} T &\equiv \frac{1}{2\pi} e^{\phi}\, \phi_{;\mu} \hat{N}^\mu = \frac{1}{2\pi}\mathcal{N}_{;\mu}\hat{N}^\mu = \frac{1}{2\pi}\frac{1}{2\mathcal{N}}(\mathcal{N}^2)_{;\mu}\hat{N}^\mu \\ &= -\frac{1}{2\pi}K_{\nu;\mu}\frac{1}{\mathcal{N}}\,K^\nu\hat{N}^\mu = \frac{1}{2\pi}K_{\mu;\nu}\frac{1}{\mathcal{N}}\,K^\nu\hat{N}^\mu \\ &= \frac{1}{2\pi}K_{\mu;\nu}\hat{T}^\nu\hat{N}^\mu. \end{aligned} \tag{2.67}$$

In summary, recall the definition of κ in Eq. (2.7), yielding

$$\boxed{T \equiv \frac{\kappa}{2\pi}.} \tag{2.68}$$

For reference, the Unruh temperature associated with a stationary observer is just the *magnitude* of the observer's proper acceleration a^μ over 2π. As the observer's 4-velocity is given by $\hat{T}^\mu$ we easily find

$$a^\mu \equiv \hat{T}^\mu{}_{;\nu}\hat{T}^\nu = \phi^{;\mu}, \tag{2.69}$$

since $\hat{T}^\mu = \hat{T}^t K^\mu = K^\mu/\mathcal{N} = e^{-\phi}K^\mu$. Thus a^μ is perpendicular to surfaces of constant ϕ. When Verlinde's temperature is measured locally (instead of referenced to spatial infinity) it is $T_{\text{local}} = \frac{1}{2\pi}\phi^{;\mu}\hat{N}_\mu$. For this to equal the Unruh temperature at the same point, the local unit normal $\hat{N}^\mu$ to the screen must be aligned with the proper acceleration a^μ of our stationary observer there. Therefore, it trivially follows that only for surfaces of constant Newtonian potential ϕ would the holographic screens be in thermal equilibrium with stationary physical surfaces of the same shape, size and location. Hence,

$$\boxed{\text{Thermodynamic equilibrium} \quad \Rightarrow \quad \hat{N}^\mu \parallel \phi^{;\mu}} \tag{2.70}$$

Finally, we show that for surfaces of constant ϕ, we have $\delta\phi = k_1/2$. Indeed, since $\hat{T}^t = 1/\mathcal{N} = e^{-\phi}$, we have

$$\boxed{\delta\phi = -\frac{1}{\hat{T}^t}\delta\hat{T}^t = \frac{1}{2}k_1,} \tag{2.71}$$

where in the last step we have used Eqs. (2.22) and (2.23).

2.4.3 Surfaces away from horizons generally do not satisfy the first law.

So far we have assumed that the inner boundaries before and after the diffeomorphic perturbation are arbitrary. But could the perturbed boundary be chosen in a specific manner so as to cause the unwanted terms in Eq. (2.61) to vanish? As already noted, holographic screens correspond to surfaces of constant Newtonian potential $\phi = \text{constant}$. Thus, the perturbed screen relies on a specification of the constant $\delta\phi$. In section 2.4.2, we showed that $\delta\phi = \frac{1}{2}k_1$, where k_1 is a metric perturbation of which the unwanted terms in Eq. (2.61) are wholly independent. Thus, the ordinary surfaces used within Verlinde's emergent gravity program cannot generally satisfy the first law, Eq. (2.13).

One caveat to this claim comes when we consider a fully spherically symmetric scenario; where both the initial spacetime and screen are spherically symmetric, so the initial shears $\sigma_j^{(l)}$ vanish, and also the final spacetime and screen are spherically symmetric, placing further constraints on the k_j. In this case, Birkhoff's theorem [17] for spherically symmetric metrics imposes extra constraints between the metric components so that a perturbed screen may always be chosen so as to satisfy the form of the first law [22]. However, as noted above, this form will not be preserved under arbitrary metric perturbations.

2.4.4 Vanishing extrinsic curvature tensor

Our analysis has been predicated on static screens. However, there is another way to define screens, so their normal direction remains parallel to the proper acceleration of a family of locally coincident timelike observers [46]. These observers are constrained to have constant 4-acceleration along with a number of other technical assumptions [46]. A first law is then obtained for these surfaces provided they additionally have a vanishing extrinsic curvature tensor $K_{\mu\nu} = 0$ [46]. The first law obtained is of a form with energy and temperature measured locally instead of at spatial infinity, which for asymptotically-flat spacetimes are unambiguous. Finally, we note that there is no easy way in this other formalism [46] to investigate stretched horizons.

In our setting with zero shift vector $\beta^\mu = 0$, so $\hat{T}^\mu = \hat{T}^t K^\mu$, and with our hypersurfaces Σ orthogonal to $\hat{T}^\mu$, we find that $K_{\mu\nu} = 0$ implies a vanishing expansion $\theta^{(l)} = 0$. Thus, for our setting, the formalism of Ref. [46] only yields a first law on horizons.

To see that this is the case, observe that the extrinsic curvature tensor of our inner boundary equals [21]

$$K_{\mu\nu} \equiv \hat{N}_{(\lambda;\rho)} P^\lambda{}_\mu P^\rho{}_\nu. \tag{2.72}$$

Taking the trace of this yields the extrinsic curvature scalar as $K = P^{\mu\nu}\hat{N}_{\mu;\nu} = \theta^{(l)}$, where in the final step we use Eq. (2.63). Thus, for our setting, the first law of Ref. [46] appears to occur at the horizon; a result which is naively consistent with the classic 1973 result.

Let us now consider a construction for a screen surrounding a gravitating body as proposed by Ref. [46]: Construct a screen using a family of stationary timelike observers at fixed radius around a Schwarzschild black hole. It is easy to calculate the extrinsic curvature tensor for the screen and see, as noted above, that this curvature vanishes only on the horizon. Hence the screen is on the horizon and the observers are null instead of timelike observers. Next drop in a spherical shell of matter. As the shell passes the screen of observers, the horizon (where $\theta^{(l)} = 0$) discontinuously jumps, the surface gravity of the new horizon changes and the original screen of observers fall into the black hole. We must then conclude either that the construction using the methods of Ref. [46] of a screen surrounding the black hole is simply impossible (because the observers are not timelike), or it fails to continue to hold under perturbation.

Thus, although Ref. [46] purports to describe a dynamical first law for ordinary surfaces its conditions are either in general impossible to satisfy or are generally *not* preserved under perturbation.

2.5 Discussion

The implications of our results are now described for (i) stretched horizons, and (ii) ordinary surfaces.

(i) Stretched horizons have long been considered to act as black bodies [54], effectively radiating with a temperature $\kappa/(2\pi)$. Thus, our demonstration that they also satisfy the first law to an excellent approximation hardly seems surprising. Nevertheless, we do not believe that our result here should be interpreted as implying that the surfaces corresponding to stretched horizons themselves should be imbued with actual thermodynamic properties.

In particular, we may consider an alternative spacetime, identical from the stretched horizon outward, but instead of a horizon, we consider an infinitesimal shell of matter just outside what would correspond to its Schwarzschild radius were the shell to collapse further, yet still within the 'stretched horizon'. In this latter spacetime, there is no horizon and hence no Hawking radiation. Notwithstanding this, our work proves that the 'stretched horizon' still closely satisfies the first law.

We conclude from this that the laws of black hole mechanics are not sufficient in themselves to guarantee whether any particular surface is truly thermodynamic in nature. For stretched horizons, we interpret this reasoning to imply that their full thermodynamic behavior is only inherited from the presence of an underlying horizon, but is not intrinsic to stretched horizons themselves. This conclusion appears to mimic the *initial* reluctance of general relativists [11] from accepting black hole horizons as truly thermodynamic despite the deep analogy to thermodynamics uncovered in the laws of black hole mechanics. By contrast, these laws should still be considered a necessary condition.

(ii) Our analysis further rigorously shows that the family of ordinary surfaces called holographic screens will generally not obey a first law of thermodynamics, in contrast to the long-standing result for horizons [11]. (Other families would not even be in thermal equilibrium with a physical surface at the same location.) Recall that the first law is more general than thermodynamics: the 'temperature' is merely an integrating factor relating changes in energy to changes in some state variable (entropy in the case of thermodynamics). Failure of the first law means that the putative state variable is not a variable of state at all. Therefore, even in static asymptotically-flat spacetimes, where Verlinde's emergent gravity program claims to derive the full Einstein field equations, our results show that the key assumption of this program is actually inconsistent with General Relativity.

Chapter 3

Thermodynamics of dynamical black holes

The thermodynamic theory of black holes posits that stationary black holes satisfy the laws of equilibrium thermodynamics such that they have a uniform temperature given by their surface gravity and a net entropy given by the area law. However, the validity of thermodynamics beyond stationary black holes remains a wide open question. Here we study asymptotically-flat dynamical spacetimes without global symmetries. We prove that the physical energy reduces to a simpler form in such spacetimes and show how this energy changes during physical processes. That is, we generalize the first law of thermodynamics for dynamical black holes. We find that spacetime horizons necessarily behave thermodynamically even in a dynamical setting. In general, the temperature along the horizons of dynamical black holes is found to be non-uniform, with equilibrium replaced by local equilibrium behavior. The local nature of the temperature and entropy on a black hole's horizon makes rigorous long-held intuitive claims on how information is encoded on a black hole's surface and may open the door for a reappraisal of well-known entropic bounds and paradoxes associated with black holes. Finally, our results demonstrate the logical equivalence between classical general relativity and the thermodynamic nature of spacetime horizons, suggesting new insights into a quantum theory of gravity [1].

[1] Zhi-Wei Wang & Samuel L. Braunstein. Submitted.

3.1 Introduction

Despite the elegance of the theory of black hole thermodynamics, the discovery of the thermal nature of stationary, i.e., eternal, black holes has led to a number of paradoxes relating to information retrieval, the nature of black hole entropy, black hole complementarity and the equivalence principle versus the fire wall [1,18,31,32,52]. The apparent clash between information loss in a thermal theory and information preservation in quantum mechanics [32] has been widely considered to hold the key for an eventual theory of quantum gravity [48]. However, while unitary evolution should presumably hold for arbitrary dynamics, black hole thermodynamics has only been rigorously proven for infinitesimal perturbations from the eternal (stationary) setting. Therefore, rigorously generalizing black hole thermodynamics to dynamical spacetimes may give us more hints for a theory of quantum gravity. I will achieve just such a rigorous generalisation in this chapter though for uncharged non-rotating spacetimes.

The first law describes how energy, E, changes under perturbation for a thermodynamic system at a given temperature and quantitatively defines this temperature. As such, it is the fundamental law which determines if a system is thermodynamic. For stationary uncharged non-rotating black holes, this law has the form [11]

$$\delta E = \frac{1}{8\pi}\,\kappa\,\delta A + \text{volume terms.} \tag{3.1}$$

Here κ is the conventional surface gravity, A is the horizon area, and the volume terms depend on the matter content. Equation 3.1 is analogous to the conventional first law of thermodynamics, $\delta E = T\,\delta S +$ volume terms, where in the analogy, the temperature is $T = \kappa/(2\pi)$ and the net entropy, $S = A/4$, is given by the area law. Hawking's discovery that stationary black holes radiate at a temperature $\kappa/(2\pi)$ cemented this result as more than mere analogy to thermodynamics [31].

Traditional approaches to proving the first law involve perturbations of a stationary spacetime, typically following changes in the physical, globally defined energy [11, 36]. In 1973, Bardeen, Carter and Hawking proved that stationary black holes satisfy an analogue to the first law of thermodynamics [11]. Their analysis was based on the assumption that both the initial and perturbed spacetimes must be stationary [11]. Two decades later, this result was rigorously extended to include arbitrary infinitesimal perturbations by Iyer and Wald but still from initially stationary spacetimes [36]. Crucially, important quantities such as surface gravity as defined in these approaches are only determined in the stationary case (corresponding to the initial spacetime). Since surface gravity plays the role of temperature, these rigorous analyses are limited to thermal equilibrium and cannot exhibit out-of-equilibrium behaviour that might be expected for astrophysical black holes. Such approaches cannot probe dynamical surface gravities and the perturbations are generally not considered to describe physical processes [61]. Nevertheless, this stationary behavior has been used on an *ad hoc* basis to model dynamical scenarios by assuming an evolution through a succession of equilibrium states [28].

Unlike the above approaches for obtaining a first law by considering energy perturbations, Ashtekar et al. obtained a first law (and a zeroth law) of black hole thermodynamics by studying the geometric structures of equilibrium horizons which they called (weakly) isolated horizons [4–8]. The key idea of an isolated horizon is to be able to model the state of a post-collapse black hole 'after' ring down. In particular, such a horizon is specifically defined to be at equilibrium with its exterior and hence has a well-defined and provably uniform surface gravity. Although such horizons satisfy an analogue to the first law of black hole mechanics in terms of changes in the "individual horizon masses" [7], the equilibrium requirements limit the applicability of this method. Indeed, it was shown that even the Brill-Lindquist initial conditions do not correspond to a scenario involving isolated horizons, except as an approximation for large inter-black hole distance [39]. By contrast, I will show that our method may be applied to interacting black holes (corresponding exactly to the Brill-Lindquist initial conditions) at the end of this chapter.

Other approaches to obtaining a first law explicitly consider physical processes associated with a local flux of energy across the horizon [27]. Such local-flux approaches do reproduce the results for stationary black holes, but remain unproven in the dynamical setting. A further difficulty with these approaches is that one cannot unambiguously define energy locally in general relativity [40].

Despite this large body of work extending the first law, the dynamical case remains an open problem: Different approaches lead to different definitions of dynamical surface gravity [42]. Indeed, for over a decade it has been recognized that there is no consensus for how to define a horizon's dynamical surface gravity even for spherically symmetric scenarios [42].

Here we return to first principles and the original perturbative approach [11] to derive the first law of thermodynamics for dynamical spacetimes. Because we are now working in a fully dynamical setting, our perturbatively connected pair of spacetimes implicitly includes scenarios where they correspond to 'nearby' time slices of a single evolving dynamical system. Thus, our approach allows for the description of physical processes while still maintaining mathematical rigor.

The first step is to obtain an integral equation for the net energy in this dynamical system. Then we must convert this into a description of how small "changes" in the net energy may be accounted for via a differential version of energy balance. The changes here actually refer to the differences between nearby solutions of the Einstein field equations related by a small diffeomorphism of the initial configuration.

We focus specifically on generalizing the first law of black hole mechanics from a one-black hole stationary spacetime to potentially multi-horizon dynamical spacetimes (where there is no timelike Killing vector). Generalizing the integral version of the first law into a multi-horizon spacetime is relatively straightforward. To construct a differential first law, we study diffeomorphisms of this integral formulation. For simplicity we only consider uncharged, non-rotating horizons. We suppose that the spacetime is asymptotically-flat from Theorem 3.1 onwards.

3.2 Covariant energy expression for dynamical spacetimes

Since the first law studies how energy is transformed, we start with a rigorous formulation of physical energy in dynamical spacetimes. For asymptotically-flat spacetimes this is provided within the Hamiltonian formulation by the ADM mass [2]. Unfortunately, the ADM mass is non-covariant in form and difficult to deal with. Our following analysis shows, for a broad class of *dynamical* asymptotically-flat spacetimes, that the ADM mass equals a convenient covariant expression.

3.2.1 Covariant conserved energy-momentum flux tensor

We start by reviewing Komar's generalized energy formalism [38]. For an arbitrary vector field ξ^μ, we define an energy-momentum flux tensor $S^{\mu\nu}(\xi)$

$$S^{\mu\nu}(\xi) \equiv \frac{1}{2}(\xi^{\nu;\mu} - \xi^{\mu;\nu}) \equiv \xi^{[\nu;\mu]}. \tag{3.2}$$

Like the anti-symmetric electromagnetic field tensor, this tensor has a corresponding 'energy' density flux vector $J^\mu(\xi)$ given by [38]

$$J^\mu(\xi) = S^{\mu\nu}(\xi)_{;\nu} = \xi^{[\nu;\mu]}{}_{;\nu} \,. \tag{3.3}$$

The covariant divergence of $J^\mu(\xi)$ vanishes since

$$\begin{aligned}
J^\mu{}_{;\mu} &= \xi^{[\nu;\mu]}{}_{;\nu\mu} = \frac{1}{2}(\xi^{\nu;\mu})_{;\nu\mu} - \frac{1}{2}(\xi^{\mu;\nu})_{;\nu\mu} \\
&= \frac{1}{2}(\xi^{\nu;\mu})_{;\nu\mu} - \frac{1}{2}(\xi^{\nu;\mu})_{;\mu\nu} \\
&= -\frac{1}{2}R^\nu{}_{\alpha\nu\mu}\xi^{\alpha;\mu} - \frac{1}{2}R^\mu{}_{\alpha\nu\mu}\xi^{\nu;\alpha} \\
&= -\frac{1}{2}R^\nu{}_{\alpha\nu\mu}\xi^{\alpha;\mu} + \frac{1}{2}R^\mu{}_{\alpha\mu\nu}\xi^{\nu;\alpha} \\
&= \frac{1}{2}R_{\alpha\nu}\xi^{\nu;\alpha} - \frac{1}{2}R_{\alpha\mu}\xi^{\alpha;\mu} \\
&= \frac{1}{2}R_{\alpha\nu}(\xi^{\nu;\alpha} - \xi^{\alpha;\nu}) = 0.
\end{aligned} \tag{3.4}$$

Thus J^μ is a *locally* covariantly conserved quantity for arbitrary vector fields ξ^μ. Integrating Eq. (3.4) over a 4-volume, a subvolume, $\mathcal{V} \subset \mathcal{M}$ of the entire manifold, yields

$$\int_{\mathcal{V}} J^\mu{}_{;\mu}\sqrt{-g}\, d^4z = 0, \tag{3.5}$$

and applying stokes' theorem Eq. (1.23) we find

$$\int_{\partial\mathcal{V}} J^\mu \varepsilon \hat{n}_\mu \sqrt{\gamma^{(\partial\mathcal{V})}}\, d^3x = \int_{\partial\mathcal{V}} \xi^{[\nu;\mu]}{}_{;\nu}\varepsilon\hat{n}_\mu\sqrt{\gamma^{(\partial\mathcal{V})}}\, d^3x = 0, \tag{3.6}$$

where $\partial\mathcal{V}$ is the boundary of $\mathcal{V}$ and $\hat{n}^\mu$ is the outward pointing unit vector normal to $\partial\mathcal{V}$, see Fig. 3.1, and $\gamma^{(\partial\mathcal{V})}$ is the determinant of the induced metric on $\partial\mathcal{V}$. This means that the current flux into the 4-volume equals the current flux out. This is a *local* conservation law for an *arbitrary* vector field in an arbitrary dynamical spacetime.

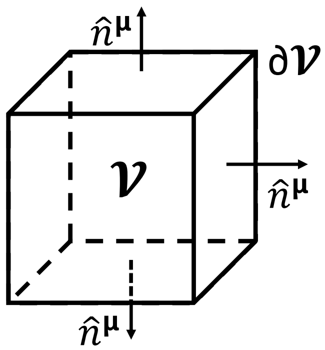

Figure 3.1: This 4-volume $\mathcal{V}$ is a subset of the entire spacetime manifold $\mathcal{M}$. Here $\partial\mathcal{V}$ is the boundary of $\mathcal{V}$, and $\hat{n}^\mu$ is the outgoing unit vector normal to the boundary $\partial\mathcal{V}$.

Next, consider a family of non-intersecting spacelike hypersurfaces with *vanishing* net flux $J^\mu\hat{N}_\mu$ out through spatial infinity ($\hat{N}^\mu$ is the spacelike unit vector normal to the boundary at spatial infinity). Consider a volume $\mathcal{V}$ consisting of the region between a pair of such hypersurfaces Σ_1, Σ_2 (see Fig. 3.2) then from Eq. (3.6) we find

$$\int_{\Sigma_1} \xi^{[\nu;\mu]}{}_{;\nu} \hat{T}_\mu \sqrt{\gamma^{(\Sigma_1)}}\, d^3x = \int_{\Sigma_2} \xi^{[\nu;\mu]}{}_{;\nu} \hat{T}_\mu \sqrt{\gamma^{(\Sigma_2)}}\, d^3x, \tag{3.7}$$

where $\hat{T}_\mu$ is the future directed timelike unit normal to the hypersurfaces.

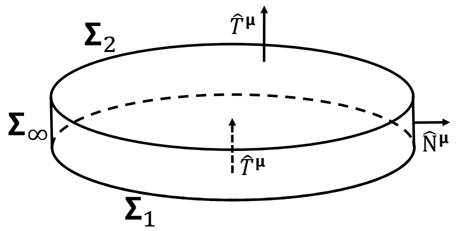

Figure 3.2: This 4-volume $\mathcal{V}$ is a region between two infinity large three dimensional hypersurfaces Σ_1, Σ_2; and Σ_1, Σ_2 and spatial infinity Σ_∞ together make up of its boundary. Here $\hat{T}^\mu$ is the timelike unit normal vector pointing to the future, and $\hat{N}^\mu$ is the spacelike outgoing unit vector normal to the spatial infinity.

Therefore, the integral of $J^\mu \hat{T}_\mu$ is a conserved quantity independent of which of the hypersurfaces is chosen and is the generalized Komar energy $E(\xi)$ [36, 38, 43, 63],

$$E(\xi) \equiv \frac{1}{4\pi} \int_{\Sigma} \xi^{[\nu;\mu]}{}_{;\nu} \hat{T}_\mu \sqrt{\gamma^{(\Sigma)}}\, d^3x, \tag{3.8}$$

where here Σ only has a boundary at spatial infinity $\partial\Sigma_\infty$, see Fig 3.3.

Now applying stokes' theorem for an anti-symmetric tensor on Σ, Eq. (1.24) to Eq. (3.8) yields

$$\boxed{E(\xi) \equiv \frac{1}{4\pi} \int_{\partial\Sigma_\infty} \xi^{[\nu;\mu]} \hat{N}_\nu \hat{T}_\mu \sqrt{\gamma^{(\partial\Sigma_\infty)}}\, d^2z\ .} \tag{3.9}$$

Note, $E(\xi)$ reduces to the Komar mass when ξ^μ is the appropriately chosen Killing vector for a stationary spacetime. We also require that the net outward flux $J^\mu \hat{N}_\mu$ vanish at spatial infinity to ensure that the integral of $J^\mu \hat{T}_\mu$ is preserved on every hypersurface. We prove below that this condition generally holds for the conventional spacelike hypersurfaces on asymptotically-flat spacetime.

Next we review the asymptotically-flat conditions and, based on these conditions, discuss the connection between the generalized Komar energy and ADM mass.

3.2.2 Conventional asymptotically-flat criteria

We first review the asymptotically-flat conditions used by York in the ADM mass definition [71]. Based on the Euclidean distance in asymptotically rectilinear coordinates, the metric

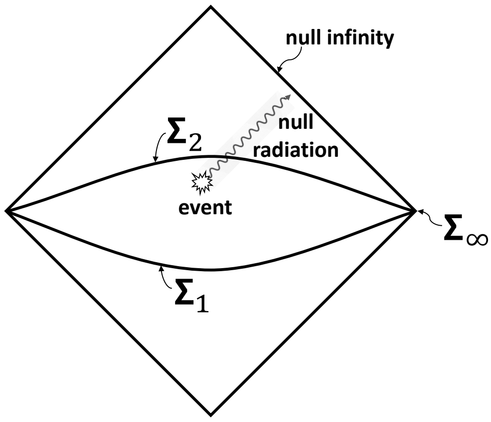

Figure 3.3: The Penrose diagram shows that gravitational and light radiation created at any finite epoch does not reach spatial infinity, Σ_∞. Thus $E(\xi)$ of Eq. (3.8) is conserved independent of the presence or absence of null (e.g., gravitational) radiation produced at any finite epoch.

and the extrinsic curvature of the hypersurface Σ at spatial infinity are assumed to take the asymptotic form

$$g_{\mu\nu} = \eta_{\mu\nu} + O\Big(\frac{1}{r}\Big), \;\; g_{ij,k} = O\Big(\frac{1}{r^2}\Big), \;\; g_{ij,kl} = O\Big(\frac{1}{r^3}\Big), \tag{3.10}$$

$$K_{ij} = O\Big(\frac{1}{r^2}\Big), \;\; K_{ij,k} = O\Big(\frac{1}{r^3}\Big), \;\; \ldots \tag{3.11}$$

where we use asymptotically rectilinear coordinates with $r^2 = (x^1)^2 + (x^2)^2 + (x^3)^2$ and $\eta_{\mu\nu} = \mathrm{diag}(-1,1,1,1)$, and the big-O notation $f = O(r^{-n})$ implies that $\lim_{r\to\infty} r^n f$ equals some non-vanishing constant. Here K_{ij} is the extrinsic curvature of Σ in $\mathcal{M}$ and as usual Greek indices run from 0 to 3, and lower-case Latin indices run from 1 to 3.

Since

$$K_{\mu\nu} = -\frac{1}{2}\mathfrak{L}_{\hat{T}}\gamma_{\mu\nu} = -\frac{1}{2}\gamma_{\mu\nu,\alpha}\hat{T}^\alpha - \frac{1}{2}\hat{T}^\alpha{}_{,\mu}\gamma_{\alpha\nu} - \frac{1}{2}\hat{T}^\alpha{}_{,\nu}\gamma_{\mu\alpha},$$

and $\hat{T}^\mu = (1,0,0,0) + O(1/r)$, the extrinsic curvature asymptotically-flat conditions of Eq. (3.11) reduce to

$$g_{ij,0} = O\Big(\frac{1}{r^2}\Big), \;\; g_{0i,j} = O\Big(\frac{1}{r^2}\Big), \;\; g_{ij,0k} = O\Big(\frac{1}{r^3}\Big), \;\; g_{0i,jk} = O\Big(\frac{1}{r^3}\Big)\;, \;\ldots . \tag{3.12}$$

Therefore, only the asymptotically-flat conditions for $g_{00,i}$, $g_{0\mu,0}$ and $g_{\mu\nu,00}$ are not specified. Parallel with Eq. (3.10), it is conventional to assume $g_{00,i} = O(r^{-2})$, $g_{0\mu,0} = O(r^{-2})$ and $g_{\mu\nu,00} = O(r^{-3})$, largely by making an analogy between temporal and spatial derivatives, hence the asymptotically-flat conditions are usually summarized as

$$g_{\mu\nu} = \eta_{\mu\nu} + O\Big(\frac{1}{r}\Big), \;\; g_{\mu\nu,\alpha} = O\Big(\frac{1}{r^2}\Big), \;\; g_{\mu\nu,\lambda\beta} = O\Big(\frac{1}{r^3}\Big), \;\; \ldots . \tag{3.13}$$

This is an alternate form to those of York, Eqs. (3.10) and (3.11) [71] for the asymptotically-flat conditions used by some authors [67]. Consequently, $\Gamma^{\alpha}_{\mu\nu}$ satisfies

$$\Gamma^{\alpha}_{\mu\nu} = \frac{1}{2} g^{\alpha\rho}(g_{\mu\rho,\nu} + g_{\nu\rho,\mu} - g_{\mu\nu,\rho}) = O\Big(\frac{1}{r^2}\Big) \;, \tag{3.14}$$

and $R_{\mu\nu} = O(r^{-3})$.

3.2.3 Covariant expression of ADM energy for dynamical spacetimes

Our covariant expression of ADM energy in dynamical spacetimes is based on some prior results obtained by Ashtekar and Hansen, so we start by introducing a representation of the ADM mass proved by them [9]. It is worth noting that Chruściel gave a straightforward proof and statement of this result in his work [23, 24].

The Ashtekar-Hansen ADM Theorem: For asymptotically matter free ($T_{\mu\nu} = o(r^{-3})$) spacetimes, the ADM 4-momentum vector p^{ADM}_{μ} may be written as [9, 23, 24]

$$p^{\mathrm{ADM}}_{\mu} \xi^{\mu}_{\infty} \equiv \frac{1}{8\pi} \int_{\partial\Sigma_\infty} \xi^{\mu} x^{\nu} R_{\mu\nu\alpha\beta} \hat{N}^{\alpha} \hat{T}^{\beta} dA. \tag{3.15}$$

Here the components of ξ^{μ}_{∞} are constant in the asymptotically rectilinear coordinate system, $\xi^{\mu} = \xi^{\mu}_{\infty} + O(1/r^n)$, for $n > 0$ (we will determine n below), $dA = \sqrt{\gamma^{(\partial\Sigma_\infty)}}\, d^2 z$, and x^{ν} is an asymptotically rectilinear coordinate (and hence the expression Eq. (3.15) is not covariant).

Note, the normalization used in Ref. [9] is different from that shown here since they use a different area measure. Ashtekar and Hansen prove this in a context where they are only interested in stationary spacetimes with a Killing vector, so it is perhaps unclear how general Eq. (3.15) is supposed to be. We therefore give a detailed proof with clearly stated assumptions. In particular, we will not assume the existence of any Killing vectors.

Proof:

It is sufficient for our purposes to use Eq. (3.15) for the ADM mass. Therefore we limit our proof to p^{ADM}_{0}, i.e., we take $\xi^{\mu} = (\partial_t)^{\mu} + O(r^{-n})$ throughout this proof. Note that rather than directly saying that the ADM mass is labeled by a vector field at spatial infinity equaling ($\xi^\mu = \partial_t$), one may assume [9, 23, 24]

$$\xi^\mu = \partial_t + O\Big(\frac{1}{r^n}\Big),\ \ \xi^\mu{}_{,\nu} = O\Big(\frac{1}{r^{1+n}}\Big),\ \ \xi^\mu{}_{,\nu\beta} = O\Big(\frac{1}{r^{2+n}}\Big),\ \ldots, \tag{3.16}$$

for some $n > 0$ to be determined.

Lemma 3.1:
$p^{\mathrm{ADM}}_\mu \xi^\mu_\infty = \frac{3}{8\pi}\int_{\partial\Sigma_\infty} \delta^{[\beta}_\lambda \delta^\alpha_\mu \delta^{\gamma]}_\nu \xi^\nu \eta^{\lambda\rho}\eta^{\mu\sigma} g_{\sigma\gamma,\rho}\hat{N}_\alpha \hat{T}_\beta dA$ equals the ADM mass.
Proof:

$$\begin{aligned}
p^{\mathrm{ADM}}_\mu \xi^\mu_\infty &= \frac{3}{8\pi}\int_{\partial\Sigma_\infty} \delta^{[\beta}_\lambda \delta^\alpha_\mu \delta^{\gamma]}_\nu \xi^\nu \eta^{\lambda\rho}\eta^{\mu\sigma} g_{\sigma\gamma,\rho}\hat{N}_\alpha \hat{T}_\beta dA \\
&= \frac{-3}{8\pi}\int_{\partial\Sigma_\infty} \Big(\delta^{[0}_\lambda \delta^i_\mu \delta^{j]}_0 \eta^{\lambda\rho}\eta^{\mu\sigma} g_{\sigma j,\rho}\hat{N}_i + O(\frac{1}{r^{2+n}})\Big) dA \\
&= \frac{-3}{8\pi}\int_{\partial\Sigma_\infty} \frac{1}{3}\Big(\delta^{[0}_\lambda \delta^{i]}_\mu \delta^j_0 + \delta^{[i}_\lambda \delta^{j]}_\mu \delta^0_0 + \delta^{[j}_\lambda \delta^{0]}_\mu \delta^i_0\Big)\eta^{\lambda\rho}\eta^{\mu\sigma} g_{\sigma j,\rho}\hat{N}_i dA \\
&= \frac{1}{16\pi}\int_{\partial\Sigma_\infty} \Big(\eta^{i\sigma}\eta^{j\rho} - \eta^{i\rho}\eta^{j\sigma}\Big) g_{\sigma j,\rho}\hat{N}_i dA \\
&= \frac{1}{16\pi}\int_{\partial\Sigma_\infty} \Big(g_{ij}{}^{,j}\hat{N}^i - g_{jj}{}^{,i}\hat{N}_i\Big) dA \\
&= M^{\mathrm{ADM}},
\end{aligned} \tag{3.17}$$

where in the first line we use $\xi^\mu = \delta^\mu_0 + O(r^{-n})$ and $\hat{T}_\mu = (-\mathcal{N}, 0, 0, 0)$.

This completes the proof of Lemma 3.1.

Since $-3!\,\delta^{[\beta}_\lambda \delta^\alpha_\mu \delta^{\gamma]}_\nu = \varepsilon^{\tau\beta\alpha\gamma}\varepsilon_{\tau\lambda\mu\nu}$ [60], $\hat{N}_{[\alpha}\hat{T}_{\beta]}dA = dS_{\alpha\beta} = \frac{1}{2}\varepsilon_{\alpha\beta\tau_1\tau_2}dx^{\tau_1}\wedge dx^{\tau_2}$ [9,23,24], and $g_{\sigma\gamma,\rho} = \Gamma_{\sigma\gamma\rho} + \Gamma_{\gamma\sigma\rho}$, the new ADM mass expression $p^{\mathrm{ADM}}_\mu \xi^\mu_\infty$ in Lemma 3.1 can then be further simplified as

$$\begin{aligned}
p^{\mathrm{ADM}}_\mu \xi^\mu_\infty &= \frac{3}{8\pi}\int_{\partial\Sigma_\infty} \frac{-1}{3!}\varepsilon^{\tau\beta\alpha\gamma}\varepsilon_{\tau\lambda\mu\nu}\xi^\nu \eta^{\lambda\rho}\eta^{\mu\sigma} g_{\sigma\gamma,\rho}\Big(\frac{1}{2}\varepsilon_{\alpha\beta\tau_1\tau_2}dx^{\tau_1}\wedge dx^{\tau_2}\Big) \\
&= \frac{1}{8\pi}\int_{\partial\Sigma_\infty} \frac{-1}{4}\varepsilon_{\tau\lambda\mu\nu}\xi^\nu \eta^{\lambda\rho}\eta^{\mu\sigma} g_{\sigma\gamma,\rho}(\varepsilon^{\tau\beta\alpha\gamma}\varepsilon_{\alpha\beta\tau_1\tau_2})dx^{\tau_1}\wedge dx^{\tau_2} \\
&= \frac{1}{8\pi}\int_{\partial\Sigma_\infty} \frac{-1}{4}\varepsilon_{\tau\lambda\mu\nu}\xi^\nu \eta^{\lambda\rho}\eta^{\mu\sigma} g_{\sigma\gamma,\rho}(2!2!)\delta^\tau_{[\tau_1}\delta^\gamma_{\tau_2]}dx^{\tau_1}\wedge dx^{\tau_2} \\
&= \frac{-1}{8\pi}\int_{\partial\Sigma_\infty} \varepsilon_{\tau\lambda\mu\nu}\xi^\nu \eta^{\lambda\rho}\eta^{\sigma\mu} g_{\sigma\gamma,\rho}dx^\tau\wedge dx^\gamma \\
&= \frac{-1}{8\pi}\int_{\partial\Sigma_\infty} \varepsilon_{\tau\lambda\mu\nu}\xi^\nu \eta^{\lambda\rho}\eta^{\sigma\mu}(\Gamma_{\sigma\gamma\rho} + \Gamma_{\gamma\sigma\rho})dx^\tau\wedge dx^\gamma \\
&= \frac{-1}{8\pi}\int_{\partial\Sigma_\infty} \varepsilon_{\tau\lambda\mu\nu}\xi^\nu \eta^{\lambda\rho}\Gamma^\mu_{\gamma\rho}dx^\tau\wedge dx^\gamma,
\end{aligned} \tag{3.18}$$

where $\Gamma_{\gamma\sigma\rho}$ in the fifth line vanishes because the symmetric indices σ and ρ are mapping to an anti-symmetric tensor.

To further simplify Eq. (3.18), we first introduce some differential tricks we will use. Since $d\sqrt{-g} = \frac{1}{2}\sqrt{-g}\, g^{\delta\beta} g_{\delta\beta,\alpha} dx^\alpha$ and $\varepsilon_{\tau\lambda\mu\nu} = \sqrt{-g}\,[\tau\lambda\mu\nu]$, we have

$$\begin{aligned} d\varepsilon_{\tau\lambda\mu\nu} &= \frac{1}{2}\varepsilon_{\tau\lambda\mu\nu} g^{\delta\beta} g_{\delta\beta,\alpha} dx^\alpha \\ &= \frac{1}{2}\varepsilon_{\tau\lambda\mu\nu} g^{\delta\beta}(\Gamma_{\delta\beta\alpha} + \Gamma_{\beta\delta\alpha}) dx^\alpha \\ &= \varepsilon_{\tau\lambda\mu\nu}\Gamma^\beta_{\beta\alpha} dx^\alpha. \end{aligned} \tag{3.19}$$

Using Leibnitz's rule for the exterior derivative, Eq. (3.18) may be simplified as

$$\begin{aligned} p^{\mathrm{ADM}}_\mu \xi^\mu_\infty &= \frac{-1}{8\pi}\int_{\partial\Sigma_\infty} \varepsilon_{\tau\lambda\mu\nu}\xi^\nu\eta^{\lambda\rho}\Gamma^\mu_{\gamma\rho} dx^\tau \wedge dx^\gamma \\ &= \frac{-1}{8\pi}\int_{\partial\Sigma_\infty} d(\varepsilon_{\tau\lambda\mu\nu}\xi^\nu\eta^{\lambda\rho}\Gamma^\mu_{\gamma\rho} x^\tau dx^\gamma) - x^\tau d(\varepsilon_{\tau\lambda\mu\nu}\xi^\nu\eta^{\lambda\rho}\Gamma^\mu_{\gamma\rho}) \wedge dx^\gamma \\ &= \frac{1}{8\pi}\int_{\partial\Sigma_\infty} x^\tau d(\varepsilon_{\tau\lambda\mu\nu}\xi^\nu\eta^{\lambda\rho}\Gamma^\mu_{\gamma\rho}) \wedge dx^\gamma \\ &= \frac{1}{8\pi}\int_{\partial\Sigma_\infty} x^\tau\eta^{\lambda\rho}\Big(\xi^\nu\Gamma^\mu_{\gamma\rho} d\varepsilon_{\tau\lambda\mu\nu}\wedge dx^\gamma + \varepsilon_{\tau\lambda\mu\nu}\Gamma^\mu_{\gamma\rho} d\xi^\nu \wedge dx^\gamma \\ &\quad + \varepsilon_{\tau\lambda\mu\nu}\xi^\nu d\Gamma^\mu_{\gamma\rho}\wedge dx^\gamma\Big) \\ &= \frac{1}{8\pi}\int_{\partial\Sigma_\infty} x^\tau\eta^{\lambda\rho}\Big(\xi^\nu\Gamma^\mu_{\gamma\rho}\varepsilon_{\tau\lambda\mu\nu}\Gamma^\beta_{\beta\alpha} dx^\alpha\wedge dx^\gamma + \varepsilon_{\tau\lambda\mu\nu}\Gamma^\mu_{\gamma\rho} O(\frac{1}{r^{1+n}})^\nu_\alpha dx^\alpha\wedge dx^\gamma \\ &\quad + \varepsilon_{\tau\lambda\mu\nu}\xi^\nu d\Gamma^\mu_{\gamma\rho}\wedge dx^\gamma\Big) \\ &= \frac{1}{8\pi}\int_{\partial\Sigma_\infty} x^\tau\eta^{\lambda\rho}\varepsilon_{\tau\lambda\mu\nu}\xi^\nu\Gamma^\mu_{\gamma\rho,\alpha} dx^\alpha\wedge dx^\gamma \\ &= \frac{1}{8\pi}\int_{\partial\Sigma_\infty} x^\tau g^{\lambda\rho}\varepsilon_{\tau\lambda\mu\nu}\xi^\nu\Gamma^\mu_{\rho[\gamma,\alpha]} dx^\alpha\wedge dx^\gamma + O\Big(\frac{1}{r}\Big), \end{aligned} \tag{3.20}$$

where stokes' theorem and the fact that the boundary of a boundary is an empty set are used in the second step, and Eq. (3.19) and the asymptotically-flat conditions are used in the fifth step.

Then since $\Gamma^\mu_{\rho[\gamma,\alpha]} = -\frac{1}{2}R^\mu{}_{\rho\gamma\alpha} + O(\frac{1}{r^4})$ and $dx^\alpha\wedge dx^\gamma = -\frac{1}{2}\varepsilon^{\alpha\gamma\tau_1\tau_2} dS_{\tau_1\tau_2}$, Eq. (3.20) may be simplified as

$$\begin{aligned} p^{\mathrm{ADM}}_\mu \xi^\mu_\infty &= \frac{1}{8\pi}\int_{\partial\Sigma_\infty} \varepsilon_{\tau\lambda\mu\nu} x^\tau\xi^\nu g^{\lambda\rho}\Big(-\frac{1}{2}R^\mu{}_{\rho\gamma\alpha} + O(\frac{1}{r^4})\Big) dx^\alpha\wedge dx^\gamma \\ &= \frac{1}{16\pi}\int_{\partial\Sigma_\infty} \varepsilon_{\mu\lambda\nu\tau}\xi^\nu x^\tau R^{\mu\lambda}{}_{\alpha\gamma} dx^\alpha\wedge dx^\gamma \\ &= \frac{1}{16\pi}\int_{\partial\Sigma_\infty} \varepsilon_{\mu\lambda\nu\tau}\xi^\nu x^\tau R^{\mu\lambda}{}_{\alpha\gamma}(-\frac{1}{2}\varepsilon^{\alpha\gamma\tau_1\tau_2} dS_{\tau_1\tau_2}) \\ &= \frac{-1}{32\pi}\int_{\partial\Sigma_\infty} \varepsilon_{\mu\lambda\nu\tau}\varepsilon^{\alpha\gamma\tau_1\tau_2}\xi^\nu x^\tau R^{\mu\lambda}{}_{\alpha\gamma} dS_{\tau_1\tau_2} \\ &= \frac{-1}{32\pi}\int_{\partial\Sigma_\infty} (-4!\delta^\alpha_{[\mu}\delta^\gamma_\lambda\delta^{\tau_1}_\nu\delta^{\tau_2}_{\tau]})\xi^\nu x^\tau R^{\mu\lambda}{}_{\alpha\gamma} dS_{\tau_1\tau_2}. \end{aligned} \tag{3.21}$$

Next as $\delta^{\alpha}_{[\mu}\delta^{\gamma}_{\lambda}\delta^{\tau_1}_{\nu}\delta^{\tau_2}_{\tau]}$ may be expanded as

$$\begin{aligned}
\delta^{\alpha}_{[\mu}\delta^{\gamma}_{\lambda}\delta^{\tau_1}_{\nu}\delta^{\tau_2}_{\tau]} &= \frac{1}{3!}\Big(\delta^{\alpha}_{[\mu}\delta^{\gamma}_{\lambda]}\delta^{\tau_1}_{[\nu}\delta^{\tau_2}_{\tau]} - \delta^{\alpha}_{[\mu}\delta^{\gamma}_{\nu]}\delta^{\tau_1}_{[\lambda}\delta^{\tau_2}_{\tau]} + \delta^{\alpha}_{[\mu}\delta^{\gamma}_{\tau]}\delta^{\tau_1}_{[\lambda}\delta^{\tau_2}_{\nu]} \\
&\quad +\delta^{\alpha}_{[\lambda}\delta^{\gamma}_{\nu]}\delta^{\tau_1}_{[\mu}\delta^{\tau_2}_{\tau]} + \delta^{\alpha}_{[\lambda}\delta^{\gamma}_{\tau]}\delta^{\tau_1}_{[\nu}\delta^{\tau_2}_{\mu]} + \delta^{\alpha}_{[\tau}\delta^{\gamma}_{\nu]}\delta^{\tau_1}_{[\lambda}\delta^{\tau_2}_{\mu]}\Big) \\
&= \frac{1}{3!}\Big(\delta^{[\alpha}_{\mu}\delta^{\gamma]}_{\lambda}\delta^{[\tau_1}_{\nu}\delta^{\tau_2]}_{\tau} - \delta^{[\alpha}_{\mu}\delta^{\gamma]}_{\nu}\delta^{[\tau_1}_{\lambda}\delta^{\tau_2]}_{\tau} + \delta^{[\alpha}_{\mu}\delta^{\gamma]}_{\tau}\delta^{[\tau_1}_{\lambda}\delta^{\tau_2]}_{\nu} \\
&\quad +\delta^{[\alpha}_{\lambda}\delta^{\gamma]}_{\nu}\delta^{[\tau_1}_{\mu}\delta^{\tau_2]}_{\tau} + \delta^{[\alpha}_{\lambda}\delta^{\gamma]}_{\tau}\delta^{[\tau_1}_{\nu}\delta^{\tau_2]}_{\mu} + \delta^{[\alpha}_{\tau}\delta^{\gamma]}_{\nu}\delta^{[\tau_1}_{\lambda}\delta^{\tau_2]}_{\mu}\Big),
\end{aligned} \tag{3.22}$$

Eq. (3.21) may be transformed into

$$\begin{aligned}
&\frac{1}{8\pi}\int_{\partial\Sigma_\infty} 3!(\delta^{\alpha}_{[\mu}\delta^{\gamma}_{\lambda}\delta^{\tau_1}_{\nu}\delta^{\tau_2}_{\tau]})\xi^{\nu}x^{\tau}R^{\mu\lambda}{}_{\alpha\gamma}dS_{\tau_1\tau_2} \\
=\;& \frac{1}{8\pi}\int_{\partial\Sigma_\infty}\Big(\delta^{[\alpha}_{\mu}\delta^{\gamma]}_{\lambda}\delta^{[\tau_1}_{\nu}\delta^{\tau_2]}_{\tau} - \delta^{[\alpha}_{\mu}\delta^{\gamma]}_{\nu}\delta^{[\tau_1}_{\lambda}\delta^{\tau_2]}_{\tau} + \delta^{[\alpha}_{\mu}\delta^{\gamma]}_{\tau}\delta^{[\tau_1}_{\lambda}\delta^{\tau_2]}_{\nu} \\
&+\delta^{[\alpha}_{\lambda}\delta^{\gamma]}_{\nu}\delta^{[\tau_1}_{\mu}\delta^{\tau_2]}_{\tau} + \delta^{[\alpha}_{\lambda}\delta^{\gamma]}_{\tau}\delta^{[\tau_1}_{\nu}\delta^{\tau_2]}_{\mu} + \delta^{[\alpha}_{\tau}\delta^{\gamma]}_{\nu}\delta^{[\tau_1}_{\lambda}\delta^{\tau_2]}_{\mu}\Big)\xi^{\nu}x^{\tau}R^{\mu\lambda}{}_{\alpha\gamma}dS_{\tau_1\tau_2} \\
=\;& \frac{1}{8\pi}\int_{\partial\Sigma_\infty}\xi^{\nu}x^{\tau}\Big(R^{\mu\lambda}{}_{\mu\lambda}dS_{\nu\tau} - R^{\mu\lambda}{}_{\mu\nu}dS_{\lambda\tau} + R^{\mu\lambda}{}_{\mu\tau}dS_{\lambda\nu} + R^{\mu\lambda}{}_{\lambda\nu}dS_{\mu\tau} \\
&+R^{\mu\lambda}{}_{\lambda\tau}dS_{\nu\mu} + R^{\mu\lambda}{}_{\tau\nu}dS_{\lambda\mu}\Big) \\
=\;& \frac{1}{8\pi}\int_{\partial\Sigma_\infty}\xi^{\nu}x^{\tau}\Big(R\,dS_{\nu\tau} - 2R^{\mu}{}_{\nu}dS_{\mu\tau} + 2R^{\mu}{}_{\tau}dS_{\mu\nu} + R^{\mu\lambda}{}_{\tau\nu}dS_{\lambda\mu}\Big) \\
=\;& \frac{1}{8\pi}\int_{\partial\Sigma_\infty}\xi^{\nu}x^{\tau}R^{\mu\lambda}{}_{\tau\nu}dS_{\lambda\mu} = \frac{1}{8\pi}\int_{\partial\Sigma_\infty}\xi^{\nu}x^{\tau}R^{\mu\lambda}{}_{\tau\nu}\hat{N}_{\lambda}\hat{T}_{\mu}dA \\
=\;& \frac{1}{8\pi}\int_{\partial\Sigma_\infty}\xi^{\mu}x^{\nu}R_{\mu\nu\alpha\beta}\hat{N}^{\alpha}\hat{T}^{\beta}dA,
\end{aligned} \tag{3.23}$$

where we use the asymptotically matter-free condition $T_{\mu\nu} = o(r^{-3})$ in the fourth step and $dS_{\lambda\mu} = \hat{N}_{[\lambda}\hat{T}_{\mu]}dA$ in the fifth step. Here the little-o notation $f = o(r^{-n})$ implies that $\lim_{r\to\infty} r^n f = 0$.

This completes the proof of the Ashtekar and Hansen ADM Theorem (for the case of the ADM mass).

To continue our discussion, instead of directly using Ashtekar and Hansen's ADM formula, we prefer to use a more general formula for the ADM mass *prior* to assuming

$T_{\mu\nu} = o(r^{-3})$. From the sixth line of Eq. (3.23), we find

$$\begin{aligned} p_\mu^{\rm ADM}\xi^\mu_\infty &= \frac{1}{8\pi}\int_{\partial\Sigma_\infty} \xi^\nu x^\tau \Big(R\hat{N}_{[\nu}\hat{T}_{\tau]} - 2R^\mu{}_\nu \hat{N}_{[\mu}\hat{T}_{\tau]} + 2R^\mu{}_\tau \hat{N}_{[\mu}\hat{T}_{\nu]}\Big) dA \\ &\quad + \frac{1}{8\pi}\int_{\partial\Sigma_\infty} \xi^\mu x^\nu R_{\mu\nu\alpha\beta}\hat{N}^\alpha \hat{T}^\beta dA \\ &= \frac{1}{16\pi}\int_{\partial\Sigma_\infty} \xi^\nu x^\tau \Big(-R\hat{T}_\nu \hat{N}_\tau + 2R^\mu{}_\nu \hat{T}_\mu \hat{N}_\tau + 2R^\mu{}_\tau \hat{N}_\mu \hat{T}_\nu\Big) dA \\ &\quad + \frac{1}{8\pi}\int_{\partial\Sigma_\infty} \xi^\mu x^\nu R_{\mu\nu\alpha\beta}\hat{N}^\alpha \hat{T}^\beta dA \\ &= \frac{1}{16\pi}\int_{\partial\Sigma_\infty} \Big(R\hat{N}_\tau x^\tau + 2R_{\mu\nu}\hat{T}^\mu \xi^\nu \hat{N}_\tau x^\tau - 2R_{\mu\tau}\hat{N}^\mu \hat{x}^\tau\Big) dA \\ &\quad + \frac{1}{8\pi}\int_{\partial\Sigma_\infty} \xi^\mu x^\nu R_{\mu\nu\alpha\beta}\hat{N}^\alpha \hat{T}^\beta dA, \end{aligned} \tag{3.24}$$

where in the first step we use $\xi^\mu \hat{N}_\mu = O(r^{-n})$ and $x^\beta \hat{T}_\beta = -t$ at spatial infinity with t finite on the hypersurface.

Since the ADM mass is the accepted definition of physical energy as seen at spatial infinity even in dynamical spacetimes, to continue our discussion with the generalized Komar energy, we prove:

Theorem 3.1: For asymptotically-flat asymptotically matter-free spacetimes, the ADM mass equals the generalized Komar energy [38]

$$E(\xi) = \frac{1}{4\pi}\int_{\partial\Sigma_\infty} \xi^{[\mu;\nu]}\hat{N}_\mu \hat{T}_\nu \, dA, \tag{3.25}$$

on hypersurfaces labeled by fixed time t provided asymptotically (in rectilinear coordinates): i) $\xi^\mu = \partial_t + o(r^{-1})$, and ii) $g_{0i,j0} = o(r^{-3})$ and $g_{ij,00} = o(r^{-3})$.

Proof:

In order to prove the ADM mass equals the generalized Komar energy, we first prove three lemmas.

Lemma 3.2: $\xi_{\mu;\nu} = \xi_{[\mu;\nu]} + \frac{1}{2}\mathfrak{L}_\xi(g_{\mu\nu})$ for any vector field ξ^μ.

Proof:

Since

$$\begin{aligned} \xi_{(\mu;\nu)} &= \frac{1}{2}(\xi_{\mu;\nu} + \xi_{\nu;\mu}) \\ &= \frac{1}{2}g_{\mu\nu;\tau}\xi^\tau + \frac{1}{2}\xi^\tau{}_{;\mu}g_{\tau\nu} + \frac{1}{2}\xi^\tau{}_{;\nu}g_{\mu\tau} \\ &= \frac{1}{2}\mathfrak{L}_\xi g_{\mu\nu}, \end{aligned} \tag{3.26}$$

we can then simplify $\xi_{\mu;\nu}$ as

$$\xi_{\mu;\nu} = \xi_{[\mu;\nu]} + \xi_{(\mu;\nu)} = \xi_{[\mu;\nu]} + \frac{1}{2}\mathfrak{L}_\xi g_{\mu\nu}. \tag{3.27}$$

This completes the proof of Lemma 3.2.

Lemma 3.3: $R_{\mu\nu\alpha\beta}\xi^\mu = \xi_{[\beta;\alpha];\nu} + \frac{1}{2}(\mathfrak{L}_\xi g_{\nu\alpha})_{;\beta} - \frac{1}{2}(\mathfrak{L}_\xi g_{\beta\nu})_{;\alpha}$ for any vector field ξ^μ.

Proof:

Since $\xi_{\nu;\alpha\beta} - \xi_{\nu;\beta\alpha} = R_{\mu\nu\alpha\beta}\xi^\mu$ [47], the right-hand-side (rhs) of Lemma 3.3 may be simplified as

$$\begin{aligned}
\text{rhs} &= \xi_{[\beta;\alpha];\nu} + \xi_{(\nu;\alpha);\beta} - \xi_{(\beta;\nu);\alpha} \\
&= \frac{1}{2}\Big(\xi_{\beta;\alpha\nu} - \xi_{\alpha;\beta\nu} + \xi_{\nu;\alpha\beta} + \xi_{\alpha;\nu\beta} - \xi_{\beta;\nu\alpha} - \xi_{\nu;\beta\alpha}\Big) \\
&= \frac{1}{2}(\xi_{\beta;\alpha\nu} - \xi_{\beta;\nu\alpha}) + \frac{1}{2}(\xi_{\nu;\alpha\beta} - \xi_{\nu;\beta\alpha}) + \frac{1}{2}(\xi_{\alpha;\nu\beta} - \xi_{\alpha;\beta\nu}) \\
&= \frac{1}{2}\Big(R_{\mu\beta\alpha\nu}\xi^\mu + R_{\mu\nu\alpha\beta}\xi^\mu + R_{\mu\alpha\nu\beta}\xi^\mu\Big) \\
&= \frac{1}{2}(R_{\mu\nu\alpha\beta}\xi^\mu - R_{\mu\nu\beta\alpha}\xi^\mu) \\
&= R_{\mu\nu\alpha\beta}\xi^\mu = \text{lhs},
\end{aligned} \tag{3.28}$$

where we have used $R_{\mu\alpha\nu\beta} + R_{\mu\nu\beta\alpha} + R_{\mu\beta\alpha\nu} = 0$ [47] in the fourth line to obtain the fifth line.

This completes the proof of Lemma 3.3.

Then by contracting ν and β, we have

$$\xi_{[\nu;\alpha]}{}^{;\nu} = R_{\mu\alpha}\xi^\mu - \frac{1}{2}(\mathfrak{L}_\xi g_{\nu\alpha})^{;\nu} + \frac{1}{2}(\mathfrak{L}_\xi g_{\beta\nu})_{;\alpha} g^{\nu\beta}. \tag{3.29}$$

Lemma 3.4: $2\xi^{[\alpha;\beta]} = -3(\xi^{[\beta;\alpha}x^{\nu]})_{;\nu} + \xi^{[\beta;\alpha]}{}_{;\nu}x^\nu + \xi^{[\nu;\beta]}{}_{;\nu}x^\alpha + \xi^{[\alpha;\nu]}{}_{;\nu}x^\beta$ for any vector field ξ^μ and coordinates x^μ.

Proof:

Recalling that the coordinates x^μ are scalar functions on the manifold we have $\delta^\beta_\nu =$

$x^{\beta}{}_{;\nu}$ and trivially $3\xi^{[\beta;\alpha}x^{\nu]} = \xi^{[\beta;\alpha]}x^{\nu} + \xi^{[\alpha;\nu]}x^{\beta} + \xi^{[\nu;\beta]}x^{\alpha}$, we may now write

$$\begin{aligned}
\xi^{[\alpha;\beta]} &= \xi^{[\alpha;\nu]}\delta^{\beta}_{\nu} = \xi^{[\alpha;\nu]}x^{\beta}{}_{;\nu} = (\xi^{[\alpha;\nu]}x^{\beta})_{;\nu} - \xi^{[\alpha;\nu]}{}_{;\nu}x^{\beta} \\
&= (3\xi^{[\beta;\alpha}x^{\nu]} - \xi^{[\beta;\alpha]}x^{\nu} - \xi^{[\nu;\beta]}x^{\alpha})_{;\nu} - \xi^{[\alpha;\nu]}{}_{;\nu}x^{\beta} \\
&= 3(\xi^{[\beta;\alpha}x^{\nu]})_{;\nu} - \xi^{[\beta;\alpha]}{}_{;\nu}x^{\nu} - \xi^{[\beta;\alpha]}x^{\nu}{}_{;\nu} - \xi^{[\nu;\beta]}{}_{;\nu}x^{\alpha} - \xi^{[\nu;\beta]}x^{\alpha}{}_{;\nu} - \xi^{[\alpha;\nu]}{}_{;\nu}x^{\beta} \\
&= 3(\xi^{[\beta;\alpha}x^{\nu]})_{;\nu} - \xi^{[\beta;\alpha]}{}_{;\nu}x^{\nu} - 4\xi^{[\beta;\alpha]} - \xi^{[\nu;\beta]}{}_{;\nu}x^{\alpha} - \xi^{[\alpha;\beta]} - \xi^{[\alpha;\nu]}{}_{;\nu}x^{\beta} \\
&= 3(\xi^{[\beta;\alpha}x^{\nu]})_{;\nu} - \xi^{[\beta;\alpha]}{}_{;\nu}x^{\nu} + 3\xi^{[\alpha;\beta]} - \xi^{[\nu;\beta]}{}_{;\nu}x^{\alpha} - \xi^{[\alpha;\nu]}{}_{;\nu}x^{\beta}\ , \qquad (3.30)
\end{aligned}$$

or equivalently,

$$2\xi^{[\alpha;\beta]} = -3(\xi^{[\beta;\alpha}x^{\nu]})_{;\nu} + \xi^{[\beta;\alpha]}{}_{;\nu}x^{\nu} + \xi^{[\nu;\beta]}{}_{;\nu}x^{\alpha} + \xi^{[\alpha;\nu]}{}_{;\nu}x^{\beta}\ , \qquad (3.31)$$

This completes the proof of Lemma 3.4.

Consequently:

$$\xi^{[\beta;\alpha]}{}_{;\nu}x^{\nu} = 2\xi^{[\alpha;\beta]} + 3(\xi^{[\beta;\alpha}x^{\nu]})_{;\nu} - \xi^{[\nu;\beta]}{}_{;\nu}x^{\alpha} - \xi^{[\alpha;\nu]}{}_{;\nu}x^{\beta}\ . \qquad (3.32)$$

Further, from Eq. (3.16), we have

$$\begin{aligned}
\mathfrak{L}_{\xi}g_{\mu\nu} &= g_{\mu\nu,\tau}\xi^{\tau} + \xi^{\tau}{}_{,\mu}g_{\tau\nu} + \xi^{\tau}{}_{,\nu}g_{\mu\tau} \\
&= g_{\mu\nu,0} + \xi^{\tau}{}_{,\mu}g_{\tau\nu} + \xi^{\tau}{}_{,\nu}g_{\mu\tau} + O\Big(\frac{1}{r^{2+n}}\Big) \\
&= g_{\mu\nu,0} + O\Big(\frac{1}{r^{1+n}}\Big). \qquad (3.33)
\end{aligned}$$

Consequently, we have $(\mathfrak{L}_{\xi}g_{\mu\nu})_{;\beta} = g_{\mu\nu,0\beta} + O(r^{-4}) + O(r^{-2-n})$ and recall that $n > 0$ and conventionally $g_{\mu\nu,0\beta} = O(r^{-3})$.

Applying Lemma 3.3 to the last term of Eq. (3.24) yields

$$\begin{aligned}
&\frac{1}{8\pi}\int_{\partial\Sigma_\infty}\xi^\mu x^\nu R_{\mu\nu\alpha\beta}\hat{N}^\alpha\hat{T}^\beta dA \\
=\ &\frac{1}{8\pi}\int_{\partial\Sigma_\infty}\xi_{[\beta;\alpha];\nu}x^\nu\hat{N}^\alpha\hat{T}^\beta dA+\frac{1}{16\pi}\int_{\partial\Sigma_\infty}x^\nu((\mathfrak{L}_\xi g_{\nu\alpha})_{;\beta}-(\mathfrak{L}_\xi g_{\beta\nu})_{;\alpha})\hat{N}^\alpha\hat{T}^\beta dA \\
=\ &\frac{1}{8\pi}\int_{\partial\Sigma_\infty}\Big(2\xi^{[\alpha;\beta]}+3(\xi^{[\beta;\alpha}x^{\nu]})_{;\nu}-\xi^{[\nu;\beta]}{}_{;\nu}x^\alpha-\xi^{[\alpha;\nu]}{}_{;\nu}x^\beta\Big)\hat{N}_\alpha\hat{T}_\beta dA \\
&+\frac{1}{16\pi}\int_{\partial\Sigma_\infty}x^\nu((\mathfrak{L}_\xi g_{\nu\alpha})_{;\beta}-(\mathfrak{L}_\xi g_{\beta\nu})_{;\alpha})\hat{N}^\alpha\hat{T}^\beta dA \\
=\ &E(\xi)+\frac{1}{8\pi}\int_{\partial\Sigma_\infty}\Big(-\xi^{[\nu;\beta]}{}_{;\nu}x^\alpha-\xi^{[\alpha;\nu]}{}_{;\nu}x^\beta\Big)\hat{N}_\alpha\hat{T}_\beta dA \\
&+\frac{1}{16\pi}\int_{\partial\Sigma_\infty}x^\nu((\mathfrak{L}_\xi g_{\nu\alpha})_{;\beta}-(\mathfrak{L}_\xi g_{\beta\nu})_{;\alpha})\hat{N}^\alpha\hat{T}^\beta dA \\
=\ &E(\xi)+\frac{1}{16\pi}\int_{\partial\Sigma_\infty}\bigg(2\Big(R_{\mu\alpha}\xi^\mu\hat{N}^\alpha\hat{T}_\beta x^\beta-R_{\mu\beta}\xi^\mu\hat{T}^\beta\hat{N}_\alpha x^\alpha\Big) \\
&+\Big((\mathfrak{L}_\xi g_{\nu\beta})^{;\nu}-(\mathfrak{L}_\xi g_{\lambda\nu})_{;\beta}g^{\nu\lambda}\Big)x^\alpha\hat{N}_\alpha\hat{T}^\beta \\
&-\Big((\mathfrak{L}_\xi g_{\nu\alpha})^{;\nu}-(\mathfrak{L}_\xi g_{\lambda\nu})_{;\alpha}g^{\nu\lambda}\Big)x^\beta\hat{N}^\alpha\hat{T}_\beta+\Big((\mathfrak{L}_\xi g_{\nu\alpha})_{;\beta}-(\mathfrak{L}_\xi g_{\beta\nu})_{;\alpha}\Big)x^\nu\hat{N}^\alpha\hat{T}^\beta\bigg)dA \\
=\ &E(\xi)+\frac{1}{16\pi}\int_{\partial\Sigma_\infty}\bigg(-2R_{\mu\beta}\xi^\mu\hat{T}^\beta\hat{N}_\alpha x^\alpha+\Big((\mathfrak{L}_\xi g_{\nu\beta})^{;\nu}-(\mathfrak{L}_\xi g_{\lambda\nu})_{;\beta}g^{\nu\lambda}\Big)x^\alpha\hat{N}_\alpha\hat{T}^\beta \\
&+\Big((\mathfrak{L}_\xi g_{\nu\alpha})_{;\beta}-(\mathfrak{L}_\xi g_{\beta\nu})_{;\alpha}\Big)x^\nu\hat{N}^\alpha\hat{T}^\beta\bigg)dA.
\end{aligned} \tag{3.34}$$

Here, in moving from the second to the third step of Eq. (3.34) we use Eq. (3.32). Next, the second term in the third line of Eq. (3.34) vanishes because the boundary of a boundary is empty, and we have also used Eq. (3.29) twice in the fifth line. Finally, we have applied $x^\beta\hat{T}_\beta=-t+O(r^{-1})$ in the last step of Eq. (3.34) and hence the terms containing $x^\beta\hat{T}_\beta$ vanish for any $n>0$ and finite t.

Inserting Eq. (3.34) back into Eq. (3.24) and after cancellation of the $R_{\mu\nu}$ term in

Eq. (3.24) we find

$$
\begin{aligned}
& p_\mu^{\mathrm{ADM}} \xi_\infty^\mu \\
= \;& E(\xi) + \frac{1}{16\pi} \int_{\partial\Sigma_\infty} (R\hat{N}_\tau x^\tau - 2R_{\mu\tau}\hat{N}^\mu x^\tau) dA \\
& + \frac{1}{16\pi} \int_{\partial\Sigma_\infty} \Big(\Big((\mathfrak{L}_\xi g_{\nu\beta})^{;\nu} - (\mathfrak{L}_\xi g_{\lambda\nu})_{;\beta} g^{\nu\lambda} \Big) x^\alpha \hat{N}_\alpha \hat{T}^\beta + \Big((\mathfrak{L}_\xi g_{\nu\alpha})_{;\beta} - (\mathfrak{L}_\xi g_{\beta\nu})_{;\alpha} \Big) x^\nu \hat{N}^\alpha \hat{T}^\beta \Big) dA \\
= \;& E(\xi) - \frac{1}{8\pi} \int_{\partial\Sigma_\infty} (R_{\mu\tau} - \frac{1}{2} R g_{\mu\tau}) \hat{N}^\mu x^\tau dA \\
& + \frac{1}{16\pi} \int_{\partial\Sigma_\infty} \Big(\Big((\mathfrak{L}_\xi g_{\nu\beta})^{;\nu} - (\mathfrak{L}_\xi g_{\lambda\nu})_{;\beta} g^{\nu\lambda} \Big) x^\alpha \hat{N}_\alpha \hat{T}^\beta + \Big((\mathfrak{L}_\xi g_{\nu\alpha})_{;\beta} - (\mathfrak{L}_\xi g_{\beta\nu})_{;\alpha} \Big) x^\nu \hat{N}^\alpha \hat{T}^\beta \Big) dA \\
= \;& E(\xi) - \int_{\partial\Sigma_\infty} T_{ij} \hat{N}^i x^j dA \\
& + \frac{1}{16\pi} \int_{\partial\Sigma_\infty} \Big(\Big((\mathfrak{L}_\xi g_{\nu\beta})^{;\nu} - (\mathfrak{L}_\xi g_{\lambda\nu})_{;\beta} g^{\nu\lambda} \Big) x^\alpha \hat{N}_\alpha \hat{T}^\beta + \Big((\mathfrak{L}_\xi g_{\nu\alpha})_{;\beta} - (\mathfrak{L}_\xi g_{\beta\nu})_{;\alpha} \Big) x^\nu \hat{N}^\alpha \hat{T}^\beta \Big) dA \\
= \;& E(\xi) \\
& + \frac{1}{16\pi} \int_{\partial\Sigma_\infty} \Big(\Big((\mathfrak{L}_\xi g_{\nu\beta})^{;\nu} - (\mathfrak{L}_\xi g_{\lambda\nu})_{;\beta} g^{\nu\lambda} \Big) x^\alpha \hat{N}_\alpha \hat{T}^\beta + \Big((\mathfrak{L}_\xi g_{\nu\alpha})_{;\beta} - (\mathfrak{L}_\xi g_{\beta\nu})_{;\alpha} \Big) x^\nu \hat{N}^\alpha \hat{T}^\beta \Big) dA,
\end{aligned}
\tag{3.35}
$$

where in the last step we assume $T_{ij} = o(r^{-3})$ which is weaker than Ashtekar and Hansen's assumption $T_{\mu\nu} = o(r^{-3})$. (As an aside, we note that trivially $p_\mu^{\mathrm{ADM}} \xi_\infty^\mu = E(\xi)$ when ξ is a timelike Killing vector, due to the known equality between the ADM and Komar masses [12]. .)

Since $dA \sim O(r^2)$ and $x^i \sim O(r)$ at infinity, we have $x^i dA \sim O(r^3)$. Applying the asymptotically-flat conditions Eqs.(3.13-3.14) and Eq. (3.33) to the final integral in Eq. (3.35) then yields

$$
\begin{aligned}
& \frac{1}{16\pi} \int_{\partial\Sigma_\infty} \Big(\Big((\mathfrak{L}_\xi g_{\nu\beta})^{,\nu} - (\mathfrak{L}_\xi g_{\lambda\nu})_{,\beta} g^{\nu\lambda} \Big) x^\alpha \hat{N}_\alpha \hat{T}^\beta + \Big((\mathfrak{L}_\xi g_{\nu\alpha})_{,\beta} - (\mathfrak{L}_\xi g_{\beta\nu})_{,\alpha} \Big) x^\nu \hat{N}^\alpha \hat{T}^\beta \\
& + O\Big(\frac{1}{r^4}\Big) + O\Big(\frac{1}{r^{2+n}}\Big) \Big) dA \\
= \;& \frac{1}{16\pi} \int_{\partial\Sigma_\infty} \Big(\Big(g_{\nu\beta,0}{}^{,\nu} + \xi^{\tau,\nu}{}_{,\nu} g_{\tau\beta} + \xi^\tau{}_{,\beta\tau} - g_{\lambda\nu,0\beta} g^{\nu\lambda} - \xi^\tau{}_{,\tau\beta} - \xi^\tau{}_{,\tau\beta} + O\Big(\frac{1}{r^{3+n}}\Big) \Big) \hat{T}^\beta x^\alpha \hat{N}_\alpha \\
& + \Big(g_{\nu\alpha,0\beta} + \xi^\tau{}_{,\nu\beta} g_{\alpha\tau} + \xi^\tau{}_{,\alpha\beta} g_{\nu\tau} - g_{\beta\nu,0\alpha} - \xi^\tau{}_{,\beta\alpha} g_{\nu\tau} - \xi^\tau{}_{,\nu\alpha} g_{\beta\tau} + O\Big(\frac{1}{r^{3+n}}\Big) \Big) x^\nu \hat{N}^\alpha \hat{T}^\beta \Big) dA \\
= \;& \frac{1}{16\pi} \int_{\partial\Sigma_\infty} \Big(\Big(g_{\nu\beta,0}{}^{,\nu} - g_{\lambda\nu,0\beta} g^{\nu\lambda} \Big) \hat{T}^\beta x^\alpha \hat{N}_\alpha + \Big(g_{\nu\alpha,0\beta} - g_{\beta\nu,0\alpha} \Big) x^\nu \hat{N}^\alpha \hat{T}^\beta + O\Big(\frac{1}{r^{1+n}}\Big) \Big) dA \\
= \;& \frac{1}{16\pi} \int_{\partial\Sigma_\infty} \Big((g_{\nu0,0\lambda} - g_{\lambda\nu,00}) g^{\nu\lambda} x^\alpha \hat{N}_\alpha + (g_{ij,00} - g_{0i,0j}) x^i \hat{N}^j + O\Big(\frac{1}{r^{1+n}}\Big) \Big) dA \\
= \;& \frac{1}{16\pi} \int_{\partial\Sigma_\infty} \Big((g_{i0,0j} - g_{ij,00}) \eta^{ij} x^\alpha \hat{N}_\alpha + (g_{ij,00} - g_{0i,0j}) x^i \hat{N}^j + O\Big(\frac{1}{r^{1+n}}\Big) \Big) dA \\
= \;& \frac{1}{16\pi} \int_{\partial\Sigma_\infty} \Big((g_{i0,0j} - g_{ij,00}) (\eta^{ij} x^\alpha \hat{N}_\alpha - x^i \hat{N}^j) + O\Big(\frac{1}{r^{1+n}}\Big) \Big) dA,
\end{aligned}
\tag{3.36}
$$

Note, since $\hat{T}_\mu \hat{N}^\mu = 0$ and from Eq. (1.8), we have $\hat{N}^\mu = (0, \hat{N}^i)$ at spatial infinity.

To ensure that Eq. (3.36) vanishes and hence the generalized Komar energy equals the ADM mass, we require $n > 1$ and the stronger asymptotically-flat conditions $g_{ij,00} = o(r^{-3})$ and $g_{0j,i0} = o(r^{-3})$ (condition ii of the theorem).

This completes the proof of Theorem 3.1.

3.2.4 Conservation of generalized Komar energy

In fulfilment of a promise made after Eq. (3.9), I now prove that the net flux of $J^\mu \hat{N}_\mu$ vanishes at spatial infinity under the traditional asymptotically-flat conditions. Therefore such conditions ensure that $J^\mu \hat{T}_\mu$ is preserved on every hypersurface and that the generalized Komar energy is preserved on every spacelike hypersurface labeled by a finite t (see Fig. 3.2). Indeed, using Eq. (3.29) and the conventional asymptotically-flat conditions Eqs. (3.13) and (3.14) the net flux $J^\mu \hat{N}_\mu$ out through spatial infinity may be written as

$$\begin{aligned} & \int dt \int_{\partial\Sigma_\infty} J^\mu \hat{N}_\mu \sqrt{\gamma^{(\partial\Sigma_\infty)}} d^2 z = \int dt \int_{\partial\Sigma_\infty} \xi^{[\nu;\mu]}{}_{;\nu} \hat{N}_\mu dA \\ = & \int dt \int_{\partial\Sigma_\infty} \Big(R_{\alpha\mu}\xi^\alpha + \frac{1}{2}(\mathfrak{L}_\xi g_{\nu\lambda})_{;\mu} g^{\lambda\nu} - \frac{1}{2}(\mathfrak{L}_\xi g_{\mu\lambda})^{;\lambda} \Big) \hat{N}^\mu dA \\ = & \int dt \frac{1}{2} \int_{\partial\Sigma_\infty} O\Big(\frac{1}{r^3}\Big) dA = 0, \end{aligned} \tag{3.37}$$

between any pair of hypersurfaces Σ_2 and Σ_1 separated by a finite Δt.

3.2.5 Discussion about Theorem 3.1

We emphasize that the vector field ξ has enormous freedom: it is arbitrary on the hypersurface except asymptotically at spatial infinity. In particular, there is no requirement that ξ be a Killing vector. In fact, other than for illustrative examples, *none* of the results in this chapter rely on any Killing vectors.

The little-o notation $f = o(r^{-n})$ implies that $\lim_{r\to\infty} r^n f = 0$, so condition ii places a bound on temporal derivatives in the metric at spatial infinity. Indeed, this condition only modestly strengthens the conventional characterization of asymptotic flatness where instead it is generically assumed that $g_{\mu\nu,\alpha\beta} = O(r^{-3})$, though this latter behavior of the temporal derivatives of the metric is partly based on an analogy with that of the spatial derivatives for stationary spacetimes [71].

When $\xi = \partial_t$ happens to be a timelike Killing vector $E(\xi)$ recovers the conventional Komar mass and for such stationary spacetimes the Komar and ADM mass are known to be equal [12]. Indeed, many textbooks only consider the Komar mass to be a satisfactory notion of mass for stationary asymptotically-flat spacetimes (e.g., Ref. [60]). Thus, Theorem 3.1 significantly generalizes the simple covariant form for physical energy from stationary to dynamical (Killing vector free) spacetimes. Hence, under the conditions of Theorem 3.1, we may freely exchange the notions of ADM mass and generalized Komar energy.

3.3 Differential "first law" for dynamical spacetimes

3.3.1 Integral expression for net generalized energy

Let us now see how far we can replicate Bardeen at el.'s analysis of the first law of black hole mechanics in terms of our covariant expression for physical energy in an asymptotically-flat *dynamical* spacetime instead of merely for a stationary spacetime as originally studied. Some of the arguments will bare initial similarities to those of Chapter 2, where a stationary spacetime is required.

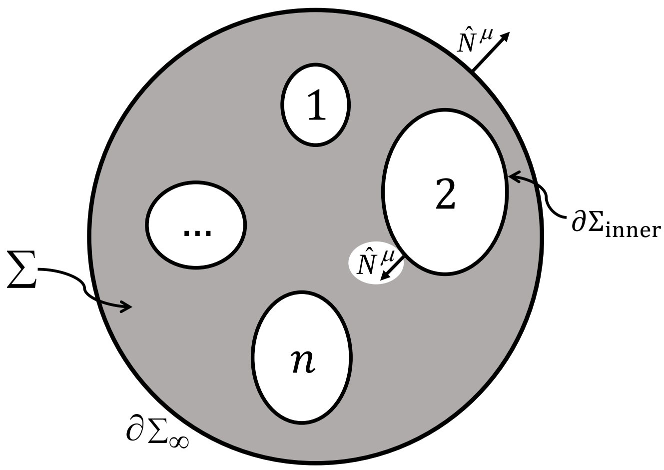

Figure 3.4: Schematic of the spacelike three-dimensional hypersurface of interest, Σ. There is an outer boundary $\partial\Sigma_\infty$ at spatial infinity and an inner boundary $\partial\Sigma_{\text{inner}}$, consisting of the sum of n individual boundaries of regions which are excised from Σ. Here $\hat{N}^\mu$ is the spacelike 4-vector in Σ normal to the boundaries $\partial\Sigma$ (note the direction convention on the inner boundary).

Consider a manifold with multiple regions whose interiors are excised from the hypersurface Σ (see Fig. 3.4). Recall that permuting the order of a pair of covariant derivatives acting on an arbitrary 4-vector ξ^μ may be expressed in terms of the Riemann curvature tensor as [47]

$$\xi^\mu{}_{;\alpha\beta} - \xi^\mu{}_{;\beta\alpha} = -R^\mu{}_{\nu\alpha\beta}\xi^\nu \ . \tag{3.38}$$

Contracting the indices μ and α reduces this to an expression in terms of the Ricci tensor

$$\xi^\mu{}_{;\mu\beta} - \xi^\mu{}_{;\beta\mu} = -R_{\nu\beta}\xi^\nu \ . \tag{3.39}$$

Consequently, for an arbitrary ξ^μ, we may write

$$J_\beta(\xi) = \xi_{[\mu;\beta]}{}^{;\mu} = R_{\mu\beta}\xi^\mu + \xi^\mu{}_{;\mu\beta} - \xi_{(\mu;\beta)}{}^{;\mu}. \tag{3.40}$$

Integrating Eq. (3.40) over a three-dimensional spacelike hypersurface Σ whose outer boundary is at spatial infinity $\partial\Sigma_\infty$ and inner boundary at $\partial\Sigma_{\text{inner}}$, yields

$$\begin{aligned}
& \int_\Sigma (R_{\mu\beta}\xi^\mu + \xi^\mu{}_{;\mu\beta} - \xi_{(\mu;\beta)}{}^{;\mu})\hat{T}^\beta \sqrt{\gamma^{(\Sigma)}}\, d^3x \\
= & \int_\Sigma \xi_{[\mu;\beta]}{}^{;\mu}\hat{T}^\beta \sqrt{\gamma^{(\Sigma)}}\, d^3x \\
= & \int_{\partial\Sigma_\infty} \xi_{[\mu;\beta]}\hat{N}^\mu\hat{T}^\beta \sqrt{\gamma^{(\partial\Sigma_\infty)}}\, d^2z - \int_{\partial\Sigma_{\text{inner}}} \xi_{[\mu;\beta]}\hat{N}^\mu\hat{T}^\beta \sqrt{\gamma^{(\partial\Sigma_{\text{inner}})}}\, d^2z \\
= & \ 4\pi E(\xi) - \int_{\partial\Sigma_{\text{inner}}} \xi_{[\mu;\beta]}\hat{N}^\mu\hat{T}^\beta \sqrt{\gamma^{(\partial\Sigma_{\text{inner}})}}\, d^2z \ ,
\end{aligned} \tag{3.41}$$

where we have used Theorem 3.1 in the last step. Therefore, we have

$$E(\xi) \ = \ \frac{1}{4\pi}\int_\Sigma (R_{\mu\beta}\xi^\mu + \xi^\mu{}_{;\mu\beta} - \xi_{(\mu;\beta)}{}^{;\mu})\hat{T}^\beta \sqrt{\gamma^{(\Sigma)}}\, d^3x + \frac{1}{4\pi}\int_{\partial\Sigma_{\text{inner}}} \xi_{[\mu;\beta]}\hat{N}^\mu\hat{T}^\beta dA \ , \tag{3.42}$$

where $dA = \sqrt{\gamma^{(\partial\Sigma_{\text{inner}})}}\, d^2z$. Next, we define the integrand of the boundary terms in Eq. (3.42) to be a generalized surface gravity

$$\boxed{\kappa(\xi) \equiv \xi_{[\mu;\beta]}\hat{N}^\mu\hat{T}^\beta.} \tag{3.43}$$

This definition reduces to the traditional surface gravity definition Eq. (2.7) when ξ^μ is a suitable Killing vector. Then Eq. (3.42) becomes

$$\boxed{E(\xi) = \frac{1}{4\pi}\int_{\partial\Sigma_{\text{inner}}} \kappa(\xi) dA + \frac{1}{4\pi}\int_\Sigma (R_{\mu\beta}\xi^\mu + \xi^\mu{}_{;\mu\beta} - \xi_{(\mu;\beta)}{}^{;\mu})\hat{T}^\beta \sqrt{\gamma^{(\Sigma)}}\, d^3x.} \tag{3.44}$$

Recall that Theorem 3.1 relates the ADM mass to this generalized Komar energy $E(\xi)$, provided that ξ is suitably constrained at spatial infinity; otherwise this vector field ξ

is arbitrary. Bardeen, Carter and Hawking find a similar expression except there ξ is required to be a Killing vector [11].

We emphasize that, at this stage, Eq. (3.44) is very general. The inner boundaries are arbitrary closed surfaces in dynamical spacetimes and ξ remains an arbitrary vector field (subject to the conditions of Theorem 3.1). As one changes ξ (subject to its asymptotic constraint), the generalized surface gravity $\kappa(\xi)$ will also change, only the sum of the terms on the right-hand-side of Eq. (3.44) have any physical meaning at this point. Next, I investigate changes in the physical energy $\delta E(\xi)$ in order to seek an expression which will be analogous to the first law of thermodynamics. I will prove that, for a suitable choice of ξ, an inner boundary with special properties will allow changes in the energy to mimic the first law of thermodynamics. The choice of ξ will then be further authenticated by showing its generalized surface gravity $\kappa(\xi)$ agrees with a calculation of the tunneling temperature [45, 66].

3.3.2 Energy changes under diffeomorphic perturbations

The above generalization, in a generic spacetime, for net energy within a hypersurface might appear to suggest that a temperature and entropy may take the form

$$T = \frac{\kappa(\xi)}{(2\pi)}, \qquad S = \frac{A}{4}. \tag{3.45}$$

However, such surface quantities need to behave thermodynamically. As in our discussion in Chapter 2, the net energy $E(\xi)$ should admit changes which behave analogously to

$$\delta E(\xi) = T\delta S, \tag{3.46}$$

(ignoring work or volume terms) so that the temperature would be acting as an integrating factor relating changes in the (state function) energy to changes in the entropy. In other words, we must show that such changes lead to the expected form of the first-law of thermodynamics. Here we follow in the footsteps of Bardeen et al.'s original analysis and consider changes corresponding to parametric differences between diffeomorphicly nearby solutions. In particular, we will consider two nearby configurations corresponding to the metrics

$$g_{\mu\nu}, \qquad g'_{\mu\nu} = g_{\mu\nu} + h_{\mu\nu}, \tag{3.47}$$

where $h_{\mu\nu} \equiv \delta g_{\mu\nu} = -g_{\mu\sigma}g_{\nu\tau}\delta g^{\sigma\tau}$, i.e., $\delta g^{\sigma\tau} = -h^{\sigma\tau}$.

As in Chapter 2 and without loss of generality, we may assume that for the two diffeomorphicly related configurations, the hypersurfaces Σ and Σ' are described by identical sets of coordinates; this is always possible due to “gauge” freedom in the choice of coordinate systems [11]. Henceforth we label both by Σ; similarly, for their boundaries $\partial\Sigma$. Further, from Eq. (1.8) we also have

$$\delta\Big(\frac{\hat{T}_\mu}{\mathcal{N}}\Big) = 0\ . \tag{3.48}$$

Using the Einstein field equations we start by rewriting Eq. (3.44) as

$$\begin{aligned}
E(\xi) \quad &= \\
&\int_\Sigma 2\,T_{\mu\beta}\xi^\mu\hat{T}^\beta\sqrt{\gamma^{(\Sigma)}}d^3x && (3.49a)\\
&+\int_\Sigma (\frac{1}{8\pi}R\,g_{\mu\beta})\xi^\mu\hat{T}^\beta\sqrt{\gamma^{(\Sigma)}}d^3x && (3.49b)\\
&+\frac{1}{4\pi}\int_\Sigma (\xi^\mu{}_{;\mu\beta}-\xi_{(\mu;\beta)}{}^{;\mu})\hat{T}^\beta\sqrt{\gamma^{(\Sigma)}}d^3x && (3.49c)\\
&+\frac{1}{4\pi}\int_{\partial\Sigma_{\mathrm{inner}}}\kappa(\xi)\,dA\ . && (3.49d)
\end{aligned}$$

We will successively investigate diffeomorphisms of each of the subterms Eq. (3.49a) through Eq. (3.49d).

Since $\mathcal{N}\sqrt{\gamma^{(\Sigma)}} = \sqrt{-g}$ on the hypersurface and $\delta(\frac{\hat{T}_\mu}{\mathcal{N}}) = 0$, the variation of the energy-momentum term Eq. (3.49a) may be computed as

$$\begin{aligned}
&\int_\Sigma \delta\Big(2\,T_{\mu\beta}\xi^\mu\hat{T}^\beta\sqrt{\gamma^{(\Sigma)}}\Big)\,d^3x = \int_\Sigma \delta\Big(2\,\xi^\mu\,\frac{\hat{T}_\beta}{\mathcal{N}}T_\mu{}^\beta\mathcal{N}\sqrt{\gamma^{(\Sigma)}}\Big)\,d^3x\\
= \quad &\int_\Sigma\Big(\Big(2\sqrt{-g}\,\delta T_\mu{}^\beta + 2\,T_\mu{}^\beta\delta\sqrt{-g}\Big)\xi^\mu\frac{\hat{T}_\beta}{\mathcal{N}} + 2\,T_\mu{}^\beta\hat{T}_\beta\,\delta\xi^\mu\sqrt{\gamma^{(\Sigma)}}\Big)\,d^3x\\
= \quad &\int_\Sigma\Big(\Big(2\sqrt{-g}\,\delta T_\mu{}^\beta + 2\,T_\mu{}^\beta(\frac{1}{2}\sqrt{-g}\,g^{\lambda\nu}\delta g_{\lambda\nu})\Big)\xi^\mu\frac{\hat{T}_\beta}{\mathcal{N}} + 2\,T_\mu{}^\beta\hat{T}_\beta\,\delta\xi^\mu\sqrt{\gamma^{(\Sigma)}}\Big)\,d^3x\\
= \quad &\int_\Sigma\Big(\Big(2\,\delta T_\mu{}^\beta + \,T_\mu{}^\beta h_\nu{}^\nu\Big)\xi^\mu + 2\,T_\mu{}^\beta\,\delta\xi^\mu\Big)\hat{T}_\beta\sqrt{\gamma^{(\Sigma)}}\,d^3x
\end{aligned} \tag{3.50}$$

where we have used $\delta\sqrt{-g} = \frac{1}{2}\sqrt{-g}\,g^{\mu\nu}\delta g_{\mu\nu}$ in the second line.

Next, we consider the variation of the Ricci scalar term Eq. (3.49b)

$$\begin{aligned}
&\frac{1}{8\pi}\int_\Sigma \delta(R\sqrt{\gamma^{(\Sigma)}}\,\xi^\beta\,\hat{T}_\beta)\,d^3x = \frac{1}{8\pi}\int_\Sigma \delta\Big(R\mathcal{N}\sqrt{\gamma^{(\Sigma)}}\,\xi^\beta\,\frac{\hat{T}_\beta}{\mathcal{N}}\Big)\,d^3x\\
= \quad &\frac{1}{8\pi}\int_\Sigma\Big(\delta(R\sqrt{-g})\,\xi^\beta\,\frac{\hat{T}_\beta}{\mathcal{N}} + R\,\hat{T}_\beta\delta\xi^\beta\sqrt{\gamma^{(\Sigma)}}\Big)\,d^3x\\
= \quad &-\frac{1}{8\pi}\int_\Sigma\Big((R_{\mu\nu}-\frac{1}{2}g_{\mu\nu}R)h^{\mu\nu} - (g^{\mu\nu}\delta\Gamma^\alpha{}_{\mu\nu} - g^{\mu\alpha}\delta\Gamma^\lambda{}_{\lambda\mu})_{;\alpha}\Big)\xi^\beta\hat{T}_\beta\sqrt{\gamma^{(\Sigma)}}\,d^3x\\
&+\frac{1}{8\pi}\int_\Sigma R\,\hat{T}_\beta\delta\xi^\beta\sqrt{\gamma^{(\Sigma)}}\,d^3x
\end{aligned} \tag{3.51}$$

where in the last step we use Eq. (2.39).

Now recalling Lemma 2.1 in Chapter 2 $-(g^{\mu\nu}\delta\Gamma^{\alpha}{}_{\mu\nu} - g^{\mu\alpha}\delta\Gamma^{\lambda}{}_{\lambda\mu})_{;\alpha} = 2\,h^{\mu}{}_{[\mu;\nu]}{}^{;\nu}$. Since no Killing vector is involved in this result, it is generally true and can be directly applied to Eq. (3.51). Using Lemma 2.1, the variation of the Ricci scalar term Eq. (3.49b) becomes

$$-\frac{1}{8\pi}\int_{\Sigma}\Big((R_{\mu\nu}-\frac{1}{2}g_{\mu\nu}R)h^{\mu\nu}+2h^{\mu}{}_{[\mu;\nu]}{}^{;\nu}\Big)\,\xi^{\beta}\,\hat{T}_{\beta}\,\sqrt{\gamma^{(\Sigma)}}\,d^3x+\frac{1}{8\pi}\int_{\Sigma}R\,\hat{T}_{\beta}\delta\xi^{\beta}\sqrt{\gamma^{(\Sigma)}}\,d^3x\,. \tag{3.52}$$

Lemma 2.2 in Chapter 2 is also a general result, applying it to Eq. (3.52), the variation of Eq. (3.49b) reduces to

$$\begin{aligned}&-\int_{\Sigma}T_{\mu\nu}h^{\mu\nu}\,\xi^{\beta}\,\hat{T}_{\beta}\,\sqrt{\gamma^{(\Sigma)}}\,d^3x-\frac{1}{4\pi}\int_{\Sigma}(\xi^{\beta}h_{\mu}{}^{[\mu;\nu]}-\xi^{\nu}h_{\mu}{}^{[\mu;\beta]})_{;\nu}\,\hat{T}_{\beta}\,\sqrt{\gamma^{(\Sigma)}}\,d^3x\\&-\frac{1}{4\pi}\int_{\Sigma}\Big(\mathfrak{L}_{\xi}(h_{\mu}{}^{[\mu;\beta]})+h_{\mu}{}^{[\mu;\beta]}\xi^{\nu}{}_{;\nu}\Big)\,\hat{T}_{\beta}\,\sqrt{\gamma^{(\Sigma)}}\,d^3x+\frac{1}{8\pi}\int_{\Sigma}R\,\hat{T}_{\beta}\delta\xi^{\beta}\sqrt{\gamma^{(\Sigma)}}\,d^3x\,.\end{aligned} \tag{3.53}$$

Then applying stokes' theorem Eq. (1.24) to Eq. (3.53) yields

$$\begin{aligned}\int_{\Sigma}\Big(R\,\delta\xi^{\beta}-T_{\mu\nu}h^{\mu\nu}\xi^{\beta}\Big)\hat{T}_{\beta}\sqrt{\gamma^{(\Sigma)}}\,d^3x\quad-\quad&\frac{1}{4\pi}\int_{\partial\Sigma}\Big(\xi^{\beta}h_{\mu}{}^{[\mu;\nu]}-\xi^{\nu}h_{\mu}{}^{[\mu;\beta]}\Big)\hat{N}_{\nu}\hat{T}_{\beta}\sqrt{\gamma^{(\partial\Sigma)}}\,d^2z\\&-\frac{1}{4\pi}\int_{\Sigma}\Big(\mathfrak{L}_{\xi}(h_{\mu}{}^{[\mu;\beta]})+h_{\mu}{}^{[\mu;\beta]}\xi^{\nu}{}_{;\nu}\Big)\hat{T}_{\beta}\sqrt{\gamma^{(\Sigma)}}\,d^3x\,.\end{aligned} \tag{3.54}$$

Splitting the boundary into the inner boundary and the outer boundary and replacing $\sqrt{\gamma^{(\partial\Sigma)}}\,d^2z$ by dA, Eq. (3.54) becomes

$$\begin{aligned}&\frac{1}{4\pi}\int_{\partial\Sigma_{\rm inner}}\Big(\xi^{\beta}h_{\mu}{}^{[\mu;\nu]}-\xi^{\nu}h_{\mu}{}^{[\mu;\beta]}\Big)\hat{N}_{\nu}\hat{T}_{\beta}dA-\frac{1}{4\pi}\int_{\partial\Sigma_{\infty}}\Big(\xi^{\beta}h_{\mu}{}^{[\mu;\nu]}-\xi^{\nu}h_{\mu}{}^{[\mu;\beta]}\Big)\hat{N}_{\nu}\hat{T}_{\beta}dA\\&+\int_{\Sigma}\Big(R\,\delta\xi^{\beta}-T_{\mu\nu}h^{\mu\nu}\xi^{\beta}\Big)\hat{T}_{\beta}\sqrt{\gamma^{(\Sigma)}}\,d^3x-\frac{1}{4\pi}\int_{\Sigma}\Big(\mathfrak{L}_{\xi}(h_{\mu}{}^{[\mu;\beta]})+h_{\mu}{}^{[\mu;\beta]}\xi^{\nu}{}_{;\nu}\Big)\,\hat{T}_{\beta}\,\sqrt{\gamma^{(\Sigma)}}\,d^3x\,.\end{aligned} \tag{3.55}$$

Note, the direction convention of $\hat{N}^{\mu}$ on the inner boundaries and outer boundary (see Fig 3.4).

In order to further simplify these terms we next consider the diffeomorphic changes in more detail.

3.3.3 Diffeomorphic conditions

Bardeen et al.'s surface gravity measures the acceleration of a freely falling particle with respect to the time at spatial infinity (or the coordinate time) [11]. We choose ξ^{μ} as $\mathcal{N}\hat{T}^{\mu}$

(including their first derivatives) at the inner boundary, so our surface gravity has the same physical meaning and agrees with their surface gravity definition for non-dynamical spacetimes [11]. Since we are only interested in the non-rotating case, to ensure that the mass definition $E(\xi)$ has the same meaning before and after the diffeomorphic variation we impose the constraint $\delta\xi^\mu = 0$.

Recall that by “gauge” freedom the sets of coordinates of Σ and $\partial\Sigma$ are chosen to be unchanged by the diffeomorphism. As a consequence the relevant tangent spaces at any coordinate point will be unchanged under the diffeomorphism [11]. Thus, as $\hat{T}_\mu dx^\mu = 0$ for all tangent vectors dx^μ at any point in Σ, we must have $\delta\hat{T}_\mu \parallel \hat{T}_\mu$, or equivalently $\delta\hat{T}_{[\mu}\hat{T}_{\nu]} = 0$, so

$$\delta\hat{T}_\mu = \delta(\ln\mathcal{N})\,\hat{T}_\mu \equiv k_1\hat{T}_\mu. \tag{3.56}$$

Since $\xi^\mu = \mathcal{N}\hat{T}^\mu$ and $\delta\xi^\mu = 0$ on the inner boundary, $\delta\hat{T}^\mu$ may be simplified there as

$$\delta\hat{T}^\mu = -\frac{\delta\mathcal{N}}{\mathcal{N}^2}\xi^\mu = -\delta(\ln\mathcal{N})\hat{T}^\mu = -k_1\hat{T}^\mu, \tag{3.57}$$

which also ensures $\delta(\hat{T}^\mu\hat{T}_\mu) = 0$.

Similarly, since $\hat{N}_\mu dx^\mu = 0$ for all tangent vectors dx^μ in $\partial\Sigma$ at any point in $\partial\Sigma$, combined with $\hat{T}^\mu\hat{N}_\mu = 0$, we find $\delta\hat{N}_\mu \parallel \hat{N}_\mu$, and so must have

$$\delta\hat{N}_\mu = k_2\hat{N}_\mu. \tag{3.58}$$

We now extend our use of the “gauge” freedom to extend this equality of coordinates for our nearby solutions slightly away from the inner boundary $\partial\Sigma_{\text{inner}}$, though still within the hypersurface Σ. Indeed, Bardeen et al. [11] used such freedom on the (future) null horizon for stationary black holes. Here, we make an analogous construction for the outgoing spacelike tangent vectors along $\hat{N}^\mu$ from each point on the boundary $\partial\Sigma_{\text{inner}}$. In particular, for an infinitesimal ‘distance’ along the tangent vectors $dx^\mu \propto \hat{N}^\mu$ from the inner boundary, we use gauge freedom to ensure that these vectors satisfy $\delta(dx^\mu) = 0$. In other words, gauge freedom allows us to choose the covariant vectors $\hat{N}^\mu$ normal to the inner boundary $\partial\Sigma_{\text{inner}}$ to remain parallel to themselves under the diffeomorphism. Consequently

$$\delta\hat{N}^\mu = -k_2\hat{N}^\mu, \tag{3.59}$$

a condition similar to that given by Bardeen et al. [11]. Combining Eqs.(3.58) and (3.59) we find

$$\delta\hat{N}_\mu = -g_{\mu\nu}\,\delta\hat{N}^\nu, \tag{3.60}$$

at the inner boundary $\partial\Sigma_{\text{inner}}$.

Next, we introduce a full tetrad basis which we will use at the inner boundary $\partial\Sigma_{\text{inner}}$. Along with the unit normals $\hat{T}^\mu$ and $\hat{N}^\mu$ to $\partial\Sigma_{\text{inner}}$ we add two orthogonal unit spacelike tangent vectors fully within $\partial\Sigma_{\text{inner}}$ labeled as $\hat{U}^\mu$ and $\hat{V}^\mu$. The 'projector' onto $\partial\Sigma_{\text{inner}}$ is thus given by

$$P^{\mu\nu} \equiv (\hat{U}\otimes\hat{U} + \hat{V}\otimes\hat{V})^{\mu\nu} = \hat{U}^\mu\hat{U}^\nu + \hat{V}^\mu\hat{V}^\nu. \tag{3.61}$$

Hence

$$g^{\mu\nu} = -\hat{T}^\mu\hat{T}^\nu + \hat{N}^\mu\hat{N}^\nu + P^{\mu\nu}. \tag{3.62}$$

Further, since the 2-dimensional spacelike tangent space at each point in $\partial\Sigma_{\text{inner}}$ is preserved under the diffeomorphism (as already noted), we may immediately write

$$\delta\hat{U}^\mu = k_3\hat{U}^\mu + k_4\hat{V}^\mu, \qquad \delta\hat{V}^\mu = k_5\hat{U}^\mu + k_6\hat{V}^\mu. \tag{3.63}$$

Up until now, we have only used the "gauge" freedom to make the inner boundaries $\partial\Sigma_{\text{inner}}$ and hypersurface Σ have identical coordinates for these nearby solutions. In this way, the diffeomorphism effectively "commutes" through the integral sign in our calculation of changes in the total energy.

Next, the direction of the *covariant* vector $\hat{U}_\mu$ is uniquely determined by

$$\hat{U}_\mu\hat{T}^\mu = \hat{U}_\mu\hat{N}^\mu = \hat{U}_\mu\hat{V}^\mu = 0. \tag{3.64}$$

However, as $\hat{T}^\mu$ and $\hat{N}^\mu$ are parallel to their diffeomorphism we may also write

$$\hat{U}_\mu\delta\hat{T}^\mu = \hat{U}_\mu\delta\hat{N}^\mu = 0 \tag{3.65}$$

From which it then follows that

$$\delta\hat{U}_\mu\,\hat{T}^\mu = \delta\hat{U}_\mu\,\hat{N}^\mu = 0. \tag{3.66}$$

Since a similar constraint applies to $\delta\hat{V}_\mu$ we see that both $\delta\hat{U}_\mu$ and $\delta\hat{V}_\mu$ remain in the span$\{\hat{U}_\mu, \hat{V}_\mu\}$. Then by considering $\delta(\hat{U}_\mu\hat{U}^\mu) = 0$, $\delta(\hat{U}_\mu\hat{V}^\mu) = 0$, $\delta(\hat{V}_\mu\hat{V}^\mu) = 0$ and $\delta(\hat{V}_\mu\hat{U}^\mu) = 0$, one may write

$$\delta\hat{U}_\mu = -k_3\hat{U}_\mu - k_5\hat{V}_\mu, \qquad \delta\hat{V}_\mu = -k_4\hat{U}_\mu - k_6\hat{V}_\mu. \tag{3.67}$$

Hence, $\delta P^{\mu\nu}$ may be explicitly computed to be

$$\begin{aligned}
\delta P^{\mu\nu} &= 2k_3\,\hat{U}^\mu\hat{U}^\nu + 2k_6\,\hat{V}^\mu\hat{V}^\nu(k_4+k_5)(\hat{U}^\mu\hat{V}^\nu + \hat{U}^\nu\hat{V}^\mu) \\
&= (k_3+k_6)(\hat{U}^\mu\hat{U}^\nu + \hat{V}^\mu\hat{V}^\nu) + (k_3-k_6)(\hat{U}^\mu\hat{U}^\nu - \hat{V}^\mu\hat{V}^\nu) \\
&\quad + (k_4+k_5)(\hat{U}^\mu\hat{V}^\nu + \hat{U}^\nu\hat{V}^\mu) \\
&= (k_3+k_6)P^{\mu\nu} + (k_3-k_6)(\hat{U}^\mu\hat{U}^\nu - \hat{V}^\mu\hat{V}^\nu) + (k_4+k_5)(\hat{U}^\mu\hat{V}^\nu + \hat{U}^\nu\hat{V}^\mu),
\end{aligned} \tag{3.68}$$

Similarly,

$$\begin{aligned}\delta P_{\mu\nu} &= -2k_3\,\hat{U}_\mu\hat{U}_\nu - 2k_6\,\hat{V}_\mu\hat{V}_\nu - (k_4+k_5)(\hat{U}_\mu\hat{V}_\nu + \hat{U}_\nu\hat{V}_\mu)\\ &= -(k_3+k_6)P_{\mu\nu} - (k_3-k_6)(\hat{U}_\mu\hat{U}_\nu - \hat{V}_\mu\hat{V}_\nu) - (k_4+k_5)(\hat{U}_\mu\hat{V}_\nu + \hat{U}_\nu\hat{V}_\mu).\end{aligned} \tag{3.69}$$

So the key diffeomorphic conditions become

$$\begin{aligned}&\delta\hat{T}^\mu = -k_1\hat{T}^\mu, \quad \delta\hat{T}_\mu = k_1\hat{T}_\mu\\ &\delta\hat{N}^\mu = -k_2\hat{N}^\mu, \;\; \delta\hat{N}_\mu = k_2\hat{N}_\mu\\ &\delta P^{\mu\nu} = \text{Eq.}(3.68), \quad \delta P_{\mu\nu} = \text{Eq.}(3.69)\end{aligned} \tag{3.70}$$

with $k_1 = \delta(\ln\mathcal{N})$.

3.3.3.1 Reduction to the first law

We firstly simplify the inner boundary term of Eq. (3.55).

From Eqs. (3.27) and (3.43), the surface gravity may be written as

$$\kappa \equiv \xi_{[\mu;\nu]}\hat{T}^\nu\hat{N}^\mu \quad = \quad \xi_{\mu;\nu}\hat{T}^\nu\hat{N}^\mu - \frac{1}{2}\mathfrak{L}_\xi g_{\mu\nu}\hat{T}^\nu\hat{N}^\mu. \tag{3.71}$$

Given the outgoing null normal of the inner boundary (see Fig. 3.4) $l_\mu = \hat{T}_\mu + \hat{N}_\mu$ and extending $\hat{N}^\mu$ as a vector field inside Σ away from $\partial\Sigma$, consistent with $\hat{T}_\mu\hat{N}^\mu = 0$ and $\hat{N}^\mu\hat{N}_\mu = 1$. The expansion of outgoing null normal congruences on the inner boundary may be written as

$$\begin{aligned}\theta^{(l)} &= P^{\mu\nu}l_{\mu;\nu} = P^{\mu\nu}(\hat{T}_\mu + \hat{N}_\mu)_{;\nu}\\ &= P^{\mu\nu}\hat{T}_{\mu;\nu} + (g^{\mu\nu} + \hat{T}^\mu\hat{T}^\nu - \hat{N}^\mu\hat{N}^\nu)\hat{N}_{\mu;\nu}\\ &= P^{\mu\nu}\hat{T}_{\mu;\nu} + \hat{N}^\mu{}_{;\mu} + \hat{N}_{\mu;\nu}\hat{T}^\mu\hat{T}^\nu\\ &= \frac{1}{\sqrt{-g}}(\sqrt{-g}\hat{N}^\mu)_{,\mu} + P^{\mu\nu}\hat{T}_{\mu;\nu} + \hat{N}_{\mu;\nu}\hat{T}^\mu\hat{T}^\nu\end{aligned} \tag{3.72}$$

where we have used Eq. (3.62) in moving from the first to the second line.

From Eq. (3.72), the variation of $\theta^{(l)}$ may be simplified as

$$\begin{aligned}
\delta\theta^{(l)} &= \delta(\frac{1}{\sqrt{-g}})(\sqrt{-g}\hat{N}^{\mu})_{,\mu} + \frac{1}{\sqrt{-g}}(\delta\sqrt{-g}\hat{N}^{\mu})_{,\mu} + \frac{1}{\sqrt{-g}}(\sqrt{-g}\delta\hat{N}^{\mu})_{,\mu} \\
&\quad +\delta(P^{\mu\nu}\hat{T}_{\mu;\nu}) + \delta(\hat{N}_{\mu;\nu}\hat{T}^{\mu}\hat{T}^{\nu}) \\
&= -\frac{1}{2}g^{\tau\nu}\delta g_{\tau\nu}\frac{1}{\sqrt{-g}}(\sqrt{-g}\hat{N}^{\mu})_{,\mu} + \frac{1}{\sqrt{-g}}(\frac{1}{2}\sqrt{-g}\,g^{\tau\nu}\delta g_{\tau\nu}\hat{N}^{\mu})_{,\mu} + (\delta\hat{N}^{\mu})_{;\mu} \\
&\quad +\delta(P^{\mu\nu}\hat{T}_{\mu;\nu}) + \delta\Big((\hat{N}_{\mu}\hat{T}^{\mu})_{;\nu}\hat{T}^{\nu} - \hat{T}_{\mu;\nu}\hat{N}^{\mu}\hat{T}^{\nu}\Big) \\
&= -\frac{1}{2}h_{\nu}{}^{\nu}\frac{1}{\sqrt{-g}}(\sqrt{-g}\hat{N}^{\mu})_{,\mu} + \frac{1}{2}h_{\nu}{}^{\nu}\frac{1}{\sqrt{-g}}(\sqrt{-g}\hat{N}^{\mu})_{,\mu} + \frac{1}{2}(h_{\nu}{}^{\nu})_{,\mu}\hat{N}^{\mu} + (\delta\hat{N}^{\mu})_{;\mu} \\
&\quad +\delta(P^{\mu\nu}\hat{T}_{\mu;\nu}) - \delta(\xi_{\mu;\nu}\hat{N}^{\mu}\hat{T}^{\nu}\frac{1}{\mathcal{N}}) \\
&= \frac{1}{2}(h_{\nu}{}^{\nu})_{,\mu}\hat{N}^{\mu} + (\delta\hat{N}^{\mu})_{;\mu} + \delta(P^{\mu\nu}\hat{T}_{\mu;\nu}) - \delta\Big(\frac{\kappa}{\mathcal{N}} + \frac{1}{2\mathcal{N}}\mathfrak{L}_{\xi}g_{\mu\nu}\hat{T}^{\nu}\hat{N}^{\mu}\Big)\,, \qquad (3.73)
\end{aligned}$$

where we have used $\delta\sqrt{-g} = \frac{1}{2}\sqrt{-g}\,g^{\mu\nu}\delta g_{\mu\nu}$ in the first step, $g^{\tau\nu}\delta g_{\tau\nu} = h_{\nu}{}^{\nu}$ in the second step, and Eq. (3.71) in the third step. Or equivalently,

$$\begin{aligned}
-\frac{1}{2}h_{\nu}{}^{\nu;\mu}\hat{N}_{\mu} &= (\delta\hat{N}^{\mu})_{;\mu} + \delta(P^{\mu\nu}\hat{T}_{\mu;\nu}) - \delta\Big(\frac{\kappa}{\mathcal{N}} + \frac{1}{2\mathcal{N}}\mathfrak{L}_{\xi}g_{\mu\nu}\hat{T}^{\nu}\hat{N}^{\mu}\Big) - \delta\theta^{(l)} \\
&= (\delta\hat{N}^{\mu})_{;\mu} + \delta\Big(\frac{1}{2\mathcal{N}}P^{\mu\nu}\mathfrak{L}_{\xi}g_{\mu\nu}\Big) - \frac{\delta\kappa}{\mathcal{N}} + \frac{k_1\kappa}{\mathcal{N}} - \delta\Big(\frac{1}{2\mathcal{N}}\mathfrak{L}_{\xi}g_{\mu\nu}\hat{T}^{\nu}\hat{N}^{\mu}\Big) - \delta\theta^{(l)}, \qquad (3.74)
\end{aligned}$$

where we have used $\hat{T}_{\mu} = \xi_{\mu}/\mathcal{N}$ and $k_1 = \delta\ln\mathcal{N}$.

The other terms in the inner boundary integral in Eq. (3.55) involve $h_{\mu}{}^{\nu;\mu}\hat{N}_{\nu}$, which may be simplified as

$$\begin{aligned}
\frac{1}{2}h_{\mu}{}^{\nu;\mu}\hat{N}_{\nu} &= \frac{1}{2}(h_{\mu}{}^{\nu}\hat{N}_{\nu})^{;\mu} - \frac{1}{2}h_{\mu}{}^{\nu}\hat{N}_{\nu}{}^{;\mu} = \frac{1}{2}(h_{\mu\nu}\hat{N}^{\nu})^{;\mu} - \frac{1}{2}h^{\mu\nu}\hat{N}_{\nu;\mu} \\
&= \frac{1}{2}(\delta g_{\mu\nu}\hat{N}^{\nu})^{;\mu} + \frac{1}{2}\delta g^{\mu\nu}\hat{N}_{\mu;\nu} \\
&= \frac{1}{2}\big(\delta(g_{\mu\nu}\hat{N}^{\nu}) - g_{\mu\nu}\delta\hat{N}^{\nu}\big)^{;\mu} + \frac{1}{2}\delta(-\hat{T}^{\mu}\hat{T}^{\nu} + \hat{N}^{\mu}\hat{N}^{\nu} + P^{\mu\nu})\hat{N}_{\mu;\nu} \\
&= -(\delta\hat{N}^{\mu})_{;\mu} + k_1\hat{N}_{\mu;\nu}\hat{T}^{\mu}\hat{T}^{\nu} + \frac{1}{2}\delta P^{\mu\nu}\hat{N}_{\mu;\nu} \\
&= -(\delta\hat{N}^{\mu})_{;\mu} + k_1\Big((\hat{N}_{\mu}\hat{T}^{\mu})_{;\nu}\hat{T}^{\nu} - \hat{T}_{\mu;\nu}\hat{N}^{\mu}\hat{T}^{\nu}\Big) + \frac{1}{2}\delta P^{\mu\nu}\hat{N}_{\mu;\nu} \\
&= -(\delta\hat{N}^{\mu})_{;\mu} - k_1\Big(\xi_{\mu;\nu}\hat{N}^{\mu}\hat{T}^{\nu}\frac{1}{\mathcal{N}}\Big) + \frac{1}{2}\delta P^{\mu\nu}\hat{N}_{\mu;\nu} \\
&= -(\delta\hat{N}^{\mu})_{;\mu} - \frac{k_1\kappa}{\mathcal{N}} - \frac{k_1}{2\mathcal{N}}\mathfrak{L}_{\xi}g_{\mu\nu}\hat{T}^{\nu}\hat{N}^{\mu} + \frac{1}{2}\delta P^{\mu\nu}\hat{N}_{\mu;\nu}. \qquad (3.75)
\end{aligned}$$

where we used Eq. (3.60) in the third line and Eq. (3.71) in the last step.

Using $\xi^{\nu}\hat{N}_{\nu} = \mathcal{N}\hat{T}^{\nu}\hat{N}_{\nu} = 0$ on $\partial\Sigma_{\text{inner}}$ and Eqs. (3.74) and (3.75) the first integral of

Eq. (3.55) may now be simplified as

$$\begin{aligned}
&\frac{1}{4\pi}\int_{\partial\Sigma_{\rm inner}}\Big(\xi^\beta {h_\mu}^{[\mu;\nu]}-\xi^\nu {h_\mu}^{[\mu;\beta]}\Big)\hat{N}_\nu\hat{T}_\beta dA=\frac{1}{4\pi}\int_{\partial\Sigma_{\rm inner}}-{h_\mu}^{[\mu;\nu]}\hat{N}_\nu\mathcal{N}dA\\
=&\frac{1}{4\pi}\int_{\partial\Sigma_{\rm inner}}\frac{1}{2}\Big({h_\mu}^{\nu;\mu}-{h_\mu}^{\mu;\nu}\Big)\hat{N}_\nu\mathcal{N}dA\\
=&\frac{1}{4\pi}\int_{\partial\Sigma_{\rm inner}}\Big[-\delta\kappa+\Big(\delta\Big(\frac{1}{2\mathcal{N}}P^{\mu\nu}\mathfrak{L}_\xi g_{\mu\nu}\Big)-\delta\Big(\frac{1}{2\mathcal{N}}\mathfrak{L}_\xi g_{\mu\nu}\hat{T}^\nu\hat{N}^\mu\Big)-\delta\theta^{(l)}\\
&-\frac{k_1}{2\mathcal{N}}\mathfrak{L}_\xi g_{\mu\nu}\hat{N}^\mu\hat{T}^\nu+\frac{1}{2}\delta P^{\mu\nu}\hat{N}_{\mu;\nu}\Big)\mathcal{N}\Big]dA.
\end{aligned}\tag{3.76}$$

Next, we define the shears of l^μ as [33]

$$\sigma_+^{(l)}=(\hat{U}^\mu\hat{U}^\nu-\hat{V}^\mu\hat{V}^\nu)l_{\mu;\nu},\qquad \sigma_\times^{(l)}=(\hat{U}^\mu\hat{V}^\nu+\hat{U}^\nu\hat{V}^\mu)l_{\mu;\nu}\,. \tag{3.77}$$

Thus we have for the final term in Eq. (3.76)

$$\begin{aligned}
\frac{1}{2}\delta P^{\mu\nu}\hat{N}_{\mu;\nu} &= \frac{1}{2}\delta P^{\mu\nu}(\hat{N}_\mu+\hat{T}_\mu)_{;\nu}-\frac{1}{2}\delta P^{\mu\nu}(\frac{1}{\mathcal{N}}\xi_\mu)_{;\nu}=\frac{1}{2}\delta P^{\mu\nu}l_{\mu;\nu}-\frac{1}{2\mathcal{N}}\delta P^{\mu\nu}\xi_{\mu;\nu}\\
&= \frac{1}{2}\Big((k_3+k_6)P^{\mu\nu}+(k_3-k_6)(\hat{U}^\mu\hat{U}^\nu-\hat{V}^\mu\hat{V}^\nu)\\
&\quad+(k_4+k_5)(\hat{U}^\mu\hat{V}^\nu+\hat{U}^\nu\hat{V}^\mu)\Big)l_{\mu;\nu}-\frac{1}{4\mathcal{N}}\delta P^{\mu\nu}\mathfrak{L}_\xi g_{\mu\nu}\\
&= \frac{1}{2}(k_3+k_6)\theta^{(l)}+\frac{1}{2}(k_3-k_6)\sigma_+^{(l)}+\frac{1}{2}(k_4+k_5)\sigma_\times^{(l)}-\frac{1}{4\mathcal{N}}\delta P^{\mu\nu}\mathfrak{L}_\xi g_{\mu\nu}.
\end{aligned}\tag{3.78}$$

We need both the generalized Komar mass and the regular ADM mass definitions to simplify the boundary term at infinity of Eq. (3.55) [60]. With the asymptotically-flat conditions discussed above, the boundary term at infinity of Eq. (3.55) may be simplified as

$$\begin{aligned}
&\frac{1}{8\pi}\int_{\partial\Sigma_\infty}({h^\mu}_{\mu;\nu}-{h^\mu}_{\nu;\mu})\hat{N}^\nu\mathcal{N}dA=\frac{1}{8\pi}\int_{\partial\Sigma_\infty}(h_{\mu\alpha;\nu}g^{\mu\alpha}-h_{\nu\alpha;\mu}g^{\mu\alpha})\hat{N}^\nu\mathcal{N}dA\\
=&\frac{1}{8\pi}\int_{\partial\Sigma_\infty}\Big(h_{\mu\alpha,\nu}-\Gamma^\lambda_{\mu\nu}h_{\lambda\alpha}-\Gamma^\lambda_{\alpha\nu}h_{\mu\lambda}-h_{\nu\alpha,\mu}+\Gamma^\lambda_{\mu\nu}h_{\lambda\alpha}+\Gamma^\lambda_{\alpha\mu}h_{\nu\lambda}\Big)g^{\mu\alpha}\hat{N}^\nu\mathcal{N}dA\\
=&\frac{1}{8\pi}\int_{\partial\Sigma_\infty}\Big(h_{\mu\alpha,\nu}-h_{\nu\alpha,\mu}+O\Big(\frac{1}{r^3}\Big)\Big)g^{\mu\alpha}\hat{N}^\nu\mathcal{N}dA\\
=&\frac{1}{8\pi}\int_{\partial\Sigma_\infty}\Big[\Big(h_{\mu\alpha,\nu}-h_{\nu\alpha,\mu}\Big)\eta^{\mu\alpha}\hat{N}^\nu+O\Big(\frac{1}{r^3}\Big)\Big]dA\\
=&\frac{1}{8\pi}\int_{\partial\Sigma_\infty}\Big(h_{\mu\alpha,\nu}-h_{\nu\alpha,\mu}\Big)\eta^{\mu\alpha}\hat{N}^\nu dA+O\Big(\frac{1}{r}\Big)\Big|_{\partial\Sigma_\infty}.
\end{aligned}\tag{3.79}$$

where we have applied $h_{\mu\nu}\simeq O(1/r)$ and $\Gamma^\lambda_{\mu\nu}h_{\lambda\alpha}\simeq O(1/r^3)$ to the second line and used $dA\simeq O(r^2)$ in the last step.

Next, we discuss the variation of the area element at infinity $dA = \sqrt{\gamma^{(\partial\Sigma_\infty)}}\, d^2z = O(r^2)$. Since $\gamma^{(\partial\Sigma_\infty)} = |\frac{\partial x^\mu}{\partial z^A}\frac{\partial x^\nu}{\partial z^B} g_{\mu\nu}| = |\frac{\partial x^\mu}{\partial z^A}\frac{\partial x^\nu}{\partial z^B} g_{\mu\nu}|$ and $\delta|g| = |g|\, g^{\mu\nu}\delta g_{\mu\nu}$ [47], $\delta(dA)$ at spatial infinity may be simplified as

$$\begin{aligned}
\delta(dA) &= \frac{\delta\gamma^{(\partial\Sigma_\infty)}}{2\sqrt{\gamma^{(\partial\Sigma_\infty)}}}\, d^2z = O\Big(\frac{1}{r^2}\Big)\delta\Big|\frac{\partial x^\mu}{\partial z^A}\frac{\partial x^\nu}{\partial z^B} g_{\mu\nu}\Big| \\
&= O\Big(\frac{1}{r^2}\Big)\Big|\frac{\partial x^\mu}{\partial z^A}\frac{\partial x^\nu}{\partial z^B} g_{\mu\nu}\Big|\,\Big(\frac{\partial z^A}{\partial x^\mu}\frac{\partial z^B}{\partial x^\nu} g^{\mu\nu}\Big)\Big(\frac{\partial x^\mu}{\partial z^C}\frac{\partial x^\nu}{\partial z^D}\delta g_{\mu\nu}\Big) \\
&= O\Big(\frac{1}{r^2}\Big)O(r^4)\,\Big(\frac{\partial z^A}{\partial x^\mu}\frac{\partial z^B}{\partial x^\nu} g^{\mu\nu}\Big)\Big(\frac{\partial x^\mu}{\partial z^C}\frac{\partial x^\nu}{\partial z^D} h_{\mu\nu}\Big) \\
&= O(r^2)\, O\Big(\frac{1}{r}\Big) \\
&= O(r).
\end{aligned} \tag{3.80}$$

Thus $\delta(\eta^{\mu\alpha}\hat{N}^\nu dA) = O(r)$, and since $h_{\mu\alpha,\nu} \simeq O(1/r^2)$, we may pull out the diffeomorphic variation and simplify Eq. (3.79) as

$$\begin{aligned}
&\frac{1}{8\pi}\int_{\partial\Sigma_\infty}\Big(h_{\mu\alpha,\nu} - h_{\nu\alpha,\mu}\Big)\eta^{\mu\alpha}\hat{N}^\nu dA \\
=\ &\frac{1}{8\pi}\int_{\partial\Sigma_\infty}\delta\Big((g_{\mu\alpha,\nu} - g_{\nu\alpha,\mu})\eta^{\mu\alpha}\hat{N}^\nu dA\Big) \\
=\ &\frac{1}{8\pi}\,\delta\!\int_{\partial\Sigma_\infty} -g_{00,\nu}\hat{N}^\nu dA + \frac{1}{8\pi}\sum_{i,j=1}^{3}\delta\!\int_{\partial\Sigma_\infty}(g_{ii,j} - g_{ji,i})\hat{N}^j dA,
\end{aligned} \tag{3.81}$$

where we have used the asymptotically-flat condition $g_{i0,0} = o(r^{-2})$ in the second line and split the metric into a temporal part and a spatial part in the third line. Then applying $g_{00} = -\mathcal{N}^2 + \beta_k\beta^k = -\mathcal{N}^2 + o(r^{-2})$ and the ADM mass definition [60] to Eq. (3.81) yields

$$\begin{aligned}
&\frac{1}{8\pi}\,\delta\!\int_{\partial\Sigma_\infty}\Big(-(\xi^\mu\xi_\mu)_{,\nu}\hat{N}^\nu + o\Big(\frac{1}{r^3}\Big)\Big)dA - 2\,\delta M^{\mathrm{ADM}} \\
=\ &\frac{1}{8\pi}\,\delta\!\int_{\partial\Sigma_\infty} -2\xi_{\mu;\nu}\xi^\mu\hat{N}^\nu dA - 2\,\delta M^{\mathrm{ADM}} \\
=\ &\frac{1}{4\pi}\,\delta\!\int_{\partial\Sigma_\infty}\Big(\xi_{[\nu;\mu]} - \frac{1}{2}\mathfrak{L}_\xi(g_{\mu\nu})\Big)\hat{T}^\mu\hat{N}^\nu dA - 2\,\delta M^{\mathrm{ADM}} \\
=\ &\delta E(\xi) - 2\,\delta M^{\mathrm{ADM}} - \frac{1}{8\pi}\,\delta\!\int_{\partial\Sigma_\infty}\mathfrak{L}_\xi(g_{\mu\nu})\hat{T}^\mu\hat{N}^\nu dA \\
=\ &\delta E(\xi) - 2\,\delta M^{\mathrm{ADM}} - \frac{1}{8\pi}\,\delta\!\int_{\partial\Sigma_\infty}\Big(g_{\mu\nu,\tau}\xi^\tau + \xi^\tau{}_{,\mu}g_{\tau\nu} + \xi^\tau{}_{,\nu}g_{\mu\tau}\Big)\hat{T}^\mu\hat{N}^\nu dA \\
=\ &\delta E(\xi) - 2\,\delta M^{\mathrm{ADM}} \\
=\ &-\delta E.
\end{aligned} \tag{3.82}$$

Where we have again applied the asymptotically-flat condition $g_{i0,0} = o(r^{-2})$ to the fifth line, and the last line holds because Theorem 3.1 proves the two mass definitions equal each other under the conditions assumed. Since the equation is covariant, the conclusion will be generally correct although we only prove it in York's asymptotically rectilinear coordinates [70].

Next substituting Eqs. (3.76), (3.78) and (3.82) into Eq. (3.55), we find that the variation of Eq. (3.49b) becomes

$$\begin{aligned}
&-\delta E - \frac{1}{4\pi}\int_{\partial\Sigma_{\text{inner}}} \delta\kappa\, dA \\
&+\frac{1}{4\pi}\int_{\partial\Sigma_{\text{inner}}}\Big(\delta\Big(\frac{1}{2\mathcal{N}}(P^{\mu\nu}-\hat{T}^{\nu}\hat{N}^{\mu})\mathfrak{L}_{\xi}g_{\mu\nu}\Big) - \frac{1}{4\mathcal{N}}\Big(\delta P^{\mu\nu} + 2k_1\hat{N}^{\mu}\hat{T}^{\nu}\Big)\mathfrak{L}_{\xi}g_{\mu\nu} - \delta\theta^{(l)} \\
&+\frac{1}{2}\Big((k_3+k_6)\theta^{(l)} + (k_3-k_6)\sigma^{(l)}_{+} + (k_4+k_5)\sigma^{(l)}_{\times}\Big)\Big)\mathcal{N}dA \\
&+\int_{\Sigma}\Big(R\,\delta\xi^{\beta} - T_{\mu\nu}h^{\mu\nu}\xi^{\beta}\Big)\hat{T}_{\beta}\sqrt{\gamma^{(\Sigma)}}d^3x - \frac{1}{4\pi}\int_{\Sigma}\Big(\mathfrak{L}_{\xi}(h_{\mu}{}^{[\mu;\beta]}) + h_{\mu}{}^{[\mu;\beta]}\xi^{\nu}{}_{;\nu}\Big)\hat{T}_{\beta}\sqrt{\gamma^{(\Sigma)}}d^3x.
\end{aligned} \tag{3.83}$$

Then Eq. (3.49c) may be transformed into

$$\begin{aligned}
&\frac{1}{4\pi}\int_{\Sigma}(\xi^{\mu}{}_{;\mu\beta} - \xi_{(\mu;\beta)}{}^{;\mu})\hat{T}^{\beta}\sqrt{\gamma^{(\Sigma)}}d^3x \\
=\ &\frac{1}{4\pi}\int_{\Sigma}\Big(\xi^{\mu}{}_{;\mu}{}^{;\beta} - \xi^{(\mu;\beta)}{}_{;\mu}\Big)\frac{\hat{T}_{\beta}}{\mathcal{N}}\mathcal{N}\sqrt{\gamma^{(\Sigma)}}d^3x \\
=\ &\frac{1}{4\pi}\int_{\Sigma}\Big(\xi^{\mu;\nu\beta} - \xi^{(\mu;\beta)\nu}\Big)g_{\mu\nu}\frac{\hat{T}_{\beta}}{\mathcal{N}}\sqrt{-g}\,d^3x,
\end{aligned} \tag{3.84}$$

where $\mathcal{N}\sqrt{\gamma^{(\Sigma)}} = \sqrt{-g}$. Since Eq. (3.48) and $\delta\sqrt{-g} = \frac{1}{2}\sqrt{-g}\,g^{\mu\nu}\delta g_{\mu\nu} = \frac{1}{2}h_{\nu}{}^{\nu}\sqrt{-g}$, the variation of Eq. (3.49c) is

$$\begin{aligned}
&\frac{1}{4\pi}\int_{\Sigma}\Big(\delta\Big(\xi^{\mu;\nu\beta} - \xi^{(\mu;\beta)\nu}\Big)g_{\mu\nu} + \Big(\xi^{\mu;\nu\beta} - \xi^{(\mu;\beta)\nu}\Big)h_{\mu\nu}\Big)\frac{\hat{T}_{\beta}}{\mathcal{N}}\sqrt{-g}d^3x \\
&+\frac{1}{4\pi}\int_{\Sigma}\Big(\xi^{\mu;\nu\beta} - \xi^{(\mu;\beta)\nu}\Big)g_{\mu\nu}\frac{\hat{T}_{\beta}}{2\mathcal{N}}h_{\nu}{}^{\nu}\sqrt{-g}d^3x \\
=\ &\frac{1}{8\pi}\int_{\Sigma}\Big(2\delta\Big(\xi^{\mu;\nu\beta} - \xi^{(\mu;\beta)\nu}\Big)g_{\mu\nu}\hat{T}_{\beta} + 2\Big(\xi^{\mu;\nu\beta} - \xi^{(\mu;\beta)\nu}\Big)h_{\mu\nu}\hat{T}_{\beta} \\
&+\Big(\xi^{\mu;\nu\beta} - \xi^{(\mu;\beta)\nu}\Big)g_{\mu\nu}\hat{T}_{\beta}h_{\nu}{}^{\nu}\Big)\sqrt{\gamma^{(\Sigma)}}d^3x.
\end{aligned} \tag{3.85}$$

Further, the variation of Eq. (3.49d) is trivially

$$\frac{1}{4\pi}\int_{\partial\Sigma_{\text{inner}}}\delta\kappa\, dA + \frac{1}{4\pi}\int_{\partial\Sigma_{\text{inner}}}\kappa\,\delta(dA)\ . \tag{3.86}$$

Based on Eqs. (3.50), (3.83), and (3.86), we find the variation of the total gravitational energy Eq. (3.49) may be written

$$\begin{aligned}
\delta E \;=\; & -\delta E + \frac{1}{4\pi}\int_{\partial\Sigma_{\mathrm{inner}}} \kappa\,\delta dA \\
& + \frac{1}{4\pi}\int_{\partial\Sigma_{\mathrm{inner}}} \Big(\delta\Big(\frac{1}{2\mathcal{N}}(P^{\mu\nu} - \hat{T}^{\nu}\hat{N}^{\mu})\mathfrak{L}_{\xi}g_{\mu\nu}\Big) - \frac{1}{4\mathcal{N}}\Big(\delta P^{\mu\nu} + 2k_1\hat{N}^{\mu}\hat{T}^{\nu}\Big)\mathfrak{L}_{\xi}g_{\mu\nu} - \delta\theta^{(l)} \\
& + \frac{1}{2}\Big((k_3+k_6)\theta^{(l)} + (k_3-k_6)\sigma_{+}^{(l)} + (k_4+k_5)\sigma_{\times}^{(l)}\Big)\Big)\mathcal{N}dA \\
& - \frac{1}{4\pi}\int_{\Sigma}\Big(\mathfrak{L}_{\xi}(h_{\mu}{}^{[\mu;\beta]}) + h_{\mu}{}^{[\mu;\beta]}\xi^{\nu}{}_{;\nu}\Big)\,\hat{T}_{\beta}\,\sqrt{\gamma^{(\Sigma)}}\,d^3x \\
& + \int_{\Sigma}\Big(\Big(2\,\delta T_{\mu}{}^{\beta} + T_{\mu}{}^{\beta}h_{\nu}{}^{\nu}\Big)\xi^{\mu} + 2\,T_{\mu}{}^{\beta}\,\delta\xi^{\mu} + \Big(R\,\delta\xi^{\beta} - T_{\mu\nu}h^{\mu\nu}\xi^{\beta}\Big)\Big)\hat{T}_{\beta}\sqrt{\gamma^{(\Sigma)}}\,d^3x \\
& + \text{Eq. (3.85)}\,. \qquad (3.87)
\end{aligned}$$

Or equivalently,

$$\boxed{\begin{aligned}
&\delta E = \frac{1}{8\pi}\int_{\partial\Sigma_{\mathrm{inner}}} \kappa\;\delta(dA) \\
& + \frac{1}{8\pi}\int_{\partial\Sigma_{\mathrm{inner}}} \Big(\delta\Big(\frac{1}{2\mathcal{N}}(P^{\mu\nu} - \hat{T}^{\nu}\hat{N}^{\mu})\mathfrak{L}_{\xi}g_{\mu\nu}\Big) - \frac{1}{4\mathcal{N}}\Big(\delta P^{\mu\nu} + 2k_1\hat{N}^{\mu}\hat{T}^{\nu}\Big)\mathfrak{L}_{\xi}g_{\mu\nu} - \delta\theta^{(l)} \\
& + \frac{1}{2}\Big((k_3+k_6)\theta^{(l)} + (k_3-k_6)\sigma_{+}^{(l)} + (k_4+k_5)\sigma_{\times}^{(l)}\Big)\Big)\mathcal{N}dA \\
& - \frac{1}{8\pi}\int_{\Sigma}\Big(\mathfrak{L}_{\xi}(h_{\mu}{}^{[\mu;\beta]}) + h_{\mu}{}^{[\mu;\beta]}\xi^{\nu}{}_{;\nu}\Big)\,\hat{T}_{\beta}\,\sqrt{\gamma^{(\Sigma)}}\,d^3x \\
& + \frac{1}{2}\int_{\Sigma}\Big(\Big(2\,\delta T_{\mu}{}^{\beta} + T_{\mu}{}^{\beta}h_{\nu}{}^{\nu}\Big)\xi^{\mu} + 2\,T_{\mu}{}^{\beta}\,\delta\xi^{\mu} + \Big(R\,\delta\xi^{\beta} - T_{\mu\nu}h^{\mu\nu}\xi^{\beta}\Big)\Big)\hat{T}_{\beta}\sqrt{\gamma^{(\Sigma)}}\,d^3x \\
& + \frac{1}{16\pi}\int_{\Sigma}\Big(2\delta\Big(\xi^{\mu;\nu\beta} - \xi^{(\mu;\beta)\nu}\Big)g_{\mu\nu}\hat{T}_{\beta} + 2\Big(\xi^{\mu;\nu\beta} - \xi^{(\mu;\beta)\nu}\Big)h_{\mu\nu}\hat{T}_{\beta} \\
& + \Big(\xi^{\mu;\nu\beta} - \xi^{(\mu;\beta)\nu}\Big)g_{\mu\nu}\hat{T}_{\beta}h_{\nu}{}^{\nu}\Big)\sqrt{\gamma^{(\Sigma)}}d^3x. \qquad (3.88)
\end{aligned}}$$

3.3.4 First law for dynamical spacetimes

Now we begin to discuss when Eq. (3.88) will reduce to an analogue of the first law of thermodynamics.

Firstly, note that the 'volume terms' (integrated over Σ) in Eq. (3.88) are exactly those one would find in the absence of an inner boundary; though in that case the domain of integration would be larger. They represent the change in global energy due to perturbations in the volume exterior to the inner boundary. Therefore their presence is independent of any thermodynamic behavior of the inner boundary. So we may write Eq. (3.88) as

$$\delta E(\xi) = \frac{1}{8\pi}\int_{\partial\Sigma_{\rm inner}} \kappa(\xi)\,\delta(dA)$$
$$-\frac{1}{16\pi}\int_{\partial\Sigma_{\rm inner}} [2\,\delta\theta^{(l)} - (k_3+k_6)\,\theta^{(l)} - (k_3-k_6)\,\sigma_+^{(l)} - (k_4+k_5)\,\sigma_\times^{(l)}]\,\mathcal{N}\,dA$$
$$+\frac{1}{32\pi}\int_{\partial\Sigma_{\rm inner}} \Big[\delta\Big(\frac{2}{\mathcal{N}}(P^{\mu\nu} - \hat{T}^\nu\hat{N}^\mu)\mathfrak{L}_\xi g_{\mu\nu}\Big)\mathcal{N} - (\delta P^{\mu\nu} + 2k_1\hat{N}^\mu\hat{T}^\nu)\mathfrak{L}_\xi g_{\mu\nu}\Big]\,dA$$
$$+\text{ volume terms}, \tag{3.89}$$

We have expressed Eq. (3.89) in terms of three surface integrals on the inner boundaries $\partial\Sigma_{\rm inner}$ and volume terms integrated over the remaining hypersurface Σ. Were only the first of these surface integrals present in Eq. (3.89) then this equation would reduce to a form of the first law of thermodynamics

$$\delta E(\xi) \;=\; \frac{1}{8\pi}\int_{\partial\Sigma_{\rm inner}} \kappa(\xi)\,\delta(dA) + \text{volume terms}\,. \tag{3.90}$$

Note, in the conventional first law, the extra work terms involve the energy required to change some conserved charges (charge or angular momentum). For us, we explicitly study uncharged and non-rotating spacetimes, i.e., our inner boundaries have no conserved charges upon which work can be done. In other words the additional terms in the inner boundary must certainly vanish if we wish to recover thermodynamic behavior there. Therefore, if the first law holds, we expect it looks like Eq. (3.90).

For the second of the surface integrals in Eq. (3.89) to vanish, the inner boundaries must correspond to surfaces with *vanishing* expansion $\theta^{(l)}$, which also implies that the shears, $\sigma_j^{(l)}$, vanish [33]. This by itself is a weaker condition than is ordinarily used in defining, for example, an apparent or trapping horizon. [27] We call such surfaces *weak future horizons*.

The third of the surface integrals in Eq. (3.89) vanishes whenever the metric is “quasi-static” on the inner boundary in the sense that $\mathfrak{L}_\xi g_{\mu\nu} = 2\,\xi_{(\mu;\nu)} = 0$ at $\partial\Sigma_{\rm inner}$. Recall that *on the horizon* we have $\xi^\mu = \mathcal{N}\hat{T}^\mu$ and due to the inclusion of the lapse function $\mathcal{N}$, the time scales involved are those of an observer at spatial infinity. Indeed, one signature feature of a black hole is that dynamics in close proximity to the horizon appears frozen to such observers. Therefore this is a rather weak constraint on the dynamics of the horizon and we call such horizons *weakly quasi-static*. Formally, this condition corresponds to ξ^μ locally satisfying the Killing equation *on the horizon*. The existence of such an ‘approximate’ (local) Killing vector is one of the fundamental assumptions in Jacobson's derivation of general relativity from the thermodynamic behavior of generic horizons [37].

For weakly quasi-static horizons the generalized surface gravity $\kappa(\xi)$ reduces to the form

$$\kappa(\xi) = \xi_{\mu;\nu}\hat{T}^{\nu}\hat{N}^{\mu} = \mathcal{N}\,\hat{T}_{\mu;\nu}\hat{T}^{\nu}\hat{N}^{\mu}. \tag{3.91}$$

This may be interpreted as the force-per-unit-mass applied at spatial infinity to hold an observer 'stationary' at the inner boundary, corresponding to an accelerated observer with 4-velocity $\hat{T}^{\mu}$. Equivalently, this is just the magnitude of the proper acceleration of such an observer, or more precisely, the limit of a family of such observers as their proper distance to the inner boundary is taken to zero, when rescaled by the lapse function $\mathcal{N}$ in order to account for measurements referenced to spatial infinity. Thus, up to the rescaling, we see that the generalized surface gravity is precisely the Unruh temperature considered by Jacobson [37].

We interpret Eq. (3.90) as *analogous* to the first law of thermodynamics with surfaces having: i) a *local* temperature $T = \kappa(\xi)/(2\pi)$. and ii) an entropy given by a local version of the conventional area law. To fully cement the thermodynamic nature of this analogy we need to prove that weakly quasi-static horizons do have a temperature $T = \kappa(\xi)/(2\pi)$. I will do this in next section.

3.4 Horizon temperature and local entropy in dynamical spacetimes

3.4.1 Hawking temperature from quantum tunneling

Theorem 3.2: For non-rotating quasi-static weak future horizons the Parikh-Wilczek tunneling temperature [45] is given by

$$T_{\text{tunneling}} = \frac{1}{2\pi}\,\xi_{\mu;\nu}\hat{N}^{\mu}\hat{T}^{\nu} = \frac{\kappa(\xi)}{2\pi}. \tag{3.92}$$

Proof:

We now consider the Hawking temperature calculated directly using the quantum tunneling formalism developed by Parikh and Wilczek [45]. We adapt this method to a scenario where the horizon is without general spherical symmetry. Indeed, there can be multiple horizons. For the purposes of illustration, we shall consider a spacetime with n

black holes whose j^{th} horizon is located at $r = R_j(\theta, \phi)$. The lack of spherical symmetry allows for the possibility that the local tunneling temperature need not be uniform across the horizon.

Consider a *massless* field mode whose amplitude is slowly varying

$$\Psi_\omega(x^\mu) \propto \exp\Big(i\omega \int^{x^\mu} \ell_\nu \, dx^\nu\Big), \tag{3.93}$$

where, following Parikh and Wilczek, the Hamilton-Jacobi equation reduces to $g^{\mu\nu}\ell_\mu\ell_\nu = 0$. In other words, ℓ^μ is just a null vector. Being null, we can normalize this vector arbitrarily. We choose a normalization where $\ell_0 = -1$ so that the parameter ω in Eq. (3.93) is the mode's frequency.

In the Parikh-Wilczek approach, the temperature comes from the integral inside the exponent of Eq. (3.93) across a pole that occurs at the horizon. The integration is performed (non-radially) along a spacelike path *orthogonal* to the j^{th} black hole's horizon, so that the Boltzmann factor for the j^{th} black hole reduces to

$$\exp\Big(\frac{-\hbar\omega}{k_B T_j}\Big) = \left|\exp\Big[-\pi\omega \, \mathrm{Res}_j\Big(\int^{x^\mu} \ell_\nu \, dx^\nu\Big)\Big]\right|^2 , \tag{3.94}$$

where $\mathrm{Res}_j(f(z))$ denotes the residue, here evaluated at the j^{th} horizon, i.e., at $r = R_j(\theta, \varphi)$. (For clarity, we include $\hbar$ and k_B in this section, but set them again to unity in our final step below.)

Denoting the direction of the path as it crosses normally to the horizon by the unit spacelike vector $\hat{N}^\mu$, where with our coordinates for the j^{th} horizon $\hat{N}_\mu \propto \partial_\mu r$, then $\ell_\mu = l_\mu/\mathcal{N} = (\hat{T}_\mu + \hat{N}_\mu)/\mathcal{N}$, since the normalization is given by $\hat{T}_0 = -\mathcal{N}$ and $\hat{N}_0 = 0$. Assuming that the integral $\int^{x^\mu} \ell_\nu dx^\nu$ along this path has a simple Laurent expansion we may extract the reciprocal of the Residue by taking the directional derivative along the path of the reciprocal of $\ell_\mu \hat{N}^\mu = 1/\mathcal{N}$. This recipe uniquely pulls out the contribution from the pole, yielding

$$\frac{1}{\mathrm{Res}_j\Big(\int^{x^\mu} \ell_\nu \, dx^\nu\Big)} = \Big(\frac{1}{\ell_\mu \hat{N}^\mu}\Big)_{;\nu} \hat{N}^\nu = \mathcal{N}_{;\nu} \hat{N}^\nu \ , \tag{3.95}$$

and hence the temperature will be given by

$$
\begin{aligned}
T_j(\theta,\varphi) &= \frac{\hbar}{2\pi k_B}\left(\frac{1}{\ell_\mu \hat{N}^\mu}\right)_{;\nu}\hat{N}^\nu\Big|_{r=R_j(\theta,\varphi)} \\
&= \frac{\hbar}{2\pi k_B}\,\mathcal{N}_{;\nu}\hat{N}^\nu\Big|_{r=R_j(\theta,\varphi)} \\
&= \frac{\hbar}{2\pi k_B}\,\frac{-1}{2\mathcal{N}}(\xi^\mu\xi_\mu)_{;\nu}\hat{N}^\nu\Big|_{r=R_j(\theta,\varphi)} \\
&= \frac{-\hbar}{2\pi k_B}\,\xi_{\mu;\nu}\hat{T}^\mu\hat{N}^\nu\Big|_{r=R_j(\theta,\varphi)} \\
&= \frac{\hbar}{2\pi k_B}\,\xi_{\nu;\mu}\hat{T}^\mu\hat{N}^\nu\Big|_{r=R_j(\theta,\varphi)} \\
&= \frac{\kappa}{2\pi}\Big|_{r=R_j(\theta,\varphi)}\,,
\end{aligned} \tag{3.96}
$$

recall that on the horizon $\xi^\mu\xi_\mu = -\mathcal{N}^2$, $\xi^\mu = \mathcal{N}\hat{T}^\mu$, and $\xi_{\mu;\nu} = -\xi_{\nu;\mu}$ since ξ^μ is a local-Killing vector (what we call weakly quasi-static) there. The final identification here of the local tunneling temperature as the local surface gravity comes from Eq. (3.91).

This completes the proof of Theorem 3.2.

In summary so far, we have obtained a first law of black hole mechanics generalized to dynamical spacetimes. The above calculation of the Hawking temperature as given by $\kappa/(2\pi)$ shows that our result is not merely an analogy, but a proof of an exact thermodynamic relation. We summary these results as Theorem 3:

Theorem 3.3: Under the assumptions of Theorem 3.1, Eq. (3.90) is a statement of the first law of thermodynamics in an extended dynamical setting: It describes real thermodynamic behavior, not mere analogy, for non-rotating weakly quasi-static weak future horizons.

We should emphasise, that at least in the non-rotating case, the relevant horizon is determined solely by the vanishing of the expansion; in other words it is the 'weak future horizon' which possesses the thermodynamic properties, without the further conditions required to define, for example, a trapping horizon or an apparent horizon.

Another point we should emphasise is that we never assume the temperature is a constant during all the above calculations in this chapter. In fact, we will show below that the temperature of dynamical black holes generally is not uniform.

3.4.2 Non-uniform temperature

Here we consider an exact calculation, to explicitly show that our surface gravity (and hence equivalently the Hawking temperature) will be in general non-uniform along the horizon even in the quasi-static case.

The specific scenario that we consider will involve interacting black holes on a spacelike hypersurface Σ where all black holes are assumed to be instantaneously stationary (and hence the quasi-static condition will be globally satisfied on the entire hypersurface). Within the $3+1$ split formalism [25] it is sufficient to construct a metric of the form

$$ds^2 = -\mathcal{N}^2 dt^2 + \gamma_{ij}(dx^i + \beta^i dt)(dx^j + \beta^j dt), \tag{3.97}$$

where γ_{ij} is the spatial metric on the hypersurface of interest.

The simplest such 'initial conditions' [25] are time-symmetric, thus all first-order time derivatives vanish and we have a conformally-flat spatial metric $\gamma_{ij} = \psi^4 \eta_{ij}$ (where $\eta_{\mu\nu}$ is the Minkowski metric) and define $\chi \equiv \psi \mathcal{N}$. Under these conditions [25], the vacuum Einstein field equations on the initial hypersurface Σ reduce to $\beta^i = 0$ and

$$^{(3)}\nabla^2 \psi = 0, \qquad ^{(3)}\nabla^2 \chi = 0, \tag{3.98}$$

where $^{(3)}\nabla^2$ is the flat-spatial Laplacian. Assuming an asymptotically-flat spacetime with n singularities one finds the Brill-Lindquist initial conditions [19, 25]

$$\psi = 1 + \sum_{i=1}^{n} \frac{\mu_i}{2|\vec{r} - \vec{r}_i|}, \qquad \chi = 1 - \sum_{i=1}^{n} \frac{\mu_i}{2|\vec{r} - \vec{r}_i|}. \tag{3.99}$$

For a single black hole (i.e., $n = 1$) Eqs. (3.97) and (3.99) would correspond to a Schwarzschild black hole in isotropic coordinates, and μ_1 would be its mass. For $n > 1$ black holes, the total gravitating mass at spatial infinity for this metric is easily calculated to be $\sum_{i=1}^{n} \mu_i$.

Consider the j^{th} horizon. The location of the weak future horizon may be determined by computing the expansion of outgoing null 4-vectors normal to the horizon. (This is done explicitly for the case $n = 2$, of a binary pair of black holes, in Ref. [39].)

Following the results of the previous section, the Hawking temperature on the horizon is given by the expression

$$T_j = \left. \frac{r^2 \mathcal{N}_{,r} - R_{,\theta} \mathcal{N}_{,\theta} - \csc^2(\theta) R_{,\varphi} \mathcal{N}_{,\varphi}}{2\pi\, r\, \psi^2 \sqrt{(R_{,\theta})^2 + \csc^2(\theta)(R_{,\varphi})^2 + r^2}} \right|_{r = R_j(\theta,\phi)} . \tag{3.100}$$

An explicit calculation for a binary pair of black holes, $n = 2$, yields the Hawking temperature of the $j = 1$ black hole as (see Chapter 4 for details):

$$T_1 = \frac{1}{8\pi\mu_1}\left[1 - \frac{\mu_2}{r_{12}} + \frac{3\mu_2(\mu_2 - \mu_1\cos\theta)}{4r_{12}^2} + O\Big(\frac{1}{r_{12}^3}\Big)\right], \tag{3.101}$$

where r_{12} is the inter-black hole 'distance' (measured in isotropic coordinates), and with a similar result for T_2 taking $\mu_1 \leftrightarrow \mu_2$.

Equation (3.101) shows that the temperature varies along the black hole horizons (see Fig 3.5). Such a failure of the zeroth law, while maintaining the first law, implies that thermal equilibrium of the horizons is replaced by local equilibrium.

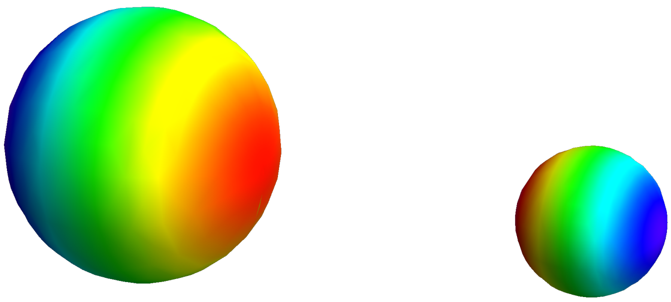

Figure 3.5: Failure of the zeroth law of black hole mechanics, illustrated through an example of a pair of interacting black holes. We plot an exaggerated profile of the temperatures from Eq. (3.101) (the transition from red to blue denotes increasing temperatures). Generically, interacting black holes are non-equilibrium objects; though they do satisfy local equilibrium.

3.4.3 Local equilibrium and the elemental area law

Since the temperatures on the horizon may be non-uniform, each element of area must correspond to a true entropy, or more precisely $dS = dA/4$, for Eq. (3.90) to be a true first law. We therefore conclude that the conventional area law applies to each element of area. We call this extension the *elemental area law*. This extension gives rigorous meaning to Bekenstein's [13] long-standing conjecture of how information is encoded on a black hole horizon (see Fig. 3.6).

Integrating the elemental area law yields the conventional stationary form of the area law, $S = A/4$, but now rigorously obtained for dynamical (Killing vector free) black holes. As noted by Wald [62] the area corresponds to a Noether invariant. He proposed a natural extension of this invariant for arbitrary diffeomorphically invariant theories. He then conjectured that this would correspond to the entropy of dynamical (i.e., non-stationary) black holes in such theories. Therefore, at least for the case of general relativity, our result proves Wald's conjecture about the nature of dynamical black hole entropy. Note, that

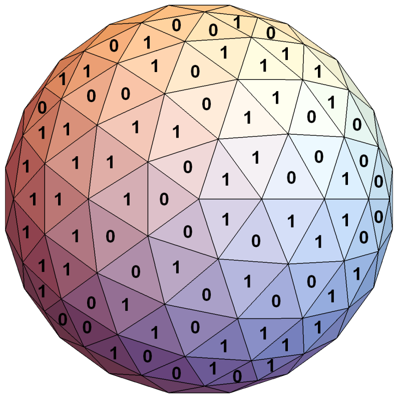

Figure 3.6: Bekenstein's heuristic picture for information encoding on a black hole's horizon: each (Planck area)/$(4\ln 2)$ is shown encoding a single bit. This picture is given rigorous meaning by the elemental area law.

the Noether charge formalism is silent about the entropic content of each element of area.

3.5 Cosmological constant

Our assumption of an asymptotically-flat spacetime excludes a cosmological constant. However, such a constant provides a potentially simple way to explain the phenomenon of dark energy. Recent observations [50], suggest that the dark energy density increases with time, which would be inconsistent with a cosmological constant model. Nevertheless, we may ask whether cosmological effects may undermine our results. In this section, we use the stationary theory to show that a cosmological contribution grows with the size of the black hole, but remains negligible except for black holes roughly 10^{11} times more massive than the largest observed black hole in our universe.

Now consider the Schwarzschild metric for a stationary black hole in a spacetime with a cosmological constant:

$$g_{\mu\nu} = -\Big(1 - \frac{2M}{r} - \frac{\Lambda r^2}{3}\Big)dt^2 + \frac{1}{1 - \frac{2M}{r} - \frac{\Lambda r^2}{3}}dr^2 + r^2 d\Omega^2, \tag{3.102}$$

where $d\Omega^2 = d\theta^2 + \sin^2\theta\, d\phi^2$, M is the black hole mass and Λ is the cosmological constant.

The location of the horizon is given by the vanishing of the expansion of the outgoing null normal congruences, which yields the condition that

$$1 - \frac{2M}{r} - \frac{\Lambda r^2}{3} = 0. \tag{3.103}$$

To lowest non-trivial order the location of the black hole horizon is therefore given by

$$r_{\text{horizon}} = 2M\left(1 + \frac{4\Lambda M^2}{3}\right) + O(\Lambda^2 M^4). \tag{3.104}$$

From the area law, we may obtain the black hole entropy as

$$S = \frac{A}{4} = 4\pi M^2\Big(1+\frac{8\Lambda M^2}{3}\Big) + O(\Lambda^2 M^4). \tag{3.105}$$

The tunneling temperature of the horizon satisfies

$$\frac{1}{T} = 4\pi\,\mathrm{Res}\Bigg(\frac{1}{1-\frac{2M}{r}-\frac{\Lambda r^2}{3}}\Bigg)\Bigg|_{r=r_{\text{horizon}}} = 8\pi M\,\Big(1+\frac{16\Lambda M^2}{3}\Big) + O(\Lambda^2 M^4). \tag{3.106}$$

Using the above results, we verify the form of the stationary first law as

$$T\,dS = dM + O(\Lambda^2 M^4). \tag{3.107}$$

In summary, the cosmologically perturbed temperature and entropy of our black hole are given by

$$\begin{aligned} T^{(\Lambda)} &= \Big(1-\frac{16\Lambda M^2}{3}\Big)T^{(\Lambda=0)} + O(\Lambda^2 M^4), \\ S^{(\Lambda)} &= \Big(1+\frac{8\Lambda M^2}{3}\Big)S^{(\Lambda=0)} + O(\Lambda^2 M^4). \end{aligned} \tag{3.108}$$

Now taking the observational value of the cosmological constant as

$$\Lambda^{\text{obs}} = 1.105610^{-52}m^{-2} \simeq \frac{1}{(10^{10}\,\text{ly})^2}, \tag{3.109}$$

where ly denotes a light year. We find that a cosmological effect on black hole thermodynamics is utterly negligible provided the black hole's Schwarzschild radius is much smaller than around 10^{10} ly. Equivalently, a cosmological effect only begins to become non-trivial for black holes which are roughly 10^{11} times more massive than the largest observed black hole in the universe.

3.6 Discussion

Based on classical general relativity, we have extended the first law of black hole mechanics from stationary to dynamical spacetimes. We used global methods based on a rigorous and well-defined physical energy. This allowed us to overcome ambiguities which have previously been found in attempted generalizations of black hole mechanics to dynamical scenarios [42]. In this classical setting, for horizons defined by vanishing expansion, we may apply Hawking's original theorem [33] for the second law of black hole mechanics which additionally assumes the null energy condition. In the quantum setting, where this condition is violated due to quantum effects (leading to Hawking radiation), to regain the second law one must presumably include both the entropic contribution from the horizon's area and from the radiation [14]. Such an analysis is beyond the methods of classical general relativity, which has been our prime focus here.

For a universe with multiple well-separated black holes and hence with negligible interactions our result becomes

$$\delta E = \sum_i T_i \, \delta S_i + \text{volume terms}. \tag{3.110}$$

One might worry that: i) this is not really a thermodynamic relation, when there are differing temperatures, and ii) this relation is actually a weaker statement than could be obtained from local-flux arguments, which would yield a distinct relation for each black hole. However, we should recall that the flexibility in choosing the diffeomorphisms means that Eq. (3.110) is actually an infinity of relations. In particular, consider a series of diffeomorphisms which largely affect only each black hole in turn. In this way, we would obtain $\delta E = T_i \, \delta S_i$ + volume terms, for each black hole. This illustrates that our monolithic formulation of the first law represents a valid thermodynamic relation for a multi-component system, and that it certainly embodies no less information than obtained from (less rigorous) local methods.

In the usual equilibrium setting, one quarter of a black hole's area is interpreted as the entropy of the entire black hole, i.e., the area law $S_{\text{black hole}} = A/4$. By contrast, our derivation of the elemental area law, $dS = dA/4$, clearly shows that this entropy is a *local property of the horizon.* Therefore, the entropy associated with the horizon may not necessarily be that of the entire black hole. Indeed, since no signal from the black hole interior can escape, one might assign it a temperature, $T_{\text{interior}} = 0$, with respect to an observer at spatial infinity (the interior is unobservable to any external observer, so such an assignment is not inconsistent). Such an assignment, however, means the first law would be unchanged by the addition of the vanishing term $T_{\text{interior}} \, \delta S_{\text{interior}}$. In that case, the entropy of the entire black hole becomes

$$S_{\text{black hole}} = \frac{A}{4} + S_{\text{interior}}. \tag{3.111}$$

Our analysis does not prove this relation, but it does show its consistency with black hole thermodynamics. Of course, the presence of such an additional term would require a wholesale reappraisal of black hole information bounds and paradoxes.

Finally, in 1995, Jacobson derived the equations of general relativity by assuming that generic horizons satisfy local thermodynamic equilibrium [37]. Here, we have proved that

Jacobson's assumptions themselves follow from classical general relativity. Combined, these converse results demonstrate the one-to-one logical equivalence between classical general relativity and the thermodynamic nature of spacetime horizons. While the two formulations yield the same classical theory of gravity, their vastly different sets of underlying variables likely yield very distinct quantizations. Working on the premise that thermodynamics is fundamental, Jacobson conjectured that general relativity is an emergent theory and inferred that the canonical route to the quantization of gravity may be incorrect. The one-to-one equivalence between a metric and thermodynamic formulation of gravity casts doubt on Jacobson's paradigm of gravity as an emergent phenomenon, and hence the question remains open as to the appropriate quantization of gravity. Regardless, the nature of quantum gravity is likely to be further informed by the richer class of behaviors accessible to dynamical spacetimes as we shall now see in Chapter 4.

Chapter 4

Non-equilibrium entropic forces: From molecular chains to black holes

The tendency of systems to become more disordered, leading to higher entropy, is a universal phenomenon in nature. Often, a system's physical drive to increase its entropy can be recast in terms of an 'entropic force.' The power of entropic force formulations arises from the ability to characterize the behavior of thermodynamic systems in terms of numbers of possible states or configurations, rather than requiring a detailed calculation of the underlying, complex and sometimes unknown physical laws. We extend the concept of entropic forces from equilibrium scenarios to the non-equilibrium thermodynamic regime. We illustrate non-equilibrium entropic forces for two scenarios: a microscopic, molecular system, and a macroscopic, black hole system. In both cases, the entropic force vanishes at equilibrium. In the latter scenario, extrapolating the force to coalescing black holes provides the first direct test for the thermodynamic nature of black hole entropy from future precision measurements of gravitational waveforms. We further show that this entropic force cannot be encoded in the Einstein field equations defining classical general relativity. Thus, its observation would constitute the first unequivocal signature of quantum gravity. By contrast, its absence would compel us to reject the statistical nature of black hole entropy initiating a major reassessment of the theoretical foundations of physics.

4.1 Non-equilibrium entropic forces for molecule chains

The prototypical example of an entropic force involves a flaccid negligible-weight long-chain molecule. In isolation, there is no energetic preference to any configuration, nor to the distance ℓ separating the two ends of the molecule. However, when placed in a bath at a temperature T, an entropic force $TdS/d\ell$ is experienced at the molecule's free end [26], see Fig. 4.1(a). Calculating the entropic force in this way requires only a count of all possible configurations of the molecular chain and the corresponding configurational entropy S as a function of the end-to-end separation. Because this force persists even for systems in equilibrium, we call it an equilibrium entropic force.

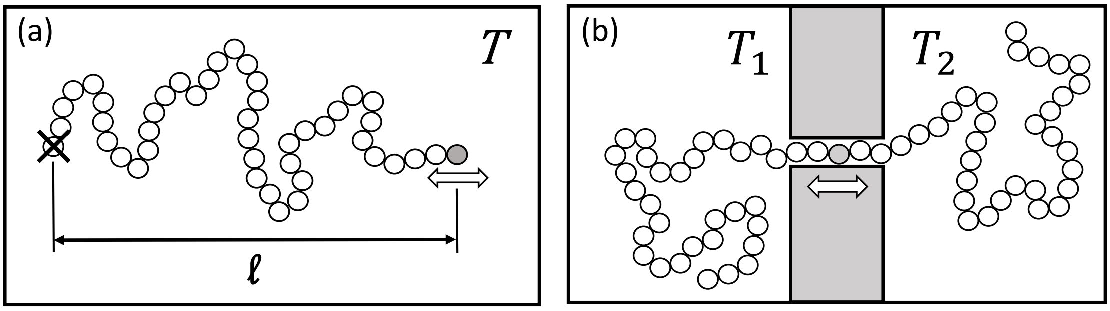

Figure 4.1: Examples of a system experiencing: (a) an equilibrium entropic force: A flaccid negligible-weight long-chain molecule with end-to-end separation ℓ sits in a bath at temperature T. One end of the molecule is fixed so that the force is experienced at the free end; (b) a non-equilibrium entropic force: Here our long-chain molecule is threaded across a nanopore between two chambers held at respective temperatures T_i.

Consider now the entropic force on the same molecule if it is threaded through a frictionless nanopore between two chambers, with bath temperatures T_1 and T_2, see Fig. 4.1(b). Provided both ends dangling into their respective chambers are not too short, then the *net* configurational entropy of the molecule is nearly independent of the proportion in each chamber [69]. Under these circumstances, a small shift of the molecule through the nanopore in the direction of chamber 2 leads to an increase dS_2 in the corresponding configurational entropy S_2, with a nearly equal [69] decrease $dS_1 \simeq -dS_2$ in chamber 1. The almost zero net change in entropy (dS_1+dS_2) means that no net entropic force will be observed on the molecule when $T_1 = T_2$ [69]. In contrast, in the non-equilibrium scenario, when the chambers are held at different temperatures, we find a non-equilibrium entropic force proportional to $\sum_{j=1}^{2} T_j dS_j \simeq (T_2 - T_1)dS_2$, driving the molecule towards the chamber with the *higher* temperature at an approximately constant rate proportional to the temperature difference. A potential application of this setup is considered in the Discussion. In summary, non-equilibrium entropic forces can arise purely due to non-equilibrium environments, even when the overall entropy of a system is conserved.

4.2 Temperature and entropy of dynamical black holes

4.2.1 An entropic force is absent for equilibrium black holes

For entropic forces to apply, the system concerned must be thermodynamic, possessing both a temperature and an entropy. That stationary black holes meet this condition is a standard result [11] and is explained in chapter 2. By the second law of black hole mechanics and the area law, any physical change on a black hole will lead to $\frac{1}{4}\delta A = \Delta S \geq 0$. Therefore, simply shifting the location of a black hole with regard to any other gravitating body (an in principle physically reversible operation), implies $\delta S \equiv 0$. If this black hole were at equilibrium at temperature T, we observe rather straightforwardly that $T\delta S = 0$. This black hole will therefore not experience an equilibrium entropic force, despite contrary claims in the literature [41, 57] (we point out the technical error in these studies below).

This is the scenario for equilibrium black holes. As we have seen in chapter 3, a black hole interacting with any other gravitating body is in fact not an equilibrium system. So we now briefly review the key results of chapter 3 and then go on to consider the entropic force for interacting (non-equilibrium) black holes.

4.2.2 Dynamical black hole thermodynamics

The non-equilibrium entropic force among black holes I will discuss is based on our recent rigorous extension of black hole thermodynamics from stationary spacetimes [11] to dynamical spacetimes (Chapter 3 of this thesis) [66]. In that extension a fully dynamical spacetime was analyzed and a thermodynamic horizon, the weak future horizon, was identified as the thermodynamic surfaces when these surfaces are weakly quasi-static in the non-rotating case. We briefly summarize the results of Chapter 3 here.

Consider a spacelike three-dimensional hypersurface Σ (corresponding to a constant time slice). We may define a future-directed timelike unit vector $\hat{T}^\mu$ normal to Σ at each point with $\hat{T}^\mu \hat{T}_\mu = -1$.

Next, consider a generic closed compact 2-surface residing in Σ. At each point on this surface we may define an outward pointing spacelike unit vector $\hat{N}^\mu$ normal to the surface and tangent to Σ, so $\hat{T}^\mu \hat{N}_\mu = 0$. Such a surface corresponds to a *weak future horizon* [66] if the expansion of outgoing null normal congruences of geodesics vanishes on the surface. Let us decompose the outgoing and incoming null geodesics as

$$\ell^\mu = \frac{1}{\sqrt{2}}(\hat{T}^\mu + \hat{N}^\mu), \qquad n^\mu = \frac{1}{\sqrt{2}}(\hat{T}^\mu - \hat{N}^\mu). \tag{4.1}$$

Trivially $\ell^\mu \ell_\mu = n^\mu n_\mu = 0$ and satisfy the conventional normalization $\ell^\mu n_\mu = -1$.

From the projector onto the tangent space of this 2-surface

$$P_{\mu\nu} \equiv g_{\mu\nu} + \hat{T}_\mu \hat{T}_\nu - \hat{N}_\mu \hat{N}_\nu = g_{\mu\nu} + \ell_\mu n_\nu + n_\mu \ell_\nu, \tag{4.2}$$

we have the conventional definition of the expansion as

$$\theta^{(\ell)} \equiv P^\mu{}_\nu \nabla_\mu \ell^\nu. \tag{4.3}$$

As already noted, weakly quasi-static non-rotating surfaces with vanishing expansion, carry a *local* temperature given by

$$T = \frac{\hbar\,\kappa}{2\pi k_B}, \tag{4.4}$$

where k_B and $\hbar$ are the Boltzmann and Planck constants respectively, and the surface gravity κ is given by Eq. (3.91) [66]. Further, each element of area $d\mathcal{A}$ of this horizon carries thermodynamic entropy

$$dS = \frac{k_B\,d\mathcal{A}}{4\hbar G}, \tag{4.5}$$

where G is the gravitational constant (setting $c = 1$). A result we called the *elemental area law* [66].

This local temperature was further found to equal the local Parikh-Wilczek tunneling temperature of the Hawking radiation [66]. Therefore the theoretical results strongly support the notion that the entropy and temperature of Eqs. (4.4) and (4.5) correspond to actual thermodynamic quantities and hence that black hole horizons are true thermodynamic entities.

4.2.3 n instantaneously stationary black holes

To date, no direct test exists for either entropy or temperature for black holes. Hawking radiation could provide evidence for temperature but is too weak to measure for astrophysical black holes [31,56]. That said, experiments on laboratory analogues of black hole horizons provide support for the underlying radiation mechanism [10]. Unlike temperature, we know of no proposal for testing the thermodynamic nature of entropy, either in real black holes or their laboratory analogues. In what follows, we show that non-equilibrium entropic forces between black holes can provide just such a test.

To extend the concept of non-equilibrium entropic forces to black holes, we begin with an example of dynamic, interacting black holes. We will first show that the equilibrium entropic force vanishes, before deriving the non-equilibrium force, which will, as in the molecular case, arise from a non-uniform temperature distribution.

Because the effect we are examining is both small and exotic, it is vital here to use exact analytic calculations over those from numerical relativity; at least for the purposes of proof of its existence. Consider the analytically describable Brill-Lindquist initial conditions [19] consisting of n instantaneously stationary black holes in an asymptotically-flat spacetime. The time symmetry around these initial conditions ensures a solution satisfying both the quasi-static and non-rotating conditions needed [66] in the dynamical expression of the first law for black holes. We therefore have a hypersurface with a rigorous non-equilibrium thermodynamic behavior [19, 66].

These initial conditions studied by Brill and Lindquist corresponds to the $t = 0$ hypersurface of the otherwise static metric

$$ds^2 = -\mathcal{N}^2 dt^2 + \psi^4 \eta_{ij} dx^i dx^j, \tag{4.6}$$

where $i, j \in \{1, 2, 3\}$, and the lapse function is given by $\mathcal{N} = \chi/\psi$ with

$$\chi = 1 - \sum_{i=1}^{n} \frac{G\mu_i}{2|\vec{r} - \vec{r}_i|}, \qquad \psi = 1 + \sum_{i=1}^{n} \frac{G\mu_i}{2|\vec{r} - \vec{r}_i|}. \tag{4.7}$$

Here $\vec{r}_i$ label the 'location' of the black holes in isotropic coordinates and the parameters μ_i are *related* to the black hole masses. The relationship between the mass parameters μ_i and the physical masses involve the gravitational interaction energy. The difficult problem of disentangling the mass parameters μ_i from the physical masses was solved by Brill and Lindquist utilizing Einstein-Rosen bridges [19] yielding

$$m_i \equiv \mu_i \left(1 + \sum_{j \neq i}^{n} \frac{G\mu_j}{2r_{ij}}\right), \tag{4.8}$$

where $r_{ij} \equiv |\vec{r}_i - \vec{r}_j|$. In fact, beyond the analytic tractability of the Brill-Lindquist system, it is this ability to express the results in terms of the physical masses which is crucial for correctly computing the entropic forces as we shall see below.

4.2.4 Location of the black hole horizons for binary black holes

Next, we seek to locate the horizons. From the results of chapter 3 (discussions after Eq. (3.90)), we know that the horizons with thermodynamic behaviors are defined by $\theta^{(l)} = 0$. For simplicity, we limit our analysis here to the case of $n = 2$ black holes, which we take to be aligned along the z-axis. Without loss of generality, we take $\vec{r}_1 = 0$ and write

$$1 - \chi = \psi - 1 = \frac{G\mu_1}{2\,r} + \frac{G\mu_2}{2\sqrt{r^2 - 2\,r\,r_{12}\cos\theta + r_{12}^2}}. \tag{4.9}$$

By azimuthal symmetry, the horizon for the first black hole (labeled by $\vec{r}_1 = 0$) will correspond to a 2-surface $r = \Delta_1(\theta)$ for some function $\Delta_1(\theta)$. A pair of (generally) independent tangent vectors within this surface are given by

$$\frac{dx^\mu}{d\theta} = (0, \partial_\theta\Delta_1, 1, 0), \qquad \frac{dx^\mu}{d\phi} = (0, 0, 0, 1). \tag{4.10}$$

We may now construct a future-directed null normal geodesic ℓ_μ to this 2-surface. The null-geodesic condition requires

$$ds^2 = -\frac{\ell_t^2}{\mathcal{N}^2} + \frac{1}{\psi^4}\left(\ell_r^2 + \frac{1}{r^2}\ell_\theta^2 + \frac{1}{r^2\sin^2\theta}\ell_\phi^2\right) = 0. \tag{4.11}$$

Normality to the 2-surface requires

$$\ell_\theta = -(\partial_\theta\Delta_1)\ell_r, \qquad \ell_\phi = 0. \tag{4.12}$$

Combining Eqs. (4.11) and (4.12) implies

$$\ell_r = \pm\psi^2\sigma\,\frac{\ell_t}{\mathcal{N}}, \tag{4.13}$$

where we define

$$\sigma^{-2} \equiv 1 + \frac{(\partial_\theta\Delta_1)^2}{r^2} > 1. \tag{4.14}$$

Therefore the outgoing and ingoing future-directed null normal geodesic tangent vectors may be written as

$$\begin{aligned} \ell_\mu &= \frac{1}{\sqrt{2}}(-\mathcal{N}, \psi^2\sigma, -\psi^2\sigma\partial_\theta\Delta_1, 0) \\ n_\mu &= \frac{1}{\sqrt{2}}(-\mathcal{N}, -\psi^2\sigma, \psi^2\sigma\partial_\theta\Delta_1, 0), \end{aligned} \tag{4.15}$$

respectively.

The expansion may now be explicitly computed as

$$\theta^{(\ell)} = \frac{f\sigma^3}{\sqrt{2}r^4\psi^3}, \tag{4.16}$$

where $f = f(r,\theta)$ is given by

$$f=[r^2+(\partial_\theta\Delta_1)^2](4r^2\partial_r\psi-4\partial_\theta\psi\,\partial_\theta\Delta_1-\psi\partial_\theta\Delta_1\cot\theta)+r\psi[2r^2+3(\partial_\theta\Delta_1)^2-r\partial_\theta^2\Delta_1]. \quad (4.17)$$

Solving $\theta^{(\ell)}=0$ order-by-order gives the location of the weak future horizon of the first black hole as $r=\Delta_1(\theta)$, where

$$\begin{aligned}
\Delta_1(\theta) \quad = \quad & \frac{G\mu_1}{2}-\frac{G^2\mu_1\mu_2}{4r_{12}}+\frac{G^3\mu_1\mu_2(\mu_2-\mu_1\cos\theta)}{8r_{12}^2}+\frac{G^4\mu_1\mu_2}{r_{12}^3}\left(\frac{5\mu_1^2(1-3\cos^2\theta)}{224}\right.\\
& \left.-\frac{\mu_2(\mu_2-3\mu_1\cos\theta)}{16}\right)+\frac{G^5\mu_1\mu_2}{r_{12}^4}\left(\frac{\mu_2^2(\mu_2-6\mu_1\cos\theta)}{32}-\frac{\mu_1^2\mu_2(9-41\cos^2\theta)}{224}\right.\\
& \left.+\frac{7\mu_1^3\cos\theta(3-5\cos^2\theta)}{832}\right)+O(G^6) \qquad (4.18)
\end{aligned}$$

with a similar result for $\Delta_2(\theta)$ from $\mu_1\leftrightarrow\mu_2$. We note that this result has previously been obtained to lower order in Ref. [39].

4.2.5 Black hole entropy for binary black holes

Now that we have the locus of the i^{th} horizon in Eq. (4.18) we may compute its area and consequently its entropy. In order to do this we must first compute the induced metric on this surface; the area then reduces to the 2-volume of the corresponding compact manifold.

The key step in obtaining the induced metric on the inner boundary is simply a coordinate transformation. For simplicity we work in spherical coordinates. We wish to transform from $x^i\equiv(r,\theta,\phi)$ to $z^A\equiv(\theta,\phi)$ restricted to the surface $r=\Delta_1(\theta)$. We then use the standard rule for coordinate transformations applied to the metric as

$$P_{AB}=\frac{\partial x^i}{\partial Z^A}\frac{\partial x^j}{\partial Z^B}\gamma_{ij}, \qquad (4.19)$$

where $\gamma_{ij}=\psi^4\eta_{ij}$ and P_{AB} is the induced metric, for $A,B\in\{2,3\}$ (in comparison with the projector $P_{\mu\nu}$ for which $\mu,\nu\in\{0,1,2,3\}$).

The area on the 2-D manifold is $\mathcal{A}\equiv\int\sqrt{|((P_{AB}))|}\,d\theta d\phi$. For the $i=1$ black hole we find

$$\begin{aligned}
d\mathcal{A}_1 \equiv & \sqrt{|((P^{(1)}_{AB}))|}\\
= & 4G^2m_1^2\sin\theta\left\{1+\frac{G^2m_1m_2\cos\theta}{2\,r_{12}^2}+\frac{G^3m_1m_2}{16\,r_{12}^3}\Big[m_1-4(m_1+3m_2)\cos\theta+3m_1\cos 2\theta\Big]\right.\\
& +\frac{G^4\,m_1m_2}{64\,r_{12}^4}\Big[-2m_1^2-11m_1m_2+5m_1^2\cos 3\theta+(11m_1^2+56m_1m_2+48m_2^2)\cos\theta\\
& \left.-m_1(6m_1+33m_2)\cos 2\theta\Big]\right\}+O(G^7), \qquad (4.20)
\end{aligned}$$

where we have replaced the mass parameters μ_i by their physical counterparts m_i using Eq. (4.8). Integration yields the area of the i$^{\text{th}}$ horizon as

$$\mathcal{A}_i = 16\pi\, G^2 m_i^2 + O(G^7). \tag{4.21}$$

Finally, from the (elemental) area law Eq. (4.5) [66] the entropy of the i$^{\text{th}}$ black hole is given by

$$S_i = \frac{4\pi k_B G m_i^2}{\hbar} + O(G^6). \tag{4.22}$$

This result proves that the entropy of each black hole is indeed independent of the separation r_{12} between the black holes, at least to the high-order expansion we explicitly computed. This result is consistent with our intuition about the statistical mechanical basis of black hole entropy, which depends only on the black hole's internal state and is independent of the external environment, including the presence and location of any external matter. Were these black holes at equilibrium, the constancy of the entropy in Eq. (4.22) would imply a vanishing equilibrium entropic force.

There have been previous attempts [41, 57] at computing an entropic force due to the mutual tidal disturbance between black holes, but they have invariably expressed the entropy in terms of the mass parameters μ_i; which as already noted include the interaction energy. These attempts have incorrectly found a non-zero equilibrium entropic force which roughly mimics the Newtonian force law. However, as shown above such an equilibrium force does not exist.

4.2.6 Hawking temperature for binary black holes

Since the surface gravity κ and the temperature along the horizon satisfy the relation [66]

$$T(\theta,\phi) \equiv \frac{\hbar\, \kappa(\theta,\phi)}{2\pi k_B}, \tag{4.23}$$

we may obtain the temperature by calculating the surface gravity of each black hole. From Eq. (3.91), we know that the surface gravity for a quasi-static non-rotating spacetime is given by

$$\kappa(\theta,\phi) \equiv \xi_{\mu;\beta}\hat{T}^{\beta}\hat{N}^{\mu}, \tag{4.24}$$

where $\xi^{\mu} = \mathcal{N}\hat{T}^{\mu}$.

Then the resulting temperature, T_1, for the first black hole is [66]

$$
\begin{aligned}
&T_1(\theta) = \\
&\frac{\hbar}{8\pi G m_1 k_B} - \frac{\hbar m_2}{16\pi m_1 r_{12} k_B} + \frac{\hbar G(-3m_1 m_2\cos\theta + m_1 m_2 + m_2^2)}{32\pi m_1 r_{12}^2 k_B} \\
&+\frac{\hbar G^2}{64\pi m_1 r_{12}^3 k_B}\Big(3m_1^2 m_2\cos\theta - 3m_1^2 m_2\cos(2\theta) - 2m_1^2 m_2 + 12 m_1 m_2^2\cos\theta - 3m_1 m_2^2 - m_2^3\Big) \\
&+\frac{\hbar G^3}{1024\pi m_1 r_{12}^4 k_B}\Big(-39 m_1^3 m_2\cos\theta + 24 m_1^3 m_2\cos(2\theta) - 25 m_1^3 m_2\cos(3\theta) \\
&+16 m_1^3 m_2 - 216 m_1^2 m_2^2\cos\theta + 178 m_1^2 m_2^2\cos(2\theta) + 134 m_1^2 m_2^2 - 240 m_1 m_2^3\cos\theta \\
&+48 m_1 m_2^3 + 8m_2^4 - 16 m_1^2 m_2^2\sin^2(\theta)\Big) \\
&+\frac{\hbar G^4}{57344\pi m_1 r_{12}^5 k_B}\Big(1092 m_1^4 m_2\cos\theta - 1092 m_1^4 m_2\cos(2\theta) + 700 m_1^4 m_2\cos(3\theta) \\
&-735 m_1^4 m_2\cos(4\theta) - 637 m_1^4 m_2 + 16512 m_1^3 m_2^2\cos\theta - 11088 m_1^3 m_2^2\cos(2\theta) \\
&+8096 m_1^3 m_2^2\cos(3\theta) - 6832 m_1^3 m_2^2 + 26208 m_1^2 m_2^3\cos\theta - 22344 m_1^2 m_2^3\cos(2\theta) \\
&-15064 m_1^2 m_2^3 + 13440 m_1 m_2^4\cos\theta - 2240 m_1 m_2^4 - 224 m_2^5\Big) + O(G^5).
\end{aligned}
\tag{4.25}
$$

with a similar result for T_2 under the replacement $m_1 \leftrightarrow m_2$.

4.3 Non-equilibrium entropic force for black holes

4.3.1 Non-equilibrium entropic force due to non-uniform temperature

The non-uniform temperature of Eq. (4.25) suggests the possibility for a non-equilibrium entropic force. To test whether the non-uniform temperature of Eq. (4.25) gives rise to a non-equilibrium entropic force in the quasi-static setting of this example, we explicitly calculate this force on the i^{th} black hole using a continuum version of $\sum_j T_j dS_j/dr_i^{\text{phys}}$ via

$$
\begin{aligned}
\overrightarrow{F}_i^{\text{phys}} &= \sum_{j=1}^{2}\int T_j(\theta_j,\phi_j)\,\frac{\partial}{\partial \overrightarrow{r}_i^{\text{phys}}}\left(\frac{k_B\, d\mathcal{A}_j(\theta_j,\phi_j)}{4\hbar G}\right) \\
&= \sum_{j=1}^{2}\int T_j(\theta_j,\phi_j)\,\frac{d\overrightarrow{r}_i}{d\overrightarrow{r}_i^{\text{phys}}}\cdot\frac{\partial}{\partial \overrightarrow{r}_i}\left(\frac{k_B\, d\mathcal{A}_j(\theta_j,\phi_j)}{4\hbar G}\right),
\end{aligned}
\tag{4.26}
$$

where $d\mathcal{A}_j(\theta_j,\phi_j)$ is the area element of the weak future horizon for the j^{th} black hole given explicitly in Eq. (4.20), and r_i^{phys} is the physical distance which is generally different from the parameter r_i in the metric.

As the effect is strongest on the side of the black hole horizons nearest each other, the most relevant physical distance between two black holes should be the physical distance between the near points of the two horizons, see Fig. 4.2. This shows the transformation between physical distance and parametric distance is independent of the angle, hence can be pulled out of the integral. Further, since Eq. (4.26) only depends on $\overrightarrow{r}_i$ via the quantity r_{12}, the entropic force simplifies to

$$\begin{aligned}\overrightarrow{F}_i^{\text{phys}} &= \frac{dr_{12}}{dr_{12}^{\text{phys}}}\sum_{j=1}^{2}\int T_j(\theta_j,\phi_j)\,\frac{\partial}{\partial \overrightarrow{r}_i}\left(\frac{k_B\,d\mathcal{A}_j(\theta_j,\phi_j)}{4\hbar G}\right)\\ &= \frac{dr_{12}}{dr_{12}^{\text{phys}}}\overrightarrow{F}_i^{\text{parametric}},\end{aligned} \tag{4.27}$$

where $\overrightarrow{F}_i^{\text{parametric}}$ is a 'parametric' entropic force, with regard to changes in coordinate r_i.

Here we first calculate the parametric entropic force. Inserting the area element and the non-uniform temperature, the lowest orders of the parametric entropic force for the first black hole are

$$\begin{aligned}\overrightarrow{F}_1^{\text{parametric}} =& \frac{G^4(m_1^3m_2^2+m_1^2m_2^3)}{8r_{12}^5}\hat{r}_{12} + \frac{G^5(-5m_1^4m_2^2-34m_1^3m_2^3-5m_1^2m_2^4)}{32r_{12}^6}\hat{r}_{12}\\ &+\frac{3G^6(27m_1^5m_2^2+343m_1^4m_2^3+343m_1^3m_2^4+27m_1^2m_2^5)}{448r_{12}^7}\hat{r}_{12}\\ &+\frac{G^7\left(-175m_1^6m_2^2-4294m_1^5m_2^3-7875m_1^4m_2^4-4294m_1^3m_2^5-175m_1^2m_2^6\right)}{1120r_{12}^8}\hat{r}_{12}\\ &+O(G^8),\end{aligned} \tag{4.28}$$

with a similar result for $\overrightarrow{F}_2^{\text{parametric}}$ by replacing $m_1 \leftrightarrow m_2$.

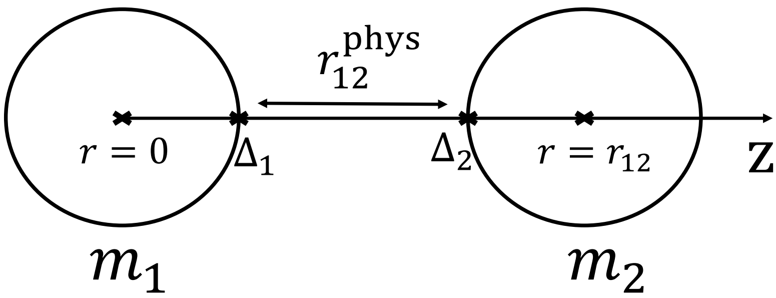

Figure 4.2: Here we take the 'location' variable r_1 of black hole 1 as 0 and r_{12}^{phys} represents the physical distance between the near points of horizons for binary black holes. Note that $r = \Delta_1$ is the horizon radius of black hole 1 and Δ_2 is the analogous horizon radius of black hole 2.

Next we calculate the transformation parameter $dr_{12}/dr_{12}^{\text{phys}}$. To do this, we first express the physical distance r_{12}^{phys} in terms of r_{12}. The physical distance between the near points of the binary black hole horizons may be calculated by

$$
\begin{aligned}
r_{12}^{\text{phys}} &= \int_{\Delta_1}^{r_{12}-\Delta_2} \psi^2 dr \\
&= \int_{\Delta_1}^{r_{12}-\Delta_2} \left(1+\frac{G\mu_1}{2r}+\frac{G\mu_2}{2\sqrt{r^2-2r\,r_{12}\cos\theta+r_{12}^2}}\right)^2 dr \\
&= \int_{\Delta_1}^{r_{12}-\Delta_2} \left(1+\frac{G\mu_1}{2r}+\frac{G\mu_2}{2\sqrt{r^2-2r\,r_{12}+r_{12}^2}}\right)^2 dr \\
&= \int_{\Delta_1}^{r_{12}-\Delta_2} \left(1+\frac{G\mu_1}{2r}+\frac{G\mu_2}{2(r_{12}-r)}\right)^2 dr \\
&= \int_{\Delta_1}^{r_{12}-\Delta_2} \left(1+\frac{G\mu_1}{r}+\frac{G\mu_2}{r_{12}-r}+\frac{G^2\mu_1^2}{4r^2}+\frac{G^2\mu_1\mu_2}{2r(r_{12}-r)}+\frac{G^2\mu_2^2}{4(r_{12}-r)^2}\right) dr \\
&= \left(r+G\mu_1\ln r-G\mu_2\ln(r_{12}-r)-\frac{G^2\mu_1^2}{4r}+\frac{G^2\mu_1\mu_2}{2r_{12}}(\ln r-\ln(r_{12}-r))+\frac{G^2\mu_2^2}{4(r_{12}-r)}\right)\Bigg|_{\Delta_1}^{r_{12}-\Delta_2} \\
&= \Bigg((r_{12}-\Delta_2-\Delta_1)+G\mu_1\left(\ln(r_{12}-\Delta_2)-\ln\Delta_1\right)-G\mu_2\ln\Delta_2+G\mu_2\ln(r_{12}-\Delta_1) \\
&\quad -\frac{G^2\mu_1^2}{4(r_{12}-\Delta_2)}+\frac{G^2\mu_1^2}{4\Delta_1}+\frac{G^2\mu_1\mu_2}{2r_{12}}(\ln(r_{12}-\Delta_2)-\ln\Delta_2) \\
&\quad -\frac{G^2\mu_1\mu_2}{2r_{12}}(\ln\Delta_1-\ln(r_{12}-\Delta_1))+\frac{G^2\mu_2^2}{4\Delta_2}-\frac{G^2\mu_2^2}{4(r_{12}-\Delta_1)}\Bigg),
\end{aligned}
\tag{4.29}
$$

where we have used $\theta = 0$ in the third line. As we discussed before, here Δ_1, Δ_2, μ_1 and μ_2 are also functions of r_{12}.

From the result of Eq. (4.29), the variation of r_{12}^{phys} with r_{12} may be calculated by

$$
\begin{aligned}
\frac{dr_{12}^{\text{phys}}}{dr_{12}} =& 1-\frac{d\Delta_2}{dr_{12}}-\frac{d\Delta_1}{dr_{12}}+G\frac{d\mu_1}{dr_{12}}(\ln(r_{12}-\Delta_2)-\ln\Delta_1) \\
&+G\mu_1\left(\frac{1}{r_{12}-\Delta_2}(1-\frac{d\Delta_2}{dr_{12}})-\frac{1}{\Delta_1}\frac{d\Delta_1}{dr_{12}}\right) \\
&-G\frac{d\mu_2}{dr_{12}}(\ln\Delta_2-\ln(r_{12}-\Delta_1))-G\mu_2\left(\frac{1}{\Delta_2}\frac{d\Delta_2}{dr_{12}}-\frac{1}{r_{12}-\Delta_1}(1-\frac{d\Delta_1}{dr_{12}})\right) \\
&-\frac{G^2\mu_1}{2(r_{12}-\Delta_2)}\frac{d\mu_1}{dr_{12}}+\frac{G^2\mu_1^2}{4(r_{12}-\Delta_2)^2}(1-\frac{d\Delta_2}{dr_{12}})+\frac{G^2\mu_1}{2\Delta_1}\frac{d\mu_1}{dr_{12}}-\frac{G^2\mu_1^2}{4\Delta_1^2}\frac{d\Delta_1}{dr_{12}} \\
&+\frac{G^2}{2}\left(\frac{\mu_2}{r_{12}}\frac{d\mu_1}{dr_{12}}+\frac{\mu_1}{r_{12}}\frac{d\mu_2}{dr_{12}}-\frac{\mu_1\mu_2}{r_{12}^2}\right)(\ln(r_{12}-\Delta_2)-\ln\Delta_2) \\
&+\frac{G^2\mu_1\mu_2}{2r_{12}}\left(\frac{1}{r_{12}-\Delta_2}(1-\frac{d\Delta_2}{dr_{12}})-\frac{1}{\Delta_2}\frac{d\Delta_2}{dr_{12}}\right) \\
&-\frac{G^2}{2}\left(\frac{\mu_2}{r_{12}}\frac{d\mu_1}{dr_{12}}+\frac{\mu_1}{r_{12}}\frac{d\mu_2}{dr_{12}}-\frac{\mu_1\mu_2}{r_{12}^2}\right)(\ln\Delta_1-\ln(r_{12}-\Delta_1)) \\
&-\frac{G^2\mu_1\mu_2}{2r_{12}}\left(\frac{1}{\Delta_1}\frac{d\Delta_1}{dr_{12}}-\frac{1}{r_{12}-\Delta_1}(1-\frac{d\Delta_1}{dr_{12}})\right) \\
&+\frac{G^2\mu_2}{2\Delta_2}\frac{d\mu_2}{dr_{12}}-\frac{G^2\mu_2^2}{4\Delta_2^2}\frac{d\Delta_2}{dr_{12}}-\frac{G^2\mu_2}{2(r_{12}-\Delta_1)}\frac{d\mu_2}{dr_{12}}+\frac{G^2\mu_2^2}{4(r_{12}-\Delta_1)^2}(1-\frac{d\Delta_1}{dr_{12}}).
\end{aligned}
\tag{4.30}
$$

Since μ_1 and μ_2 are the mass parameters and the dependence of μ_1 and μ_2 on r_{12} comes from the gravitational potential between these two black holes, one has $\frac{d\mu_1}{dr_{12}} = \frac{d\mu_2}{dr_{12}}$. Therefore, Eq. (4.30) may be further simplified to

$$\begin{aligned}
\frac{dr_{12}^{\text{phys}}}{dr_{12}} =& 1 - \frac{d\Delta_2}{dr_{12}} - \frac{d\Delta_1}{dr_{12}} + G\frac{d\mu_1}{dr_{12}}(\ln\frac{(r_{12}-\Delta_1)(r_{12}-\Delta_2)}{\Delta_1\Delta_2}) \\
&+G\mu_1(\frac{1}{r_{12}-\Delta_2}(1-\frac{d\Delta_2}{dr_{12}}) - \frac{1}{\Delta_1}\frac{d\Delta_1}{dr_{12}}) - G\mu_2(\frac{1}{\Delta_2}\frac{d\Delta_2}{dr_{12}} - \frac{1}{r_{12}-\Delta_1}(1-\frac{d\Delta_1}{dr_{12}})) \\
&+\frac{G^2\mu_1}{2}\frac{d\mu_1}{dr_{12}}\frac{r_{12}-\Delta_2-\Delta_1}{\Delta_1(r_{12}-\Delta_2)} + \frac{G^2\mu_1^2}{4(r_{12}-\Delta_2)^2}(1-\frac{d\Delta_2}{dr_{12}}) - \frac{G^2\mu_1^2}{4\Delta_1^2}\frac{d\Delta_1}{dr_{12}} \\
&+\frac{G^2}{2}(\frac{\mu_1+\mu_2}{r_{12}}\frac{d\mu_1}{dr_{12}} - \frac{\mu_1\mu_2}{r_{12}^2})(\ln\frac{(r_{12}-\Delta_1)(r_{12}-\Delta_2)}{\Delta_1\Delta_2}) \\
&+\frac{G^2\mu_1\mu_2}{2r_{12}}(\frac{1}{r_{12}-\Delta_2}(1-\frac{d\Delta_2}{dr_{12}}) - \frac{1}{\Delta_2}\frac{d\Delta_2}{dr_{12}}) - \frac{G^2\mu_1\mu_2}{2r_{12}}(\frac{1}{\Delta_1}\frac{d\Delta_1}{dr_{12}} - \frac{1}{r_{12}-\Delta_1}(1-\frac{d\Delta_1}{dr_{12}})) \\
&+\frac{G^2\mu_2}{2}\frac{d\mu_1}{dr_{12}}\frac{r_{12}-\Delta_1-\Delta_2}{\Delta_2(r_{12}-\Delta_1)} - \frac{G^2\mu_2^2}{4\Delta_2^2}\frac{d\Delta_2}{dr_{12}} + \frac{G^2\mu_2^2}{4(r_{12}-\Delta_1)^2}(1-\frac{d\Delta_1}{dr_{12}}).
\end{aligned} \tag{4.31}$$

Or equivalently,

$$\begin{aligned}
\frac{dr_{12}^{\text{phys}}}{dr_{12}} =& 1 - \frac{d\Delta_2}{dr_{12}} - \frac{d\Delta_1}{dr_{12}} + \Big(G\frac{d\mu_1}{dr_{12}} + \frac{G^2}{2}(\frac{\mu_1+\mu_2}{r_{12}}\frac{d\mu_1}{dr_{12}} - \frac{\mu_1\mu_2}{r_{12}^2})\Big)(\ln\frac{(r_{12}-\Delta_1)(r_{12}-\Delta_2)}{\Delta_1\Delta_2}) \\
&+G\mu_1(\frac{1}{r_{12}-\Delta_2}(1-\frac{d\Delta_2}{dr_{12}}) - \frac{1}{\Delta_1}\frac{d\Delta_1}{dr_{12}}) - G\mu_2(\frac{1}{\Delta_2}\frac{d\Delta_2}{dr_{12}} - \frac{1}{r_{12}-\Delta_1}(1-\frac{d\Delta_1}{dr_{12}})) \\
&+\frac{G^2\mu_1}{2}\frac{d\mu_1}{dr_{12}}\frac{r_{12}-\Delta_2-\Delta_1}{\Delta_1(r_{12}-\Delta_2)} + \frac{G^2\mu_1^2}{4(r_{12}-\Delta_2)^2}(1-\frac{d\Delta_2}{dr_{12}}) - \frac{G^2\mu_1^2}{4\Delta_1^2}\frac{d\Delta_1}{dr_{12}} \\
&+\frac{G^2\mu_1\mu_2}{2r_{12}}(\frac{1}{r_{12}-\Delta_2}(1-\frac{d\Delta_2}{dr_{12}}) - \frac{1}{\Delta_2}\frac{d\Delta_2}{dr_{12}}) - \frac{G^2\mu_1\mu_2}{2r_{12}}(\frac{1}{\Delta_1}\frac{d\Delta_1}{dr_{12}} - \frac{1}{r_{12}-\Delta_1}(1-\frac{d\Delta_1}{dr_{12}})) \\
&+\frac{G^2\mu_2}{2}\frac{d\mu_1}{dr_{12}}\frac{r_{12}-\Delta_1-\Delta_2}{\Delta_2(r_{12}-\Delta_1)} - \frac{G^2\mu_2^2}{4\Delta_2^2}\frac{d\Delta_2}{dr_{12}} + \frac{G^2\mu_2^2}{4(r_{12}-\Delta_1)^2}(1-\frac{d\Delta_1}{dr_{12}}).
\end{aligned} \tag{4.32}$$

Inserting all the relevant expressions into Eq. (4.32), we find

$$\begin{aligned}
\frac{dr_{12}^{\text{phys}}}{dr_{12}} =& 1 + \frac{(m_1+m_2)G}{r_{12}} + \frac{(m_1^2 - 8m_2m_1 + m_2^2)\,G^2}{4r_{12}^2} + \frac{3\,(m_2m_1^2 + m_2^2m_1)\,G^3}{4r_{12}^3} \\
&- \frac{(37m_2m_1^3 + 42m_2^2m_1^2 + 37m_2^3m_1)\,G^4}{56r_{12}^4} \\
&+ \frac{(401m_2m_1^4 + 2210m_2^2m_1^3 + 2210m_2^3m_1^2 + 401m_2^4m_1)\,G^5}{1456r_{12}^5} + O(G^6) \\
&+ A\ln\frac{B}{C},
\end{aligned} \tag{4.33}$$

where A, B and C are each functions of m_1, m_2, r_{12}, and G. Here

$$A=-\frac{m_1m_2\,G^7}{2048r_{12}^{12}}\Bigg(11m_1m_2\Big(m_1^4+10m_2m_1^3+20m_2^2m_1^2+10m_2^3m_1+m_2^4\Big)^2G^5$$
$$40m_1m_2\Big(m_1^7+16m_2m_1^6+86m_2^2m_1^5+191m_2^3m_1^4+191m_2^4m_1^3+86m_2^5m_1^2$$
$$+16m_2^6m_1+m_2^7\Big)r_{12}G^4+36m_1m_2\Big(3m_1^6+38m_2m_1^5+150m_2^2m_1^4+234m_2^3m_1^3$$
$$+150m_2^4m_1^2+38m_2^5m_1+3m_2^6\Big)r_{12}^2G^3-128m_1m_2\Big(2m_1^5+20m_2m_1^4+55m_2^2m_1^3$$
$$+55m_2^3m_1^2+20m_2^4m_1+2m_2^5\Big)r_{12}^3G^2+560m_1m_2\Big(m_1^4+8m_2m_1^3+15m_2^2m_1^2$$
$$+8m_2^3m_1+m_2^4\Big)r_{12}^4G-192\Big(m_1^5+15m_2m_1^4+50m_2^2m_1^3+50m_2^3m_1^2+15m_2^4m_1$$
$$+m_2^5\Big)r_{12}^5\Bigg), \tag{4.34}$$

$$B=\Bigg(m_1m_2\Big(107m_1^3+975m_2m_1^2+455m_2^2m_1+455m_2^3\Big)G^5-26m_1m_2\Big(12m_1^2$$
$$+21m_2m_1+28m_2^2\Big)r_{12}G^4+364m_1m_2(m_1+3m_2)r_{12}^2G^3-1456m_1m_2r_{12}^3G^2$$
$$+1456m_1r_{12}^4G-2912r_{12}^5\Bigg)\Bigg(m_1m_2\Big(455m_1^3+455m_2m_1^2+975m_2^2m_1+107m_2^3\Big)G^5$$
$$-26m_1m_2\left(28m_1^2+21m_2m_1+12m_2^2\right)r_{12}G^4+364m_1m_2(3m_1+m_2)r_{12}^2G^3$$
$$-1456m_1m_2r_{12}^3G^2+1456m_2r_{12}^4G-2912r_{12}^5\Bigg), \tag{4.35}$$

and

$$C=\Bigg(G^2m_1m_2\Big(m_1\Big(455m_1^3+455m_2m_1^2+975m_2^2m_1+107m_2^3\Big)G^4-26m_1\Big(28m_1^2$$
$$+21m_2m_1+12m_2^2\Big)r_{12}G^3+364m_1(3m_1+m_2)r_{12}^2G^2-1456m_1r_{12}^3G+1456r_{12}^4\Big)$$
$$\Big(m_2\left(107m_1^3+975m_2m_1^2+455m_2^2m_1+455m_2^3\right)G^4-26m_2\Big(12m_1^2+21m_2m_1$$
$$+28m_2^2\Big)r_{12}G^3+364m_2(m_1+3m_2)r_{12}^2G^2-1456m_2r_{12}^3G+1456r_{12}^4\Big)\Bigg). \tag{4.36}$$

Since $r_{12}^{\rm phys}$ and r_{12} are in one-to-one correspondence, we have

$$\frac{dr_{12}}{dr_{12}^{\rm phys}} = \frac{1}{\frac{dr_{12}^{\rm phys}}{dr_{12}}}. \tag{4.37}$$

For simplicity, we will write $\alpha \equiv \frac{dr_{12}}{dr_{12}^{\rm phys}}$ to represent this distance transformation connection from now on. Numerical calculations show that α is always positive outside the black hole horizon and approaches 1 as r_{12} approaches ∞. To explicitly show how α depends on r_{12}, we plot α versus r_{12} when $m_1 = m_2 = m$ in Fig.4.3. In this case, $r_{12} \approx 1.0775\,G\,m$ is the 'kissing distance', which is the closest the black holes can

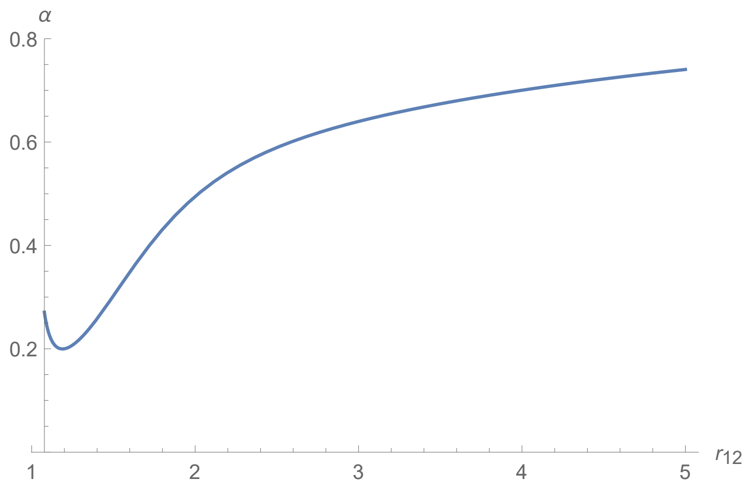

Figure 4.3: This figure shows how α changes with r_{12}, where $\alpha \equiv \frac{dr_{12}}{dr_{12}^{\rm phys}}$, for $m_1 = m_2 = 1$, $G = 1$.

approach in isotropic coordinates while remaining separate, so the figure begins from $r_{12} = 1.0775\, G\, m$.

Based on the above analysis and Eq. (4.28), the entropic force may be written as

$$\begin{aligned}\overrightarrow{F}_1^{\rm phys} = \alpha\Big(&\frac{G^4(m_1^3m_2^2+m_1^2m_2^3)}{8r_{12}^5}+\frac{G^5(-5m_1^4m_2^2-34m_1^3m_2^3-5m_1^2m_2^4)}{32r_{12}^6}\\ &+\frac{3G^6(27m_1^5m_2^2+343m_1^4m_2^3+343m_1^3m_2^4+27m_1^2m_2^5)}{448r_{12}^7}\\ &+\frac{G^7\left(-175m_1^6m_2^2-4294m_1^5m_2^3-7875m_1^4m_2^4-4294m_1^3m_2^5-175m_1^2m_2^6\right)}{1120r_{12}^8}\Big)\hat{r}_{12}\\ &+O(G^8),\end{aligned} \tag{4.38}$$

We first focus on the leading term of the entropic force to do some qualitative analysis. From Eq. (4.38), the leading term is

$$\overrightarrow{F}_1^{\rm phys} = \frac{\alpha\, G^4(m_1^3m_2^2+m_1^2m_2^3)}{8r_{12}^5}\hat{r}_{12}+O(G^5), \tag{4.39}$$

Since $\alpha > 0$, the entropic force in Eq. (4.39) is repulsive. This is due to the distribution of temperatures, with the higher temperatures on the far sides of the neighboring black holes, and is consistent with the intuition gained from the behavior of the molecular chain in Fig. 4.1(b).

The form of Eq. (4.39) is intriguing, especially if we reparameterize it in terms of the mass ratio $\rho = m_2/m_1$ and define the approximate 'Kissing distance' $r_{\rm Kiss} = G(m_1+m_2)/2$. Now the magnitude of the entropic force of Eq. (4.39) reduces to

$$F^{\rm phys} = +\frac{4\,\alpha}{G}\left(\frac{\rho}{(1+\rho)^2}\right)^2\left(\frac{r_{\rm Kiss}}{r_{12}}\right)^5+O(G^5), \tag{4.40}$$

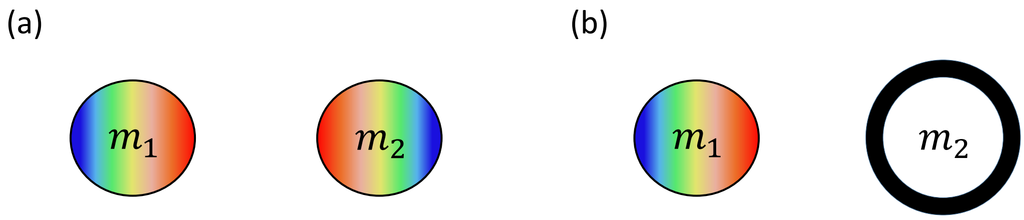

Figure 4.4: Non-equilibrium entropic force between: (a) A pair of black holes. (b) An identical black hole (m_1) and a shell of matter with the same mass as black hole m_2 in (a). Both scenarios give rise to repulsive non-equilibrium entropic forces. The temperature profiles are identical for the left black hole in both the (a) and (b) scenarios, but give rise to different forces. (The color provides a highly exaggerated rendition of the temperature profile along the black holes' horizons: from red to blue denotes a transition from lower to higher temperatures.)

revealing that the acceleration due to the entropic force:

i) vanishes in the test-particle limit, where $\rho \to 0$;

ii) vanishes in the Rindler limit, where $\rho \to \infty$; and

iii) conserves total momentum, since $\sum_i \overrightarrow{F}_i^{\,\text{phys}} = 0$.

If we use the same parameterization for the *form* of the Newtonian force law

$$F^{\,\text{Newton}} = -\frac{Gm_1m_2}{r_{12}^2} = -\frac{4}{G}\frac{\rho}{(1+\rho)^2}\left(\frac{r_{\text{Kiss}}}{r_{12}}\right)^2, \tag{4.41}$$

we see that the entropic force is remarkably short range and relatively weak. For the case that $m_1 = m_2$, the magnitude of the entropic force is at most about 2.03% of the Newtonian form at the 'Kissing point', decaying rapidly at larger separations. Note that this value is obtained by a continued fraction approximation of the power series in Eq. (4.38), where continued fraction approximations are a special case of Padé approximations [15].

The remote contribution (from black hole 2) to the force described by Eq. (4.26) suggests that this force cannot be encoded in the Einstein field equations. To pin down this conjecture we construct a modified scenario in section 4.3.2, in which the entropic force results from contributions of a single black hole. In this modified scenario, the second black hole is replaced with a shell of matter, designed so that the spacetime metric everywhere outside the shell is identical to that of the original scenario, and consequently the temperature profile along the surface of the black hole will also be identical (see Fig. 4.4(b)). Second, by construction this shell is assumed to be in a pure quantum state. This ensures that it carries no entropy which may contribute to an entropic force. It has been shown in chapter 2 of this thesis that the only spacetime surfaces which carry thermodynamic properties are those associated with spacetime horizons [64]. The entropic force between a black hole of mass m_1 and this matter shell of mass m_2 then becomes

$$\overrightarrow{F}_1^{\text{phys}} = \frac{\alpha\, G^4 m_1^3 m_2^2}{8\, r_{12}^5}\, \hat{r}_{12} + O(G^5). \tag{4.42}$$

Despite the fact that, outside the shell, the spacetime metric is identical in both scenarios, the magnitude of the entropic force on the black hole, and hence its acceleration, is *different*. This comparison therefore proves that the spacetime metric is insufficient to predict motion due to entropic effects. We conclude that this entropic force cannot be encoded in the conventional Einstein field equations of general relativity.

4.3.2 Replacing black holes by ordinary matter

In this section we consider more carefully the claim that the entropic force is not already implicitly present in the solutions to the Einstein field equations. Our strategy will be to show that we can obtain the same solutions to the field equations, but in a scenario where there are no horizons and hence no entropic forces. We shall achieve this by replacing our black holes by shells of ordinary matter, which will be assumed to have zero or negligible regular thermodynamic entropy.

To obtain such solutions, we must write down the full Einstein field equations, including the presence of matter. In the $3+1$ formalism [30], assuming $\beta_i = 0$, we obtain

$$\begin{aligned} {}^{(3)}\nabla^2\psi &= -2\pi G E \psi^5 \\ {}^{(3)}\nabla^2\chi &= +2\pi G (E + 2F)\psi^4\chi, \end{aligned} \tag{4.43}$$

on the initial hypersurface, where the energy-momentum variables are defined $E \equiv T_{00}$ and $F \equiv \gamma^{ij} T_{ij}$.

For a single black hole spacetime with mass M, the black hole and its entropy carrying horizon may be replaced with a thin shell of ordinary matter. Solutions to Eqs. (4.43) are readily obtained as

$$1 - \chi = \psi - 1 = \begin{cases} \dfrac{GM}{2\,r}, & r > R \\ \dfrac{GM}{2\,R}, & r \le R, \end{cases} \tag{4.44}$$

where R is the shell radius and substituting these expressions into Eqs. (4.43) yields

$$\begin{aligned} E &= \frac{M\,\delta(r-R)}{4\pi R^2(1+\frac{GM}{2R})^5} \\ F &= \frac{\psi-\chi}{2\,\chi}\,E = \frac{GM}{2R\,(1-\frac{GM}{2R})}\,E. \end{aligned} \tag{4.45}$$

These are both positive (and finite) when the shell lies outside the original Schwarzschild horizon, $R > GM/2$.

Now consider the solutions involving n black holes replaced by shells of ordinary matter. Based on the result above for a single black hole this is straightforward to generalize to this case yielding

$$\begin{aligned} &1-\chi(\vec{x}) = \psi(\vec{x}) - 1 \\ &= \begin{cases} \displaystyle\sum_{j=1}^{n} \frac{G\mu_j}{2\,|\vec{x}-\vec{r}_j|}, & \forall i,\ |\vec{x}-\vec{r}_i| > R_i \\ \displaystyle\frac{G\mu_i}{2R_i} + \sum_{j\neq i}^{n} \frac{G\mu_j}{2\,|\vec{x}-\vec{r}_j|}, & |\vec{x}-\vec{r}_i| \leq R_i, \end{cases} \end{aligned} \tag{4.46}$$

These describe n shells centered at $\vec{x} = \vec{r}_i$ with respective radii R_i. For the $i^{\rm th}$ shell

$$\begin{aligned} E_i(\vec{x}) &= \frac{\mu_i\,\delta(|\vec{x}-\vec{r}_i|-R_i)}{4\pi R_i^2\,\psi(\vec{x})^5} \\ F_i(\vec{x}) &= \frac{\psi(\vec{x})-\chi(\vec{x})}{2\,\chi(\vec{x})}\,E_i(\vec{x}). \end{aligned} \tag{4.47}$$

Again, these are finite and non-negative provided the shells lie outside the location of the future horizon for the black holes they are respectively replacing. (We note, that one can trivially generalize these solutions to involve b black holes and $n-b$ shells of ordinary matter.)

Thus replacing ψ and χ in Eq. (4.7) by the corresponding terms of Eqs. (4.46) and (4.47) yields the exact same solutions to the Einstein field equations in the entire region of the initial hypersurface outside the radii corresponding to the solutions with the shells. The former solutions have black holes and their horizon entropy, the latter have only the entropy associated with the ordinary matter of the shells, which without loss of generality we may take to be zero or negligible. Consequently, the entropic forces present in the former solutions are absent in the latter. We conclude from this that any entropic forces are *not* encoded in the conventional Einstein field equations.

4.3.3 Influence on the gravitational wave form

Our calculations of this non-equilibrium entropic force are exact to the order computed and cannot be canceled by higher-order contributions. Although the entropic force is short range and comparatively weak, it would still have an in-principle observable effect on black hole collision processes. For two merging black holes with similar mass, our following rough calculations show that the effect of this force on gravitational waves is a roughly 1.02% frequency shift in the peak chirp frequency before final coalescence. Finally, it would not be surprising if the underlying phenomenon might be enhanced or diminished in more general scenarios than those considered here.

The gravitational wave emission by a nearly Newtonian binary star system may be roughly approximated by the Newtonian force law [40]. Label the two gravitating bodies by m_1 and m_2, their total mass and reduced mass are

$$M = m_1 + m_2, \quad m = \frac{m_1 m_2}{M}. \tag{4.48}$$

If we assume their orbit is circular with distance between the star's centers r_{12}, then the orbital angular velocity ω is given by Kepler's law [53]

$$\omega = \sqrt{\frac{M}{r_{12}^3}}. \tag{4.49}$$

Note that the gravitational wave's angular frequency is 2ω.

To estimate the effect of our non-equilibrium entropic force F^{phys} on the wave's angular frequency, we use Newton's second law $\overrightarrow{F} = m\overrightarrow{a}$ to roughly calculate its effect on the angular velocity of the gravitating body:

$$\frac{Mm}{r_{12}^2} - F^{\text{phys}} = {\omega'}^2 r_{12}\, m, \tag{4.50}$$

which may be simplified as

$$\begin{aligned} \omega' &= \sqrt{\frac{M}{r_{12}^3} - \frac{F^{\text{phys}}}{r_{12}\, m}} \\ &= \sqrt{\frac{M}{r_{12}^3}}\sqrt{1 - \frac{r_{12}^2 F^{\text{phys}}}{M\, m}}. \end{aligned}$$

For simplicity, we consider two black holes with equal mass at the 'Kissing distance' ($r_{12} \approx \frac{1.0775M}{2}$). From the discussion following Eq. (4.41), we know $F^{\text{phys}} \approx 0.0203\, F^{\text{Newton}} = 0.0203\frac{Mm}{r_{12}^2}$ for this case. Then ω' reduces to

$$\omega' = \sqrt{\frac{M}{r_{12}^3}}\sqrt{1 - 0.0203}.$$

For this case, the non-equilibrium entropic force would cause the frequency of the gravitational waves to shift by roughly 1.02%. Since this effect decreases rapidly as r_{12} increases, it may only have an observable effect on the very final stage of black hole coalescence.

4.4 Discussion

Entropic forces are generically studied in equilibrium settings. Here we have described for the first time an interesting variant exists for non-equilibrium environments and which vanishes at equilibrium. We called such variants non-equilibrium entropic forces.

Our canonical example consisted of a non-equilibrium entropic force driving a long-chain molecule through a nanopore from a cooler to a warmer thermal bath. The temperature difference between the baths allows one to sensitively control the rate at which the molecule is driven through the nanopore. The non-equilibrium entropic force in this scenario could be generalized to achieve the simultaneous controlled passage of many molecular strands through parallel nanopores in a similar two-chamber setup, with a variety of plausible applications. For example, such a highly parallel architecture might be combined with electrically sensing an amino acid's identity [58] at each nanopore exit for high-throughput label-free amplification-free sequencing of biological macromolecules. A key advantage of this approach is its power to overcome the bottleneck associated with attaining sufficiently slow motion of a molecule through the nanopore for accurate sensing [58].

In the interacting black hole setting, we found that the potential for observation of a non-equilibrium entropic force would allow for confirmation of the long-standing conjecture about the thermodynamic nature of spacetime horizons; not just in a laboratory setting with analogue black holes, but with the actual astrophysical objects themselves.

In string theories of gravity, one convincingly finds that black hole horizons are truly thermodynamic. One can pass from a thermal state to a black hole by smoothly changing the string coupling [34]. Indeed, for supersymmetric and near-extremal string-theory black holes one can even obtain the exact conventional area law [20, 35, 51]. Thus, by construction, string theories of gravity imply the existence of the entropic force which we have demonstrated here between black holes and any other gravitating body. Observation of this entropic force would not only provide further support for string theories, but would also constitute the first signature of quantum gravity – beyond conventional general relativity. Conversely, convincing evidence for an absence of such a force would not only imply a failure for the full thermodynamic analogy of black holes and spacetime horizons, but would also convincingly refute one of string theories' few emphatic predictions.

Chapter 5

Conclusions

5.1 Summary of contributions

In Chapter 2, to test the thermal assumptions in Verlinde's emergent gravity program [59], we extended the analysis of the first law of black hole mechanics beyond the horizon to generic surfaces for static spacetimes [59, 64]:

1. We achieved this by generalizing the tools from Ref. [11] beyond horizons to arbitrary surfaces.

2. These new tools allowed us to obtain an integral Komar mass expression for stationary spacetimes with an arbitrary inner boundary. This integral formula naturally defines a 'surface gravity' for ordinary surfaces for static spacetimes.

3. The stretched horizons proposed by Kip Thorne are surfaces microscopically outside a horizon [54]. Our analysis shows that such stretched horizons do satisfy the first law of black hole thermodynamics to high precision.

4. Since there is no fundamental mechanics to support a physical temperature for the stretched horizon, it can only inherit the thermodynamic phenomenon from an underlying black hole horizon. Thus, that the first law still holds on the stretched horizon, in fact implies that the laws of black hole thermodynamics should only be considered as a necessary condition but not a sufficient condition for true thermodynamic behavior.

5. For surfaces further away from a horizon (so-called holographic screens) in Verlinde's emergent gravity program, we find the first law fails. Thus eroding the basis of this program.

In Chapter 3, we extended the analysis of the black hole thermodynamics, especially the first law of black hole mechanics, from stationary to dynamical spacetimes [66]:

1. Theorem 3.1 provides a simple, covariant expression to compute the physical energy (the ADM mass) in a broad class of dynamical asymptotically-flat spacetimes. By dynamical spacetime, we follow the conventional definition, i.e., the absence of a timelike Killing vector (indeed, we do not assume any kind of Killing vector).

2. Theorem 3.3 rigorously extends the first law of black hole thermodynamics to the dynamical setting (with no other requirements beyond the black holes having non-rotating horizons that appear quasi-static to an observer at spatial infinity — what we dub *weakly quasi-static*).

3. We emphasize that this first law allows for the description of physical processes and requires no Killing vectors. In particular, the behavior of the matter outside the horizon is totally unconstrained. Therefore the thermodynamic behavior we find for horizons is truly ubiquitous. *That the thermodynamic behavior of horizons is not swamped by the potentially wildly "dirty" and dynamical environments we study is perhaps the biggest surprise of our work.*

4. The extended first law implies an unambiguous definition for a surface gravity for the horizons of dynamical spacetimes, which reduces to the conventional definition for stationary spacetimes.

5. We derive a general formula for the Parikh-Wilczek tunneling temperature of the Hawking radiation for non-rotating weakly quasi-static weak future horizons (our Theorem 3.2). We find that it is identical to the temperature obtained from our dynamical surface gravity. This proves that our dynamical first law is not mere analogy, but a real thermodynamic relation.

6. Using our rigorous formalism and an analytic calculation of an explicit example, we show that the temperatures associated with interacting black holes are, in general, non-uniform. Therefore the zeroth law of black hole thermodynamics generally fails in the dynamical setting.

7. We show that, in order to allow for non-uniform temperatures, the conventional area law must generalize to a local or elemental form. This provides a rigorous meaning to Bekenstein's early intuition that every element of area of a horizon represents a fundamental unit of information [13].

8. In 1993, Wald *conjectured* a "dynamical entropy" for dynamical black holes within any diffeomorphically invariant theory — which includes general relativity (Ref. [62]). Our work proves that this conjecture holds true within general relativity.

9. We show that the thermodynamic surface for dynamic horizons is not the long-studied trapping horizon, but a much simpler structural element which we dub the *weak future horizon*.

10. We prove that the underlying assumptions made by Jacobson in his derivation of general relativity from a presumed thermodynamic behavior of horizons (Ref. [37]) are themselves derivable from general relativity itself. Our work shows that the thermodynamic nature of horizons and general relativity are logically equivalent (i.e., if $A \Rightarrow B$ and $B \Rightarrow A$ then $A \equiv B$). Therefore either approach leads to identical classical theories of gravity. The same may not be true when these approaches are used to obtain a quantum theory of gravity. In particular, the canonical quantization of the vastly differing underlying variables in these two theories may lead to very distinct theories of quantum gravity.

In Chapter 4, we characterized an entirely new class of entropic forces which vanish at equilibrium, but are non-vanishing only out of equilibrium. We called such variants non-equilibrium entropic forces. This effect is discussed for both molecular chains and black holes. We found [65]:

1. For a long-chain molecule threaded through a frictionless nanopore between two chambers with different bath temperatures, a non-equilibrium entropic force exists and will drive the molecule to the chamber with a higher temperature.

2. For binary black holes in a quasi-static spacetime, the entropy of each black hole is independent of the distance between the two black holes. This result supports the intuition that black hole entropy should be independent of the exterior environments and implies that any equilibrium entropic force between black holes generally vanishes.

3. We also calculate an expansion of the non-uniform temperature for such black holes and we find that this non-equilibrium temperature induces a non-equilibrium entropic force between dynamical black holes. Our calculations of this non-equilibrium entropic force are exact to the order computed and cannot be canceled by higher-order contributions.

4. Replacing one of the black holes in a binary black hole scenario by a shell of ordinary matter, we find that despite the spacetime metric outside the shell being unchanged, the magnitude of the entropic force in this new scenario is *different* from that in the binary black holes case. This comparison therefore proves that the spacetime metric is insufficient to predict motion due to entropic effects. We conclude that this entropic force cannot be encoded in the conventional Einstein field equations of general relativity. So if it exists, it would be the first signature of quantum gravity – beyond conventional general relativity.

5. Although the non-equilibrium entropic force is short range and comparatively weak, we show that it should still have an in-principle observable effect on black hole collision processes. For binary black holes that are twice the 'Kissing radius' away from each other, this effect would be roughly 1.03% shift in the peak chirp frequency of gravitational radiation during the very final stage of coalescence.

6. String theory is the only theory that is consistent with both quantum mechanics and gravity [68] but with no current experimental support. It fundamentally supports the entropy-area law of black holes [20, 35, 51] and thus implies the existence of the entropic force between black holes and any other gravitating body. Therefore, the non-equilibrium entropic force proposed in this chapter also provides a fresh way to test the string theory.

5.2 Critical analysis

In Chapter 2 we proved that the thermodynamic assumptions in Verlinde's emergent gravity program are inconsistent with Einstein's general relativity. One implication of this result is that Verlinde's emergent gravity program must reassess its thermodynamic assumptions. Because there is no one-to-one correspondence between Verlinde's emergent gravity program and general relativity, this program cannot directly inherit the achievements obtained by general relativity. However, the possibility of considering gravity as having an entropic origin is still very inspiring.

Our results in Chapter 3 also suggest that entropy should be a local property of the horizon instead of the originally believed global property of a black hole, and thus the interior of a black hole may also have an independent entropy. Since no signal from the black hole interior can escape, a temperature $T_{\text{interior}} = 0$ *with respect to an observer at spatial infinity* could be assigned to the interior of the black hole. Such an assignment means that the first law would be unchanged by the addition of the vanishing term $T_{\text{interior}}\,\delta S_{\text{interior}}$. In that case, the entropy of the entire black hole becomes

$$S_{\text{black hole}} = \frac{A}{4} + S_{\text{interior}}. \tag{5.1}$$

Our analysis does not prove this relation, but it does show its consistency with black hole thermodynamics. Of course, the presence of such an additional term would require a wholesale reappraisal of black hole information bounds and paradoxes.

In Chapter 4, our canonical example of non-equilibrium entropic force consisted of a long-chain molecule passing through a nanopore between a cooler and a warmer thermal bath. The temperature difference between the baths allows one to sensitively control the rate at which the molecule is driven through the nanopore. The non-equilibrium entropic force in this scenario could be generalized to achieve the simultaneous controlled passage of many molecular strands through parallel nanopores in a similar two-chamber setup, with a variety of plausible applications. For example, such a highly parallel architecture might be combined with electrically sensing an amino acid's identity [58] at each nanopore exit for high-throughput label-free amplification-free sequencing of biological macromolecules. A key advantage of this approach is its power to overcome the bottleneck associated with attaining sufficiently slow motion of a molecule through the nanopore for accurate sensing [58].

In string theories of gravity, one convincingly finds that black hole horizons are truly thermodynamic. One can pass from a thermal state to a black hole by smoothly changing the string coupling [34]. Indeed, for supersymmetric and near-extremal string-theory black holes one can even obtain the exact conventional area law [20, 35, 51]. Thus, by construction, string theories of gravity imply the existence of the entropic force which we have demonstrated in Chapter 4 between black holes and any other gravitating body. Observation of this entropic force would not only provide further support for string theories, but would also constitute the first signature of quantum gravity – beyond conventional general relativity. Conversely, convincing evidence for an absence of such a force would not only imply a failure for the full thermodynamic analogy of black holes and spacetime horizons, but would also convincingly refute one of string theories' few emphatic predictions.

5.3 Future work

In the analysis of thermodynamics for dynamical black holes, we choose $\xi^\mu \equiv \mathcal{N}\hat{T}^\mu = (1, \beta^i)$ on the horizons and require that the diffeomorphic variation $\delta\xi^\mu$ vanishes there. This requirement implies that $\delta\beta^i = 0$ which only holds for black holes with zero angular momentum. Therefore this condition limits our analysis to non-rotating black holes. Since astrophysical black holes have non-zero angular momentum, extending our analysis to conditions where $\delta\xi^\mu \neq 0$ should be of keen future interest.

Except for this issue related to $\delta\xi^\mu$, to extend our analysis to dynamical spacetimes with rotation, a covariant form for the ADM momentum and ADM angular momentum are presumably also needed. Our preliminary calculations show that instead of transforming the ADM momentum into a covariant expression, we may need to derive a covariant energy-momentum expression from first principles. This will be one of our main future research efforts.

Appendix A

List of publications

4. Non-equilibrium entropic forces: From molecular chains to black holes.
Zhi-Wei Wang & Samuel L. Braunstein.
In preparation.

3. Black hole mechanics for dynamical spacetimes
Zhi-Wei Wang & Samuel L. Braunstein.
Submitted.

2. Surfaces away from horizons are not thermodynamic
Zhi-Wei Wang & Samuel L. Braunstein.
Nature Communications **9**, 2977 (2018).

1. Higher-dimensional performance of port-based teleportation
Zhi-Wei Wang & Samuel L. Braunstein.
Scientific Reports **6**, 33004 (2016).

References

[1] Ahmed Almheiri, Donald Marolf, Joseph Polchinski, and James Sully. Black holes: complementarity or firewalls? *Journal of High Energy Physics*, 2013(2):62, 2013.

[2] Richard Arnowitt, Stanley Deser, and Charles W Misner. Dynamical structure and definition of energy in general relativity. *Physical Review*, 116(5):1322, 1959.

[3] Richard Arnowitt, Stanley Deser, and Charles W Misner. Energy and the criteria for radiation in general relativity. *Physical Review*, 118(4):1100, 1960.

[4] Abhay Ashtekar, Christopher Beetle, Olaf Dreyer, Stephen Fairhurst, Badri Krishnan, Jerzy Lewandowski, and Jacek Wiśniewski. Generic isolated horizons and their applications. *Physical Review Letters*, 85(17):3564, 2000.

[5] Abhay Ashtekar, Christopher Beetle, and Stephen Fairhurst. Isolated horizons: a generalization of black hole mechanics. *Classical and Quantum Gravity*, 16(2):L1, 1999.

[6] Abhay Ashtekar, Christopher Beetle, and Jerzy Lewandowski. Mechanics of rotating isolated horizons. *Physical Review D*, 64(4):044016, 2001.

[7] Abhay Ashtekar, Christopher Beetle, and Jerzy Lewandowski. Geometry of generic isolated horizons. *Classical and Quantum Gravity*, 19(6):1195, 2002.

[8] Abhay Ashtekar, Stephen Fairhurst, and Badri Krishnan. Isolated horizons: Hamiltonian evolution and the first law. *Physical Review D*, 62(10):104025, 2000.

[9] Abhay Ashtekar and Richard O Hansen. A unified treatment of null and spatial infinity in general relativity. i. universal structure, asymptotic symmetries, and conserved quantities at spatial infinity. *Journal of Mathematical Physics*, 19(7):1542–1566, 1978.

[10] C. Barcelo. Analogue black-hole horizons. *Nature Physics*, 15:210–213, 2019.

[11] James M Bardeen, Brandon Carter, and Stephen W Hawking. The four laws of black hole mechanics. *Communications in mathematical physics*, 31(2):161–170, 1973.

[12] Robert Beig. Arnowitt-deser-misner energy and g00. *Physics Letters A*, 69(3):153–155, 1978.

[13] Jacob D Bekenstein. Black holes and the second law. *Lettere Al Nuovo Cimento*, 4(15):737–740, 1972.

[14] Jacob D Bekenstein. Generalized second law of thermodynamics in black-hole physics. *Physical Review D*, 9(12):3292, 1974.

[15] Carl M Bender and Steven A Orszag. *Advanced mathematical methods for scientists and engineers I: Asymptotic methods and perturbation theory.* Springer Science & Business Media, 2013.

[16] Edmund Bertschinger. Symmetry transformations, the einstein-hilbert action and gauge invariance. *Massachusetts Institute of Technology, Department of Physics*, 2002.

[17] George David Birkhoff and Rudolph Ernest Langer. *Relativity and modern physics.* Harvard University Press, 1923.

[18] Samuel L Braunstein. Black hole entropy as entropy of entanglement, or it's curtains for the equivalence principle. *arXiv preprint arXiv:0907.1190*, 2009.

[19] Dieter R Brill and Richard W Lindquist. Interaction energy in geometrostatics. *Physical Review*, 131(1):471, 1963.

[20] Curtis G Callan Jr and Juan M Maldacena. D-brane approach to black hole quantum mechanics. *Nuclear Physics B*, 472(3):591–608, 1996.

[21] Sean M Carroll. An introduction to general relativity: spacetime and geometry. *Addison Wesley (p251, p450, p456)*, 101:102, 2004.

[22] Yi-Xin Chen and Jian-Long Li. First law of thermodynamics on holographic screens in entropic force frame. *Physics Letters B*, 700(5):380–384, 2011.

[23] Piotr T Chrusciel. A remark on the positive-energy theorem. *Classical and Quantum Gravity*, 3(6):L115, 1986.

[24] Piotr T Chrusciel. Lectures on energy in general relativity, kraków, march-april 2010. *Preprint, available at http://homepage. univie. ac. at/piotr. chrusciel*, pages 28–30, 2012.

[25] Gregory B Cook. Initial data for numerical relativity. *Living reviews in relativity*, 3(1):5, 2000.

[26] Pierre-Gilles De Gennes. *Scaling concepts in polymer physics.* Cornell university press, 1979.

[27] V. Faraoni. Cosmological and black hole apparent horizons. *Lecture Notes in Physics*, 907, 2015.

[28] John L Friedman, Kōji Uryū, and Masaru Shibata. Thermodynamics of binary black holes and neutron stars. *Physical Review D*, 65(6):064035, 2002.

[29] Gary W Gibbons and Stephen W Hawking. Cosmological event horizons, thermodynamics, and particle creation. *Physical Review D*, 15(10):2738, 1977.

[30] Eric Gourgoulhon. *3+1 formalism in general relativity: bases of numerical relativity*, volume 846. Springer Science & Business Media, 2012.

[31] Stephen W Hawking. Particle creation by black holes. *Communications in mathematical physics*, 43(3):199–220, 1975.

[32] Stephen W Hawking. Breakdown of predictability in gravitational collapse. *Physical Review D*, 14(10):2460, 1976.

[33] Stephen W Hawking and George Francis Rayner Ellis. *The large scale structure of space-time.* Cambridge university press (pp82-83 p334), 1973.

[34] Gary T. Horowitz. Quantum states of black hole. *In Black Holes and Relativistic Stars (Editor: Wald, R. M.)*, University of Chicago Press (pp.241-266), 1998.

[35] Gary T. Horowitz and Andrew Strominger. Counting states of near-extremal black holes. *Physical Review Letters*, 77(12):2368, 1996.

[36] Vivek Iyer and Robert M Wald. Some properties of the noether charge and a proposal for dynamical black hole entropy. *Physical review D*, 50(2):846, 1994.

[37] Ted Jacobson. Thermodynamics of spacetime: the einstein equation of state. *Physical Review Letters*, 75(7):1260, 1995.

[38] Arthur Komar. Covariant conservation laws in general relativity. *Physical Review*, 113(3):934, 1959.

[39] Badri Krishnan. Isolated horizons in numerical relativity. *Ph.D. Thesis*, 2002.

[40] Charles W Misner, Kip S Thorne, and John Archibald Wheeler. *Gravitation*. Macmillan, 1973.

[41] A Morozov. Black hole motion in entropic reformulation of general relativity. *JETP letters*, 91(12):615–619, 2010.

[42] Alex B Nielsen and Jong Hyuk Yoon. Dynamical surface gravity. *Classical and Quantum Gravity*, 25(8):085010, 2008.

[43] Emmy Noether. Invariant variation problems. *Transport Theory and Statistical Physics*, 1(3):186–207, 1971.

[44] Attilio Palatini. Deduzione invariantiva delle equazioni gravitazionali dal principio di hamilton. *Rendiconti del Circolo Matematico di Palermo (1884-1940)*, 43(1):203–212, 1919.

[45] Maulik K Parikh and Frank Wilczek. Hawking radiation as tunneling. *Physical Review Letters*, 85(24):5042, 2000.

[46] Federico Piazza. Gauss-codazzi thermodynamics on the timelike screen. *Physical Review D*, 82(8):084004, 2010.

[47] Eric Poisson. *A relativist's toolkit: the mathematics of black-hole mechanics*. Cambridge university press, 2004.

[48] John Preskill. Stephen Hawking (1942–2018). *Science*, 360(6385):156–156, 2018.

[49] Richard H Price and Kip S Thorne. Membrane viewpoint on black holes: properties and evolution of the stretched horizon. *Physical Review D*, 33(4):915, 1986.

[50] Guido Risaliti and Elisabeta Lusso. Cosmological constraints from the hubble diagram of quasars at high redshifts. *Nature Astronomy*, 3(3):272, 2019.

[51] Andrew Strominger and Cumrun Vafa. Microscopic origin of the bekenstein-hawking entropy. *Physics Letters B*, 379(1-4):99–104, 1996.

[52] Leonard Susskind, Larus Thorlacius, and John Uglum. The stretched horizon and black hole complementarity. *Physical Review D*, 48(8):3743, 1993.

[53] Kip S Thorne and Roger D Blandford. *Modern Classical Physics: Optics, Fluids, Plasmas, Elasticity, Relativity, and Statistical Physics.* Princeton University Press, 2017.

[54] Kip S Thorne, Richard H Price, and Douglas A MacDonald. *Black holes: the membrane paradigm.* Yale university press, 1986.

[55] William G Unruh. Notes on black-hole evaporation. *Physical Review D*, 14(4):870, 1976.

[56] William G Unruh. Experimental black-hole evaporation? *Physical Review Letters*, 46(21):1351, 1981.

[57] Maurice HPM van Putten. Entropic force in black hole binaries and its newtonian limits. *Physical Review D*, 85(6):064046, 2012.

[58] Bala Murali Venkatesan and Rashid Bashir. Nanopore sensors for nucleic acid analysis. *Nature nanotechnology*, 6(10):615, 2011.

[59] Erik Verlinde. On the origin of gravity and the laws of newton. *Journal of High Energy Physics*, 2011(4):29, 2011.

[60] Robert M Wald. *General Relativity.* University of Chicago press, 1984.

[61] Robert M Wald. *Black holes and thermodynamics. In Black hole physics (Editor: De, S. V. & Zhang, Z.).* Springer Science & Business Media, 1992.

[62] Robert M Wald. Black hole entropy is the noether charge. *Physical Review D*, 48(8):R3427, 1993.

[63] Robert M Wald and Andreas Zoupas. General definition of "conserved quantities" in general relativity and other theories of gravity. *Physical Review D*, 61(8):084027, 2000.

[64] Zhi-Wei Wang and Samuel L Braunstein. Surfaces away from horizons are not thermodynamic. *Nature communications*, 9(1):2977, 2018.

[65] Zhi-Wei Wang and Samuel L Braunstein. Non-equilibrium entropic forces: From molecular chains to black holes. In preparation.

[66] Zhi-Wei Wang and Samuel L Braunstein. Black hole mechanics for dynamical spacetimes. Submitted.

[67] Steven Weinberg. *Gravitation and cosmology: principles and applications of the general theory of relativity.* Wiley, New York, 1972.

[68] Edward Witten. Symmetry and emergence. *Nature Physics*, 14(2):116, 2018.

[69] Hong-Qing Xie and Cheng-Hung Chang. Chemical potential formalism for polymer entropic forces. *Communications Physics*, 2(1):24, 2019.

[70] James W York Jr. Kinematics and dynamics of general relativity. *Sources of gravitational radiation*, pages 83–126, 1979.

[71] James W York Jr. Energy and momentum of the gravitational field. In *Essays in General Relativity*, pages 39–58. Elsevier, 1980.

[72] WH Zurek and Kip S Thorne. Statistical mechanical origin of the entropy of a rotating, charged black hole. *Physical review letters*, 54(20):2171, 1985.