The Bright Side of Black Holes:
Radiation from Black Hole Accretion Disks

A dissertation presented

by

Yucong Zhu

to

The Department of Astronomy

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Astronomy & Astrophysics

Harvard University

Cambridge, Massachusetts

May 2015
The Bright Side of Black Holes: 
Radiation from Black Hole Accretion Disks

Abstract

An understanding of radiation is paramount for connecting observations of accretion disks with the theory of black holes. In this thesis, we explore via radiative transfer postprocessing calculations the observational signatures of black holes. We investigate disk spectra by analyzing general relativistic magnetohydrodynamic (GRMHD) simulations of accretion disks. For the most part there are no surprises – the resulting GRMHD spectrum is very close to the analytic Novikov & Thorne (1973) prediction from decades past, except for a small modification in the case of spinning black holes, which exhibit a high-energy power-law tail that is sourced by hot Comptonized gas from within the plunging region of the accretion flow. These conclusions are borne out by both 1D and 3D radiative transfer calculations of the disk. Significant effort was spent in developing from scratch the 3D radiative code that we used for the analysis. The code is named HERO (Hybrid Evaluator for Radiative Objects) and it is a new general purpose grid-based 3D general relativistic radiative solver.
Contents

Abstract iii

Acknowledgments ix

Dedication x

1 Introduction 1

1.1 Why Care About Black Holes? ................................. 1

1.2 Mathematical Basis and Understanding ......................... 3

1.2.1 Kerr Metric .............................................. 3

1.2.2 Conservation Laws ......................................... 5

1.2.3 Metric Singularities ....................................... 6

1.2.4 Ergosphere ................................................ 9

1.2.5 Circular Orbits ............................................ 11

1.3 Observational Evidence for Black Holes ......................... 15

1.3.1 Spin Fitting Techniques .................................... 19

1.3.2 Zoology of Disk States ..................................... 22

1.4 Disk physics .................................................. 27

1.4.1 Classic Thin Disk Model ................................... 28

1.4.2 Relativistic Disk Model (Novikov & Thorne 1973) ........... 32

1.4.3 Open problems in Accretion Physics ....................... 34
## CONTENTS

1.5 Numerical Simulations ................................................. 36
1.5.1 Shearing Boxes ..................................................... 38
1.5.2 Global Simulations .................................................. 39
1.5.3 Future Directions ..................................................... 40
1.6 Including Radiation ..................................................... 41
1.6.1 Radiation Hydrodynamics .......................................... 42
1.7 Chapter Summaries ..................................................... 45

2 The Eye of the Storm:  
Light from the Inner Plunging Region of Black Hole Accretion Discs  49

2.1 Introduction ........................................................... 50
2.2 GRMHD Simulations ................................................... 57
2.3 Annuli Spectra .......................................................... 58
2.3.1 Assumptions in the TLUSTY model ............................... 61
2.4 Slicing the GRMHD disc into Annuli ................................. 62
2.4.1 Flux profile .......................................................... 64
2.4.2 Vertical gravity profile .............................................. 68
2.4.3 Column density profile .............................................. 69
2.5 Ray Tracing .............................................................. 73
2.6 Results ................................................................. 76
2.6.1 Power law tail ....................................................... 78
2.6.2 Quantitative effect on spin ........................................ 83
2.7 Discussion .............................................................. 91
2.7.1 Other signatures of the plunging region ......................... 95
2.7.2 How does our choice of cooling function influence the results? 96
2.7.3 Equation of state .................................................... 98
2.8 Summary ............................................................... 99
CONTENTS

2.9 Luminosity Matching Model .................................................. 101
2.10 Generalized Novikov & Thorne Model ................................. 102
2.11 Interpolation Methods ....................................................... 105
2.12 An Alternative GRMHD Luminosity Profile ......................... 108
  2.12.1 Obtaining the GRMHD dissipation profile ...................... 110
  2.12.2 Net result of the luminosity calculation ....................... 110

3 Thermal Stability in Turbulent Accretion Discs ........................ 113
  3.1 Introduction ..................................................................... 114
  3.2 Physical model ............................................................... 117
    3.2.1 Radial structure ......................................................... 118
    3.2.2 Vertical structure ..................................................... 119
  3.3 Disc solutions ............................................................... 126
    3.3.1 Classic unmixed disc .................................................. 128
    3.3.2 Convective solutions ............................................... 131
    3.3.3 Convective and turbulent disc solutions ..................... 133
    3.3.4 Radial structure of solutions ..................................... 139
  3.4 Discussion ................................................................... 144
    3.4.1 Use of logarithmic temperature gradient .................... 145
    3.4.2 Impact of $\Sigma_{\text{scale}}$ .......................................... 148
    3.4.3 Scaling of Critical $\dot{M}$ with Black Hole Mass ............ 149
    3.4.4 Choosing $\zeta$ – comparison with simulations .......... 151
    3.4.5 $\zeta$ from observations ............................................ 158
    3.4.6 Radiative Outer Zone .............................................. 158
    3.4.7 Complete stabilization from turbulence ..................... 162
  3.5 Summary ................................................................... 165
  3.6 Acknowledgments ........................................................... 166
CONTENTS

3.7 Solving for 8 unknowns ................................................. 167

4 HERO - A 3D General Relativistic Radiative Postprocessor for Accretion Discs around Black Holes 170
   4.1 Introduction .......................................................... 171
   4.2 Radiative Solver .................................................... 176
      4.2.1 Short Characteristics ........................................ 178
      4.2.2 Implementation of Short Characteristics .................. 183
      4.2.3 Long Characteristics ........................................ 185
      4.2.4 Acceleration Schemes ........................................ 188
      4.2.5 Raytracing ..................................................... 190
      4.2.6 Frequency Discretization .................................... 191
      4.2.7 Angular Discretization ....................................... 191
   4.3 Numerical Tests .................................................... 194
      4.3.1 1D Plane-parallel Grey Atmosphere ......................... 195
      4.3.2 Convergence Tests ............................................ 197
      4.3.3 Multiray Temperature Solution ............................... 198
      4.3.4 Test of Spectral Hardening .................................. 200
      4.3.5 Effect of a Heating Source ................................. 201
      4.3.6 2D Solutions and Ray Defects ............................... 204
      4.3.7 3D Solutions .................................................. 209
      4.3.8 GR Solutions .................................................. 216
   4.4 Summary .............................................................. 228
   4.5 Acknowledgements ................................................ 229
   4.6 Ray Defects .......................................................... 230
      4.6.1 Mathematical Origin of Ray Defects ....................... 233
      4.6.2 Ray Defect Correction Schemes ............................. 235
CONTENTS

4.7 Analytic 1D Atmosphere Spectrum ............................................. 238

5 HEROIC - A Comptonization Module for the HERO radiative code 242

5.1 Introduction ................................................................. 243
5.2 Radiative Transfer Solution .................................................. 246
  5.2.1 Kompaneets-Ray ....................................................... 248
  5.2.2 Quadratic Variation of the Source Function ......................... 250
5.3 Numerical Tests ............................................................ 253
  5.3.1 Kompaneets ............................................................. 253
  5.3.2 Escape Time Distributions ......................................... 256
5.4 Application – Accretion Disk ............................................... 264
  5.4.1 Numerical Disk Setup .................................................. 265
  5.4.2 Results ................................................................. 267
5.5 Discussion ........................................................................ 274
5.6 Summary ........................................................................... 280

6 Summary and Future Directions ................................................. 282

6.1 Summary ........................................................................... 282
6.2 Future Directions ................................................................ 284
  6.2.1 Astrophysical Applications ............................................. 284
  6.2.2 Future Plans for HERO .................................................. 285

References .............................................................................. 288
Acknowledgments

I thank my advisor, Ramesh Narayan, for his boundless insights, and endless patience.

I thank my friends, for bringing so much joy to my life.

I thank my family, for all their love, support, and encouragement throughout the years.
For my fellow travelers in space and time.
Chapter 1

Introduction

1.1 Why Care About Black Holes?

Our universe can be thought of as a gigantic all encompassing house – every material thing that we know of (matter and energy) lives inside of it. Just like a human domicile, the universe burns fuel (stars) to keep warm, releasing tremendous amounts of radiative energy before burning out, leaving only a dark ash behind; in the case of stars, the ash is comprised of compact objects such as white dwarfs, neutron stars, or black holes, depending on the mass of the stellar progenitor.

Black holes are perhaps the most exciting of these three destinations of stellar evolution. They are more than just ash – when conditions are right (e.g., under the influence of gas accretion), a black hole can become a phoenix. It rises from the ashes to light up the universe anew with unimaginable luminosity, producing at the same time tremendous jets that are powerful enough to shape and govern the evolution of galaxies!
Even more remarkable is the efficiency at which black holes convert matter into energy. Due to their extreme gravity, the accretion process (and the subsequent dissipation of gravitational potential energy) leads to the extraction of rest mass energy at an efficiency of 1-100 times higher\textsuperscript{1} than that of even nuclear fusion! Hence black holes can be thought of as the true powerplants of our universe, and therefore crucial to understanding the evolution and fate of our universe (e.g. AGN feedback on galactic dynamics, dominating the epoch of reionization, and dictating the physics of the most violent explosions in the universe).

Perhaps the most interesting aspect of black holes lies in their extremal nature. Space and time become intimately and infinitely mixed at the surface of a black hole, which provides the ultimate testbed for our physical theories of gravity. It is only near the event horizon\textsuperscript{2} of a black hole where general relativity (our current best-guess on the inner workings of gravity) can truly be put to the test.

Black holes are the ultimate triumph of modern astrophysics. They were originally predicted from purely mathematical arguments, only to be later verified observationally in the 1970’s by X-ray observations of accreting stellar binaries and in the 1990s by high angular resolution optical and infrared observations of galactic nuclei. They have a profound influence on our universe and are also objects of immense mathematical simplicity and beauty.

\textsuperscript{1}The mass-energy conversion efficiency depends on the spin of the black hole, whereby high spin leads to higher efficiencies

\textsuperscript{2}The event horizon is the surface at which gravity becomes so strong that light itself cannot escape its wretched pull
1.2 Mathematical Basis and Understanding

The ubiquity and simplicity of black holes in our universe make them truly marvelous objects of study. The complete physics of each and every black hole (BH) can be distilled down to just three numbers: the black hole’s mass, charge and spin. This miracle implies that all physical theories that involve black hole physics\(^3\) must connect in some way to these three numbers! However, due to the tendency for charged objects to neutralize, we do not believe astrophysical black holes to hold any significant charge. Thus, to get a complete description of any black hole (ranging from the supermassive ones dominating the centers of galaxies down to their more humble stellar mass brethren), we simply need to measure two quantities: its mass and spin.

1.2.1 Kerr Metric

In the framework of general relativity, spacetime near a black hole can be described by the Kerr metric, which is a solution to the vacuum Einstein field equations (Kerr 1963). As mentioned above, it is completely described with two free parameters: \(M_\ast\) (the mass of the black hole), and \(J_\ast\) (its angular momentum). For convenience, these quantities are commonly rescaled according to:

\[
M = \frac{GM_\ast}{c^2},
\]

\[
a = \frac{J_\ast}{M_\ast c},
\]

\(^3\)For instance, the link between BH spin and jet power (Blandford & Znajek 1977), Gamma Ray Bursts and spinning BHs (MacFadyen & Woosley 1999), a theory for quasi periodic oscillations (Abramowicz & Kluzniak 2001), etc...
so that they have units of length. With these rescalings and expressing the metric in
Boyer-Lindquist \((t, r, \theta, \phi)\) coordinates, the Kerr metric simply becomes:

\[
\begin{align*}
\text{ds}^2 &= g_{tt}dt^2 + g_{t\phi}dtd\phi + g_{rr}dr^2 + g_{\theta\theta}d\theta^2 + g_{\phi\phi}d\phi^2, \\
&= -\left(1 - \frac{2Mr}{\rho^2}\right)dt^2 + \frac{4Mar\sin^2\theta}{\rho^2}dtd\phi + \frac{\rho^2}{\Delta}dr^2 + \frac{\rho^2}{\rho^2}d\theta^2 + \left(\frac{r^2 + a^2 + 2Ma^2r\sin^2\theta}{\rho^2}\right)\sin^2\theta d\phi^2,
\end{align*}
\]

(1.3)

with metric functions:

\[
\begin{align*}
g_{tt} &= -\frac{1 - 2Mr}{\rho^2}, \\
g_{t\phi} &= -\frac{4Mar\sin^2\theta}{\rho^2}, \\
g_{rr} &= \frac{\rho^2}{\Delta}, \\
g_{\theta\theta} &= \frac{\rho^2}{\rho^2}, \\
g_{\phi\phi} &= \left(r^2 + a^2 + \frac{2Ma^2r\sin^2\theta}{\rho^2}\right)\sin^2\theta,
\end{align*}
\]

(1.4)

and inverse metric functions:

\[
\begin{align*}
g^{tt} &= -\frac{(r^2 + a^2)^2 - a^2\Delta\sin^2\theta}{\rho^2\Delta}, \\
g^{t\phi} &= -\frac{2Mar}{\rho^2\Delta}, \\
g^{rr} &= \Delta/\rho^2, \\
g^{\theta\theta} &= 1/\rho^2, \\
g^{\phi\phi} &= \frac{\Delta - a^2\sin^2\theta}{\Delta\rho^2\sin^2\theta},
\end{align*}
\]

(1.5)

where we have:

\[
\Delta \equiv r^2 - 2Mr + a^2, \\
\rho^2 \equiv r^2 + a^2\cos^2\theta.
\]

(1.6)  (1.7)

Several interesting properties of the Kerr metric are apparent by inspection of the line

element \(\text{ds}^2\). We see that the Kerr metric is:
CHAPTER 1. INTRODUCTION

- stationary: it does not depend explicitly on time $t$.
- axisymmetric: it does not depend explicitly on $\phi$.
- invariant under simultaneous inversion of $t$ and $\phi$ (i.e. $t \rightarrow -t$ and $\phi \rightarrow -\phi$). A rotating object also produces this symmetry, motivating the link between rotation and the Kerr metric.
- Minkowski (i.e. special-relativistic flat space) in the limit $r \rightarrow \infty$, hence Kerr spacetime is asymptotically flat.
- Schwarzchild in the limit $a \rightarrow 0$.
- Minkowskii in the limit $M \rightarrow 0$ (even with $a \neq 0$).

We explore the implications of these properties and draw additional insights from the mathematical structure of the metric in the next few subsections.

1.2.2 Conservation Laws

From Noether’s theorem (Noether 1918), one expects conservation laws to be associated with symmetries within a system. The Kerr metric is independent of both time $t$ and azimuth $\phi$ and is hence symmetric in translations about these coordinates. This allows one to define a set of orthogonal “Killing” vectors that point along directions that leave the metric unchanged, i.e.

$$\eta^\mu = \delta^\mu_t,$$

$$\xi^\mu = \delta^\mu_\phi.$$  \hspace{1cm} \text{(1.8)}  \hspace{1cm} \text{(1.9)}
CHAPTER 1. INTRODUCTION

A Killing vector is a general way of characterizing the symmetry in a given coordinate system. It has the property that its inner product with a particle/photon’s four-momentum \( p_\mu \) is conserved along geodesics. This gives rise to the notion of energy and angular momentum conservation corresponding to the two symmetries/Killing vectors present:

\[
\begin{align*}
\text{energy:} & \quad E = -\eta^\mu p_\mu = -p_t \\
\text{angular momentum:} & \quad \mathcal{L} = \xi^\mu p_\mu = p_\phi.
\end{align*}
\] (1.10)

For massive particles the four-momentum corresponds to \( p^\mu = m u^\mu \), and the conserved quantities \( E, \mathcal{L} \) correspond to the per unit mass energy and angular momentum respectively. For massless photons, an affine invariant can be chosen such that the four-momentum becomes \( p^\mu = u^\mu \), leading to the conserved quantities \( E, \mathcal{L} \) representing the photon’s energy and angular momentum as measured at infinity.

The two Killing vectors \( \eta^\mu \) and \( \xi^\mu \) also span all available Killing vector fields present in the Kerr metric. Any other Killing vector can be expressed as a linear combination of these two. \(^4\):

1.2.3 Metric Singularities

The metric as defined in Eq. 1.3 becomes singular on the submanifolds

\[
0 = \rho^2 = r^2 + a^2 \cos^2 \theta,
\] (1.12)

\(^4\)A third symmetry also exists within the Kerr metric, though it is generated via a more complex \textbf{Killing tensor field} \( K^{\mu\nu} \). This gives rise to a third conserved quantity, the Carter constant \( \mathcal{C} = K^{\mu\nu} p_\mu p_\nu = (p_\theta)^2 + (p_\phi)^2 / \sin^2(\theta) = \text{const} \), which describes the mix of poloidal and toroidal momentum that is conserved during particle motion.
or

\[ 0 = \Delta = r^2 - 2Mr + a^2. \]  

(1.13)

Of these two, only the first condition \( \rho^2 \to 0 \) corresponds to a true curvature singularity (i.e. as verified by evaluating the scalar curvature invariant \( R^{\alpha\beta\gamma\delta}R_{\alpha\beta\gamma\delta} \to \infty \)). It corresponds to the coordinates \( r = 0 \) and \( \theta = \pi/2 \), although mapping these coordinates to the usual Boyer-Lindquist/spherical-polar sense leaves some confusion since \( r = 0 \) also corresponds to a coordinate singularity. Under a transformation to a new coordinate system (i.e. Kerr-Schild which does not suffer from the \( r = 0 \) coordinate singularity), the topological nature of the curvature singularity is revealed to be a thin ring. The radius of the ring-singularity is set by the spin of the black hole, vanishing to a point-singularity in the limit of a nonrotating black hole \( (a = 0) \).

\( \Delta \to 0 \) corresponds to a coordinate singularity that can be dealt with by means of a coordinate transformation. Kerr coordinates is one such one that regularizes the \( \Delta \to 0 \) condition, and is motivated by using null-worldlines to define a new time-basis (i.e. in a similar fashion as Eddington-Finkelstein coordinates).

An interesting feature of the \( \Delta = 0 \) submanifold is its quadratic nature. Two solutions exist for the condition \( \Delta = 0 \), corresponding to “inner” and “outer” singularities and are located at

\[ 0 = r^2 - 2Mr + a^2 = (r - r_+)(r - r_-) \]  

(1.14)

\[ \Rightarrow r_{\pm} = M \pm \sqrt{M^2 - a^2}. \]  

(1.15)

The interpretation of this \( \Delta = 0 \) surface can be motivated by considering the Schwarzschild \( a = 0 \) case. Here we have \( r_+ = 2M \), which corresponds to the location
of the classic black hole event horizon. Thus, we identify $r_{\pm}$ as the outer and inner horizons of a spinning black hole, with the outer horizon $r_+$ serving as the boundary that separates the causally connected exterior Universe from the Kerr interior.

Alternatively, a more mathematically rigorous way of interpreting the $r = r_{\pm}$ hypersurfaces is to consider the radial surface normal four-vector, i.e. in Boyer-Lindquist $(t, r, \theta, \phi)$:

$$n_\alpha = (0, 1, 0, 0).$$

(1.16)

By examining the inner product of $n_\alpha$, we can characterize the normal vector. Note that

$$n_\alpha n^\alpha = n_\alpha n_\beta g^{\alpha\beta} = g^{rr} = \frac{\Delta}{\rho^2}$$

(1.17)

and therefore the normal vectors along the $\Delta = 0$ surface have $n_\alpha n^\alpha = 0$ (null-like). Hence $r = r_{\pm}$ are horizons since the light cone collapses to the radial vector here. The two horizons also separate space into three distinct subregions, whose properties can be further explored by consideration of the tangent vector field along constant radius surfaces (Misner, Thorne, & Wheeler 1973; D’Inverno 1992):

- $r > r_+$: This region represents the exterior of the black hole, becoming flat spacetime in the limit $r \to \infty$. Here, constant radius surfaces have tangent fields that are timelike, and therefore non-descending worldlines are accessible.

- $r_- < r < r_+$: This is the region between the two horizons. Tangent vectors for constant radius surfaces are spacelike, and therefore all valid worldlines must descend towards smaller radius. Eventually all material in this zone is channeled inwards and falls through the inner horizon.

- $r < r_-$: This is the innermost region containing the ring singularity as discussed
above. Tangent bundles on radial surfaces are once again timelike and the worldline restriction on radial monotonicity is lifted here.

Finally note that whenever \( a > M \), the equation \( \Delta = 0 \) has no real solution, implying the absence of a horizon and leaving the Kerr singularity at \( \rho^2 = 0 \) “not covered” by a horizon, resulting in a “naked singularity” that is exposed and accessible to the rest of the Universe. Due to the strange nature of spacetime around the singularity (i.e. causality paradoxes), this situation is considered unphysical, and it is believed that all astrophysical black holes in the universe are constrained to have \( a \leq M \).\(^5\)

1.2.4 Ergosphere

An interesting property of the Kerr spacetime is that relativistic “frame-dragging” effects due to black hole rotation prevents the existence of stationary observers close to the black hole. This property can be easily deduced by considering the four-normal of a stationary observer, i.e.

\[
\eta^\mu = (1, 0, 0, 0). \tag{1.18}
\]

To check whether this time-stationary worldline is valid, we proceed like in the previous section on horizons – we compute its inner product to classify whether it is either timelike

\(^5\)The limiting case of \( a = M \) is typically referred to as an extremal black hole.
CHAPTER 1. INTRODUCTION

or spacelike to determine if it is a valid observer worldline:

\[ \eta_{\mu} \eta^\mu = g_{\mu \nu} \eta^\mu \eta^\nu = g_{tt} = -1 + \frac{2Mr}{\rho^2} = -\frac{1}{\rho^2} (r^2 - 2Mr + a^2 \cos^2 \theta) \]
\[ = -\frac{1}{\rho^2} (r - r_{E+}) (r - r_{E-}), \quad (1.19) \]

where the function roots are located at

\[ r_{E\pm} = M \pm \sqrt{M^2 - a^2 \cos^2 \theta}. \quad (1.20) \]

Once again, due to the quadratic nature of Eq. 1.19, the two roots subdivide space into three regions. Of particular interest is the subregion \( r_{E-} < r < r_{E+} \) known as the “ergosphere”, where \( g_{tt} = \eta_{\mu} \eta^\mu > 0 \) indicating that the stationary observer has an unphysical spacelike worldline. Thus, in this region a stationary observer with fixed \((r, \theta, \phi)\) cannot exist. This ergosphere constitutes a portion of space where the rotational influence of the black hole is so violent that it is impossible to remain fixed in space – in the ergosphere, all material is forced to rotate in the same sense as the black hole.

The manifold spanned by \( r_{E\pm} \) also enjoys the distinction of being infinite redshift surfaces. That is, any source located on this surface emits light signals that are suppressed by a factor:

\[ \nu_{\text{obs}} = \sqrt{\frac{g_{tt}^{\text{emit}}}{g_{tt}^{\text{obs}}}} \nu_{\text{emit}}, \quad (1.21) \]

and since \( g_{tt}^{\text{emit}} = 0 \), the photon is redshifted to oblivion when measured by another nonlocal observer.

Finally, notice that the topology of the twin \( g_{tt} \) null surfaces are simply ellipsoids in Boyer-Lindquist coordinates. The boundary of the ergoregion given by Eq. 1.20 also
satisfies $r_{E-} < r_+ < r_{E+}$, yielding the topology for the Kerr black hole shown in Fig. 1.1:

**Figure 1.1**: Schematic of the characteristic boundaries that split the Kerr black hole. In order from large to small radius, the key boundaries are: the outer surface of the ergosphere $r_{E+}$, the outer horizon $r_+$, the inner horizon $r_-$, the inner surface of the ergosphere $r_{E-}$, and the central ring singularity.

### 1.2.5 Circular Orbits

Of particular importance to accretion disks is the free-streaming motion of fluid in orbit about the black hole. Suppose the particle has trajectory $x^\mu(\lambda)$ where $\lambda$ is an
affine parametrization of the particle’s path, then in the framework of GR, this free “geodesic” motion with say velocity \( u^\mu = dx^\mu / d\tau \) is governed by trajectories with vanishing acceleration, i.e.

\[
0 = a^\mu = u^\nu \nabla_\nu u^\mu \\
= \frac{d^2x^\alpha}{d\tau^2} + \Gamma^\alpha_{\beta\gamma} \frac{dx^\beta}{d\tau} \frac{dx^\gamma}{d\tau}
\]

where \( \Gamma^\alpha_{\beta\gamma} \) are the connection coefficients of the coordinate system, given by

\[
\Gamma^\alpha_{\beta\gamma} = \frac{1}{2} g^{\alpha\kappa} \left( \frac{\partial g_{\beta\kappa}}{\partial x^\gamma} + \frac{\partial g_{\gamma\kappa}}{\partial x^\beta} - \frac{\partial g_{\beta\gamma}}{\partial x^\kappa} \right).
\]

By consideration of the constants of motion \( \mathcal{E}, \mathcal{L} \) as described in 1.2.2, and for the case of motion in the equatorial plane \( (\theta = \pi/2) \), Bardeen, Press, & Teukolsky (1972) deduced simpler and decoupled forms corresponding to geodesic motion. The radial component of the trajectory obeys:

\[
\rho^2 \frac{dr}{d\lambda} = \pm V_r^{1/2},
\]

where \( V_r \) defines an effective potential given by

\[
V_r = T^2 - \Delta \left[ \mu r^2 + (\mathcal{L} - a\mathcal{E})^2 \right]
\]

\[
T = \mathcal{E}(r^2 + a^2) - \mathcal{L}a,
\]

and \( \mu = u \cdot u \) is a switch that toggles between the case of massive particles \( (\mu = 1) \) and photons \( (\mu = 0) \). For circular motion, we must have \( dr/d\lambda = 0 \) both instantaneously and for all subsequent times (since \( r \) is constant for circular orbits), which from Eq.1.25

---

\(^6\)The affine parameter \( \lambda \) is typically chosen as the proper time \( \tau \) for massive particles and the proper path length for massless particles.

\(^7\)see Appendix D of Chung (2010) for a full listing of nonzero connection coefficients in the Kerr metric.
CHAPTER 1. INTRODUCTION

translates to the following conditions for the effective potential:

\[ V_r = 0 \quad \text{and} \quad \frac{\partial V_r}{\partial \lambda} = 0 \]  \hspace{1cm} (1.28)

For a given radius \( r \), the above constraints can be solved simultaneously, which sets the required values of \( E \) and \( L \). However, circular orbits do not exist for all values of \( r \) since the solutions for \( E \) and \( L \) are real valued only under the condition

\[ r^{3/2} - 3Mr \pm 2aM^{1/2} \geq 0. \]  \hspace{1cm} (1.29)

The limiting case of equality in Eq. 1.29 yields infinite energy per rest mass, corresponding to the photon orbit, and is given by the root

\[ r_{ph} = 2M \left\{ 1 + \cos \left[ \frac{2}{3} \cos^{-1} \left( \frac{\pm a}{M} \right) \right] \right\}. \]  \hspace{1cm} (1.30)

If one further demands that the circular orbit be stable, then we must also have \( V_r'' \leq 0 \). This translates into the condition

\[ r^2 - 6Mr \pm 8aM^{1/2}r^{1/2} - 3a^2 \geq 0 \]  \hspace{1cm} (1.31)

or \( r > r_{ms} \) where \( r_{ms} \) is the marginally stable orbit (alternatively, the “Innermost Stable Circular Orbit”– ISCO). It is located at

\[ r_{ms} = M \left\{ 3 + Z_2 \mp [(3 - Z_1)(3 + Z_1 + 2Z_2)]^{1/2} \right\}, \]  \hspace{1cm} (1.32)

\[ Z_1 \equiv 1 + (1 - a^2/M^2)^{1/3} \left[ (1 + a/M)^{1/3} + (1 - a/M)^{1/3} \right] \]

\[ Z_2 \equiv (3a^2/M^2 + Z_1^2)^{1/2}. \]

The ISCO is particularly relevant for accretion disks since it acts as a boundary that separates two regimes of the accretion flow. When \( r > r_{ms} \), the turbulent flow quickly circularizes and follows Keplerian orbits about the black hole. Material is viscously
CHAPTER 1. INTRODUCTION

transported inwards and upon reaching $r = r_{\text{ms}}$ suddenly switches from circular to nearly radial free-fall. This transition leads to an inner truncation of the accretion disk at the location of $r_{\text{ms}}$, and provides a strong observational signature for accretion flows about black holes.

Figure 1.2: All characteristic radii of the Kerr metric as a function of the black hole spin parameter $a$. From top to bottom, we show the location of the ISCO ($r_{\text{ms}}$ – Eq. 1.32), the equatorial photon orbit ($r_{\text{ph}}$ – Eq. 1.30), the location where the ergosphere begins ($r_{E^+}$ – Eq. 1.20), and the outer and inner horizons ($r_{\pm}$ – Eq. 1.15). Negative $a$ corresponds to retrograde orbits relative to the spin direction of the black hole.

In Figure 1.2, we summarize all characteristic radii present in the Kerr metric as a function of black hole spin $a$. Qualitatively, all features become more compact about the black hole as the spin is increased. This trend is the basis for all observational methods
that measure black hole spin (see §1.3.1 for details).

### 1.3 Observational Evidence for Black Holes

Black holes have been in the forefront of the imagination of both scientists and the general public for nearly a century. Karl Schwarzschild originally proposed their existence as a mathematical curiosity in 1916, and it was only much later that their link as the final end product of astrophysical stellar collapse was made by Oppenheimer & Snyder in 1939. Nevertheless, the long journey of gathering sufficient evidence to serve as conclusive proof of black holes was only recently realized through careful observations of stellar binary systems (Bolton 1972; Webster & Murdin 1972; McClintock & Remillard 1986).

Multiple lines of evidence are used to argue the case of black hole existence. For instance in some binary systems, the observed millisecond flickering rate of several X-ray binaries suggests the presence of a very compact primary object with size scales comparable to the predicted size of black holes. Another hint comes from radial velocity measurements of stellar companion – for some systems, the inferred lower limit on the mass of the primary object exceeds $3M_\odot$, the maximal limit allowed for neutron stars. This combination of small spatial size and high mass rule out any other astrophysical object except for a black hole. Similar mass and size constraint arguments have also been made to argue for the existence of our Galaxy’s central supermassive black hole (Genzel, Hollenbach & Townes 1994; Kormendy & Richstone 1995; Genzel, Eisenhauer & Gillenssen 2010), as well as for many extragalactic supermassive black holes (e.g. M32 – van der Marel et al. 1997; M31 – Tonry 1984; NGC4526 – Miyoshi et al. 1995; and
CHAPTER 1. INTRODUCTION

many others – Kormendy & Ho 2014).

Aside from the mass argument, other ways to confirm the existence of black holes include making use of the various relativistic properties associated with the Kerr metric as mentioned in the previous section. The event horizon is an obvious defining characteristic that is observed qualitatively in many cases – black hole systems are found to emit far less light when compared to identical neutron star counterpart systems (Narayan, Garcia, & McClintock 1997; Garcia et al. 2001; Narayan, & McClintock 2008; Broderick, Loeb & Narayan 2009; Broderick et al. 2015). The higher accretion luminosity for neutron stars is primarily due to the hard surface onto which the accreting material pile ups and radiates.

More recently, the Event Horizon Telescope collaboration is embarking on direct imaging project of the supermassive black hole at the center of our Galaxy. The idea is to take advantage of the exquisite \(10 \mu\text{as}\) angular resolution offered by radio interferometry to construct an image of the synchrotron emission from gas surrounding our central black hole. The expectation is that we will be able to glimpse the “shadow” of the black hole – in other words, since the emission is optically thin, a distinct ring-like enhancement near the black hole’s photon orbit is expected to be seen (see bottom row of Figure 1.3 for a few examples of what the optically thin emission should look like).

Finally, since a black hole is completely specified by only two parameters (its mass \(M\) and spin \(a\)), measuring these quantities constitute the ultimate goal of black hole

---

8Interestingly, this hard surface pile up of material also leads to the periodic onset of thermonuclear bursts in neutron star sources. These bursts are also conspicuously absent in black hole systems (Narayan & Heyl 2002; Remillard et al. 2006a), which is another line of evidence for the event horizon.
CHAPTER 1. INTRODUCTION

observations. Having these numbers will also inform us if our understanding of black hole physics is correct (and by extension, test the theory of relativity in the strong gravity limit). For binary systems, the task of measuring mass is relatively straightforward. So long as one can accurately measure the inclination, period, and radial velocity for a companion orbiting a black hole, one can immediately obtain the black hole mass as direct consequence of Newtonian gravity. To date, over 20 mass measurements have been made for stellar mass binary black holes (Remillard & McClintock 2006b), and over \( \sim 50 \) for supermassive black holes (Gültekin et al. 2009).

Despite this success in the mass sector, spin has proven to be a much more difficult quantity to pin down. The spin of the black hole only imprints its signature through the general relativistic frame dragging effect, which only occurs at very short distances (see discussion on Ergosphere in §1.2.4). Thus to probe black hole spin, we must rely on observations of fluid that is very close to the BH horizon – namely the part of the accretion disk orbiting close-in (i.e. see figure 1.3, which shows a few example raytraced images of the accretion disk). Currently, there are two main techniques (see §1.3.1 for details) that make use of the light from accretion disks to pin down BH spin. Both techniques make use of the same underlying principle, which is to measure the location of the innermost stable circular orbit (ISCO). As shown in Fig. 1.2, the ISCO size, which can be determined by measuring the size of the dark central void, is a monotonic function of black hole spin. It is through this ISCO size monotonicity relation that we infer the spin (McClintock et al. 2011; McClintock, Narayan & Steiner 2014).
Figure 1.3: Raytraced images of a geometrically thin accretion disk viewed at $i = 60^\circ$ inclination angle for various choices of BH spin parameter $a_\times \equiv a/M$. These images are computed via a GPU accelerated version of the radiative transfer code as described in Chapter 4. For illustrative purposes, the top panels show the equivalent image of the disk in flat space using linear light geodesics. Notice that as $a_\times$ becomes more positive, the light emission becomes more compact. The center panels show lensed images of an optically thick disk in the Kerr metric, while the bottom panels correspond to an optically thin disk.
1.3.1 Spin Fitting Techniques

Table 1.1: Spin measurements yielded by continuum fitting.

<table>
<thead>
<tr>
<th>Black hole</th>
<th>$a_*$</th>
<th>$i$ (°)</th>
<th>$L/L_{\text{Edd}}$</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>A0620-00</td>
<td>0.12 ± 0.19</td>
<td>51.0 ± 0.9</td>
<td>0.11</td>
<td>Gou et al. (2010)</td>
</tr>
<tr>
<td>H1743-322$^c$</td>
<td>0.2 ± 0.3$^c$</td>
<td>75 ± 3</td>
<td>0.03-0.3</td>
<td>Steiner et al. (2012)</td>
</tr>
<tr>
<td>LMC X-3</td>
<td>&lt; 0.3</td>
<td>67 ± 2</td>
<td>0.1-0.7</td>
<td>Davis et al. (2006b)</td>
</tr>
<tr>
<td>XTE J1550</td>
<td>0.34$^{+0.2}_{-0.28}$</td>
<td>74.7 ± 3.8</td>
<td>0.05-0.30</td>
<td>Steiner et al. (2011)</td>
</tr>
<tr>
<td>GRO J1655$^d$</td>
<td>0.70 ± 0.1</td>
<td>70.2 ± 1.2</td>
<td>0.04-0.1</td>
<td>Shafee et al. (2006)</td>
</tr>
<tr>
<td>4U 1543$^d$</td>
<td>0.80 ± 0.1</td>
<td>20.7 ± 1.5</td>
<td>0.06-0.1</td>
<td>Shafee et al. (2006)</td>
</tr>
<tr>
<td>M33 X-7</td>
<td>0.84 ± 0.05</td>
<td>74.6 ± 1.0</td>
<td>0.07-0.11</td>
<td>Liu et al. (2008, 2010)</td>
</tr>
<tr>
<td>LMC X-1</td>
<td>0.92$^{+0.05}_{-0.07}$</td>
<td>36.4 ± 2.0</td>
<td>0.15-0.17</td>
<td>Gou et al. (2009)</td>
</tr>
<tr>
<td>GRS 1915$^d$</td>
<td>&gt; 0.95</td>
<td>61.5-68.6</td>
<td>0.2-0.3</td>
<td>McClintock et al. (2006)</td>
</tr>
<tr>
<td>Cygnus X-1</td>
<td>&gt; 0.95</td>
<td>27.1 ± 0.8</td>
<td>0.018-0.026</td>
<td>Gou et al. (2011)</td>
</tr>
</tbody>
</table>

Note – The spin uncertainties correspond to the 68 per cent (1σ) level of confidence, whereas the inequalities are to the 3σ level.

$^c$ No reliable mass estimate is available for this source

$^d$ The quoted spin errors have not been rigorously computed, and have been arbitrarily doubled from the published estimates since these are among the first systems for which continuum fitting was applied.)
CHAPTER 1. INTRODUCTION

For a black hole binary, we measure the size of the ISCO by looking at its accretion disk. Objects inside the ISCO (hereafter referred to as the ‘plunging zone’) cannot remain in stable circular orbits, and are forced to quickly plunge into the black hole in just a few free-fall times. Since the plunging timescale is short, it is thought that fluid inside the plunging zone does not have time to radiate (Page & Thorne 1974), which means that if we could resolve an image of the accretion disk, then we would see a dark void corresponding to the plunging region in the center! In practice, we cannot resolve the accretion disks around black holes, and thus our only information comes in the form of X-ray spectra. The two main spin determination techniques seek to use different aspects of this spectral information to infer the size of the ISCO.

The first method is called the ‘iron line method’, pioneered by Fabian et al. (1989); Tanaka et al. (1995), uses the shape of a strong X-ray emission line due to iron (in particular, the 6.4 keV Fe Kα transition) to constrain the location of the ISCO. The basic picture is a cold disk illuminated by a very hot corona. The hot corona irradiates the disk with hard X-rays exciting fluorescence lines, of which the Fe Kα line is the strongest (Miller 2007). The line shape contains information about the disk geometry, which when taken with the assumption of no plunging region emission (Reynolds & Fabian 2008), leads to a measurement of the ISCO size. Figure 1.4 shows how the line profile varies as a function of the black hole spin parameter.

9Although it has been proposed to use radio VLBI observations to test GR by resolving the apparent shape of SgA∗!
CHAPTER 1. INTRODUCTION

Figure 1.4: Iron line profiles for various choices of the black hole spin parameter $a$ for a disk viewed at $40^\circ$ inclination. Here, the emissivity profile of the disk was set obeying the power law $F \propto r^{-3}$. Notice that as the black hole spins up, the ISCO shrinks, resulting in a stronger red-wing to the line profile. (Figure credit: Dauser et al. 2010)

This line modeling is complicated since the precise shape of the observed line profile depends on four physical processes: 1) the emissivity profile for Fe Kα line as a function of disk radius (which depends on both the degree of disk ionization and how the corona irradiates the disk), 2) light-bending changing the apparent sizes of emitting regions, 3) doppler boosting of the line, and 4) gravitational redshifting. The key piece of physics used to infer the size of the ISCO is 4) – the gravitational redshifting produces a long red tail in the line profile which cannot be due to processes 1-3. The maximum extent of this red tail is set by fluorescing fluid that feels the strongest gravitational redshift (i.e. the fluid right at the ISCO since the plunging region is assumed to be dark). This allows one to determine the location of the ISCO (and hence the spin) from purely the line profile.
CHAPTER 1. INTRODUCTION

The second method (called the 'continuum fitting method') uses the spectral shape of the thermal continuum emission in the accretion disk to constrain the size of the ISCO (originally applied by Zhang et al. 1997a with further refinements by Li et al. 2005; Davis & Hubeny 2006; see table 2.3 for results from recent work). The idea of continuum fitting is analogous to the operation in stellar astrophysics where one determines the size of a star given only the spectrum. Given the temperature (which one gets from the emission peak frequency), distance and flux, one can determine the star’s emitting area (and hence size) without actually needing to resolve the star. For the purpose of BH spin measurements, we determine the size of the accretion disk’s inner dark region (the size of the ISCO) by observing the disk’s thermal spectra. The main drawback of the continuum fitting method is that for it to work, we first need accurate estimates of the BH distance (to turn fluxes into areas), disk inclination (to turn the area into an ISCO radius), and BH mass (to get spin from the ISCO radius via figure 1.2). Luckily, the techniques needed for measuring distance (Reid et al. 2011), mass, and inclination (Cantrell et al. 2010; Orosz et al. 2011a,b) for binary systems are well known, and to date have been successfully applied to about ten black hole binary systems (McClintock et al. 2011). In table 1.1 below, we list a few recent measurements of black hole spins as derived from the continuum fitting method.

1.3.2 Zoology of Disk States

In all previous discussions, we ignored the fact that the process of accretion is a time dependent phenomenon. However, long term observations of black hole binary systems reveal a rich variety of different accretion phases that each source cycles through.
CHAPTER 1. INTRODUCTION

Figure 1.5 shows a few example spectra and power density spectra associated with the 3 primary states that stellar-mass black holes are observed to fall under (see McClintock & Remillard 2006; Remillard & McClintock 2006b for an in-depth primer). Quantitatively, four parameters are used to distinguish between the states, as listed below:

- the disk fraction $f$, which is the ratio of the thermal disk flux to the total (both unabsorbed) over the 2-20keV band
- the power law index $\Gamma$ of the high energy component of the flux
- the fractional Root-Mean-Squared (RMS) power $r$ in the power density spectrum integrated from 0.1-10Hz compared to the average source count rate
- the integrated RMS power $q_{\text{max}}$ of any detected quasi-periodic-oscillation (QPO) feature that is seen in the power density spectrum

The “thermal” state is characterized by strong continuum emission in the low energy bands that is attributed to heat radiation from an optically thick accretion disk. The overall fluctuations in the flux have a low RMS and are featureless, indicating a steady accretion flow. There is also usually a weak high energy power law tail in the 2-20keV band.
Figure 1.5: Examples of the three spectral states for black hole binaries. Absorption corrected Spectral Energy Distributions (SED) are shown in the left panels, and their corresponding power density spectra are shown on the right. The SEDs are decomposed with a three component fit: thermal emission (red, solid), power-law (blue, dashed), and relativistically broadened Fe K\(\alpha\) line (black, dotted). Figure Credit: Remillard & McClintock 2006b
CHAPTER 1. INTRODUCTION

Strong power law emission with slope $\Gamma \sim 1.7$ in the hard X-ray bands characterizes the “hard” state of black hole binaries. Here, the fluctuations in the power spectrum are significant ($r > 0.1$), and the prevailing picture is a hot accretion flow with negligible thermal disk (Yuan & Narayan 2014). The hard state is also associated with the presence of a quasi-steady radio jet, which produces strong correlations between X-ray and radio flux of the system.

We also have the “steep power law” state, which shares many similarities to the thermal state, but with a much stronger power law component and a steeper photon index $\Gamma \sim 2.5$. The differences between SPL and thermal are a fainter thermal disk component in the former (SPL typical values $f \sim 50\%$), and a less variable but higher sloped photon index in the latter. Perhaps the most exciting feature is the frequent detection of strong QPOs in the SPL state. Typically, systems are found in the SPL state when the luminosity approaches Eddington.

Finally, there is the quiescent state corresponding to extremely faint luminosities (typically, more than 3 orders of magnitude fainter than the other states), where the spectrum is nonthermal. The quiescent state is important since it allows for robust mass determinations for these binary systems. The low disk luminosities allow the emission from the secondary star to become prominent, which leads to accurate radial velocity measurements.
Table 1.2: Nomenclature and definitions for black hole states.

<table>
<thead>
<tr>
<th>Black Hole State: (State Alias)</th>
<th>Quantitative Classification Criteria$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thermal (High/Soft)</td>
<td>Disk fraction $f &gt; 75%$ QPOs weak or absent: $q_{\text{max}} &lt; 0.005$ Power continuum level $r &lt; 0.075$</td>
</tr>
<tr>
<td>H1743-322 (Low/Hard)</td>
<td>Disk fraction $f &lt; 20%$ Power law index $1.4 &lt; \Gamma &lt; 2.1$ Power continuum level $r &gt; 0.1$</td>
</tr>
<tr>
<td>Steep Power Law (Very High)</td>
<td>Either $f &lt; 80%$ and with QPOs: $q_{\text{max}} &gt; 0.01$ or $f &lt; 50%$ with QPOs absent Significant power law component, $\Gamma &gt; 2.4$ Power continuum level $r &lt; 0.15$</td>
</tr>
<tr>
<td>Quiescent</td>
<td>Extremely low luminosity $L/L_{\text{Edd}} &lt; 10^{-3}$ Nonthermal power law emission $\Gamma \sim 1.5 - 2$</td>
</tr>
<tr>
<td>Intermediate</td>
<td>Any remaining systems that do not fall in above categories (typically during state transitions)</td>
</tr>
</tbody>
</table>

$^a$ See parameter definition list in §1.3.2
CHAPTER 1. INTRODUCTION

Table 1.2 summarizes all the states of accretion in black hole binary systems. The thermal state (associated with a steady accretion flow) is the only one that is well understood from a theoretical standpoint (see the next section §1.4 for an overview of the classic picture of disk physics). The complex time variability and multi-component nature of the other states have made them difficult to analyze, motivating the need for better accretion disk models. An active area of research is the study of the disk corona, which is thought to be responsible for the power law emission at high energies. Open questions remain about the corona’s geometry, strength, and generating mechanism that we hope will be elucidated through better numerical modeling of the disk and beyond (i.e. jets, winds, and corona).

1.4 Disk physics

Since a black hole does not radiate\textsuperscript{10}, our only source of information comes from its interaction on observable background/companion objects. In particular, the case of a nearby massive companion star reaching the end of its life turns out to be the ideal environment for detecting black holes. The star releases strong winds, overflows its Roche lobe, and undergoes mass transfer onto the black hole.

This gradual inspiraling of material donated by the companion star forms an accretion disk around the black hole. The disk has a few key properties that make it useful for probing black hole physics: 1) it radiates incredibly brightly due to the high temperatures induced by viscous processes in the disk making it easy to detect

\textsuperscript{10}Ignoring Hawking radiation since its emission rate is negligible for astrophysical black holes
observationally; and 2) it reaches deep in the gravitational potential of the black hole, with an inner edge at $r_{\text{ms}}$ as discussed in §1.2.5 and thereby probing the multitude of relativistic effects active in this region (frame dragging, doppler boosting, gravitational redshifting, light bending, etc). Thus, an understanding of the accretion process is vitally important if one wishes to connect black hole theory with observations.

1.4.1 Classic Thin Disk Model

It has been more than four decades since the landmark work of Shakura & Sunyaev (1973) was published, giving birth to the whole subfield of accretion disk physics. Since then many refinements have been made to the models, becoming ever more accurate and precise (e.g. Pringle & Rees 1972; Lynden-Bell & Pringle 1974; Novikov & Thorne 1973; Page & Thorne 1974; Pringle 1981, also see frank2002, kato08 for a thorough primer). The standard assumptions made for modeling the accretion disk are:

- The disk is stationary and axisymmetric
- The disk is geometrically thin ($h/r \ll 1$)
- Disk self gravity is negligible compared to that of the black hole
- No magnetic fields
- Radial and vertical structures can be decoupled

Given these simplifications, the goal is to solve for both the radial and vertical structure of the disk obeying the usual physical constraints for gaseous astrophysical systems. For a specific black hole configuration (i.e. black hole mass $M$, spin $a$, and
accretion rate $\dot{M}$), as well as for a particular location in accretion disk $r > r_{\text{ms}}$, the game is to simultaneously solve a system of equations that govern the physics of accretion. For simplicity and to close the set of equations, one can simply assume a crude “one-zone” model for the vertical structure and solve for the radial dependence of all disk quantities. As we will see later, due to the polynomial nature of the various physics equations that set the system, the final solution simplifies and takes the form of a series of power-law scalings for each of the physical quantities.

**Radial Structure**

Amazingly, it is possible to deduce the radial energetics of the system with only very few assumptions. Although viscous angular momentum transport ultimately governs the diffusion of material within the accretion disk, a precise understanding of viscosity within the disk is not needed provided that the resultant radial transport of material is slow (i.e. whenever the viscous radial advection timescale greatly exceeds the vertical thermal diffusivity timescale). Under this assumption, the flux of energy out of the disk does not depend on the precise nature of viscosity – it is simply the rate at which gravitational potential energy is dissipated and released via local radiative cooling of the accreting matter.

We illustrate the exercise in the Newtonian limit for simplicity. Given a disk with mass accretion rate $\dot{M}$ and assuming the gas follows Keplerian circular orbits with orbital frequency $\Omega_K = \sqrt{GM/r^3}$, the system of equations that must be solved to yield
the energy flux is simply:

\[
\frac{d\dot{M}}{dr} = 0, \quad \text{(conservation of mass)} \tag{1.33}
\]

\[
\dot{M} \frac{d(\Omega_K r^2)}{dr} = \frac{d(2\pi r^2 W)}{dr}, \quad \text{— of angular momentum} \tag{1.34}
\]

\[
F = \frac{d(3\Omega_K W)}{dr}, \quad \text{— of energy} \tag{1.35}
\]

where \( W \) represents the vertically integrated shear stress \( t_{r\phi} \) of the disk

\[
W(r) = \int t_{r\phi}(r,z)dz. \tag{1.36}
\]

Solving Eqs. 1.33-1.36 for \( F \) yields the classic disk solution:

\[
F(r) = \frac{3GM\dot{M}}{8\pi r^3} \left(1 - \sqrt{\frac{r}{r_{in}}} \right), \tag{1.37}
\]

where the constant of integration is set by \( r_{in} = r_{ms} \), corresponding to the innermost radius of the disk where the flux vanishes.

**Vertical Structure**

Now, armed with the radial flux profile, one can solve for various other physical quantities that relate to the vertical structure of the disk. The usual approach is to introduce height integrated quantities to collapse the vertical structure, making the system of equations tractable algebraically (e.g. defining a vertically integrated column density \( \Sigma \equiv 2H\rho \), where \( H \) is some characteristic height of the disk and \( \rho \) is the central gas density). The key equations relating to vertical structure are:

\[
\frac{dP_{tot}}{dz} = -\rho g_z, \quad \text{(hydrostatic balance)} \tag{1.38}
\]

\[
F = -\frac{4a c T^3}{3\kappa \rho} \frac{\partial T}{\partial z}, \quad \text{(radiative flux)} \tag{1.39}
\]

\[
\dot{M} = -2\pi r \int v_r \rho dz \quad \text{(mass accretion)}, \tag{1.40}
\]
CHAPTER 1. INTRODUCTION

where \( P_{\text{tot}} = P_{\text{gas}} + P_{\text{rad}} \) is the total pressure, \( g_z \sim \Omega_K H \) is the vertical disk gravity, \( T \) is the disk temperature, \( \kappa \) is the opacity law (typically a combination of free-free and electron scattering), and \( v_r \) is the characteristic radial advection velocity of the gas.

Closure Equations

A few more equations are needed to close the system of equations and yield a complete disk solution:

\[
\text{Equation of State : } \quad P_{\text{tot}}(\rho, T) \quad (1.41) \\
\text{Opacity Law : } \quad \kappa(\rho, T) \quad (1.42) \\
\alpha\text{-Prescription for Stress : } \quad t_{r\phi} = \alpha P_{\text{tot}} \quad (1.43)
\]

The last closure relation with a constant \( \alpha \) is simply an ansatz that was originally motivated by a combination of dimensional analysis and algebraic convenience. Luckily, it has been borne out by recent investigations of computer simulations of turbulence (Stone et al. 2008; Hirose, Blaes & Krolik 2009; Hirose, Krolik & Blaes 2009; Jiang et al. 2012, 2013). It appears to hold remarkably well for a whole host of accretion disk simulation parameters, with \( \alpha \sim 0.01 - 0.1 \) depending on the details of simulations.

Finally, combining all the vertically integrated forms of the physics equations along with the three closure relations results allows one to solve for all relevant accretion disk quantities. For the sake of brevity, we only list below the relativistic solution worked out by Novikov & Thorne (1973) for the disk properties and Page & Thorne (1974) for the flux profile. Please refer to Shakura & Sunyaev (1973) for Newtonian disk solutions.

31
1.4.2 Relativistic Disk Model (Novikov & Thorne 1973)

Although the previous discussion invoked Newtonian expressions for the physics equations, the relativistic case follows the same structure. The only difference is the presence of a few additional metric factors that slightly modify the behaviour close to the black hole. Novikov & Thorne (1973) worked out solutions for three regimes of the disk that we reproduce below using the scalings $m = M/M_\odot$ and $\dot{m} = \dot{M}c^2/L_{Edd}$:

Outer region: $P = P_{\text{gas}}, \kappa = \kappa_{ff}$ (free-free opacity)

$$F = [7 \times 10^{26} \text{ erg cm}^{-2} \text{ s}^{-1}] (m^{-1} \dot{m}) r^{-3} B^{-1} C^{-1/2} Q$$

$$\Sigma = [4 \times 10^5 \text{ g cm}^{-2}] (\alpha^{-4/5} m^{1/5} \dot{m}^{7/10}) r^{-3/4} A^{1/10} B^{-4/5} C^{1/2} D^{-17/20} E^{-1/20} Q^{7/10}$$

$$H = [4 \times 10^2 \text{ cm}] (\alpha^{-1/10} m^{9/10} \dot{m}^{3/20}) r^{9/8} A^{19/20} B^{-11/10} C^{1/2} D^{-23/40} E^{-19/40} Q^{3/20}$$

$$\rho_0 = [4 \times 10^2 \text{ g cm}^{-3}] (\alpha^{-7/10} m^{-7/10} \dot{m}^{11/20}) r^{-15/8} A^{-17/20} B^{3/10} D^{-11/40} E^{17/40} Q^{11/20}$$

$$T = [2 \times 10^8 \text{ K}] (\alpha^{-1/5} m^{-1/5} \dot{m}^{3/10}) r^{-3/4} A^{-1/10} B^{-1/5} D^{-3/20} E^{1/20} Q^{3/10} \quad (1.44)$$

Middle region: $P = P_{\text{gas}}, \kappa = \kappa_{es}$ (electron scattering opacity)

$$F = [7 \times 10^{26} \text{ erg cm}^{-2} \text{ s}^{-1}] (m^{-1} \dot{m}) r^{-3} B^{-1} C^{-1/2} Q$$

$$\Sigma = [9 \times 10^4 \text{ g cm}^{-2}] (\alpha^{-4/5} m^{1/5} \dot{m}^{3/5}) r^{-3/5} B^{-4/5} C^{1/2} D^{-4/5} Q^{3/5}$$

$$H = [1 \times 10^3 \text{ cm}] (\alpha^{-1/10} m^{9/10} \dot{m}^{1/5}) r^{21/20} A^{6/5} B^{1/2} C^{-3/5} D^{1/2} E^{1/2} Q^{1/5}$$

$$\rho_0 = [4 \times 10^1 \text{ g cm}^{-3}] (\alpha^{-7/10} m^{-7/10} \dot{m}^{2/5}) r^{-33/20} A^{-1} B^{3/5} D^{-1/5} E^{1/2} Q^{2/5}$$

$$T = [7 \times 10^8 \text{ K}] (\alpha^{-1/5} m^{-1/5} \dot{m}^{2/5}) r^{-9/10} B^{-2/5} D^{-1/5} E^{2/5} Q^{2/5} \quad (1.45)$$
CHAPTER 1. INTRODUCTION

Inner region: \( P = P_{\text{rad}}, \kappa = \kappa_{\text{es}} \)

\[
F = [7 \times 10^{26} \text{ erg cm}^{-2} \text{s}^{-1}] (m^{-1} \dot{m}) r^{-3} \mathcal{B}^{-1} \mathcal{C}^{-1/2} \mathcal{Q}
\]

\[
\Sigma = [5 \text{ g cm}^{-2}] (\alpha^{-1} m^{-1} r^{3/2} \mathcal{A}^{-2} \mathcal{B}^{3} \mathcal{C}^{1/2} \mathcal{E} \mathcal{Q}^{-1}
\]

\[
H = [1 \times 10^{5} \text{ cm}] (\dot{m}) \mathcal{A}^{2} \mathcal{B}^{-3} \mathcal{C}^{1/2} \mathcal{D}^{-1} \mathcal{E}^{-1} \mathcal{Q}
\]

\[
\rho_{0} = [2 \times 10^{-5} \text{ g cm}^{-3}] (\alpha^{-1} m^{-1} \dot{m}^{-2}) r^{3/2} \mathcal{A}^{-4} \mathcal{B}^{6} \mathcal{D} \mathcal{E}^{2} \mathcal{Q}^{-2}
\]

\[
T = [5 \times 10^{7} \text{ K}] (\alpha^{-1/4} m^{-1/4}) r^{-3/8} \mathcal{A}^{-1/2} \mathcal{B}^{1/2} \mathcal{E}^{1/4}
\]

where the radial functions are defined as: (in terms of dimensionless radius \( x = \sqrt{r/M} \) and dimensionless spin \( a_{\ast} = a/M \):

\[
\mathcal{A} = 1 + a_{\ast}^{2} x^{-4} + 2 a_{\ast}^{2} x^{-6}
\]

\[
\mathcal{B} = 1 + a_{\ast} x^{-3}
\]

\[
\mathcal{C} = 1 - 3 x^{-2} + 2 a_{\ast} x^{-3}
\]

\[
\mathcal{D} = 1 - 2 x^{-2} + a_{\ast}^{2} x^{-4}
\]

\[
\mathcal{E} = 1 + 4 a_{\ast}^{2} x^{-4} - 4 a_{\ast}^{2} x^{-6} + 3 a_{\ast}^{4} x^{-8}
\]

\[
\mathcal{Q} = \frac{1 + a_{\ast} x^{-3}}{x(1 - 3 x^{-2} + 2 a_{\ast} x^{-3})^{1/2}} \left[ x - x_{0} - \frac{3}{2} a_{\ast} \ln \left( \frac{x}{x_{0}} \right) - \frac{3(x_{1} - a_{\ast})^{2}}{x_{1}(x_{1} - x_{2})(x_{1} - x_{3})} \ln \left( \frac{x - x_{1}}{x_{0} - x_{1}} \right) \right]
\]

\[
\frac{3(x_{2} - a_{\ast})^{2}}{x_{2}(x_{2} - x_{1})(x_{2} - x_{3})} \ln \left( \frac{x - x_{2}}{x_{0} - x_{2}} \right) - \frac{3(x_{3} - a_{\ast})^{2}}{x_{3}(x_{3} - x_{1})(x_{3} - x_{2})} \ln \left( \frac{x - x_{3}}{x_{0} - x_{3}} \right) \right] (1.47)
\]

where the cofactors are set by:

\[
x_{0} = \sqrt{r_{\text{ms}}/M}
\]

\[
x_{1} = 2 \cos \left[ (\cos^{-1} a_{\ast} - \pi)/3 \right]
\]

\[
x_{2} = 2 \cos \left[ (\cos^{-1} a_{\ast} + \pi)/3 \right]
\]

\[
x_{3} = -2 \cos \left[ (\cos^{-1} a_{\ast})/3 \right]
\]

(1.48)
CHAPTER 1. INTRODUCTION

Due to its prevalence in the black hole accretion disk literature, we will make extensive
use of this model in the remaining chapters of this thesis.

1.4.3 Open problems in Accretion Physics

Although the theory of accretion discs has been thoroughly explored over the past several
decades (i.e. above, we have outlined the physics for the family of “thin” accretion disks),
several of the assumptions fundamental to the classic disk model have been called into
question. One issue is the fate of the material that reaches the ISCO – the conventional
assumption is that the gas plunges into the black hole sufficiently rapidly so as to leave
no signature of its final descent. However, it is possible for this gas to be viscously
or magnetically coupled to the gas in or above the accretion disk proper, resulting to
modified emission profiles – we explore this idea extensively in Chapters 2 and 5. The
validity of the $\alpha$-prescription has also been called into question in recent years, especially
the assumption of constancy with all radii. Recent work by Penna et al. (2013) suggests
that more realistic models of accretion disks should have $\alpha$ vary with radius, with large
radii having lower values of $\alpha$. There is also a question of how the magnetic structure
(neglected in the classic disk models) can affect the overall structure and appearance of
the disk. Recent simulations have shown that magnetic buoyancy can act to lift material
away from the disk midplane (Jiang et al. 2012), and can result in modifications to the
emission profile of the disk (Davis et al. 2010).

The case of sub-Eddington accretion is another hotly debated topic in the literature.
Here, the disk is governed by an Advection Dominated Accretion Flow (ADAF) solution,
which was initially explored by the likes of Lightman, Eardley, and Rees early in the
1970’s. Interest in this topic was later revived in the 1990’s (Narayan & Yi 1994; Abramowicz et al. 1995). In recent years, various groups have proposed different mechanisms that can achieve the radiatively inefficient property of the ADAF such as the advection dominated inflow-outflow solution (ADIOS – Blandford & Begelman 1999) or the convection dominated accretion flow (CDAF – Narayan, Igumenshchev & Abramowicz 2000), which has led to an ongoing debate as to which of the models is the most physically realistic. It is also an open question (both observationally and theoretically) as to how these systems transition from the standard thin accretion disk state to the ADAF state – the current state of affairs is described in Yuan & Narayan 2014.\textsuperscript{11}

Perhaps the least understood aspect of disk models relates to the case of super-Eddington accretion, where the radiative output of the disk can become so extreme as to limit/halt the supply of infalling matter. This phase of accretion is crucial for understanding the AGN population and the quasar luminosity function since it acts as a limiter on black hole growth. Recent observations of supermassive quasars at high redshift (e.g. $1.2 \times 10^{10} M_\odot$ at redshift $z = 6.3$; Wu et al. 2015) suggest that the Eddington limit can be broken (otherwise, how could such massive black holes exist at such early times?). This is also borne out in recent numerical simulations of accreting black holes (Sadowski & Narayan 2015b) suggests that the apparent luminosity can reach as high as thousands of Eddington units. However, there is no consensus yet on how far the Eddington limit can be exceeded and is currently an area of active research.

\textsuperscript{11}One idea (posited by Narayan and collaborators) is that at low accretion rates, say $\dot{m} \sim 0.01$, the innermost parts of the usual thin accretion disk puff up into the ADAF state, with a transition radius being set according to $\dot{m}$.
CHAPTER 1. INTRODUCTION

Finally, there is a question of how jets are formed, and where the energy that powers the jet is being extracted from. One idea that is that the black hole rotational energy could be tapped to power a jet. The Blandford & Znajek (1977) mechanism is the leading model for explaining how jets can be launched as a consequence of a black hole’s spin interaction with its embedded magnetic field. This jet formation mechanism has been verified in numerical simulations of accretion (Tchekhovskoy et al. 2011; Tchekhovskoy & McKinney 2012). Alternative ideas include magnetically driven winds in the disk collimating to become a jet (Blandford & Payne 1982), or radiatively driven jets (Sądowski & Narayan 2015b).

Ultimately, due to the complexity and nonlinearity of the physics involved (i.e. electromagnetism, relativity, radiative transport, fluid-dynamics, turbulence), computer simulations give perhaps the best shot at understanding the true nature of the accretion process.

1.5 Numerical Simulations

The physics of accretion is a messy business; the precise details are the result of the interplay between the physics of general relativity (due to proximity to a compact object), hydrodynamics (since the accretion flow is a gaseous fluid), magnetism (the magneto-rotational instability is the chief driver of turbulence, and hence viscosity), and radiation (how we ultimately observe the disks). Finally there is the issue of accretion disks being multidimensional objects, coupled with the fact that disk MHD turbulence has to be studied in three dimensions with high resolving power. All this contributes to a significant computational expense. Another factor adding to the difficulty is the large
range of dynamical scales in the problem – although most of the action and energetics
are concentrated close to the black hole with a rather short characteristic timescale, the
gas in the disk flows in from much farther out, evolving on a much long viscous time at
large radii. Due to these challenges, it is only within the last decade that it has been
possible to tackle this difficult problem computationally (and even still, most current
codes only consider a subset of the physics!).

To date, there are a wide variety of relativistic hydrodynamic codes available. A
partial list includes: ATHENA (Stone et al. 2008), Cosmos++ (Anninos et al. 2005),
ECHO (Del Zanna et al. 2007), HARM (Gammie et al. 2003), HERACLES (Gonzáles,
Audit, & Huynh 2007), KORAL (Sadowski et al. 2013), RAISHIN (Mizuno et al. 2006),
ZEUS (Hayes & Norman 2003) etc... Typically, the codes fall into one of two categories:
1) local shearing box simulations, which are designed to focus on resolving the onset of
the turbulent dynamo at small scales, or 2) global simulations that seek to capture the
full three-dimensional structure and evolution of disk material.

The approach taken for handling fluid dynamics also spans multiple paradigms.
The most common approaches include artificial viscosity schemes (Wilson 1972), and
Godunov-type methods that make use of approximate or exact Riemann solvers. In
either case, finite difference representations are used for handling the general relativistic
hydrodynamic equations. The advantage of the artificial viscosity technique is that it is
more straightforward to implement, can be easily extended to include additional physics,
and is less expensive than Godunov schemes. However, Godunov schemes enjoy the
distinction of being fully conservative by design (and thus potentially more accurate) even
in the case of ultrarelativistic flows. They also require less tuning since a viscosity term
does not have to be set. However the tradeoff is that the system then becomes subject
to numerical viscosity, which depends sensitively on the simulation setup (i.e. spatial grid resolution and accuracy of the Riemann solver). Aside from these two primary approaches, other schemes include smooth-particle hydrodynamics (Lucy 1977; Gingold & Monaghan 1977; Springel, Yoshida & White 2001) and spectral methods (Canuto et al. 1988), though they are less well-developed in the field of relativistic accretion.

1.5.1 Shearing Boxes

For the problem of simulating numerical accretion disks, a decision must be made as to whether the problem should be tackled piecewise locally, or in a global sense. The former case lends itself to “shearing box” simulations, where the accretion flow is studied in detail locally and mimics the shear present in a small rectangular patch of disk. All disk structure is handled via appropriate choice of boundary conditions, ignoring the remaining large scale structure of the disk. The most obvious advantage of this approach is that by focusing only on a small patch, high spatial resolutions can be achieved, allowing the turbulent process to be completely resolved.

Perhaps the most significant breakthrough resulting from shearing-box simulations has been the discovery and exploration of the magneto-rotational instability (MRI) in accretion disks (Balbus & Hawley 1991, 1998). This instability is able to amplify any poloidal seed magnetic field, growing the field strength exponentially until it reaches a nonlinear turbulent saturation state. It resolved a longstanding problem in accretion disk theory, explaining the mechanism for the turbulent fluctuations that give rise to the viscous transport of material in discs. Prior to the MRI, all other mechanisms (e.g. convection, gas-dynamical viscosity) were found to be far too weak to account for the
observed properties of accretion disks. Generally, shearing box simulations reproduce viscosity laws that resemble the $\alpha$-prescription as mentioned in §1.4.1, with typical values $\alpha \sim 0.05$ (Blackman et al. 2008; Guan et al. 2009; Hawley et al. 2011; Sorathia et al. 2012) although it has been found to depend weakly on the net initial magnetic field (Hawley et al. 1995; Pessah et al. 2007) and also the numerical setup/resolution (Fromang & Papaloizou 2007; Davis et al. 2010; Bai & Stone 2013).

Recently, shearing boxes have been useful for investigating properties of radiatively dominated discs. By including a radiation scheme known as Flux Limited Diffusion (FLD), Turner and Hirose were able to study the evolution of vertical structure within accretion disks. However recently, tension has mounted among different groups for the topic of disk stability in the radiation dominated limit (Jiang vs Hirose).

### 1.5.2 Global Simulations

Simulating the accretion problem globally is a far more computationally taxing problem due to the huge range of dynamical scales involved. Compared to shearing-box simulations, sacrifices must be made for the sake of computational expediency, such as much lower spatial resolutions, dropping some of the physics (e.g. ignoring radiation), or reducing dimensionality. However, global simulations are necessary since they are the only method that can capture the full effects of relativity on the accretion disk as well as connecting the turbulent evolution of the magnetic field with properties of the disk at large.

The earliest works for simulating accretion discs in a global sense were that of Wilson, who pioneered the field when he numerically considered the hydrodynamical
problem of spherical accretion of material with non-zero angular momentum. Wilson found that the centrifugal barrier induced by the material’s angular momentum led to the formation of a fat non-accreting torus. Interest in the field of numerical global accretion has grown over time due to the recent trend of exponential growth in computational power. The last decade has seen the rise of a wide variety of magneto-hydrodynamic (MHD) codes tailored for the problem of accretion. The focus on MHD is well-motivated since magnetic fields play such a pivotal role in accretion – acting as both a generator for the turbulent motions (via MRI) and as a mechanism for launching and confining jets (Blandford-Znajek).

This has spurred a decade of furious activity in the field, with early work carried out in nonrelativistic frameworks although taking advantage of a pseudo-Newtonian potential to simulate some of the relativistic effects (Armitage 1998; Hawley & Krolik 2001; Igumenshchev et al. 2003). General-relativistic formulations of these simulations were developed soon after (Koide et al 1999; De Villiers & Hawley 2003; Gammie et al. 2003; Fragile et al. 2007) and most of the recent discoveries on disk/jet physics are based on these codes.

1.5.3 Future Directions

To date, most numerical simulations of accretion have either ignored radiation altogether or have implemented extremely crude/unrealistic models of the radiation. This is a deliberate choice and reflects the difficulty of the physics involved. Since the radiation field is a seven dimensional quantity\textsuperscript{12}; it is quite complex to model and poses a huge

\textsuperscript{12}\textup{Radiation involves three spatial, two angular, one time, and one frequency dimension}
challenge in terms of computational resources (see §1.6 for more details). It is only in the past few years that radiatively coupled MHD simulations have become feasible, initially without including relativity (Ohsuga et al. 2009; Ohsuga & Mineshige 2011; Jiang, Stone & Davis 2014). The current push in the numerical accretion community is to tackle the full GRMHD and radiation problem (Sadowski et al. 2013; McKinney et al. 2014).

1.6 Including Radiation

So far, we have seen a decade of success in the realm of general relativistic magneto hydrodynamic (GRMHD) codes (Gammie et al. 2003; DeVilliers & Hawley 2003; Shafee et al. 2008; Noble et al. 2009; Penna et al. 2010; McKinney et al. 2012; Narayan et al. 2012). However, as already mentioned, radiation has typically been neglected for the sake of computational speed. Although there are several prior examples of radiation hydrodynamic codes in the literature, none were capable of handling the full general relativistic problem in three-dimensions. Nevertheless, due to the exponential growth of computing power in recent times, we are now poised to tackle this problem in full-force (i.e. including radiation). Presently, radiation hydrodynamics is an active area of research being pursued by several accretion disk research groups (and our group is no exception!); it is the final unexplored frontier in numerical accretion.

Generally, codes that model radiation fall under four broad categories, distinguished by their radiative transport algorithms. These four numerical treatments of radiation transport are:

- **Diffusion** – Radiation is treated as simply another independent relativistic fluid.
CHAPTER 1. INTRODUCTION

The radiation energy density is the key parameter that dictates its flow and transport.

- **Discrete Ordinates** – The radiation field is interpreted as a finite collection of rectilinear light rays (i.e. light is modeled to travel along a finite set of angles).

- **Spherical Harmonics** – Decompose the radiation field into various spherical harmonic moments. These moments can then be easily evolved in time independently as the propagation of spherical waves.

- **Monte Carlo** – Introduce and track a large number of individual simulated photons.

1.6.1 Radiation Hydrodynamics

Historically, radiation hydrodynamics has been successfully applied in several other astrophysical contexts, such as the cores of collapsing supermassive stars (to understand the ignition mechanism for supernovae – Nordhaus et al 2010), star formation within the ISM (Haworth & Harries 2012), and models of the solar corona (e.g. flares – Fisher 1986). However, the added complexity and computational cost of marrying radiation, fluid mechanics, and general relativity has resulted in a paucity of codes capable of tackling the problem of BH accretion. To date, there are only a handful of groups concurrently developing codes to handle radiative hydrodynamics, either in a local shearing box (Hirose, Blaes & Krolik 2009; Jiang et al. 2012), or in a global fully realized accretion disk (Farris et al. 2008; Ohsuga & Mineshige 2011; Fragile et al. 2012).
CHAPTER 1. INTRODUCTION

Figure 1.6: The radiation temperature distribution in the ‘hohlraum’ problem (radiation leaking into a cavity with a central rectangular block) as computed by Brunner (2002). The panels represent results calculated via the four paradigms: a) Diffusion, b) Discrete Ordinates, c) Spherical Harmonics, and d) Monte Carlo.

For the purposes of relativistic hydrodynamics, the radiative diffusion model is the natural choice as it is in essence a fluid treatment of radiation (making it easy to merge with the fluid treatment of gas). Additionally, the other methods suffer from various defects that make them unattractive for our purposes (see Fig. 1.6 for a visual
CHAPTER 1. INTRODUCTION

comparison of the different methods). Discrete ordinates has the drawback of ray-defects and having poor computational scaling in higher dimensions (since the number of rays needed per cell grows too quickly with increasing dimension). For the problem of accretion around a compact object, the method of spherical harmonics cannot be used; In a curved spacetime, it is unclear how one should go about decomposing the radiation field into its various moments. Finally, monte carlo methods are too slow, and suffer from greatly uneven error bounds across different simulation cells due to the nature of Poisson noise (i.e. photon starved areas experience large numerical errors).

However, the diffusion paradigm is not without its own drawbacks; it has difficulty in resolving sharply shadowed regions (sharp gradients are the bane of any diffusive treatment – see the distinct lack of shadow in panel (a) in Fig. 1.6). We hope to overcome this defect by use of the ‘M1 closure scheme’, which is a higher order treatment of diffusion that allows for anisotropic radiation fields and hence shadows (for details, see Dubroca & Feugeas 1999; Gonzáles, Audit, & Huynh 2007; Sądowski et al. 2013 – see Fig. 1.7 for a comparison between M1 closure and standard isotropic diffusion).
CHAPTER 1. INTRODUCTION

Figure 1.7: The radiation temperature distribution in a shadowing test, where highly beamed light shines on an optically thick sphere. Panel a) is the result from the simple diffusive treatment of radiation, whereas b) is the result from M1 closure (as calculated by González, Audit, & Huynh 2007). Notice that the M1 closure scheme is able to accurately handle shadowing, despite being a diffusive treatment. Panel c) shows the results from Sadowski et al. (2013), also using M1.

1.7 Chapter Summaries

Chapter 2 compares spectral calculations of GRMHD accretion disks from (Penna et al. 2010) to those of the classic (Novikov & Thorne 1973) disk models. The GRMHD disk simulations capture all but the radiation physics for magnetized flow around a black
CHAPTER 1. INTRODUCTION

hole. We apply radiative transfer post-processing to get simulated images of the discs making use of the 1D radiative transfer solutions of TLUSTY (Hubeny & Lanz 1995). This is an advance over previous work on computing GRMHD disk spectra (Kulkarni et al. 2011), which only assumed (modified) blackbody annuli spectra. The following are the key results from our investigation:

(1). The GRMHD based accretion discs have hotter spectra than the standard Novikov & Thorne (1973) discs. The GRMHD discs produce more luminosity everywhere, and the contrast becomes most apparent inside the ISCO.

(2). The increased luminosity of the GRMHD discs compared to the classic NT discs induces a modest systematic bias in the derived spins of these GRMHD discs. For black holes of spin \( a_* = 0, 0.7, 0.9 \) the spin deviation is \( \Delta a_* \sim 0.15, 0.07, 0.03 \) in the worst cases (corresponding to inclination angles of \( 75^\circ \)).

(3). The GRMHD discs around spinning black holes exhibit a weak high-energy power-law tail. This power-law tail arises from the combined emission of the hot plunging region gas. The strength of this plunging region power-law increases with the system’s inclination angle.

In Chapter 3, we develop a simple one-zone model for black hole accretion discs that include both convective and turbulent vertical energy transport channels. We find that the action of mixing from convection and turbulence provides a stabilizing effect on the
disk, pushing the threshold for thermal instability up towards higher accretion rates. In some cases, we find that turbulent mixing pushes the threshold for instability far above 10 per cent Eddington – even inducing complete stability in the most extreme cases. This is in agreement with previous studies that show convective mixing protects against the onset of thermal instability. However since MRI-induced turbulence is much more vigorous than convective turbulence, it is not surprising that turbulently mixed discs experience a much stronger version of this stabilizing effect.

Chapter 4 describes HERO, a new general relativistic radiative transfer code. It is primarily designed to model the radiation field for accretion flows around black holes. The unique features of HERO are: 1) it uses a hybrid short/long characteristics radiative solver; and 2) it is implemented in a general relativistic framework that takes into account the effects of light bending, doppler beaming, and gravitational redshifting. HERO is written as a post-processing code decoupled from the hydrodynamic evolution of the fluid. It computes the time-independent radiation field assuming a given fixed background fluid structure. To verify that HERO produces physically correct answers, we have performed a comprehensive set of tests designed to examine the code’s convergence properties, accuracy, and capability to handle multidimensional relativistic problems. We also confirm the well known result that 2D and 3D problems with compact sources suffer from significant ray defects in the far field when analyzed with the short characteristics method. We present an approximate fix which mitigates the effects in the case of a 3D spherical grid. However, for accurate results, we find that it is necessary to switch to a long characteristics solver which is unaffected by ray-defects.

Chapter 5 covers the Comptonization module that we develop for HERO. Comptonization is handled through a quadratic Kompaneets based solver that
CHAPTER 1. INTRODUCTION

evolves light rays by means of a modified Comptonized source function. We test our
Comptonization module by comparing with exact solutions of light diffusing through
purely scattering media. We find excellent agreement between HERO and the exact
solutions for both 1D and 3D test problems. We also apply our code to the astrophysical
problem of radiation emerging from an accretion disk. The self consistent temperature
and radiation fields throughout the accretion disk are calculated taking into account
relativistic effects, returning disk radiation, and Compton cooling. The complete solution
results in a two-phase disk structure, with a hot spherical corona surrounding a cooler
thermal disk. Finally, performing raytracing on the solution yields integrated disk
spectra with slightly higher spectral hardening factors compared to previous 1D based
estimates. We also verify the Zhu et al. (2012) result of a high-energy power law tail
feature emerging in the spectrum of a black hole accretion disk.
Chapter 2

The Eye of the Storm:
Light from the Inner Plunging
Region of Black Hole Accretion Discs

This thesis chapter originally appeared in the literature as
Y. Zhu, S.W. Davis, R. Narayan, A.K. Kulkarni, R.F. Penna,

Abstract

It is generally thought that the light coming from the inner plunging region of black hole accretion discs contributes negligibly to the disc’s overall spectrum, i.e. the plunging fluid is swallowed by the black hole before it has time to radiate. In the standard disc
model used to fit X-ray observations of accretion discs, the plunging region is assumed to be perfectly dark. However, numerical simulations that include the full physics of the magnetized flow predict that a small fraction of the disc’s total luminosity emanates from this plunging region. In this work, we investigate the observational consequences of this neglected inner light. We compute radiative transfer based disc spectra that correspond to 3D general relativistic magnetohydrodynamic simulated discs (which produce light inside their plunging regions). In the context of black hole spin estimation, we find that this neglected inner light only has a modest effect (this bias is less than typical observational systematic errors). For rapidly spinning black holes, we find that the combined emission from the plunging region produces a weak power-law tail at high energies. This indicates that infalling matter is the origin for some of the ‘coronal’ emission observed in the thermal dominant and steep power-law states of X-ray binaries.

2.1 Introduction

The ubiquity and simplicity of black holes in our universe make them truly marvelous objects of study. The complete physics of each and every black hole (BH) can be distilled down to just two numbers: the black hole’s mass $M$ and angular momentum $J$, the latter of which is usually expressed as a dimensionless spin parameter $a_* = J/(GM^2/c)$. This implies that all physical theories that involve black holes, e.g. the link between BH spin and jet power (Blandford & Znajek 1977), Gamma Ray Bursts and spinning BHs (MacFadyen & Woosley 1999), models of quasi-periodic oscillations (Abramowicz

\footnote{In principle, BH charge is an independent parameter; however, we do not expect astrophysical black holes to retain any significant charge.}
CHAPTER 2. LIGHT FROM THE PLUNGING REGION OF ACCRETION DISCS

& Kluźniak 2001), must connect in some way to these two numbers.

For binary systems, the task of measuring mass is relatively straightforward; so long as one can obtain the period, orbital velocity, orbital geometry (i.e. inclination and eccentricity), and mass for a companion orbiting a black hole, one can immediately obtain the black hole mass using only Newtonian gravity (for a recent example, see Orosz et al. 2011b). The game of measuring black hole masses has been played as early as 1972 for Cygnus X-1 (Webster & Murdin 1972; Bolton 1972), and to date we have robust mass estimates for about 20 stellar mass binary black hole systems (Remillard & McClintock 2006b; Orosz et al. 2007, 2009, 2011a,b; Cantrell et al. 2010), and ~50 supermassive black holes (Gültekin et al. 2009).

Despite this success in measuring black hole mass, spin has been a more difficult quantity to obtain. The spin of an object only makes itself known at very short distances through a general relativistic effect known as frame dragging. Thus to probe black hole spin, we must rely on observations of matter that is very close to the BH horizon. In practice, the only available probe is the accretion disc orbiting the black hole, which transitions from nearly circular orbits to plunging at a special location known as the innermost stable circular orbit (ISCO). By observing the light given off by the accretion disc, it is possible to determine this transition radius. When the ISCO is expressed in terms of the gravitational radius \( r_g = GM/c^2 \), it has a well-defined monotonic dependence on black hole spin (e.g. Shapiro & Teukolsky 1983). Therefore, a measurement of the ISCO size yields the BH spin if the mass is independently known.

For a black hole binary system, we measure the size of the ISCO by modelling the light given off by the accretion disc. Objects inside the ISCO (hereafter referred to as
the ‘plunging zone’) cannot remain in stable circular orbits, and are forced to quickly plunge into the black hole in just a few free-fall times. Since the plunging time-scale is short, it is thought that fluid inside the plunging zone does not have time to radiate (Page & Thorne 1974), which means that if we could resolve an image of the accretion disc, then we would see a dark void corresponding to the plunging region in the centre. In practice, we cannot resolve images\(^2\) of accretion discs around black holes, and thus our only information comes in the form of X-ray spectra.

One method of estimating BH spin\(^3\) works by fitting the spectral shape of the thermal continuum emission from the accretion disc to thereby estimate the radius of the ISCO (this method was pioneered by Zhang et al. 1997a; see Table 2.3 for some recent results).

This ‘continuum fitting’ method uses the colour temperature of the thermal component, the distance to the source, and the received X-ray flux to obtain a characteristic emitting area for the disc. Given a model for the radial dependence of the disc emission, this characteristic emitting area determines the ISCO radius. The main drawback of the continuum fitting method is that we first need accurate estimates of the BH distance (to turn fluxes into areas), disc inclination (to turn the projected area into an ISCO radius), and BH mass (to get spin from the monotonic \(r_{\text{ISCO}}/r_g\) vs. \(a_*\) relation). Luckily, the techniques needed for measuring distance (Reid et al. 2011), mass, \(^2\)Although it has been recently proposed to use radio VLBI observations to test GR by resolving the apparent shape of SgrA* (Doeleman et al. 2009).

\(^3\)Another commonly applied method is known as the ‘iron line’ method, which estimates BH spin through spectral modelling of the Fe K\(\alpha\) fluorescence line (Fabian et al. 1989, see Miller 2007 for a recent review).
and inclination (Orosz et al. 2011b; Cantrell et al. 2010) for binary systems are well developed, and to date have been successfully applied to more than half a dozen black hole binary systems (McClintock et al. 2011).

The continuum fitting method has evolved through a sequence of progressively more complex disc models. The simplest model assumes multitemperature blackbody emission from the disc (Mitsuda et al. 1984). Including relativistic effects such as Doppler beaming, gravitational redshifting, and light bending yielded the next generation of models (called \textit{kerrbb} – Li et al. 2005). The current generation of disc models (named \textit{bhspec} – Davis et al. 2005; Davis & Hubeny 2006) frees itself from the blackbody assumption, and instead obtains the disc spectra by means of radiative transfer calculations. All of the above mentioned models use the prescription of Novikov & Thorne (1973, hereafter NT) as the underlying disc model\(^4\), which assumes a perfectly dark void inside the ISCO. The advance that we make in this work is to apply radiative transfer calculations to compute the spectrum for a general relativistic magneto-hydrodynamic (GRMHD) simulated disc that produces emission from inside the ISCO.

\(^{4}\)This is the relativistic generalization of the standard Shakura & Sunyaev (1973) model for viscous geometrically thin but optically thick accretion discs.
Figure 2.1: Comparison of dimensionless luminosity profiles from the GRMHD simulations of Penna et. al 2011 (solid) with the standard disc model (dotted – Novikov & Thorne 1973; Page & Thorne 1974). Results from models A,B,C (defined in Table 2.1), are shown. The dashed line depicts an alternate measure of the GRMHD luminosity (described in §2.7.2 and Appendix 2.12). Note that the luminosity in the standard Novikov & Thorne (1973) disc model goes to zero at the ISCO.

Currently, a crucial assumption in the continuum fitting enterprise is the perfect darkness of the plunging zone. This scenario applies only to the idealized case of a razor thin unmagnetized disc (see Novikov & Thorne (1973)). For finite thickness discs, it has been argued that as long as the disc remains geometrically thin (with disc aspect ratios $h/r \ll 1$), the disc can be well approximated by the NT model (Paczyński 2000;
Afshordi & Paczyński 2003; Shafee et al. 2008a). However, recent work with magnetized discs (Shafee et al. 2008b; Noble et al. 2009, 2010, 2011; Penna et al. 2010) suggests that there may be departures from NT even in the limit of thin discs. An observational difference between these magnetized discs and the classic NT discs is the appearance of non-negligible luminosity inside the plunging region (see Fig. 2.1 for the result from Penna et al. (2010), where the extra plunging region luminosity constitutes ∼4 per cent of the total). Recent models of unmagnetized discs that include the physics of energy advection (Abramowicz et al. 1988; most recently Sadowski 2009, Sadowski et al. 2011) also share the feature of nonzero luminosity inside the ISCO, but in these advective models, the extra light only becomes significant when the accretion rate approaches the Eddington rate.

The primary goal of the present project is to investigate the importance of the neglected light from the inner plunging region of accretion discs. Since GRMHD discs exhibit the phenomenon that we wish to investigate (i.e. nonzero light from the plunging region), we use them as the basis for all our investigations. We analyse the GRMHD simulations of Penna et al. (2010) in this work. To connect the GRMHD simulations with observables, we compute realistic (i.e. radiative transfer based) disc spectra from them. Essentially, we adapt the ideas pioneered in bhspec (Davis et al. 2005; Davis & Hubeny 2006) on to the domain of GRMHD discs. One limitation of the GRMHD simulations is that they do not inherently model the radiation field of the disc fluid. The way we put the photons back in (for the purposes of computing disc spectra) is to do a radiative transfer post-processing step. To get the integrated disc spectrum corresponding to a particular simulation run, we take the following three steps: 1) We slice the GRMHD disc into many individual annuli; 2) For each annulus, we compute the local spectrum
through a radiative transfer calculation; 3) By means of ray tracing, we sum up the light from every annulus to get the overall spectrum of the accretion disc.

Several previous studies have also used GRMHD simulations to consider the impact of emission from the plunging region on disc spectra (e.g. Beckwith et al. 2008a; Noble et al. 2009, 2011; Kulkarni et al. 2011), but all assumed (modified) blackbody emission. Since the plunging region is precisely where the fluid becomes tenuous and optically thin, here the blackbody assumption is likely to be strongly violated. Our work extends the analysis of Kulkarni et al. (2011, hereafter K2011), which also made use of the GRMHD discs from Penna et al. (2010). The advance that we make beyond Kulkarni et al. (2011) is that we obtain disc spectra by means of radiative transfer calculations (this allows our work to be free from the blackbody assumption which underpins all previous work on GRMHD disc spectral modelling).

Our paper is organized as follows: In §2.2 we describe the setup and properties of the simulated GRMHD discs that we use to generate disc spectra. We then detail the physics of the radiative transfer calculation in §2.3, and list the simulation quantities needed to generate the spectrum for each annulus. In §2.4, we explain how these simulation quantities are extracted, and we give a brief overview of the ray tracing process in §2.5. The implications from the additional plunging region light are discussed in §2.6, and in §2.7 we briefly touch upon the limitations of our disc model. We summarize the key results in §2.8. Appendices 2.9–2.12 expand upon the technical details of various aspects of this work.
2.2 GRMHD Simulations

The simulation numerically evolves the 3D GRMHD equations in the Kerr spacetime via the code HARM (Gammie et al. 2003; McKinney 2006; McKinney & Blandford 2009). The code works in dimensionless units \((G = c = k_B = 1)\), and we assume that the fluid follows an ideal gas equation of state \(U = (\Gamma - 1)P_{\text{gas}}\) where \(U\) is the gas internal energy density, and \(P_{\text{gas}}\) is the gas pressure. We choose the adiabatic index to be \(\Gamma = 4/3\), which corresponds to an ultrarelativistic equation of state. We start off the simulation with a torus of gas that is initially in pure hydrodynamic equilibrium (DeVilliers et al. 2003; Gammie et al. 2003) threaded by a weak \((100 < P_{\text{gas}}/P_{\text{mag}} < 1000)\) 4-loop poloidal magnetic field structure (see Penna et al. 2010 for details on the field topology). The torus is also set up such that its orbital spin axis is aligned with the spin axis of the black hole.

There are three primary degrees of freedom for the torus: 1) the initial gas entropy, 2) the initial magnetic field strength and geometry, and 3) the initial angular momentum profile. The choice of initial gas entropy controls the overall thickness of the disc, whereas the choice of initial magnetic field strength/geometry controls the strength of the MRI turbulence (Beckwith et al. 2008b). The angular momentum profile sets the radial extent of the starting torus. We evolve the system in time via a conservative Godunov scheme, with an additional caveat that we cool the gas via the following prescription:

\[
\frac{dU}{d\tau} = -U \frac{\ln (K/K_i)}{\tau_{\text{cool}}},
\]

where \(U\) is the internal energy of the gas, \(K = P/\rho^\Gamma\) is the gas entropy constant, \(K_i = P_i/\rho_i^\Gamma\) is the initial entropy constant, and we set \(\tau_{\text{cool}} = 2\pi/\Omega_k\) for the cooling.
CHAPTER 2. LIGHT FROM THE PLUNGING REGION OF ACCRETION DISCS

time-scale ($\Omega_k = \sqrt{GM/r^3}$). After a gas element heats up from dissipation, energy is removed according to Eq. 2.1 such that the gas returns to its initial entropy\(^5\)(this acts to preserve the initial aspect ratio of the disc). In the absence of a full radiative transfer calculation, the cooling function is a substitute for the radiative energy loss that we expect from a geometrically thin, optically thick disc.

The initial conditions of the specific simulations that we consider are listed in Table 2.1. All models are run using a fixed spherical polar mesh with $N_r = 256$ (radial cells), $N_\theta = 64$ (poloidal cells), and $N_\phi = 32$ (toroidal cells), except for model E, which has twice the number of poloidal and toroidal cells. The cell spacing is chosen such that resolution is concentrated near the horizon and the disc midplane (see Penna et al. 2010 for details). We find that the increase in resolution from model D to E has the effect of increasing $\alpha$. Models A-C correspond to thin discs and most of our reported results focus on these models.

2.3 Annuli Spectra

We obtain the overall GRMHD accretion disc spectrum by summing up the light contributions from all annuli that comprise the disc. Most previous work on disc spectra assume a modified blackbody prescription for the local spectrum emitted by each

\(^5\)The cooling function in Eq. 2.1 is not invoked when $K < K_i$ (i.e. the cooling prescription does not add any heat to the fluid)
### Table 2.1: GRMHD disc parameters

<table>
<thead>
<tr>
<th>Model</th>
<th>$a_*$</th>
<th>Target h/r</th>
<th>$L/L_{Edd}$</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.9</td>
<td>0.05</td>
<td>0.35</td>
<td>0.22</td>
</tr>
<tr>
<td>B</td>
<td>0.7</td>
<td>0.05</td>
<td>0.32</td>
<td>0.10</td>
</tr>
<tr>
<td>C</td>
<td>0.0</td>
<td>0.05</td>
<td>0.37</td>
<td>0.04</td>
</tr>
<tr>
<td>D</td>
<td>0.0</td>
<td>0.1</td>
<td>0.70</td>
<td>0.04</td>
</tr>
<tr>
<td>E</td>
<td>0.0</td>
<td>0.1</td>
<td>0.71</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Note – The BH mass was set to be ($M_{BH} = 10M_\odot$). $L/L_{Edd}$ and $\alpha$ are derived quantities that are measured from the simulations (see §2.4). Kulkarni et al. (2011) also used Models A, B, and C in their work.
CHAPTER 2. LIGHT FROM THE PLUNGING REGION OF ACCRETION DISCS

annulus. The blackbody is modified such that the specific intensity is given by:

\[ I_\nu(T) = f^{-4}B_\nu(fT), \]  

(2.2)

where \( B_\nu(T) \) is the standard Planck function and \( f \) is spectral hardening constant (often assumed to be 1.7 for X-ray binaries). Following Davis & Hubeny (2006), we improve on this assumption by solving the radiative transfer equation. The emergent spectrum for each annulus is obtained through the stellar atmospheres code TLUSTY (Hubeny & Lanz 1995). The code simultaneously solves for the vertical structure of a plane-parallel atmosphere and its angle dependent radiation field, and is independent of any blackbody assumption. The atmosphere is modelled as a series of 1D vertical cells that are in radiative, hydrostatic, and statistical equilibrium (which allows for departures from local thermodynamic equilibrium in the atomic populations). In each cell, we obtain: the spectrum as a function of viewing angle \( I_\nu(\theta) \), density \( \rho \), gas temperature \( T \), vertical height above the midplane \( z \), and the particle composition \( (n_e, n_{\text{HI}}, n_{\text{HII}}, \text{etc}...) \).

We are primarily interested in each annulus’ emergent spectrum (i.e. \( I_\nu(\theta) \) for the surface cell). To obtain a unique solution in the stellar atmospheres problem, we need to specify the following 3 boundary conditions:

(1). The radiative cooling flux \( F = \sigma_{\text{SB}}T_{\text{eff}}^4 \) (where \( T_{\text{eff}} \) is the annulus effective temperature).

(2). The vertical tidal gravity experienced by the fluid (parametrized by \( Q = g_\perp/z \) where \( g_\perp \) is the vertical acceleration and \( z \) is the height above the midplane).
(3). The column density to the midplane \( m = \Sigma/2 \) (where \( \Sigma \) is the total column density of the annulus).

In the framework of TLUSTY, the problem of computing the disc spectrum simply becomes one of obtaining the radial profiles of \( T_{\text{eff}}(r) \), \( Q(r) \), and \( m(r) \) for the GRMHD disc. We compute the spectrum for each disc annulus using the corresponding values of \( T_{\text{eff}} \), \( Q \), and \( m \).

### 2.3.1 Assumptions in the TLUSTY model

In our annulus problem, we make the assumption of uniform viscous heating per unit mass, and we ignore the effects of magnetic pressure support, fluid convection, and heat conduction. We allow for deviations from LTE by explicitly computing the ion populations for H, He, C, N, O, Ne, Mg, Si, S, Ar, Ca, and Fe assuming solar abundances. In the non-LTE calculations, only the lowest energy level is considered for each ionization state, except for the cases of H and He\(^+\), which are modelled with 9 and 4 levels respectively. The treatment of bound-free transitions include all outer shell ionization processes, collisional excitations, and an approximate treatment for Auger (inner shell) processes. Free-bound processes modelled include radiative, dielectric, and three-body recombination. Free-free transitions are modelled for all listed elements, whereas bound-bound transitions are not modelled. Comptonization is included through an angle-averaged Kompaneets treatment (Hubeny et al. 2001). Finally, to reduce the number of independent vertical cells in the problem, we assume that the annulus is symmetric about the midplane.

One drawback of the algorithm used in TLUSTY (lambda-iteration and complete
linearization, see Hubeny & Lanz 1995 for details) is that TLUSTRY requires an initial guess for the conditions in each vertical cell. The code often fails to converge when the initial guess is poor. However, if TLUSTRY finds a converged solution, it is robust to the choice of initial guess. Often, manual intervention (in the form of providing better initial guesses) is needed to ensure convergence, which means that the process of generating annuli spectra cannot be completely automated. However, since only three parameters are necessary to uniquely specify an annulus, we have simply manually precomputed a grid of annuli spanning the full range of parameter space for the case of stellar mass black hole systems. The grid spans \( \log_{10} T_{\text{eff}} \in \{5.5, 5.6, \ldots, 7.3\} \), \( \log_{10} Q \in \{-4.0, -3.9, \ldots, 9.0\} \), and \( \log_{10} m \in \{0.5, 0.6, \ldots, 2.8, 2.9, 3.0, 4.0, 5.0, 6.0\} \). We then interpolate on this grid to obtain any needed annuli spectra (see Appendix 2.11 for details on the interpolation process).

### 2.4 Slicing the GRMHD disc into Annuli

Since we wish to describe the accretion disc from the 3D GRMHD simulations as a series of annuli, we take out the poloidal and toroidal structure through an averaging process. All quantities are first subject to azimuthal averaging (since we model the annuli as azimuthally symmetric structures), followed by vertical averaging (i.e. averaging over \( \theta \)) weighted by rest mass density. For each annulus in the GRMHD simulation, we must identify \( T_{\text{eff}}(r), Q(r), \) and \( \Sigma(r) \) before we can call on TLUSTRY to provide the local spectrum.

We obtain \( T_{\text{eff}}(r) \) directly from the simulation luminosity profile (§2.4.1), \( Q(r) \) from the Kerr metric and the fluid velocities (§2.4.2), and \( \Sigma(r) \) from the fluid velocities and
accretion rate (§2.4.3). Since the simulations are dimensionless, we must first choose a black hole mass $M$ and a disc luminosity $L / L_{\text{Edd}}$ to dimensionalize the radial profiles of $T_{\text{eff}}(r)$, $Q(r)$, and $\Sigma(r)$. For all the simulation runs, we have chosen a fiducial value of $M = 10 M_\odot$, and we set $L / L_{\text{Edd}}$ so the disc thickness in the radiative model matches that of the GRMHD simulation (see §2.4.1).

One complicating factor is that the simulated disc is only reliable for a short range of radii. Beyond a certain critical radius $r > r_{ie}$ (where $r_{ie}$ stands for the radius of inflow equilibrium – see Penna et al. 2010 for a detailed discussion), the simulation has not been run long enough for the fluid to reach its equilibrium configuration. The relevant physical time-scale for reaching inflow equilibrium is the accretion time $t_{\text{acc}} = r / u^r$, where $u^r$ is the fluid’s radial velocity. The problem with fluid outside of $r_{ie}$ is that this fluid still has memory of its initial conditions (i.e. fluid outside of $r_{ie}$ has $t_{\text{acc}} > t_{\text{sim}}$, where $t_{\text{sim}}$ is the total runtime of the simulation). Beyond the inflow equilibrium point, we instead turn to an analytic disc model that is matched to the simulation disc (we use a generalized NT model that allows for nonzero stresses at the ISCO, see Appendices 2.9 and 2.10 for details). The analytic model also requires us to specify a disc viscosity parameter $\alpha$ (defined as the ratio between the vertically integrated stress and pressure, see Novikov & Thorne 1973). Table 2.1 lists the values of $L / L_{\text{Edd}}$ and $\alpha$ corresponding to the different simulation runs, where both parameters are determined by matching the GRMHD disc at $r_{ie}$ to the analytic generalized NT disc (see §2.4.1 and §2.4.3).
2.4.1 Flux profile

From the simulations, we have extracted in the Boyer-Lindquist \((t,r,\theta,\phi)\) frame, the fluid four-velocity \((u^t, u^r, u^\phi, u^\theta)\), and luminosity \(L_{BL}\), where the luminosity is computed from the cooling function \(dU/d\tau\) (see Eq. 2.1) in the following fashion:

\[
L_{BL}(r) = \frac{1}{t_f - t_i} \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi} \int_{r=r_H}^{r} \int_{t=t_i}^{t_f} \frac{dU}{d\tau} u_t \sqrt{-g} dt dr d\theta d\phi,
\]

(2.3)

where \(t_i\) and \(t_f\) are the initial and final times considered for the time integral, \(r_H\) is the black hole horizon, \(u_t\) is the fluid’s conserved specific energy, and \(\sqrt{-g}\) is the determinant of the Kerr metric. We do the integration for only the bound fluid, and then obtain the comoving flux by transforming from the Boyer-Lindquist flux (see Kulkarni et al. (2011)):

\[
\sigma_{SB} T_{eh}^4(r) = F_{com}(r) = \frac{F_{BL}(r)}{-u_t} = \frac{1}{-4\pi r u_t} \frac{dL_{BL}(r)}{dr}.
\]

(2.4)

However, since the simulations are converged only for a short radial extent, we extrapolate the flux profile to large radii by matching the GRMHD flux with an analytic disc model at a matching radius \(r_{ie}\). The extrapolation model we use is the same as Kulkarni et al. (2011), which is a generalized Page & Thorne (1974) model (see Appendix 2.9) that does not assume zero torque at the disc inner boundary, and hence allows for nonzero ISCO luminosities.

To determine the point at which we can no longer trust the simulations (in other words, to locate \(r_{ie}\)), we look at the radial profile of the GRMHD accretion rate (Figure 2.2). Beyond a certain radius, we see that the accretion rate begins to significantly deviate from a constant value (the deviation from a constant accretion rate only occurs in fluid that has not reached inflow equilibrium). We pick the matching point to be the outermost point where we have well behaved accretion, though the final luminosity...
profile is fairly insensitive\(^6\) to our choice of matching radius (i.e. after varying the matching location by ±20 per cent, the luminosity profile changes by less than 10 per cent).

\[
\text{Figure 2.2: The GRMHD simulation accretion rates normalized to } \dot{M} \text{ at the horizon are shown. We identify the region left of the vertical dashed line as converged, so the line marks the boundary } r_{ie} \text{ where we match the GRMHD disc to an analytic model.}
\]

**Disc thickness matching**

The GRMHD luminosity that we measure (as shown in Figure 2.1) is actually in dimensionless units of \(L/\dot{M}_{\text{sim}}\) where \(\dot{M}_{\text{sim}}\) is the local accretion rate of the simulations (see Figure 2.2). To calculate \(T_{\text{eff}}\) in dimensioned units of [K], we need to first determine

\(^6\)The insensitivity of the luminosity extrapolation to the matching radius \(r_{ie}\) was also demonstrated in appendix A of K2011.
the dimensioned luminosity corresponding to the GRMHD simulations. Unfortunately, the simulations do not include radiation physics, so the simulation fluid does not have a set density scale, which means there is no direct way of measuring $\dot{M}$ and hence $L$. We thus resort to an indirect method for determining this accretion rate. We identify $L/L_{\text{Edd}}$ of the simulation with the value that matches the disc thickness in both the GRMHD simulated disc and the TLUSTY annuli\(^7\) (see Figure 2.3). However, since TLUSTY ignores magnetic support and the simulations ignore radiation pressure, the TLUSTY based and GRMHD thickness profiles have completely different shapes. We opt to match the thickness at the radius corresponding to the luminosity peak (see Figure 2.3).

Although the luminosity profile drops off fairly rapidly inside the ISCO (see Fig. 2.1), we find that the disc’s effective temperature remains roughly constant (Fig. 2.4). Therefore the falloff in disc luminosity when approaching the BH horizon is purely a geometric effect. Annuli near the horizon have less surface area, and hence they contribute less to the total luminosity.

\(^7\)Kulkarni et al. (2011) also applied this method to dimensionalize their luminosity profiles
Figure 2.3: Shown are the $h/r$ disc thickness profiles for model C, as computed by TLUSTY and our GRMHD simulation, where $h = \int \rho |z| dz / \int \rho dz$. The black dotted lines depict the TLUSTY disc thicknesses for different choices of luminosity (the four sets of dotted lines correspond to $L/L_{\text{Edd}} = 0.1, 0.2, 0.4, 0.8$). The black solid line denotes the matched thickness profile with $L/L_{\text{Edd}} = 0.37$. The black dot represents the radius where the luminosity is greatest, which we have chosen to be the radius at which we match the TLUSTY and GRMHD thickness profiles.
Figure 2.4: Here we plot the $T_{\text{eff}}$ flux profile, computed from the luminosity profiles of Figure 2.1 for the accretion rates listed in Table 2.1.

2.4.2 Vertical gravity profile

The vertical gravity parameter $g_{\perp}$, which represents the tidal vertical acceleration in the disc, is obtained by evaluating the $R_{zz}^z$ component of Riemann curvature tensor in the comoving frame. A commonly used prescription for the vertical gravity is that of Riffert & Herold (1995), which assumes that the disc fluid moves along circular geodesics; this approach becomes invalid inside the plunging zone. We thus turn to the prescription of Abramowicz et al. (1997), which relaxes the circular orbit assumption:

$$g_{\perp}(r,z) = \Omega_k^2 R_z(r) \cdot z,$$  \hspace{1cm} (2.5)
where $\Omega_k = (GM/r^3)^{1/2}$ is the Keplerian orbital frequency, and the dimensionless relativistic factor $R_z(r)$ is given by:

$$R_z(r) = \left( \frac{u_\perp^2 + a^2(u_t - 1)}{r} \right).$$

(2.6)

In Eq. 2.6 we use the four-velocity corresponding to the midplane fluid from the GRMHD simulations for $r<r_{ie}$ (as determined from Figure 2.2). For $r>r_{ie}$, we use circular geodesics as the four-velocity in Eq. 2.6. We find that the two prescriptions for vertical gravity (Riffert & Herold 1995; Abramowicz et al. 1997) give nearly identical results outside the ISCO.

For convenience, we define the function $Q(r)$ such that:

$$g_\perp(r, z) = Q(r) \cdot z,$$

(2.7)

and we interpret $Q(r)$ as the radial dependence of the vertical gravity.

### 2.4.3 Column density profile

Since the GRMHD simulation does not include radiation physics, the simulation mass density is scale-free, so we cannot directly read out the disc column mass density in physical units. We dimensionalize the simulation density by solving for $\Sigma$ in the vertically integrated mass conservation equation after having picked a constant accretion rate $\dot{M}$:

$$\dot{M} = 2\pi r \Sigma \hat{u}_r,$$

(2.8)

---

8We set $\dot{M}$ to be the value corresponding to the disc luminosity as determined in §2.4.1. They are related by $L = \eta \dot{M} c^2$ where $\eta(a_*)$ is the spin dependent accretion efficiency of the NT disc.
where $r$ is the BL radial coordinate, and $\bar{u}^r$ is the mass-averaged radial velocity for the GRMHD simulations, computed via:

$$\bar{u}^r = \frac{\int_{\theta=0}^{\pi} \hat{\rho} u^r \sqrt{-g \text{d}\theta}}{\int_{\theta=0}^{\pi} \hat{\rho} \sqrt{-g \text{d}\theta}}.$$  \hspace{0.5cm} (2.9)

**Radial velocity matching**

At large radii ($r > r_{ic}$), the simulation is no longer converged, and we resort to matching the simulation result with a disc model. Rather than directly using the Novikov & Thorne (1973) disc model, we re-solve the NT disc equations using our matched GRMHD luminosity profile (see Appendix 2.10 for details). The final matched radial velocity profile is shown in Figure 2.5.
Figure 2.5: A comparison between the mass averaged simulation radial velocity profile (dotted) and the final matched radial velocity profile (solid line). The matching point is depicted as the large coloured circle, whereas the vertical dashed line denotes the location of the ISCO. The gaps in the GRMHD radial velocity profile represent grid cells whose radial velocity is pointed outwards.
Figure 2.6: The vertically averaged $\alpha$ profile as computed from the GRMHD simulation (dotted) compared to the $\alpha$ obtained by the radial velocity matching method (solid horizontal lines, corresponding to the $\alpha$ values listed in Table 2.1). For reference, we have plotted the ISCO locations using dashed vertical lines.

We choose the disc viscosity $\alpha$ so as to ensure continuity in the radial velocity profile (Fig. 2.5). The matching values of $\alpha$ for each disc are listed in Table 2.1. The location of the radial velocity matching point is set to be the same as the luminosity matching point. To verify that our $\alpha$ value from radial velocity matching is sensible, we compare it to the directly computed value from the simulation $\alpha_{\text{sim}} = \tau_{\hat{r}\hat{\phi}}/(hP_{\text{tot}})$ where $\tau_{\hat{r}\hat{\phi}}$ is the height integrated shear stress, and $hP_{\text{tot}}$ is the height integrated total pressure (where $P_{\text{tot}} = P_{\text{mag}} + P_{\text{gas}}$). The near continuity in the simulation $\alpha$ profile with the matched value suggests that our method for determining $\alpha$ is sound.

From the radial velocity profile and from our choice of constant $\dot{M}$, the mass
continuity equation (Eq. 2.8) yields the final matched disc column density profile (Fig. 2.7).

![Graph of dimensioned column density profiles for GRMHD discs compared to Novikov & Thorne (1973) discs.](image)

**Figure 2.7:** Dimensioned column density profiles for the GRMHD discs (solid lines) compared to Novikov & Thorne (1973) discs (dotted lines). Note the sharp drop in NT column density as the fluid reaches the plunging region, in contrast to the more gradual tapering observed in the GRMHD simulations.

### 2.5 Ray Tracing

The final step in the process of generating the accretion disc’s SED is to sum the light from all the individual annuli. The observed disc image is determined by shooting a series of parallel light rays from an image plane that is perpendicular to the black hole line of sight. We partition the image plane into a sequence of polar grid cells with logarithmic radial spacing (to concentrate resolution near the black hole), and with uniform angular
spacing. To produce the ray traced images, we shoot rays for \( N_r = 300 \) radial grid cells and \( N_\phi = 100 \) polar cells. The grid is then further squashed by a factor of \( \cos i \) to match the geometry of the inclined disc.

A single ray is shot for each grid cell, and by numerically integrating the (second-order) geodesic equations for light:

\[
\frac{d^2 x^\alpha}{d\lambda^2} + \Gamma^\alpha_{\beta\gamma} \frac{dx^\beta}{d\lambda} \frac{dx^\gamma}{d\lambda} = 0, 
\]

(2.10)

where \( \lambda \) is an affine parameter along the geodesic, and \( \Gamma^\alpha_{\beta\gamma} \) are the connection coefficients, we locate the first disc midplane intersection for each light ray (we stop at the disc midplane since we have an optically thick disc – see Figure 2.11 for a plot of the optical depth profile). For simplicity, we opt to use the disc midplane as the point of light emission rather than using the annuli’s precise photosphere. Kulkarni et al. (2011) has suggested that this effect only becomes important for thick discs viewed at high inclinations angles (which is when disc self-shadowing becomes important). In this work, we do not expect the distinction between midplane and photosphere to be important since we are dealing with thin discs (where \( h/r \ll \cos i \)). This process of following the trajectories of multiple rays of light to produce an image is called ‘ray tracing’. The ray tracing code used in this work was originally developed by Scherbakov & Huang (2011), with further refinements by Kulkarni et al. (2011).

We then compute (by interpolation on the GRMHD grid cells, see Appendix 2.11 for details) the local values of: flux \( T_{\text{eff}} \), column density \( \Sigma \), vertical gravity \( Q \), and comoving light ray incidence angle \( \mu \). The local comoving spectrum is then obtained by interpolating on our grid of TLUSTY atmosphere models. We then apply the relevant Doppler boosting and gravitational redshifting to these spectra (see Fig. 2.8 for the ray
CHAPTER 2. LIGHT FROM THE PLUNGING REGION OF ACCRETION DISCS

Figure 2.8: Colour images of the inner disc \((r < 15M)\) produced via ray tracing (viewed at an inclination angle of \(i = 60^\circ\)) for models A, B, and C in Table 2.1. The colours correspond to the flux \(F_\nu\) integrated over different energy bands, where red colour corresponds to the energy band \(E < 4\) keV, green \(4\) keV \(< E < 12\) keV, and blue \(E > 12\) keV. For each value of BH spin, the colour mapping for the GRMHD and NT panels are identical, and each colour channel is normalized to the maximum flux in that channel. Note the appearance of a blue spot in the GRMHD disc plunging region. This blue spot is responsible for the appearance of a nonthermal high energy power-law feature in the disc spectra (see Figure 2.9).
CHAPTER 2. LIGHT FROM THE PLUNGING REGION OF ACCRETION DISCS

traced images). Finally, the overall disc spectrum is obtained by integrating the spectra corresponding to these light rays over the apparent disc area in the image plane. We take a fiducial BH distance of 10 kpc when generating the disc spectra.

2.6 Results

Qualitatively, we can spot a few differences in Figure 2.8 between the GRMHD (top) and NT (bottom) panels. We see that for spinning black holes, the plunging fluid emits in the hard X-rays (appearing as a (> 12 keV) blue smudge in Fig. 2.8). Even outside the plunging region, there is a noticeable increase in disc luminosity, which results in harder annuli spectra everywhere (compare the bright white Doppler spot in the upper right (Model C, GRMHD) panel, with the duller orange spot in the lower right (Model C, NT) panel.

By integrating the flux in the image plane, we arrive at the observed disc spectrum (see Fig. 2.9). For spinning black holes, the spectra corresponding to the GRMHD simulations appear to exhibit a power-law component at high energies. We discuss this effect in §2.6.1. In addition, for all models, the location of the energy peak in the spectra for the GRMHD discs is shifted towards higher energies. The harder overall spectrum of GRMHD compared to NT implies that there would be an asymmetric error in BH spin estimates (since the current fitting models are based on the cooler Novikov & Thorne (1973) disc). The quantitative size of this spin bias is discussed in §2.6.2.
Figure 2.9: GRMHD and Novikov & Thorne (1973) disc spectra for each of the 5 models listed in Table 2.1. The flux normalization corresponds to an assumed distance of 10 kpc. Note that the GRMHD spectra peak at slightly higher energies than their NT counterparts. Note also the emergence of a high energy power-law tail for the spinning black holes. The dashed spectra (labelled GRMHD2) correspond to the alternate GRMHD luminosity profile (dashed line in Figure 2.1), discussed in §2.7.2 and Appendix 2.12.
CHAPTER 2. LIGHT FROM THE PLUNGING REGION OF ACCRETION DISCS

2.6.1 Power law tail

Observationally, high-energy power-law tails are often seen in the X-ray spectra of black hole binary systems (e.g. Miyamoto et al. 1991; McClintock et al. 2001; Reis et al. 2010). The current leading theory to explain these high-energy power-law phenomena is the action of a hot optically thin corona (Zhang et al. 1997b), which produces the power-law tail by means of either Comptonization of disc photons, synchrotron radiation, and/or synchrotron self-Comptonization (Miller 2007; McClintock & Remillard 2006). Despite the many proposed coronal models (e.g. Haardt & Maraschi 1991; Dove et al. 1997; Kawaguchi et al. 2000; Liu et al. 2002), there is no agreed-upon standard (i.e. there is scarce agreement on the geometry and physical properties of the corona). Although a hot corona is often invoked to explain the existence of a high energy power-law tail, in this work we naturally recover a hot power-law tail from just the thermal disc fluid that occupies the plunging region.

To better understand the origin of the apparent high-energy power-law tail in our models, we examine the area-weighted local spectrum for several annuli in model B ($a_*=0.7$), chosen since this simulated disc exhibits the strongest tail relative to the disc continuum. From Figure 2.10, we see that the high energy power-law tail stems solely from the emission of annuli inside the plunging region (coloured in blue). Furthermore, we see that the power-law emerges from the combined emission of plunging annuli at various locations (i.e. the envelope of spectra from $r=2M-3M$ comprise the power-law). No single annulus emits the full power-law.

In our plunging region model, we find that the strength of the high energy power-law also depends on BH spin. The spinless models (C, D, E) do not appear to produce any
significant power-law component. The reason for why model B ($a_*=0.7$) produces a stronger power-law than model A ($a_*=0.9$) is simply that the plunging region in B covers a larger area in the image plane than A, leading to a larger overall plunging region flux (compare the relative sizes of the blue plunging region blobs in Figure 2.8). We also find a correlation in the strength of the power-law with disc viewing angle; edge-on discs produce the strongest power laws relative to the thermal continuum since Doppler beaming increases with inclination angle, and hence acts to preferentially boost the plunging region emission.

**Figure 2.10**: Local emergent fluxes from various annuli for model B ($a_*=0.7$), weighted by their face-on emitting areas (i.e. $A = 2\pi r \Delta r$, where $\Delta r$ is the radial width of the annulus). Plunging region annuli are coloured blue, whereas annuli beyond the ISCO are coloured red to black (for reference, the ISCO is located at $3.4M$, and the horizon is at $1.7M$). Doppler and gravitational redshift corrections have not been applied to these spectra.
We find that the spectral hardening factor for the plunging annuli appears to start growing non-linearly as the gas approaches the horizon (see Fig. 2.10 and note the rapid shift in the spectral peak locations for the blue plunging region annuli, despite that $T_{\text{eff}}$ is nearly constant in the plunging zone as per Fig. 2.4). The physical reason for this rapid spectral hardening is simply that the plunging gas quickly becomes extremely hot. Upon entering the plunging region, the fast radial plunge causes the annuli to rapidly thin out in column density. The dropping column mass allows the annuli to make the transition from being optically thick to effectively optically thin\(^9\) (see Fig. 2.11). Despite this drop in column density, the gas needs to maintain a roughly constant cooling rate (i.e. $T_{\text{eff}}$ remains constant in the plunging zone – see Fig. 2.4). A property of effectively optically thin gas is that it cannot radiate efficiently (Shapiro, Lightman, & Eardley 1976). To keep a constant cooling rate despite the rapidly dropping optical depth, the gas must heat up tremendously, leading to very hot emission.

\(^9\)An effectively optically thin medium is one that is optically thin to absorption, and optically thick to scattering.
Figure 2.11: Optical depths corresponding to model B \((a_* = 0.7)\). The total optical depth \(\tau_R = \Sigma \kappa_R\) is computed using the Rosseland mean opacity \(\kappa_R\) (including both absorption and scattering). The effective optical depth is computed as \(\tau_{\text{eff}} = \sqrt{3\tau_{\text{abs}}\tau_{\text{scat}}}\) where \(\tau_{\text{abs}}\) and \(\tau_{\text{scat}}\) are the individual absorption and scattering optical depths respectively.

The power laws produced by the plunging region in our models have photon power-law indices that range from \(\Gamma = 4 - 6\), with scattering fractions\(^{10}\) ranging from 1-15 per cent. In our model, we find that the scattering fraction (and hence strength of the power-law) increases monotonically with inclination angle, whereas the photon power-law index decreases monotonically. The scattering fraction range spanned by our models is comparable to the observed range covered by the thermal-dominant state, as

\(^{10}\)The scattering fraction denotes the fraction of thermal seed photons that undergo Comptonization to produce the power-law component.
CHAPTER 2. LIGHT FROM THE PLUNGING REGION OF ACCRETION DISCS

well as much of the range of the intermediate, and steep power-law states for black hole binaries (which ranges from 0-25 per cent – see Steiner et al. 2009b, 2011). However, even in the steep power-law state, the observed photon index range of $\Gamma = 2.2 - 2.7$ (Gou et al. 2009, 2011; Steiner et al. 2011) is much less steep than what we predict with our plunging region model. This discrepancy may simply be due to a limitation of our model, which neglects the physics of disc self-irradiation. Including this effect would result in a hotter disc that would act to boost the strength of the high energy tails. Also, self-irradiation becomes increasingly important as you approach the black hole (where light bending is most severe). This implies that including irradiation will likely result in power laws that are less steep (since the emission from the hottest, innermost annuli will be boosted the most by this self-irradiation effect). We speculate that a more complete treatment of the plunging region (that includes the physics of disc irradiation and magnetic fields) would produce high energy tails that are better matches to observations.

Since many of our approximations become poor in the plunging region (see §2.7 for a discussion of some of the pitfalls of our method), we strongly caution against reading too much into the quantitative details of the power laws of Figure 2.9 (i.e. slope, normalization). The high energy power-law result should be taken only at the qualitative level; although the plunging region luminosity is dwarfed by the disc luminosity, it starts to dominate the flux at $E > 20\text{keV}$ taking on a nonthermal shape that appears as a power-law.
2.6.2 Quantitative effect on spin

As already discussed, the thermal emission from the disc is slightly stronger and hotter in the GRMHD models compared to the equivalent NT model. This will introduce an error in BH spin estimates obtained from the continuum fitting method. To precisely quantify the effect, we generate mock observations of our GRMHD discs and we fit these simulated observations with the currently used suite of continuum-fitting disc models. For simplicity, to generate these mock observations, we use an idealized instrument with effective area that is independent of photon energy. We choose the instrument’s effective area\textsuperscript{11} to be 1000 cm\textsuperscript{2}, and use a 3h exposure time. We compute photon statistics in 1000 energy channels uniformly spaced in log $E$, ranging from $0.4 \text{ keV} < E < 50.0 \text{ keV}$.

We then fit our simulated disc observations with \textsc{bhspec} (Davis et al. 2005; Davis & Hubeny 2006), which is a set of (TLUSTY based) disc spectra often used in continuum fitting. The \textsc{bhspec} spectra are computed in much the same way as our procedure in §2.1, except that \textsc{bhspec} uses a classic Novikov & Thorne (1973) disc instead of our GRMHD disc. Whenever necessary, we also attempt to fit a power-law tail to the models using \textsc{simpl} (Steiner et al. 2009a), which is a physically motivated Comptonization model that has been quite successful in fitting many observed high energy power-law tails (Steiner et al. 2009b).

\textsuperscript{11}This is modelled after the Rossi X-Ray Timing Explorer’s total effective area in the hard energy band ($> 20 \text{ keV}$), see Jahoda et al. (2006)
### Table 2.2: Recovered BH spins from fitting of simulated disc observations.

<table>
<thead>
<tr>
<th>$i$ (°)</th>
<th>NT</th>
<th>GRMHD</th>
<th>GRMHD+PL</th>
<th>GRMHD2$^†$</th>
<th>GRMHD2+PL$^†$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Model A $- a_*=0.9, \ L/\text{L}_{\text{Edd}}=0.35, \ \alpha=0.22$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.902±0.004</td>
<td>0.909±0.018</td>
<td>--</td>
<td>0.913 ±0.017</td>
<td>--</td>
</tr>
<tr>
<td>30</td>
<td>0.909±0.005</td>
<td>0.914±0.015</td>
<td>--</td>
<td>0.916 ±0.016</td>
<td>--</td>
</tr>
<tr>
<td>45</td>
<td>0.904±0.006</td>
<td>0.912±0.013</td>
<td>--</td>
<td>0.916 ±0.013</td>
<td>0.916±0.005</td>
</tr>
<tr>
<td>60</td>
<td>0.902±0.005</td>
<td>0.919±0.012</td>
<td>0.913±0.006</td>
<td>0.921 ±0.011</td>
<td>0.911±0.004</td>
</tr>
<tr>
<td>75</td>
<td>0.897±0.007</td>
<td>0.925±0.011</td>
<td>0.916±0.005</td>
<td>0.934 ±0.010</td>
<td>0.908±0.004</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Model B $- a_*=0.7, \ L/\text{L}_{\text{Edd}}=0.32, \ \alpha=0.10$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.68±0.01</td>
<td>0.73±0.03</td>
<td>0.72±0.03</td>
<td>0.75±0.03</td>
<td>0.73±0.03</td>
</tr>
<tr>
<td>30</td>
<td>0.69±0.01</td>
<td>0.74±0.03</td>
<td>0.72±0.02</td>
<td>0.75±0.03</td>
<td>0.72±0.03</td>
</tr>
<tr>
<td>45</td>
<td>0.69±0.01</td>
<td>0.75±0.02</td>
<td>0.71±0.03</td>
<td>0.76±0.03</td>
<td>0.72±0.03</td>
</tr>
<tr>
<td>60</td>
<td>0.70±0.01</td>
<td>0.75±0.02</td>
<td>0.71±0.03</td>
<td>0.76±0.02</td>
<td>0.71±0.04</td>
</tr>
<tr>
<td>75</td>
<td>0.70±0.01</td>
<td>0.76±0.01</td>
<td>0.69±0.04</td>
<td>0.77±0.03</td>
<td>0.69±0.04</td>
</tr>
</tbody>
</table>
## CHAPTER 2. LIGHT FROM THE PLUNGING REGION OF ACCRETION DISCS

Table 2.2 (Continued)

<table>
<thead>
<tr>
<th>$i$ (°)</th>
<th>NT</th>
<th>GRMHD</th>
<th>GRMHD+PL</th>
<th>GRMHD2†</th>
<th>GRMHD2+PL†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>-0.12±0.04</td>
<td>-0.04±0.09</td>
<td>−</td>
<td>-0.01±0.09</td>
<td>−</td>
</tr>
<tr>
<td>30</td>
<td>-0.10±0.04</td>
<td>-0.03±0.09</td>
<td>−</td>
<td>0.01±0.09</td>
<td>−</td>
</tr>
<tr>
<td>45</td>
<td>-0.09±0.04</td>
<td>0.01±0.08</td>
<td>−</td>
<td>0.03±0.08</td>
<td>−</td>
</tr>
<tr>
<td>60</td>
<td>-0.08±0.04</td>
<td>0.03±0.08</td>
<td>−</td>
<td>0.07±0.09</td>
<td>−</td>
</tr>
<tr>
<td>75</td>
<td>-0.06±0.05</td>
<td>0.05±0.09</td>
<td>−</td>
<td>0.10±0.10</td>
<td>−</td>
</tr>
</tbody>
</table>

(Model C – $a_* = 0$, $L/L_{\text{Edd}} = 0.37$, $\alpha = 0.04$)

| 15      | -0.19±0.02 | -0.01±0.09 | −       | 0.04±0.09  | −          |
| 30      | -0.17±0.02 | 0.00±0.09  | −       | 0.05±0.09  | −          |
| 45      | -0.15±0.03 | 0.01±0.08  | −       | 0.08±0.09  | −          |
| 60      | -0.14±0.03 | 0.03±0.08  | −       | 0.11±0.09  | 0.08±0.10  |
| 75      | -0.12±0.04 | 0.08±0.09  | 0.04±0.09| 0.17±0.09  | 0.07±0.13  |

(Model D – $a_* = 0$, $L/L_{\text{Edd}} = 0.70$, $\alpha = 0.04$)
Table 2.2 (Continued)

<table>
<thead>
<tr>
<th>$i$ (°)</th>
<th>NT</th>
<th>GRMHD</th>
<th>GRMHD+PL</th>
<th>GRMHD2†</th>
<th>GRMHD2+PL†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Model E – $a_*$ = 0, $L/L_{Edd} = 0.71$, $\alpha = 0.08$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>-0.08±0.03</td>
<td>0.04±0.08</td>
<td>–</td>
<td>0.07±0.08</td>
<td>–</td>
</tr>
<tr>
<td>30</td>
<td>-0.08±0.03</td>
<td>0.05±0.08</td>
<td>–</td>
<td>0.10±0.08</td>
<td>–</td>
</tr>
<tr>
<td>45</td>
<td>-0.07±0.02</td>
<td>0.09±0.07</td>
<td>–</td>
<td>0.12±0.09</td>
<td>–</td>
</tr>
<tr>
<td>60</td>
<td>-0.06±0.02</td>
<td>0.10±0.08</td>
<td>–</td>
<td>0.16±0.08</td>
<td>–</td>
</tr>
<tr>
<td>75</td>
<td>-0.06±0.03</td>
<td>0.13±0.09</td>
<td>–</td>
<td>0.20±0.09</td>
<td>0.11±0.07</td>
</tr>
</tbody>
</table>

Note – All spectra were fit with BHSPEC disc spectra, where $\alpha_{fit} = 0.1$. The quoted uncertainties correspond to the systematic GRMHD luminosity profile uncertainties (determined empirically by analysing the last 5 subsequent 1000M time chunks of the GRMHD simulations). The spectral fitting statistical uncertainties were negligibly small ($\Delta a_* \sim 0.001$). This is because the disc parameters that are responsible for most of the uncertainty (mass, distance, inclination) were not allowed to vary during the fit.

† These two rightmost columns correspond to a disc model that uses an alternative measure of the GRMHD luminosity (See §2.7.2 and Appendix 2.12 for details). The luminosity profiles corresponding to GRMHD2 are shown in Figure 2.1.
CHAPTER 2. LIGHT FROM THE PLUNGING REGION OF ACCRETION DISCS

The fitting is handled through xspec, a spectral fitting software package commonly used by X-ray astronomers (Arnaud 1996). For each spectrum, we fit for only two parameters: BH spin, and mass accretion rate. We fix the mass, distance, and inclination to exactly the values used to generate the simulated observations. In the cases where a power-law fit via simpl is needed, we fit for two additional power-law parameters (essentially the normalization and the spectral slope of the power-law, which simpl encapsulates as the scattering fraction of soft disc photons and the photon power-law index). We list in Table 2.2 our spin results from this fitting exercise. The first three columns correspond to the following:

‘NT’ – We perform a bhspec fit on spectra computed from a Novikov & Thorne (1973) disc model, with disc parameters corresponding to Table 2.1. We use this as a baseline for comparing with our GRMHD fit results. For the fit, we use an energy range of 0.4–8.0 keV.

‘GRMHD’ – We perform a bhspec fit on the GRMHD disc spectra on the energy range 0.4–8.0 keV. We employ an 8.0keV upper energy cutoff in the fitted spectrum since we want to exclude the power-law spectral feature.

‘GRMHD+PL’ – In addition to continuum fitting, we also fit for the power-law. Formally, we fit a bhspec⊗simpl model over the energy range 0.4–50.0 keV (the full range of our mock observation).

The primary purpose of the ‘NT’ column is to disentangle another systematic effect
that is independent of the extra luminosity of the GRMHD simulations. The \texttt{bhspec} spectral model is tabulated only for $\alpha_{\text{fit}} = 0.1$, however in our disc models, only the $a_* = 0.7$ disc happens to have a matching $\alpha = 0.1$ (see Model B in Table 2.1). This mismatch between the fitting model $\alpha_{\text{fit}}$ and the input model $\alpha$ leads to a systematic bias in the recovered spin. The spectral dependence on $\alpha$ arises from the fact that column density scales inversely with $\alpha$, so discs with higher $\alpha$ values tend towards lower optical depths. A lower optical depth medium has gas that is hotter, which results in harder spectra. In Figure 2.12, we illustrate the influence that the choice of $\alpha$ has on the spectrum of model C ($a_* = 0$). As expected, we find that the higher $\alpha$ disc has a harder spectrum (Fig. 2.12).

In the context of spin fitting, if $\alpha_{\text{fit}} > \alpha$, we expect the fitted spins to be too low (since for any given spin, the fitting spectrum will be harder than the intrinsic disc spectrum), and vice versa for $\alpha_{\text{fit}} < \alpha$. This effect is seen in the ‘NT’ column of Table 2.2 (which differs from the fitting model only by its choice of $\alpha$), models C, D, and E all have $\alpha_{\text{fit}} > \alpha$, and consequently the recovered spins are too low. On the other hand, model A has $\alpha_{\text{fit}} < \alpha$, yielding recovered spins that are too high.
Figure 2.12: Novikov & Thorne (1973) disc spectra for model C ($a_*=0$) computed for two choices of $\alpha$. The slight shift in the location of the spectral peak causes a systematic error in the recovered BH spin if $\alpha_{\text{fit}}$ does not match the intrinsic $\alpha$ of the disc.

We interpret the difference between the ‘NT’ column and the ‘GRMHD’ column as the effect of the extra luminosity from the inner regions of the GRMHD discs, which is to systematically increase the recovered spin. The spin deviation also grows with inclination angle (see the ‘GRMHD’ column in Table 2.2), since the net effect of Doppler beaming (which becomes more important at higher inclinations) is to enhance the inner disc emission. We also find that the spin deviation becomes smaller at high spins (this is due to the fact that the ISCO-spin relation becomes steeper at high spins).

In those cases where using the SIMPL model significantly improves the fit (i.e. when the fit $\chi^2$ improved by more than a factor of 2, upon inclusion of a power-law fit component), the modelling of the power-law also leads to better agreement between
fitted and input BH spin (compare ‘GRMHD’ and ‘GRMHD+PL’ columns in Table 2.2). By fitting a power-law component, some of the hot photons get associated with the power-law, which lowers the fitted spin (since the remaining disc spectrum looks softer after removal of the hard power-law photons). However, we note that the GRMHD luminosity excess extends even beyond the ISCO (see Fig. 2.1). This explains why even fitting for the power-law does not completely eliminate the upwards spin bias (i.e. the ‘GRMHD+PL’ columns still yield spins that are higher than the ‘NT’ column in Table 2.2); although some of the excess luminosity is bound up in the nonthermal plunging region (which can be excluded by means of power-law fitting), there is still some residual thermal disc luminosity excess, that acts to bias the GRMHD spin upwards. Overall, the effect of the extra plunging region light results in ∆a* ∼ 0.1, 0.06, 0.03 corresponding to spins a* = 0, 0.7, 0.9 respectively (estimated by comparing the ‘NT’ column with the ‘GRMHD’ column in Table 2.2 for the ∼30 per cent Eddington discs).

These systematic spin errors should be thought of as an upper bound since the discs that we consider correspond to fairly high accretion rates (with L/L_{Edd} > 0.3). Previous work (Paczyński 2000; Shafee et al. 2008a) suggests that thinner discs (corresponding to lower accretion rates) better match the Novikov & Thorne (1973) model. Our results in Table 2.2 also support this claim; in comparing the spin fits from model C (thin disc) with models D and E (thick discs), we find that model C agrees best to the NT disc (i.e. that model C has the smallest difference between the ‘NT’ and ‘GRMHD’ columns). In addition, model C also exhibits the least deviation from a thermal spectrum at high energies (compare the > 20 keV emission from model C with those of D and E in Fig. 2.9).

We did not simulate lower accretion rate discs due to the rapidly increasing
computational cost associated with such simulations. Typically, continuum fitting is not applied to systems where the accretion rate exceeds 30 per cent Eddington since beyond this critical point, the continuum fitting method is no longer robust to variations in the disc accretion rate (see McClintock & Remillard 2006; Steiner et al. 2009b). One possible explanation is that disc self-shadowing becomes important beyond this critical accretion rate threshold (Li et al. 2010).

To put into perspective how significant our quoted spin deviations are, we compare our results to the spin uncertainties from actual continuum fitting exercises in the literature (See Table 2.3). Since the current observational uncertainties are significantly larger than the deviations that we find in this work, we conclude that the extra light from the inner and plunging regions of the disc do not limit our ability to make accurate spin estimates through the continuum fitting method. The current limiting factors in measuring BH spin are how accurately one can determine the distance, mass, and inclination of black hole binary systems.

2.7 Discussion

The disc model that we adopt in this work is not completely self-consistent. The annuli that we compute using TLUSTY are assumed to have reached their equilibrium configuration. However, close to the black hole there is simply insufficient time for the fluid to reach equilibrium (see Fig. 2.13 for a plot of the various time-scales). In addition, the GRMHD simulations show that the pressure in the innermost regions of the disc is magnetically dominated, yet in the TLUSTY annuli calculations, we completely ignore magnetic pressure support. Previous work by Davis et al. (2009) suggests that including
Table 2.3: Spin measurements yielded by continuum fitting.

<table>
<thead>
<tr>
<th>Black hole</th>
<th>$a_*$</th>
<th>$i$(°)</th>
<th>$L/L_{Edd}$</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>A0620-00</td>
<td>$0.12 \pm 0.19$</td>
<td>$51.0 \pm 0.9$</td>
<td>$0.11$</td>
<td>Gou et al. (2010)</td>
</tr>
<tr>
<td>H1743-322$^c$</td>
<td>$0.2 \pm 0.3$</td>
<td>$75 \pm 3$</td>
<td>$0.03-0.3$</td>
<td>Steiner et al. (2012)</td>
</tr>
<tr>
<td>LMC X-3</td>
<td>$&lt; 0.3$</td>
<td>$67 \pm 2$</td>
<td>$0.1-0.7$</td>
<td>Davis et al. (2006b)</td>
</tr>
<tr>
<td>XTE J1550</td>
<td>$0.34^{+0.2}_{-0.28}$</td>
<td>$74.7 \pm 3.8$</td>
<td>$0.05-0.30$</td>
<td>Steiner et al. (2011)</td>
</tr>
<tr>
<td>GRO J1655$^d$</td>
<td>$0.70 \pm 0.1$</td>
<td>$70.2 \pm 1.2$</td>
<td>$0.04-0.1$</td>
<td>Shafee et al. (2006)</td>
</tr>
<tr>
<td>4U 1543$^d$</td>
<td>$0.80 \pm 0.1$</td>
<td>$20.7 \pm 1.5$</td>
<td>$0.06-0.1$</td>
<td>Shafee et al. (2006)</td>
</tr>
<tr>
<td>M33 X-7</td>
<td>$0.84 \pm 0.05$</td>
<td>$74.6 \pm 1.0$</td>
<td>$0.07-0.11$</td>
<td>Liu et al. (2008, 2010)</td>
</tr>
<tr>
<td>LMC X-1</td>
<td>$0.92^{+0.05}_{-0.07}$</td>
<td>$36.4 \pm 2.0$</td>
<td>$0.15-0.17$</td>
<td>Gou et al. (2009)</td>
</tr>
<tr>
<td>GRS 1915$^d$</td>
<td>$&gt; 0.95$</td>
<td>$61.5-68.6$</td>
<td>$0.2-0.3$</td>
<td>McClintock et al. (2006)</td>
</tr>
<tr>
<td>Cygnus X-1</td>
<td>$&gt; 0.95$</td>
<td>$27.1 \pm 0.8$</td>
<td>$0.018-0.026$</td>
<td>Gou et al. (2011)</td>
</tr>
</tbody>
</table>

Note – The spin uncertainties correspond to the 68 per cent (1σ) level of confidence, whereas the inequalities are to the 3σ level.

$^c$ No reliable mass estimate is available for this source

$^d$ The quoted spin errors have not been rigorously computed, and have been arbitrarily doubled from the published estimates since these are among the first systems for which continuum fitting was applied.
CHAPTER 2. LIGHT FROM THE PLUNGING REGION OF ACCRETION DISCS

magnetic support may lead to a slight hardening of the annuli spectrum. Another problem for the innermost annuli is that they become extremely hot \( (T_{\text{gas}} > 10^8 - 10^9 \text{ K}) \), which may invalidate \textsc{tlusty}’s handling of Comptonization (i.e. an angle averaged Kompaneets treatment, which is only valid for nonrelativistic electrons, see Hubeny et al. 2001). Finally, due to the increasingly strong light bending effects near the black hole, the process of disc self-irradiation (i.e. where one part of the disc shines on another part) may also become important, especially for the innermost annuli (Li et al. 2005). The net effect of this disc self-irradiation would be to further heat up the annuli, causing additional spectral hardening. We have ignored disc self-irradiation in our spectral model since this reprocessing of light cannot be easily handled by our grid of precomputed annuli.

These issues primarily affect annuli that are very close to the black hole, which radiate in the hard X-rays \( (> 20 \text{ keV}) \). We do not expect these inconsistent annuli to affect our spin fitting results in Table 2.2 since the recovered spin is mainly constrained by photons near the peak of the spectrum (located at a few keV in Figure 2.9). The inconsistent annuli will however affect the high energy part of the spectrum, which is why we caution against reading too deeply into the quantitative details of the power laws.
Figure 2.13: Several time-scales of interest are plotted here for model B ($a_*=0.7$). We show the accretion time-scale (solid blue), the photon diffusion time-scale (dash-dotted green), the photon absorption time-scale (dotted red), and the photon scattering time-scale (dashed dark cyan). The absorption and scattering time-scales are computed via $t = 1/(\rho\kappa c)$, using midplane gas densities ($\rho$) and opacities ($\kappa$). The diffusion time-scale is computed via $t_d = \tau_{scat}^2 t_{scat}$ where $\tau_{scat}$ is the scattering optical depth to the disc midplane (i.e. we expect $\tau_{scat}^2$ scatterings to occur before a photon can diffuse out). Note that close to the black hole, the accretion time-scale becomes shorter than the photon absorption time-scale, which indicates that the fluid close to the black hole has insufficient time to reach thermal equilibrium.
CHAPTER 2. LIGHT FROM THE PLUNGING REGION OF ACCRETION DISCS

2.7.1 Other signatures of the plunging region

Observations of disc variability could potentially be used to discriminate between our plunging region model and other coronal models. Assuming that the variability is caused by Doppler beaming of hotspots moving about in the disc flow, we associate the hotspot orbital time-scale with the variability time-scale. Since the orbital time decreases monotonically as one approaches the horizon, the variability from the inner plunging region of the disc ought to be more rapid than that occurring farther out in disc. Hence, our plunging region model predicts that the hard X-ray light (originating from the innermost regions of the disc) would have faster variability than the soft X-ray light (produced farther out in the thermal disc). Furthermore, in the power-density spectrum in the hard X-ray bands, we expect most of the variability power to occur above the ISCO orbital frequency (since we associate this hard emission with the plunging region).

Polarization observations could also serve as a means to distinguish between coronal models and our plunging region model. Generally, one expects evolution of the polarized fraction and net polarization angle as the observed photon energy is varied. This is due to higher energy photons (which are typically emitted closer to the black hole) feeling stronger relativistic effects, leading to a stronger shift in their polarization vectors (Schnittman & Krolik 2009, 2010). The planar geometry of the plunging region may lead to a polarization signature that differs significantly from coronal models. If the high energy power-law originates from the plunging region, we expect that its polarization properties will vary continuously from the thermal component to the non-thermal power-law. This contrasts with some alternative coronal geometries (e.g. spherical models) that provide abrupt transitions in the polarization properties from
CHAPTER 2. LIGHT FROM THE PLUNGING REGION OF ACCRETION DISCS

one component to the next (Schnittman & Krolik 2010). In the plunging region model, even at energies where the power-law dominates, the polarization fraction and angle will be consistent with thermal state polarization models that assume a thin disc geometry (see Schnittman & Krolik 2009). Due to the complex interplay between relativistic effects, returning radiation, and viewing geometry, a detailed calculation is required to confirm the above speculation and discriminate between the models. The impending launch of NASA’s Gravity and Extreme Magnetism Small Explorer (Swank et al. 2010; GEMS) mission promises observational constraints that may be capable of discriminating between such models.

2.7.2 How does our choice of cooling function influence the results?

A criticism that might be levied against the GRMHD simulations is that our choice of cooling function (as defined in Eq. 2.1) is arbitrary. We have interpreted this cooling function as the rate of radiative cooling, despite the fact that the GRMHD simulations do not inherently model radiation physics.

We would like to test how our choice of cooling function impacts the results of §2.6. Rather than rerunning the GRMHD simulations with another prescription for cooling (i.e. by changing the functional form of Eq. 2.1), we use the simulation’s dissipative heating profile as another means to get a cooling rate. Instead of simply assuming that dissipation equals cooling locally, we consider the physics of energy advection (i.e. that the fluid releases its heat downstream). To generate a disc model that is more self-consistent with the TLUSTY annuli, we use the TLUSTY vertical
structure when computing the rate of energy advection (see Appendix 2.12 for a thorough exposition on this method of computing the new GRMHD luminosity profile). The new dissipation-based cooling function (which we label as GRMHD2) releases slightly more luminosity inside the plunging region (compare the solid and dashed lines in Figure 2.1).

We repeat the same exercise as in §2.6 with this new GRMHD2 derived luminosity profile. The spectra corresponding to GRMHD2 are shown as the dashed lines in Figure 2.9. Without fitting for the power laws, we find that the recovered spins are slightly higher owing to the extra plunging region luminosity (compare the ‘GRMHD’ and ‘GRMHD2’ columns in Table 2.2). However, the more realistic fit (i.e. including the power-law component) takes care of this extra light, yielding comparable spins estimates to before. The enhanced power laws are purely due to the extra plunging region luminosity (compare solid and dashed lines in Figures 2.1 and 2.9). The worst cases for the 30 per cent Eddington discs now have $\Delta a_* \sim 0.15, 0.07, 0.03$, for $a_* = 0, 0.7, 0.9$ respectively. Given these results, we still conclude that the extra luminosity from the GRMHD inner disc does not limit our ability to measure BH spin through the continuum fitting method (the dominant source of uncertainty is still the observational uncertainties in distance, mass, and inclination).

As in Kulkarni et al. (2011), we estimate that $\Delta a_* \leq 0.15$, which is lower than the estimate of Noble et al. (2011), who infer $\Delta a \sim 0.2 - 0.3$ for a Schwarzschild black hole. Due to the reduced sensitivity of the continuum fitting method at low spins, this amounts to a somewhat modest discrepancy, which we attribute to various differences in the setup and initial conditions of the GRMHD simulations used by the different groups. We refer the reader to Penna et al. (2010); Kulkarni et al. (2011); Noble et al. (2010, 2011); Hawley, Guan, & Krolik (2011) for further discussion.
2.7.3 Equation of state

Another inconsistency in the GRMHD simulations is that we adopt an ultrarelativistic gas equation of state (with $\Gamma = 4/3$), whereas the TLUSTY annuli calculations suggest that the gas is nonrelativistic (since $kT_{\text{gas}} \ll 511\text{keV}$), which corresponds to a choice of $\Gamma = 5/3$. However despite the difference in $\Gamma$, Noble et al. (2009, 2010, 2011) have also performed GRMHD simulations of thin discs with $\Gamma = 5/3$, and they find similar results to the $\Gamma = 4/3$ discs of Shafee et al. (2008b); Penna et al. (2010).

Figure 2.14: GRMHD vs. TLUSTY vertically averaged specific gas entropy profiles (relative to the entropy at the horizon $s_{\text{hor}}$). The dimensionless entropy is computed using an ideal gas equation of state, where $s = \left(\frac{k}{\mu}\right) \cdot \frac{1}{\Gamma - 1} \ln\left(\frac{p_{\text{gas}}}{\rho^\Gamma}\right)$. The TLUSTY lines (dotted) denote the gas entropy as computed from the TLUSTY annuli vertical structure. The TLUSTY2 lines (dashed) represent annuli that comprise the GRMHD2 disc model (described in §2.7.2 and Appendix 2.12).
In figure 2.14, we examine the relative impact that this equation of state inconsistency has on the disc’s entropy profile. A fully self-consistent model would have identical GRMHD and TLUSTY gas entropy profiles. It appears that well outside the ISCO, the assumption that the gas cools towards a fixed entropy is supported by the radiative transfer calculations of TLUSTY (Compare the flat plateaus in Figure 2.14). However Figure 2.14 shows that the GRMHD (solid) and TLUSTY (dotted) discs reach different constant entropies. The alternate disc cooling model (dashed) discussed in §2.7.2 yields annuli which are more self-consistent with the simulation entropy profile (compare dashed lines and solid lines in Figure 2.14, which now track each other fairly well outside the ISCO).

2.8 Summary

The primary goal of this work is to determine how important the neglected light from the plunging region is in the context of BH spin measurements. To answer this question, we rely on GRMHD disc simulations of Penna et al. (2010), which capture all but the radiation physics for magnetized flow around a black hole. We seek to convert these dimensionless simulations into a form that can be directly compared with observations, namely the accretion discs’ X-ray continuum spectra.

To do this, we apply radiative transfer post-processing on the simulated discs. This is an advance over previous work on computing GRMHD disc spectra (i.e. Kulkarni et al. 2011), which assumes (modified) blackbody annuli spectra. We slice the GRMHD disc into many individual annuli, and for each annulus, we apply a 1-dimensional radiative transfer calculation to solve for its emergent spectrum. We sum up this collection of local
spectra by means of ray tracing to get the full disc spectrum, which we then compare
with contemporary disc models used to measure black hole spin. The following are the
key results from this work:

(1). The GRMHD based accretion discs have hotter spectra than the standard
Novikov & Thorne (1973) discs. The GRMHD discs produce more luminosity ev-
erywhere, and the contrast becomes most apparent inside the ISCO (see Figure 2.1).

(2). The increased luminosity of the GRMHD discs compared to the classic NT discs
induces a modest systematic bias in the derived spins of these GRMHD discs. For
black holes of spin $a_\ast = 0, 0.7, 0.9$ the spin deviation is $\Delta a_\ast \sim 0.15, 0.07, 0.03$ in
the worst cases (corresponding to inclination angles of $75^\circ$). We remark that these
errors are still well within other observational uncertainties (i.e. from not precisely
knowing the system’s mass, inclination, and distance – see Table 2.3).

(3). Without needing to invoke an external corona, the GRMHD discs around spinning
black holes exhibit a weak high-energy power-law tail (Fig. 2.9). This power-law
tail arises from the combined emission of the hot plunging region gas. The strength
of this plunging region power-law increases with the system’s inclination angle.

In our spectral modelling approach, we have made many simplifications and
assumptions (see §2.7), some of which may be incorrect close to the black hole horizon.
Due to these problems, we trust the power laws presented here only to a qualitative level
we only trust that the plunging region fluid emits at much higher energies than the disc, which adds a nonthermal component to the overall spectrum at high energies.

### 2.9 Luminosity Matching Model

The functional form of the matching flux profile is given below (see Kulkarni et al. 2011, Page & Thorne 1974 – hereafter PT):

\[
f(r) = f_{\text{PT}}(r) - \left[ \frac{(E^\dagger - \Omega L^\dagger)^2 / \Omega_r}{(E^\dagger - \Omega L^\dagger)^2 / \Omega_r} \right] f_{\text{PT}}(r_{ie})
- \left[ \frac{\Omega_r}{(E^\dagger - \Omega L^\dagger)^2} \right] r_{ie} C, \quad \text{when } r > r_{ie}
\]  

(2.11)

where \( f(r) = 4\pi r F_{\text{com}}(r)/\dot{M} \) is the dimensionless flux, \( L^\dagger, E^\dagger, \Omega \) are the specific angular momentum, energy-at-infinity, angular velocity (given by Eqs. 15f-h of PT), \( f_{\text{PT}} \) is the flux from the Page & Thorne model (Eq. 15n of PT), and \( C \) is a free parameter that is determined by the torque at the inner disc boundary.

The task now is to find an appropriate value of \( C \) such that the GRMHD flux and Eq. 2.11 match at \( r_{ie} \). For simplicity, we factor out the radial dependence from the last two terms of Eq. 2.11 to get:

\[
f(r) = f_{\text{PT}}(r) - \left[ \frac{\Omega_r}{(E^\dagger - \Omega L^\dagger)^2} \right] K,
\]  

(2.12)

where \( K \equiv \left\{ \left[ (E^\dagger - \Omega L^\dagger)^2 / \Omega_r \right] f_{\text{PT}}(r_{ie}) - C \right\} \) is just another constant. To ensure continuity in the flux profile at \( r_{ie} \), we require \( K \) to be:

\[
K = \left[ \frac{(E^\dagger - \Omega L^\dagger)^2}{\Omega_r} \right]_{r_{ie}} \left[ f_{\text{GRMHD}}(r_{ie}) - f_{\text{PT}}(r_{ie}) \right].
\]  

(2.13)
Thus, our final matched flux profile is given by:

\[ \sigma_{SB} T_{\text{eff}}(r)^4 = \begin{cases} \frac{1}{4\pi r u_t} \frac{dL_{\text{BL}}(r)}{dr}, & r \leq r_{ie}, \\ \frac{\dot{M}}{4\pi r} f(r), & r > r_{ie}, \end{cases} \]

(2.14a)

where \( L_{\text{BL}}(r) \) is the Boyer-Lindquist luminosity profile measured from the GRMHD simulations, and \( f(r) \) is the dimensionless flux as shown in Eq. 2.12.

### 2.10 Generalized Novikov & Thorne Model

To get the generalized Novikov & Thorne (1973) column density and radial velocity profile far out in the disc (to extend the GRMHD simulated disc beyond \( r_{ie} \)), we solve the following set of vertical structure equations (with a one-zone model for the vertical structure):

1. Vertical Pressure Balance:

\[ \frac{dP_{\text{tot}}}{dz} = \rho g_{\perp}, \]

has the vertically integrated form of:

\[ \frac{P_{\text{tot}}}{h} = \rho Q h, \]

(2.15a)

where \( h \) is the vertical scale height, and \( Q = g_{\perp}/z \) is the prescription for the local vertical gravity (as defined in Eq. 2.7).

2. Radiative Transfer:
CHAPTER 2. LIGHT FROM THE PLUNGING REGION OF ACCRETION DISCS

From the second moment of radiative transfer equation (see §1.8 of Rybicki & Lightman 1979):

\[
\frac{dP_{\text{rad}}}{d\tau} = \frac{F_{\text{rad}}}{c},
\]

we transform the \(d\tau\) optical depth differential to a column mass differential \(dm\) via
\(d\tau = \kappa \cdot dm\), where \(\kappa\) is the Rosseland mean opacity. We integrate the column mass from the surface to the disc midplane (from \(m = 0\) to \(m = m_{\text{tot}} = \Sigma/2\)). We also assume that the flux profile linearly increases with mass away from the midplane (i.e. \(F(m) = F_{\text{tot}}(1 - m/m_{\text{tot}})\) – this linearity assumption is also used in TLUSTY).

The vertically integrated radiative diffusion equation thus becomes

\[
P_{\text{rad}} = \frac{aT^4}{3} = \frac{F_{\text{tot}} \kappa \Sigma}{4c}, \tag{2.15b}
\]

where \(a\) is the radiation constant, and \(T\) is radiation temperature. Note: The factor 2 discrepancy in Eq. 2.15b compared to the standard prescription for radiative diffusion stems from the linearity assumption in \(F(m)\).

3. Stress:

We adopt an \(\alpha\) prescription for the stress where \(t_{\phi}\) = \(\alpha p_{\text{tot}}\). Vertical integration yields

\[
W \equiv \int_{-h}^{h} t_{\phi} dz = 2h\alpha P_{\text{tot}}. \tag{2.15c}
\]

4. Viscous Heating:
CHAPTER 2. LIGHT FROM THE PLUNGING REGION OF ACCRETION DISCS

Through the energy equation for viscous heating, it is possible to link the heating flux with the vertically integrated stress (c.f. 5.6.7-12 of Novikov & Thorne 1973). The resulting expression is

\[ F_{\text{tot}} = \frac{3}{4} \Omega_k R_F(r) W, \]  

(2.15d)

with the dimensionless relativistic factor \( R_F(r) \) defined as

\[ R_F(r) = \frac{1 - 2/r_+ + a_s^2/r_+^2}{1 - 3/r_+ + 2a_s/r_+^{3/2}}. \]

5. Equation of State:

We ignore magnetic pressure in this analysis, and adopt an ideal gas equation of state, yielding

\[ P_{\text{tot}} = \frac{\rho k_B T}{\mu} + \frac{aT^4}{3}, \]  

(2.15e)

where \( \mu \) is the mean particle weight of the fluid.

6. Column Density

In this one-zone model, the column density represents the quantity

\[ \Sigma = \int_{-h}^{h} \rho dz = 2h \rho. \]  

(2.15f)

Our goal is to solve Eqs. 2.15 with unknowns \((P_{\text{tot}}, \rho, T, \Sigma, h)\), given the values of \((\kappa, \mu, \alpha, F_{\text{tot}}, Q, M, r)\). After some algebra, we obtain an expression for \( \Sigma \) that solves Eqs.
2.15. We find $\Sigma$ by solving for the real root of the following polynomial in $x$:
\[
(F_1)^2 - \left[ \frac{F_1}{4F_3} \right] x^4 - \left[ 2 \left( \frac{3}{8} \right)^{1/4} F_1 F_2 \right] x^5 + \left[ \left( \frac{3}{8} \right)^{1/2} (F_2)^2 \right] x^{10} = 0,
\]
where $x = \Sigma^{1/4}$ and the polynomial coefficients $F_1, F_2, F_3$ depend only on the given values:
\[
F_1 = \frac{W}{\alpha} = \frac{4F_{\text{tot}}^3}{3\alpha R_F \Omega_k}, \quad (2.17a)
\]
\[
F_2 = \left( \frac{\kappa F_{\text{tot}}}{2\sigma_{SB}} \right)^{1/4} \left( \frac{k_B}{\mu} \right), \quad (2.17b)
\]
\[
F_3 = \frac{c^2 Q}{\kappa^2 F_{\text{tot}}^2}. \quad (2.17c)
\]
We then solve Eq. 2.16 at different radii to obtain $\Sigma(r)$ for $(r > r_{ie})$. We pick $\kappa = \kappa_{es} = 0.3383 \text{cm}^2 \text{g}^{-1}$ and $\mu = 0.615 m_H$, which corresponds to a fluid composition of 70 per cent Hydrogen, 30 per cent Helium by mass. Given $\Sigma(r)$, we get the mass averaged radial velocity by solving for $\tilde{u}^r$ in the mass conservation equation $\dot{M} = 2\pi r \Sigma \tilde{u}^r$.

2.11 Interpolation Methods

All interpolated quantities are computed by linear interpolation in log space (i.e. for two variables $x$, and $y$, we interpolate linearly on log($x$) and log($y$)). We have chosen this interpolation scheme since it can perfectly capture all power-law scalings. For quantities
CHAPTER 2. LIGHT FROM THE PLUNGING REGION OF ACCRETION DISCS

obtained by interpolating on TLUSTY annuli (such as the annuli vertical structure), we employ trilinear interpolation over the 3 annuli parameters \( \log(T_{\text{eff}}) \), \( \log(\Sigma) \), and \( \log(Q) \) (i.e. each parameter is linearly interpolated separately in log space).

The interpolation of spectra \( I_\nu \) is handled by a more complicated method to account for the shape of the Planck function. For an optically thick medium with scattering, the resultant spectrum takes on the form of a modified blackbody (c.f. Eq. 2.2). Since the spectra doesn’t depend very sensitively on \( \Sigma \), or \( Q \), we apply the log-linear interpolation discussed above for these two annuli parameters. For \( T_{\text{eff}} \), we use a more complicated interpolation scheme. Rather than follow Davis et al. (2005) (they applied a linear interpolation on the brightness temperature, computed with a fixed spectral hardening factor of \( f = 2.0 \)), we switch between three different interpolation methods that each work for all choices of \( f \). The three methods are applied to different frequency ranges of the spectrum.

- Method 1: At low frequencies, we have that \( I_\nu \propto T_{\text{eff}} \) in the Raleigh-Jeans tail, and thus we can apply the linear interpolation method.

- Method 2: At high frequencies, we apply linear interpolation on \( 1/T_{\text{eff}} \) and \( \log(I_\nu) \). The motivation is that for a blackbody-like spectrum in the high-frequency Wien limit, the specific intensity as a function of temperature scales as \( \log(I_\nu) \propto -1/T_{\text{eff}} \).

- Method 3: For intermediate frequencies, we take a non-linear combination of the two interpolation results so that the transition between interpolation methods is smooth.
CHAPTER 2. LIGHT FROM THE PLUNGING REGION OF ACCRETION DISCS

Figure 2.15: Comparison of annuli spectra obtained through the interpolation method (solid) and an exact calculation (dashed) for the spin $a_*=0.7$ disc. The black spectra represent annuli outside the plunging region, whereas the red spectrum corresponds to a plunging region annulus.

Denoting the interpolated intensities using methods 1 and 2 as $I_1(\nu)$ and $I_2(\nu)$ respectively, the expression for the final combined interpolation method is:

$$I_\nu(\nu) = \begin{cases} 
I_1(\nu), & \nu \leq \nu_1, \\
\exp \left\{ \frac{\ln(\nu_2) - \ln(\nu)}{\ln(\nu_2) - \ln(\nu_1)} \ln \left[ I_1(\nu) \right] \\
+ \frac{\ln(\nu) - \ln(\nu_1)}{\ln(\nu_2) - \ln(\nu_1)} \ln \left[ I_2(\nu) \right] \right\}, & \nu_1 < \nu < \nu_2, \\
I_2(\nu), & \nu \geq \nu_2.
\end{cases}$$

We have set $\nu_1 = 0.07 \nu_{\text{max}}$, and $\nu_2 = 2.3 \nu_{\text{max}}$, where $\nu_{\text{max}}$ is the frequency that maximizes $I_\nu$ in the spectra of the lowest $T_{\text{eff}}$ annuli in the bracketing set used for the interpolation. These frequency boundary values were determined by minimizing the
interpolation error for a pure Planck function. For a perfect modified Planck function with constant \( f \), the chosen \( \nu_1 \) and \( \nu_2 \) correspond to a maximum interpolation error of 0.08 per cent in \( I_{\nu} \) for our grid with temperature spacings \( \Delta \log_{10}(T_{\text{eff}}) = 0.1 \). In practice, the interpolation error after interpolating over the 3 annuli parameters and emission angle is \( \sim 1 \) per cent (see Fig. 2.15), except for annuli in the plunging region, where the flux error grows to \( \sim 10 \) per cent (the assumption of a constant \( f \) is violated for annuli in the plunging region). The main advantage of this interpolation method is it provides interpolated spectra that peak very close to the correct frequencies, and the method works equally well for spectra of all spectral hardening factors.

2.12 An Alternative GRMHD Luminosity Profile

We wish to generate a disc luminosity profile that has the following desirable properties: 1) it is based on the GRMHD dissipation profile rather than cooling profile, and 2) it produces a disc whose entropy profile is self-consistent with the TLUSTY annuli. We accomplish this task by considering the physics of energy advection. The idea is to solve for the radiative cooling rate, taking into account both the dissipative heating rate (measured from the GRMHD simulations) and the heat advection rate (computed using the vertical structure from TLUSTY annuli). The energy balance equation relating these three quantities is:

\[
\dot{q}_{\text{adv}} = \dot{q}_{\text{heat}} - \dot{q}_{\text{cool}},
\]

(2.19)

where \( \dot{q}_{\text{adv}} \) is the advected heating rate per unit volume, \( \dot{q}_{\text{heat}} \) the viscous heating rate per unit volume, and \( \dot{q}_{\text{cool}} \) is the cooling rate per unit volume (assumed to be purely radiative
CHAPTER 2. LIGHT FROM THE PLUNGING REGION OF ACCRETION DISCS

– i.e. we ignore conduction and convection). We take the convention that all hatted quantities are measured in the comoving frame of the fluid. The advective heating rate is obtained by evaluating $\dot{q}_{\text{adv}} = \dot{\rho}\dot{T}(d\dot{s}/d\tau)$, where $\dot{\rho}$ is the rest mass density, $\dot{T}$ is the gas temperature, $\dot{s}$ is the specific gas entropy of the fluid, and $\tau$ is the fluid’s proper time.

Making the approximation that we have stationary axisymmetric flow that is devoid of motion perpendicular to the midplane, we have $d\dot{s}/d\tau = u^r d\dot{s}/dr$ via the chain-rule ($d\dot{s}/dr$ and $u^r$ are evaluated in BL-coordinates). Vertically integrating Eq. 2.19 while making use of $\dot{q}_{\text{adv}} = \dot{\rho}\dot{T}u^r(d\dot{s}/dr)$ yields:

$$\dot{Q}_{\text{adv}} = \dot{Q}^{\text{sim}}_{\text{heat}} - \dot{Q}_{\text{cool}}$$

$$= \int_{z=-\infty}^{z=+\infty} \dot{\rho}\dot{T}u^r \left( \frac{d\dot{s}}{dr} \right) dz,$$

where $\dot{Q}_{\text{adv}}$ is the vertically integrated heat advection rate (evaluated with TLUSTY vertical structure), $\dot{Q}^{\text{sim}}_{\text{heat}}$ is the vertically integrated viscous heating rate (measured from the GRMHD simulated discs), and $\dot{Q}_{\text{cool}}$ is the net radiative cooling flux that escapes.

For simplicity in Eq. 2.21, we adopt the following constant mass-averaged radial velocity:

$$u^r = \frac{\int_0^\pi \dot{\rho}u^r_{\text{sim}} \sqrt{-g} d\theta}{\int_0^\pi \dot{\rho} \sqrt{-g} d\theta}.$$

where $u^r_{\text{sim}}$ represents the pointwise radial velocities measured from GRMHD simulations.

Our goal is to find the value of $\dot{Q}_{\text{cool}}$ that satisfies Eq. 2.20, given that we can measure $\dot{Q}^{\text{sim}}_{\text{heat}}$ from the simulations and calculate $\dot{Q}_{\text{adv}}$ from Eq. 2.21.
2.12.1 Obtaining the GRMHD dissipation profile

Unfortunately, the GRMHD simulations that we ran were not set up to directly keep track of the numerical dissipation $\dot{Q}_{\text{heat}}^{\text{sim}}(r)$. Despite this lack of information, we can still estimate the dissipation indirectly by running the argument in Eq. 2.20 backwards; we solve for the vertically integrated dissipative heating rate

$$\dot{Q}_{\text{heat}}^{\text{sim}} = \dot{Q}_{\text{adv}}^{\text{sim}} + \dot{Q}_{\text{cool}}^{\text{sim}}.$$  \hspace{1cm} (2.23)

The ‘sim’ superscript is used to denote quantities derived solely from the GRMHD simulations (i.e. these quantities are independent of TLUSTY). Analogous to Eq. 2.21, the vertically integrated\(^{12}\) GRMHD advective heating rate is obtained by

$$\dot{Q}_{\text{adv}}^{\text{sim}} = \int_{\theta=0}^{\pi} \rho \hat{T} \left( \frac{d\hat{s}}{dr} \right) u^r \sqrt{-g} d\theta.$$  \hspace{1cm} (2.24)

For simplicity, we first apply azimuthal and time averaging to all simulation quantities used in Eq. 2.24. Since the GRMHD simulations are dimensionless ($k/\mu = 1$) and employ an ideal gas equation of state, we have that $\hat{T} = \hat{P}_\text{gas}/\hat{\rho}$, and $\hat{s} = (\Gamma - 1)^{-1} \ln(\hat{P}_\text{gas}/\hat{\rho}^\Gamma)$. The GRMHD cooling flux is simply $\dot{Q}_{\text{cool}}^{\text{sim}} = F_{\text{com}}$ where $F_{\text{com}}$ is the comoving disc flux, as given by Eq. 2.4.

2.12.2 Net result of the luminosity calculation

The final goal is to solve for the value of $\dot{Q}_{\text{cool}}$ that satisfies Eq. 2.20. The physical interpretation of this newly derived $\dot{Q}_{\text{cool}}$ is simply the cooling rate corresponding to a

\(^{12}\)To maintain consistency with the simulation cooling flux $\dot{Q}_{\text{cool}}^{\text{sim}}$ (as calculated by Eqs. 2.3 and 2.4, which only considers bound disc fluid), the integral in Eq. 2.24 is also only taken over the bound fluid.
disc with the vertical structure given by TLUSTY that is heated up according to the
GRMHD dissipation profile. Putting everything together, substituting Eq. 2.23 for $\hat{Q}_{\text{heat}}$
into Eq. 2.20 gives $\hat{Q}_{\text{cool}}$ as:

$$\hat{Q}_{\text{cool}} = \hat{Q}_{\text{cool}}^{\text{sim}} + \hat{Q}_{\text{adv}}^{\text{sim}} - \hat{Q}_{\text{adv}}. \quad (2.25)$$

$\hat{Q}_{\text{adv}}^{\text{sim}}$ and $\hat{Q}_{\text{cool}}^{\text{sim}}$ are computed from Eq. 2.24 and Eq. 2.4 respectively. $\hat{Q}_{\text{adv}}$ uses the
TLUSTY vertical structure, and is evaluated via Eq. 2.21. However, the process of
locating the correct value for $\hat{Q}_{\text{cool}}$ that solves Eq. 2.25 is complicated by two factors:

1. $\hat{Q}_{\text{adv}}$ on the right hand side is a function of $\hat{Q}_{\text{cool}}$. $\hat{Q}_{\text{adv}}$ indirectly depends on $\hat{Q}_{\text{cool}}$ through the annuli vertical structure (since $T_{\text{eff}} = [\hat{Q}_{\text{cool}}/\sigma_{\text{SB}}]^{1/4}$ is an annuli parameter). The strategy that we employ to find the correct $\hat{Q}_{\text{cool}}$ is a bisection method; we start with the two initial guesses for $\hat{Q}_{\text{cool}}$ that bracket the relation in
Eq. 2.25 (which we have empirically found to be monotonic with $\hat{Q}_{\text{cool}}$). We then
bisect on this $\hat{Q}_{\text{cool}}$ interval until Eq. 2.25 is satisfied.

2. To compute the $d\hat{s}/dr$ term, we need to know $\hat{Q}_{\text{cool}}$ (or equivalently $T_{\text{eff}}$) for two
neighboring annuli. However the process outlined above (in point 1) only lets us
solve $\hat{Q}_{\text{cool}}$ for a single annulus. The resolution to this problem is to choose a
value of $\hat{Q}_{\text{cool}}$ for the outermost annulus as a boundary condition. If there are a
total of $N$ annuli, then given $\hat{Q}_{\text{cool}}$ for the $N^{\text{th}}$ annuli, we can solve $\hat{Q}_{\text{cool}}$ for the
$(N - 1)^{\text{th}}$ annuli. This process can be iterated to obtain $\hat{Q}_{\text{cool}}$ for all remaining
interior annuli. We set $\hat{Q}_{\text{cool}} = \hat{Q}_{\text{heat}}$ as the boundary condition for our outermost
annuli since far into the disc, energy advection becomes negligible (in other words,
CHAPTER 2. LIGHT FROM THE PLUNGING REGION OF ACCRETION DISCS

\[ \dot{Q}_{\text{adv}} \to 0 \text{ far out in the disc, which implies } \dot{Q}_{\text{cool}} \to \dot{Q}_{\text{heat}} \text{ from Eq. 2.20).} \]

Running through the above two steps allows us to solve for \( \dot{Q}_{\text{cool}} \), and hence obtain a second measure of the disc luminosity (labelled as GRMHD2 in all plots and tables). Note that the GRMHD2 model is more self-consistent with TLUSTY in Fig. 2.14. The TLUSTY2 model (computed from the newly derived GRMHD2 disc luminosities) and the intrinsic GRMHD entropy profiles agree fairly well when \( r > r_{\text{ISCO}} \).
Chapter 3

Thermal Stability in Turbulent Accretion Discs

This thesis chapter originally appeared in the literature as


Abstract

The standard thin accretion disc model predicts that discs around stellar mass black holes become radiation pressure dominated and thermally unstable once their luminosity exceeds $L \gtrsim 0.02L_{\text{Edd}}$. Observationally, discs in the high/soft state of X-ray binaries show little variability in the range $0.01L_{\text{Edd}} < L < 0.5L_{\text{Edd}}$, implying that these discs in nature are in fact quite stable. In an attempt to reconcile this conflict, we investigate one-zone disc models including turbulent and convective modes of vertical energy transport. We find both mixing mechanisms to have a stabilizing effect, leading to an increase in the
CHAPTER 3. CONVECTION AND TURBULENT STABILIZATION OF DISCS

$L$ threshold up to which the disc is thermally stable. In the case of stellar mass black hole systems, convection alone leads to only a minor increase in this threshold, up to $\sim 5$ per cent of Eddington. However turbulent mixing has a much greater effect – the threshold rises up to 20 per cent Eddington under reasonable assumptions. In optimistic models with superefficient turbulent mixing, we even find solutions that are completely thermally stable for all accretion rates. Similar results are obtained for supermassive black holes, except that all critical accretion rates are a factor $\sim 10$ lower in Eddington ratio.

3.1 Introduction

Accretion discs are ubiquitous in our universe; their physics governs the production of powerful jets from Active Galactic Nuclei, the formation of planets, and the growth of black holes. This wealth of applicability has led to a rich field of study where many open problems still persist to this day (Frank, King & Raine 2002).

One longstanding puzzle relates to the stability of discs. In the standard $\alpha$-model for thin accretion discs (Shakura & Sunyaev 1973; Novikov & Thorne 1973), there is a critical accretion rate above which the disc transitions from being gas pressure dominated to radiation pressure dominated, and simultaneously switches from being thermally stable to unstable (Shakura & Sunyaev 1976; Piran 1978). The model predicts that the transition should occur at an accretion rate above a few tenths of a per cent of Eddington, depending on the mass of the central object. Above this rate, we expect limit-cycle behaviour due to the onset of both thermal instability (Shakura & Sunyaev 1976; Honma, Matsumoto & Kato 1992; Szuszkiewicz & Miller 1998; Janiuk, Czerny
\textit{CHAPTER 3. CONVECTION AND TURBULENT STABILIZATION OF DISCS}

& Siemiginowska 2002) and viscous instability (Lightman & Eardley 1974). However, observationally there is little variability and no evidence of a limit-cycle for systems with substantially larger luminosities than the theoretical limit. Indeed, Gierliński & Done (2004) show that discs around stellar mass black holes remain stable up to 50 per cent Eddington, which conflicts with the prediction of standard thin-disc theory by more than an order of magnitude.

Over the years, many ideas have been proposed to resolve this inconsistency. Piran (1978) noted that the prediction of thermal instability hinges crucially on taking the standard $\alpha$-viscosity prescription in disc models. Other prescriptions for viscosity, such as having disc stress scale with gas pressure instead of total pressure, admit solutions that are thermally stable everywhere (Kato, Fukue & Mineshige 2008). However, shearing box disc simulations with radiation (Hirose, Blaes & Krolik 2009) demonstrate that the standard $\alpha$-prescription, in which the stress scales linearly with the total pressure, is correct. Thus, we do not have freedom to modify the viscosity prescription.

Other attempts to resolve the stability paradox in the context of the $\alpha$-viscosity paradigm include modeling discs with: strong irradiation at the disc surface (Tuchman et al. 1990); a superefficient corona that rapidly siphons heat from the disc and avoids instability by keeping the disc cool (Svensson & Zdziarski 1994; Różańska et al. 1999); strong magnetic pressure support that dominates the vertical structure at low accretion rates (Zheng et al. 2011); time lags between the generation of disc stress and pressure response (Ciesielski et al. 2012) as suggested by numerical simulations (Hirose, Krolik & Blaes 2009); and finally convection (Milsom, Chen & Taam 1994; Goldman & Wandel 1995).
CHAPTER 3. CONVECTION AND TURBULENT STABILIZATION OF DISCS

In early works on convective discs, Bisnovatyi-Kogan & Blinnikov (1977) and Goldman & Wandel (1995) found convection to be superefficient, dominating over radiation in the vertical transport of energy. This superefficient convective flux strongly suppresses the radiative channel, resulting in cooler discs that are completely thermally stable (Goldman & Wandel 1995). However, other groups who performed more detailed calculations in which they evaluated the complete vertical structure (Shakura et al. 1978; Cannizzo 1992; Milsom, Chen & Taam 1994; Heinzeller, Duschl & Mineshige 2009) found a more modest effect where convection carries no more than $\sim 1/2$ of the vertical flux of energy, and has only a weak effect on the disc’s stability (Cannizzo 1992; Milsom, Chen & Taam 1994; Sadowski et al. 2011).

These later models still find that convection does provide a stabilizing force that pushes the instability threshold towards higher accretion rates, but the effect is modest and falls far short of what is needed to explain observations. In this connection, we note that turbulent convection alone is too small by a factor of 10-100 (Ruden, Papaloizou & Lin 1988; Ryu & Goodman 1992; Goldman & Wandel 1995) to produce the viscous stress present in accretion discs (Pringle & Rees 1972; King, Pringle & Livio 2007).

Disc viscosity is believed to arise from the Magneto-Rotational Instability (MRI; Balbus & Hawley 1991, 1998). One consequence of MRI turbulence is that it will itself induce a vertical transfer of energy, in a somewhat analogous fashion as convection. What effect will this have on the thermal stability of the disc? To date, no analytic model has been developed that takes into account both convection and turbulence. In this work we attempt to develop the simplest version of such a model. Based on the results of Agol et al. (2001), we expect turbulent mixing to have a significant impact on the disc’s vertical structure and hence also on disc stability. We find that turbulence
Indeed acts in a similar way to convection, i.e. turbulent discs are substantially more resilient to the onset of thermal instability. For reasonable choices of model parameters, we show that it is possible to push the instability threshold up close to the observational limits.

The organization of the paper is as follows. In §3.2, we present the governing equations and details of our turbulent disc model. We then proceed to compute three types of discs using our model: 1) a classic purely radiative disc, 2) a convective disc, and 3) a convective plus turbulent disc. In §3.3 we compare the stability properties and radial structure for these three types of disc. Section 3.4 focuses on the impact of various assumptions made in our disc model and also briefly compares our work to the results of numerical simulations. Finally, we conclude in §3.5 with a summary of the important points.

### 3.2 Physical model

We wish to build the simplest possible disc model that includes the physics of convection and turbulent vertical mixing. To this end, we separate the equations describing the disc’s vertical and radial structure, and model the vertical structure as a single homogeneous slab (see Fig. 3.1). In essence, our approach is similar to the one-zone relativistic disc model of Novikov & Thorne (1973), but including the effects of convective and turbulent mixing.

The primary effect of convection and turbulence is to add a new channel for vertical energy transport. Since the computation of convective flux requires information about
the vertical gradient of temperature, any convective model must have at least two vertically separated probe points. For convenience, we choose these two sampling points to be the disc midplane and the density scale height.

The strategy we employ is to first obtain as a function of radius quantities that are independent of the vertical disc structure. We then use these quantities as boundary conditions to uniquely specify the disc vertical structure, and hence solve for the complete disc model.

3.2.1 Radial structure

Luminosity:

For simplicity, we assume that the advected energy is negligible. This allows us to calculate the net radiative flux leaving the disc at each radius, given only the black hole mass \( M \), dimensionless spin \( a_* = J/(GM^2/c) \), and mass accretion rate \( \dot{M} \). We adopt the prescription of Page & Thorne (1974) for determining the disc’s luminosity profile. In the subsequent vertical structure equations, we make use of the total emergent disc flux

\[
F_{\text{tot}} = \sigma T_{\text{eff}}^4.
\] (3.1)

Vertical gravity:

The tidal vertical gravity in the disc is a function purely of the Kerr metric and the motion of the disc fluid. For circular orbits, the tidal vertical gravity \( g_z \) is given by


CHAPTER 3. CONVECTION AND TURBULENT STABILIZATION OF DISCS

(Riffert & Herold 1995):

\[ g_z \equiv Q \cdot z = \left( \frac{GM}{r^3} \right)^{1/2} R_z(r) \cdot z, \]  \quad (3.2)

where \( z \) is height above the midplane, and \( R_z(r) \) is a dimensionless relativistic factor given by:

\[ R_z(r) = \frac{1 - 4a_*r_*^{3/2} + 3a_*^2/r_*^2}{1 - 3/r_* + 2a_*r_*^{3/2}}. \]  \quad (3.3)

We have defined the dimensionless radius \( r_* = r/(GM/c^2) \). Note that the quantity \( Q = g_z/z \) is independent of \( z \).

Rotation rate:

For circular motion about a Kerr black hole, the orbital velocity is given by (c.f. Novikov & Thorne 1973):

\[ \Omega_k = \left( \frac{GM}{r^3} \right)^{1/2} \left( \frac{1}{1 + a_*r_*^{3/2}} \right). \]  \quad (3.4)

3.2.2 Vertical structure

Unknown variables

At every fixed radius of the accretion disc, we wish to solve for the following 8 unknowns that describe the vertical structure:

- \( T_{\text{mid}} \) – Midplane temperature
- \( P_{\text{mid}} \) – Midplane pressure (gas+radiation)
CHAPTER 3. CONVECTION AND TURBULENT STABILIZATION OF DISCS

- $T_{\text{scale}}$ – Temperature at one scale height
- $P_{\text{scale}}$ – Pressure at one scale height (gas+radiation)
- $\Sigma$ – Vertical column density
- $H$ – Vertical pressure scale height
- $\nabla = d\ln T/d\ln P$ of the ambient medium
- $\nabla_e = d\ln T/d\ln P$ of a convective element

Equations:

We employ the following 8 equations to solve for the 8 unknowns:

1. Vertical pressure balance This gives
   \[
   \frac{dP_{\text{tot}}}{dz} = \rho g_z, \]
   which has the vertically integrated form:
   \[
   \frac{P_{\text{tot}}}{H} = \rho QH = \frac{\Sigma Q}{2}, \tag{3.5}
   \]
   where $H$ is the vertical pressure scale height, $Q = g_z/z$ is defined by Eq.(3.2), and we have written the density as $\rho = \Sigma/2H$ in the spirit of a one-zone model.

2. Viscous heating:

   Through the energy equation for viscous heating, it is possible to link the disc flux $F_{\text{tot}}$ with the vertically integrated stress $W$ (See 5.6.7-12 of Novikov & Thorne 1973). The resulting expression is
   \[
   F_{\text{tot}} = \frac{3}{4} \Omega_k R_F(r) W, \tag{3.6a}
   \]
with the dimensionless relativistic factor $R_F(r)$ defined as

$$R_F(r) = \frac{1 - 2/r_s + a^2/r_s^2}{1 - 3/r_s + 2a_s/r_s^{3/2}}. \tag{3.6b}$$

Taking the $\alpha$-prescription for the stress (where $t_{\phi\phi} = \alpha P_{\text{tot}}$), we have:

$$W \equiv \int_{-H}^{H} t_{\phi\phi} dz = 2H\alpha P_{\text{mid}}. \tag{3.6c}$$

3. Midplane equation of state:

We ignore magnetic pressure in this analysis, and adopt an ideal gas plus radiation equation of state,

$$P_{\text{mid}} = \frac{\rho k_B T_{\text{mid}}}{\mu} + \frac{aT_{\text{mid}}^4}{3} = \frac{\Sigma k_B T_{\text{scale}}}{2eH\mu} + \frac{aT_{\text{scale}}^4}{3}, \tag{3.7}$$

where $k_B$ is the Boltzmann constant, and $\mu$ is the mean molecular weight of the fluid (taken to be $0.615m_H$, which corresponds to ionized gas comprised of 70 per cent Hydrogen and 30 per cent Helium by mass).

4. Equation of state at a density scale height:

We define the density scale height to be where the density falls to $e^{-1}$ of its midplane value. Thus the equation here is:

$$P_{\text{scale}} = \frac{\Sigma k_B T_{\text{scale}}}{2eH\mu} + \frac{aT_{\text{scale}}^4}{3}. \tag{3.8}$$
5. Radiative diffusion:

By integrating the second moment of the radiative transfer equation \( dP/d\tau = F_{\text{rad}} \) and using the condition of constant radiative flux, we arrive at the following expression for the vertical temperature profile (for a grey-atmosphere):

\[
T(\tau) = \left[ \frac{3}{4} \left( \frac{F_{\text{rad}}}{\sigma} \right) \left( \frac{2}{3} + \tau \right) \right]^{1/4},
\]  

(3.9a)

where \( \tau \) is the optical depth measured from the surface, \( \sigma \) is the Stefan-Boltzmann constant, and \( F_{\text{rad}} \) is the radiative flux as evaluated from the radiative diffusion equation:

\[
F_{\text{rad}} = \frac{4}{3} \sigma \frac{dT^4}{\rho \kappa} dz \approx \left( \frac{4ac Q H T^4_{\text{mid}}}{3 \kappa P_{\text{mid}}} \right) \cdot \nabla.
\]  

(3.9b)

Equation (3.9a) gives the temperature at the scale height \( T_{\text{scale}} \) by plugging in \( \tau_{\text{scale}} = \kappa \Sigma_{\text{scale}} \) for the optical depth, where \( \Sigma_{\text{scale}} \) refers to the vertical column density from the scale height to the disc surface, and \( \kappa \) is the opacity, which for simplicity we take to be the electron scattering opacity \( \kappa_{\text{es}} \). The latter is justified since we are dealing with fairly hot discs. Now, we need a way to estimate \( \Sigma_{\text{scale}} \), defined as the location where the density falls from its mid-plane value by an e-fold (cf. Eq. 3.8). In our highly simplified one-zone model, we make the approximation that \( \Sigma_{\text{scale}} = 0.1 \Sigma \), which is roughly the value measured in more detailed multi-zone treatments of the vertical structure (see §3.4.2 for more discussion).

6. Thermodynamic gradient \( \nabla \):

We calculate \( \nabla \) directly from the values of \( P_{\text{mid}}, T_{\text{mid}}, P_{\text{scale}}, T_{\text{scale}} \) via:
7. Energy transport:

Radiation, convection, and turbulence all contribute to the vertical transport of energy. Thus, the expression for the total energy flux \( F_{\text{tot}} \) is:

\[
F_{\text{tot}} = F_{\text{rad}} + F_{\text{conv}} + F_{\text{turb}},
\]

(3.11a)

where \( F_{\text{rad}} \) is the radiative flux given by Eq. (3.9c), and \( F_{\text{conv}} \) is the convective flux given by (Mihalas 1978):

\[
F_{\text{conv}} = \rho c_p \bar{v}_{\text{convect}} \Delta T.
\]

(3.11b)

Here \( c_p \) is the specific heat capacity, \( \bar{v} \) is the average vertical speed of convective blobs, and \( \Delta T \) is the typical temperature differential between a fluid parcel and the ambient medium. We evaluate these quantities using standard mixing-length theory, with mixing length \( \Lambda \) taken to be some multiple of the pressure scale height \( H \). As a default we set \( \Lambda/H = 1 \), though we also consider other values. The convective mixing-length velocity is then (Mihalas 1978):

\[
\bar{v}_{\text{convect}} = \left| \frac{Q \Lambda^2}{8} \left( \nabla - \nabla_e \right) \right|^{1/2}.
\]

(3.11c)

Henceforth, we use subscript \( e \) to denote quantities that are measured within the convective blob; \( \nabla_e \) therefore represents the effective \( d\ln T/d\ln P \) experienced by a convective/turbulent element as it moves vertically before dissolving back into the
surrounding medium. For the temperature differential, we take:

\[
\Delta T = \Lambda \left( \frac{dT}{dz} - \left. \frac{dT}{dz} \right|_e \right)
\]
\[
= \frac{\Delta T}{P} \left( \frac{dp}{dz} \right) \left( \frac{d\ln T}{d\ln P} - \left. \frac{d\ln T}{d\ln P} \right|_e \right)
\]
\[
= \frac{\Delta T_{\text{mid}}}{H} (\nabla - \nabla_e).
\] (3.11d)

Here, the specific heat capacity \( c_p \) is given by standard formulae corresponding to a monatomic gas/radiation mixture (Chandrasekhar 1967) using midplane quantities to set the gas/radiation pressure ratio.

For turbulent mixing, we follow the same prescription as our convective mixing length theory (Eq. 3.11b). We assume that the turbulent mixing flux is given by:

\[
F_{\text{turb}} = \rho c_p \bar{v}_{\text{turb}} \Delta T
\]
\[
= \rho c_p (\bar{v}_{\text{turb}} \Lambda_{\text{turb}}) \frac{T_{\text{mid}}}{H} (\nabla - \nabla_e),
\] (3.11e)

where we now invoke a turbulent velocity \( \bar{v}_{\text{turb}} \) and turbulent scale height \( \Lambda_{\text{turb}} \). To estimate the size of the product \( (\bar{v}_{\text{turb}} \Lambda_{\text{turb}}) \), recall that turbulent viscosity has the form

\[
\nu_{\text{turb}} \sim \bar{v}_{\text{turb}} \Lambda_{\text{turb}}.
\] (3.11f)

Consistency with the \( \alpha \)-prescription requires that this viscosity equal

\[
\nu_{\text{turb}} = \alpha c_s H,
\] (3.11g)

where \( \alpha \) is the viscosity coefficient introduced in Eqs. (3.6), \( c_s = \sqrt{2H P_{\text{mid}}/\Sigma} \) is the fluid sound speed, and \( H \) is the disc scale height. Comparing Eqs. 3.11f and 3.11g and further assuming that \( \Lambda_{\text{turb}} = H \), we arrive at \( \bar{v}_{\text{turb}} = \alpha c_s \). For the purpose of exploration, we allow the following more general scaling:

\[
\bar{v}_{\text{turb}} = \zeta c_s.
\] (3.11h)
where $\zeta$ is a dimensionless number $< 1$. For most of our models, we set $\zeta = \alpha$ as our fiducial value. A detailed investigation of how $\zeta$ affects the disc solution is presented in §3.4.4.

8. Effectiveness of convection and turbulence vs. radiative diffusion:

To complete our mixing-length theory, we require another set of equations relating $\nabla$ and $\nabla_{e}$. This is done by comparing the efficiency of convective and turbulent transport with radiative diffusion:

$$\gamma \equiv \frac{\text{convective/turbulent energy lost at dissolution of blob}}{\text{radiative energy lost during lifetime of blob}}. \tag{3.12a}$$

The total energy loss from the convective and turbulent element is proportional to $(\nabla - \nabla_{a})$, where $\nabla_{a}$ represents the adiabatic gradient of the surrounding fluid, i.e. $d\ln T/d\ln P$ for an ideal gas/radiation mixture. It is given by standard formulae (Chandrasekhar 1967) and only depends on the gas/radiation pressure ratio; $\nabla_{a} = 1/3$ in the gas limit and $\nabla_{a} = 1/4$ in the radiation limit. The total energy loss from the blob can be split into two components: 1) the energy released by dissolution at the end of the blob’s life (proportional to $\nabla - \nabla_{e}$), and 2) the radiative energy loss from radiative diffusion out of the blob before it dissolves (proportional to $\nabla_{e} - \nabla_{a}$). From Mihalas (1978) p.189-190, the ratio of these two components $\gamma$ can be written as:

$$\gamma = \frac{\nabla - \nabla_{e}}{\nabla_{e} - \nabla_{a}} = \left(\frac{\Sigma c_{p}(\bar{v}_{\text{convect}} + \bar{v}_{\text{turb}})}{16 H \sigma T_{\text{mid}}^{3}}\right) \left(\frac{1 + \tau_{m}^{2}/2}{\tau_{m}}\right). \tag{3.12b}$$
where $\tau_m$ represents the optical depth of the convective cell, which we take to be:

$$\tau_m = \kappa \Sigma / 2$$

(3.12c)

Solving the system of equations

Eqs. (3.5)-(3.12) form a system of 8 equations, where the only unspecified values are the 8 unknowns listed in §3.2.2. These 8 equations can be solved at each radius in the disc to yield a complete model. The set of model parameters needed are: 1) the central black hole mass $M$, 2) the black hole spin $a_*$, 3) the accretion rate $\dot{M}$, 4) the viscosity parameter $\alpha$, and 5) the two mixing parameters $\Lambda, \zeta$. Although the system of equations is highly non-linear, Appendix 3.7 describes a simple procedure to numerically solve for all unknowns. We discuss numerical results in the following section.

3.3 Disc solutions

Using the methods outlined in §3.2 and Appendix 3.7, we have calculated a wide array of disc models to understand how convection and turbulent mixing affect the overall disc structure. We compare three distinct classes of disc models (see Fig. 3.1 for a schematic):

a) Purely radiative discs with no convective/turbulent mixing (these correspond to the classic solutions of Novikov & Thorne 1973 and are identified as “No mixing” in the plots); b) Convective discs with vertical energy transport via both radiation and convection (these models are labelled as “Convect”); c) discs with vertical energy transport via radiation, convection, and turbulent mixing (labelled as “Conv+Turb”). A direct comparison of the stability properties for the 3 classes of disc models is shown in
Fig. 3.4 and discussed below in §3.3.3. For all cases, we have calculated models spanning accretion rates over a wide range, $0.001 < \dot{M}/\dot{M}_{\text{Edd}} < 1$, where we define the Eddington accretion rate to be:

$$\dot{M}_{\text{Edd}} = \frac{4\pi GM}{\kappa es\epsilon}.$$  \hfill (3.13)

Here $\epsilon = 1 - E_{\text{ISCO}}$ is the accretion efficiency of the disc, a measure of the total gravitational energy released by matter along its journey to the ISCO. The efficiency $\epsilon$ ranges from $\sim 6 - 40$ per cent depending on black hole spin $a_*$. Unless specifically noted otherwise, models were calculated with the following fiducial parameters: black hole mass $M = 10M_\odot$, black hole spin $a_* = 0$, disc radius $r_* = 12$ (twice the ISCO radius for $a_* = 0$), viscosity coefficient $\alpha = 0.1$. For models with convection (models “Convect” and “Conv+Turb”), the mixing length was set equal to the scale height $\Lambda = H$. Finally, in the case of turbulent mixing (model “Conv+Turb”), the mixing velocity was set to $\bar{v}_{\text{turb}} = \alpha c_s$, i.e. $\zeta = \alpha$. 

127
CHAPTER 3. CONVECTION AND TURBULENT STABILIZATION OF DISCS

Figure 3.1: A schematic of the vertical structure in our one-zone model. We assume a homogeneous disc where energy is transported vertically along three channels: radiative diffusion ($F_{\text{rad}}$), convection ($F_{\text{conv}}$), and additional turbulent mixing ($F_{\text{turb}}$). We assume that all heating occurs at the disc midplane so that $F_{\text{rad}}$, $F_{\text{conv}}$, and $F_{\text{turb}}$ are constant everywhere. We compute vertical gradients by comparing values at the midplane and at a density scale height. In this work, we compare three distinct classes of disc models: a) purely radiative discs (with only $F_{\text{rad}}$), b) convective discs (with $F_{\text{rad}}$ and $F_{\text{conv}}$), and c) turbulently mixed convective discs (including all 3 components $F_{\text{rad}}$, $F_{\text{conv}}$, and $F_{\text{turb}}$).

3.3.1 Classic unmixed disc

To set the stage, we first solve our disc equations without invoking either convective or turbulent flux transport. This corresponds to the classic disc solution of Novikov & Thorne 1973, where at each radius, we only need to solve for 4 unknowns: $T_{\text{mid}}$, $P_{\text{mid}}$, $\Sigma$, and $H$. We obtain the disc model by solving a reduced set of 4 equations: Eq. (3.5), Eq. (3.6), Eq. (3.7), and Eq. (3.9b). In Eq. (3.9b), we substitute $\rho = \Sigma/2H$ and write the differential $dT^4/dz$ as

$$\frac{dT^4}{dz} \approx \frac{T^4_{\text{mid}}}{H}. \quad (3.14)$$
Figure 3.2: Disc solutions in the $\dot{M} - \Sigma$ plane (“S-curve”) for a purely radiative disc with $\alpha = 0.1$ around a non-spinning black hole. The different tracks represent different disc radii measured in units of $r_* = r/(GM/c^2)$. The slope $d\dot{M}/d\Sigma$ is an indicator for stability. Solutions with $d\dot{M}/d\Sigma > 0$ are thermally and viscously stable whereas $d\dot{M}/d\Sigma < 0$ represents instability. Thus, the characteristic bend in the solutions signifies the transition point from stable to unstable discs. Among the four tracks shown, the lowest $\dot{M}$ for the bend occurs at around 2 per cent Eddington in the $r_* = 12$ track.
CHAPTER 3. CONVECTION AND TURBULENT STABILIZATION OF DISCS

The method used to solve this set of 4 equations is outlined in Appendix B of Zhu et al. (2012). In general, we find that for all choices of disc parameters $M, a_*, r_*, \alpha$, at any given radius the solutions fold back in the $(\dot{M}, \Sigma)$ plane (Fig. 3.2 shows examples). It is well known that the slope of the solution track is an indicator for the disc’s viscous and thermal stability (e.g. Bath & Pringle 1982; Kato, Fukue & Mineshige 2008). Solutions with $d\dot{M}/d\Sigma > 0$ are stable, while those with $d\dot{M}/d\Sigma < 0$ are unstable. This result is independent of the details of the heating/cooling prescription. Thus, the characteristic bend that we see in the $\dot{M} - \Sigma$ plane signifies that, at all radii, the accretion solution transitions from being stable at low $\dot{M}$ to unstable above some critical accretion rate. Disc models that include the physics of advection (Abramowicz et al. 1988; Sadowski et al. 2011) exhibit a second turnover at high $\dot{M}/\dot{M}_{\text{Edd}}$ due to rapid advective cooling. Above this $\dot{M}$, the disc is stable. The resulting track in the $\dot{M} - \Sigma$ plane has an “S” shape. Although our solutions do not capture this upper advection stabilized branch, for consistency with previous analyses on disc stability, we will still refer to our disc solution track in the $\dot{M} - \Sigma$ plane as an “S-curve”.

The focus of the present paper is the lower two segments of the S-curve, and the transition from thermal stability at lower $\dot{M}$ to instability at higher $\dot{M}$. The thermal instability threshold is located approximately where the disc transitions from being gas-pressure to radiation-pressure dominated. When the disc is gas pressure dominated $P \sim P_{\text{gas}} \sim T$, and for a fixed $\Sigma$, the cooling rate $Q_{\text{cool}} \sim \sigma T_{\text{eff}}^4 \sim T^4$ is a steeper function of temperature than the heating rate $Q_{\text{heat}} \sim Ht_{r}\phi \sim H \alpha P \sim P^2 \sim T^2$. This implies thermal stability since any positive temperature perturbation is quickly eliminated as the system responds with net cooling. However, the opposite occurs in the radiation dominated limit. Here, $P \sim P_{\text{rad}} \sim T^4$, and while the cooling still scales
as \( Q_{\text{cool}} \sim T^4 \), the heating is steeper where \( Q_{\text{heat}} \sim P^2 \sim T^8 \). This is unstable since a positive temperature perturbation is self-reinforcing.

It is also easy to understand the variation with radius of the critical \( \dot{M} \) for stability. The transition from being gas to radiation pressure dominated as one increases \( \dot{M} \) first occurs in the most luminous parts of the disc. This is indeed borne out in figure 3.2, where we see the lowest critical \( \dot{M} \) occurs in the \( r_* = 12 \) track, which lies closest to where the bulk of the disc energy is released (\( dL/dr \ln r \) is maximum at \( r_* \sim 15 \)).

### 3.3.2 Convective solutions

To obtain the class of convective disc solutions, we solve the full system of equations outlined in §3.2.2, but without including a turbulent flux component (i.e. we set \( \bar{v}_{\text{turb}} = 0 \) in Eq.3.11e). As Fig. 3.3 shows, convective disc models are qualitatively very similar to the standard nonconvective discs. Convection provides a modest stabilizing effect on the disc. Therefore the transition value of \( \dot{M} \) is a factor \( \lesssim 2 \) higher than in the corresponding purely radiative model (Fig. 3.3). Convection also causes the disc solutions to move towards higher column densities. To understand this, note from Eqs. (3.5) and (3.6), that for a fixed \( F_{\text{tot}} \), we have \( \Sigma \propto H^{-2} \) (cf. Appendix Eq. 3.21). Since convective discs are cooler than their nonconvective counterparts (e.g. only a fraction of the flux is carried by radiation, necessitating a smaller temperature gradient), they are thinner (smaller \( H \)) and hence have larger column densities.
**Figure 3.3:** Solution tracks in the $\dot{M} - \Sigma$ plane ("S-curve") at $r_* = 12$ for two choices of $\alpha = 0.1$ (left), $\alpha = 0.01$ (right), and three choices of the convective length scale: $\Lambda/H = 0.5$ (dotted), 1 (dashed), 2 (dash-dotted). The classic nonconvective disc solution is shown by thick solid lines. The net effect of increasing the mixing length is to increase the strength of convection, which tends to push the solutions out toward higher column densities. Convective cooling also increases the critical turnoff accretion rate. However, even with an optimistic choice of $\Lambda/H = 2$, the critical turnoff point is no more than $\sim 5$ per cent Eddington for any of the convective solutions.
CHAPTER 3. CONVECTION AND TURBULENT STABILIZATION OF DISCS

Another feature that we see in our convective disc solutions is that the onset of strong convection occurs strictly in the radiation dominated regime (in Fig. 3.3, convection primarily modifies the upper, radiatively dominant branch). This fact can be understood by the following argument: as the solution becomes more radiation dominated, the adiabatic gradient gets pushed to lower values (i.e. starting from $\nabla_a \sim 1/3$ in the gas limit, the gradient falls to $\nabla_a \sim 1/4$ in the radiation limit). However, a disc becomes radiation dominated only at high accretion rates, and high accretion rates necessitate larger radiative gradients to push out the increased flux (cf. Eq. 3.9c, where higher $\dot{M}$ and hence higher $F_{\text{rad}}$ implies larger $\nabla$). This divergence between adiabatic and radiative gradients (specifically, the push towards $\nabla \gg \nabla_a$) at high accretion rates is the engine that drives convection.

Even given optimistic assumptions about convection (e.g. solutions with $\Lambda/H = 2$ in Fig. 3.3), we find that the instability threshold for convective discs stays well below 10 per cent Eddington. To within a factor of 2, the instability threshold is similar to that found in nonconvective discs (compare solid with dashed lines in Fig. 3.3). Thus, we conclude that the action of convection alone results in only a modest stabilizing effect.

3.3.3 Convective and turbulent disc solutions

We now ask if MRI-induced turbulent mixing can provide a larger boost to the instability threshold. Our treatment of the turbulent velocity is given by Eq.(3.11h). For simplicity, we take as an ansatz that the proportionality constant $\zeta$ between the fluid sound speed $\zeta$ and turbulent eddy velocity is $\zeta = \alpha$. We explore the impact of varying $\zeta$ in §3.4.4.
Figure 3.4: S-curves for 3 choices of BH spin, calculated in each case at twice the ISCO radius: \( r_* = (12, 9, 5) \) for spins \( a_* = (0, 0.5, 0.9) \) respectively. The three sets of curves correspond to the three classes of disc solutions: classic disc on the left (No mixing), convective disc in the centre (Convect), and convective + turbulent disc on the right (Conv+Turb). Varying black hole spin does not have a significant impact on the location of the turnoff \( \dot{M} \). Note that the turnoff \( \dot{M} \) in the turbulent models greatly exceed those in the other two disc models. This shows that the stabilizing effect of turbulent mixing is much stronger than that of convective mixing alone. The models assume \( \Lambda/H = 1 \), \( \alpha = 0.1 \), and \( \zeta = \alpha \).
CHAPTER 3. CONVECTION AND TURBULENT STABILIZATION OF DISCS

Fig. 3.4 shows our results. We find that disc stability is strongly modified by the inclusion of turbulent mixing, which pushes the stability threshold to much higher accretion rates. Thus, turbulent mixing is a viable mechanism for stabilizing the disc at high accretion rates; it is much stronger than convection alone, which has trouble pushing the stability threshold above $\sim 5$ per cent Eddington (compare centre and rightmost tracks in Fig. 3.4). The action of turbulence is essentially a stronger version of convection – the added motion of the turbulent eddies preferentially increases gradients within the disc in the same way that convection does. The net effect is to further reduce the interior temperature of the gas, thereby requiring the disc to hit much higher accretion rates before it can transition to being radiation dominated (compare the $\dot{M}$ required to reach the same $P_{\text{rad}}/P_{\text{gas}}$ ratio for the different classes of discs in Fig. 3.5).

Although we computed solutions all the way up to the Eddington limit, we do not find any cases where convection is overwhelmingly dominant. In Fig. 3.6, we see that convection never accounts for more than $\sim 1/2$ of the total vertical energy flux. This is in line with previous studies on convective discs (Shakura et al. 1978; Cannizzo 1992; Milsom, Chen & Taam 1994; Heinzeller, Duschl & Mineshige 2009). In turbulent disc solutions, we also find that the turbulent flux $F_{\text{turb}}$ becomes dominant at high accretion rates. Fig. 3.6(b) shows the partition of turbulent vertical flux in the “Conv + Turb” model. A consequence of this dominant turbulent flux at large $\dot{M}$ is to push down the total radiative flux. This in turn causes the disc to become cooler overall, which lowers the radiation to gas pressure ratio everywhere (Figs. 3.5 and 3.7). These cooler discs must reach higher $\dot{M}$ values to hit the same ratio of $P_{\text{rad}}/P_{\text{gas}}$ and hence require higher $\dot{M}$ to become unstable (since instability is governed by exceeding a critical $P_{\text{rad}}/P_{\text{gas}}$ ratio).
In addition, there is a second effect in operation. For the purely radiative standard disc, the transition to instability happens at \( P_{\text{rad}}/P_{\text{gas}} = 1.5 \). This is no longer true for convective/turbulent models. The critical pressure ratio is 2.1 for the convective model and is as large as 7.3 for the convective and turbulent model (compare black dots in Fig. 3.5). This is not too surprising since the critical pressure ratio is achieved when the growth rate for cooling and heating balance (as a function of fluid temperature). When the convective and turbulent cooling channels are introduced, one must increase the overall cooling rate of the system to keep the same radiative flux as before. The heating rate must therefore increase to balance this increase in cooling. The higher heating rate implies higher temperatures, or higher radiation to gas pressure ratios. Thus, including non-radiative channels causes the critical \( P_{\text{rad}}/P_{\text{gas}} \) to increase.

The net result of these two effects is that models with both convection and turbulence have critical \( \dot{M} \) values well above 10 per cent Eddington. The cooler interior caused by the suppression of radiative flux, coupled with the increase in the critical \( P_{\text{rad}}/P_{\text{gas}} \) ratio, yields a strong stabilizing effect on the disc. In the models shown in Fig. 3.4, the critical \( \dot{M} \) for turbulent models is about a factor of 10 higher than that of the purely radiative model. As we show in §3.4.2 & §3.4.4, even larger changes are possible with modest changes to disc parameters.
Figure 3.5: Comparison of radiation-to-gas pressure ratios for $a_*=0$, $r_*=12$, and the three classes of disc solutions shown in Fig. 3.4. The large black dots denote the critical accretion rate for each solution. The pressure ratio for this critical point is not the same in the three classes of discs: $P_{\text{rad}}/P_{\text{gas}} = (1.5, 2.1, 7.3)$ for the radiative, convective, and turbulent disc models respectively. Note also that we require higher $\dot{M}$ to reach the same $P_{\text{rad}}/P_{\text{gas}}$ in the convective solutions as in the “no mixing” case.
CHAPTER 3. CONVECTION AND TURBULENT STABILIZATION OF DISCS

Figure 3.6: Plot of the radiative (solid), convective (dashed), and turbulent (dotted) vertical fluxes as a fraction of the total vertical flux for different accretion rates. Model parameters: $a_* = 0$, $r_* = 12$, $\alpha = 0.1$, $\Lambda/H = 1$. Panel (a) shows the fluxes in the convective disc, whereas panel (b) shows the results from the convective and turbulent disc. In both cases, radiation provides the bulk of the vertical energy transport at low accretion rates $\dot{M} < 0.01\dot{M}_{\text{Edd}}$. 
CHAPTER 3. CONVECTION AND TURBULENT STABILIZATION OF DISCS

3.3.4 Radial structure of solutions

Thus far, we have focused on the impact that turbulent energy transport has on the S-curve at a fixed radius. We now consider the variation in the three disc models with radius. In all plots below (Figs. 3.7 - 3.10), we show the results for a non-spinning black hole accreting at 10 per cent Eddington. Other disc parameters are set to their fiducial values (as listed in §3.3), though the results remain qualitatively the same regardless of the choice of disc parameters.

At large radii, the disc is gas pressure dominated, whereas radiation pressure dominates at smaller radii. Near \( r_* \sim 15 \), where the bulk of the disc luminosity is released, one finds the highest ratio of \( P_{\text{rad}}/P_{\text{tot}} \) (shown in Fig. 3.7). At yet smaller radii \( (r_* \lesssim 8) \), the disc transitions back to being gas pressure dominated. This is due to the luminosity profile (and hence disc temperature) dropping off to zero at \( r_* = 6 \) (the location of the ISCO), which arises from the assumption of zero net torque at the ISCO in the classical theory of accretion discs (Page & Thorne 1974). Recent analysis of GRMHD simulations of the disc show that although the luminosity plummets sharply at the ISCO, the tenuous plunging gas in the inner region can remain hot and stay radiation dominated (Zhu et al. 2012). This effect is beyond the scope of the present work, so we focus on radii \( r_* > 8 \).

From the radial structure of the disc, we find that as mixing becomes more important (as a result of either convection or turbulent mixing), the disc interior becomes cooler. In models with turbulent/convective modes of vertical energy transport, the hot inner radiation dominated region cools off and becomes more gas-pressure dominated (see the range \( 10 < r_* < 100 \) in Fig. 3.7, where \( P_{\text{rad}}/P_{\text{gas}} \) shifts towards the gas-limit when mixing
is introduced). This effect occurs due to the additional non-radiative channels for vertical energy transport; less overall energy now travels along the radiative channel (notice the lowering of $F_{\text{rad}}$ in panel (b) of Fig. 3.8 compared to panel (a) due to the addition of the turbulent flux channel). This added cooling from mixing also causes the disc to become thinner (compare disc thickness profiles in Fig. 3.9), which yields higher mass column densities due to the scaling $\Sigma \propto H^{-2}$ at constant flux and radius. Fig. 3.10 shows the disc column density profiles.
Figure 3.7: Radiation to gas pressure ratio as a function of disc radius for the three classes of disc models. Model parameters were set to $a_*=0$, $\dot{M}/\dot{M}_{\text{Edd}} = 0.1$, $\alpha = 0.1$, $\Lambda/H = 1$, and $\zeta = \alpha$. 
**Figure 3.8:** Radial dependence of the fractional vertical fluxes in the radiative (solid), convective (dashed), and turbulent (dotted) channels for 10 per cent Eddington solutions. The upper panel (a) corresponds to the purely convective case, and the lower panel (b) corresponds to the turbulent plus convective case. Note that the convective flux saturates at about $\sim 1/5$ of the total flux. Model parameters are the same as in Fig. 3.7.
Figure 3.9: Radial profiles of the disc scale height for the three models. The net result of convective/turbulent mixing is to cool the disc, yielding thinner discs than the purely radiative unmixed disc. Model parameters are the same as in Fig. 3.7.
Figure 3.10: Radial dependence of the vertical column mass for the three disc models considered. Including convective and turbulent mixing causes the vertical column mass to increase everywhere. Model parameters are the same as in Fig. 3.7.

3.4 Discussion

In developing our simplified disc model, we have made a number of assumptions. Below we justify our choices, and discuss their impact on the results.
3.4.1 Use of logarithmic temperature gradient

In setting the temperature gradient for our one-zone model, we opted to take the quotient of logarithmic differentials (cf. Eq. 3.10). Another possibility is to set:

$$\nabla = \frac{d \ln T}{d \ln P} \approx \left( \frac{P_{\text{mid}}}{T_{\text{mid}}} \right) \left( \frac{T_{\text{mid}} - T_{\text{scale}}}{P_{\text{mid}} - P_{\text{scale}}} \right),$$

which is similar to the prescription used by Goldman & Wandel (1995):

$$\frac{d \ln T}{d \ln z} \sim \frac{H T - T_s}{T} = 1 - \frac{T_s}{T}.$$ (3.16)

However, taking the logarithm outside of the differential biases $\nabla$ towards large values\(^1\).

In particular, when the second probe point is chosen at a scale height, the prescription corresponding to Eqs. (3.15) and (3.16) produces unphysically large values of $\nabla$. As a result, convection becomes overwhelmingly strong, with $F_{\text{conv}}/F_{\text{tot}} \to 1$ at large $\dot{M}$.

This artefact in $\nabla$ distorts the disc solutions so severely that the disc becomes thermally stable everywhere (compare left and right solution tracks in Fig. 3.11 corresponding to two prescriptions for $\nabla$). We believe this choice in evaluating $\nabla$ is why some previous one-zone estimates (e.g. Bisnovatyi-Kogan & Blinnikov 1977; Goldman & Wandel 1995) found the efficiency of convection to be exceedingly large, in contrast to later more detailed convective disc models that included the full vertical structure and find that the efficiency of convection saturates at $F_{\text{conv}}/F_{\text{tot}} \sim 1/2$ (Cannizzo 1992; Milsom, Chen & Taam 1994; Heinzeller, Duschl & Mineshige 2009). Our one-zone model calculates $\nabla$ by means of Eq. 3.10 and produces results consistent with the latter studies.

\(^1\)Mathematically, this upwards bias occurs whenever $0 < \nabla < 1$ for any two variables with scaling $(x \propto y^{\nabla})$. 

145
Figure 3.11: S-curves for two different $\nabla$ prescriptions: (a) Blue = $\nabla$ as defined by Eq. (3.10), (b) Red = alternate definition for $\nabla$ according to Eq. (3.15). For reference we also plot the standard radiative disc solution in thick black. The red (b) solutions severely overestimate the strength of convection since $\nabla$ reaches unphysically large values. We believe that these solutions correspond to the set of superefficient convective solutions found in some previous studies (e.g. Goldman & Wandel 1995 – compare with their figure 1).
Figure 3.12: S-curves for various choices of $\Sigma_{\text{scale}}/\Sigma = (0.125, 0.1, 0.075, 0.05)$ shown by (dotted, dash-dotted, dashed, solid) lines respectively. The thick black line denotes the fiducial S-curve as determined from the purely radiative standard disc. Lines in panel (a) denote the family of convective disc models, whereas those in (b) denote the family of turbulently mixed disc models. For each choice of $\Sigma_{\text{scale}}/\Sigma$, the critical $\dot{M}$ for the turbulent model (b) is much higher than for the corresponding convective model (a). For some choices of $\Sigma_{\text{scale}}/\Sigma$, the turbulent model (b) is completely stable at all $\dot{M}$. 
3.4.2 Impact of $\Sigma_{\text{scale}}$

In our disc model, one arbitrary parameter is the scale height $\Sigma_{\text{scale}}$ used as the upper probe point in defining the temperature gradient $\nabla$. In all models presented so far, we set $\Sigma_{\text{scale}} = 0.1\Sigma$. Fig. 3.12 shows the impact of varying $\Sigma_{\text{scale}}$. We find that the critical $\dot{M}$ is quite sensitive to the choice of $\Sigma_{\text{scale}}$. In general, larger values of $\Sigma_{\text{scale}}$ act to weaken the strength of mixing, resulting in smaller critical $\dot{M}$ values. However, regardless of the value picked for $\Sigma_{\text{scale}}$, the critical $\dot{M}$ for the turbulently mixed disc is always much higher than that of the convective disc (compare left and right panels in Fig. 3.12 for the same choice of $\Sigma_{\text{scale}}$).

We have looked at more detailed treatments of the vertical structure to see how good our estimate of $\Sigma_{\text{scale}}$ is. For the purely radiative standard disc, we have computed a few detailed models of the vertical structure across a wide range of accretion rates (from $0.1 - 50$ per cent Eddington) using the stellar atmospheres code TLUSTY (Hubeny & Lanz 1995; Davis et al. 2005). We find $\Sigma_{\text{scale}} \sim 0.03 - 0.15\Sigma$, with a systematic trend such that $\Sigma_{\text{scale}}/\Sigma$ is lower at large $\dot{M}$. This is simply a consequence of the disc becoming more centrally concentrated when it is cool and in the gas-pressure dominated limit.

As another reference point, a polytropic disc with equation of state $P = K\rho^{1+1/n}$, where $K$ is a constant and $n$ is the polytropic index, has a density profile given by

$$\frac{\rho(z)}{\rho(z = 0)} = \left(1 - \left(\frac{z}{H}\right)^2\right)^n,$$

where $z$ is the height above the midplane, $H$ is the total height of the disc, and $\rho_0$ is the central density. For $n = 3/2$ (corresponding to a strongly convective column), we have $\Sigma_{\text{scale}}/\Sigma \sim 0.04$. For $n = 3$ (corresponding to a radiatively cooled column), we find $\Sigma_{\text{scale}}/\Sigma \sim 0.06$. 

148
For the case of an isothermal disc, the vertical density profile is given by

$$\frac{\rho(z)}{\rho(z = 0)} = \exp \left[-\frac{z^2}{2}\right].$$

(3.18)

This corresponds to $\Sigma_{\text{scale}} = \text{erfc}(1) \cdot \Sigma/2 \approx 0.08\Sigma$. In all cases, the value of $\Sigma_{\text{scale}}$ is pretty close (i.e. within a factor of 2) to the fiducial value that we have adopted in our models ($\Sigma_{\text{scale}} \sim 0.1$). If anything, our choice is conservative since values of ($\Sigma_{\text{scale}} < 0.1\Sigma$) enhance the stability of these models (Fig. 3.12).

### 3.4.3 Scaling of Critical $\dot{M}$ with Black Hole Mass

So far, we have only treated the case where $M = 10M_\odot$. However, from eq. (2.18) of Shakura & Sunyaev (1973) we expect a weak $\dot{M}_{\text{crit}} \propto M^{-1/8}$ scaling for the critical point in the S-curve. Figure 3.13 shows the scaling for our disc models. For supermassive black holes with masses $M \sim 10^7-10^8M_\odot$, we expect a roughly tenfold reduction in the critical accretion rate threshold compared to stellar mass black holes, and this is seen in the plot. Convection and turbulence increase $\dot{M}_{\text{crit}}$ by the same factor $\sim 10$. Even with this, we see that AGN discs are stable only up to luminosities of $\sim 3$ per cent of Eddington.
Figure 3.13: S-curve for various choices of black hole masses. Solid lines denote standard discs without convective/turbulent transport, whereas dotted lines show turbulent + convective disc solutions. Except for black hole mass, all disc parameters were set to their fiducial values listed in §3.3. Note that the $M^{-1/8}$ scaling for the critical $\dot{M}$ holds even for the turbulent solutions.
3.4.4 Choosing $\zeta$ – comparison with simulations

The value that we take for $\zeta = v_{\text{turb}}/c_s$ in our model is primarily motivated by the speed of turbulent motions observed in numerical simulations. From the global GRMHD simulations of Penna et al. (2010), we find that $v_{\text{turb}}/c_s \sim 0.1 - 0.2$ for $r_* > 10$ (compare dashed and solid lines in Fig. 3.14). The effective viscosity coefficient in these simulations is measured to be $\alpha \sim 0.05 - 0.1$, consistent with our choice $\zeta = \alpha$. However shearing box simulations measure a much larger spread in the vertical advective speed (see figure 23 in Blaes et al. 2011), which can range from $v_{\text{turb}} \sim 0.001 - 0.1c_s$ depending sensitively on where the advective speed is measured (slowest at midplane, faster at surface). Note that for shearing box simulations, the effective viscosity $\alpha$ is also much lower than the corresponding values in global simulations; typical values are $\alpha \sim 0.01$ (King, Pringle & Livio 2007). Table 3.1 shows a list of inferred $\zeta$ values from various simulations in the literature, and we find the scaling is roughly $\zeta \sim 2\alpha$. Note that this scaling for $v_{\text{turb}} \propto \alpha$ is inconsistent with an isotropically turbulent $\alpha$-model since the stress would now scale as $W \propto \rho v^2 \propto v_{\text{turb}}^2 \propto \alpha^2 \neq \alpha$. The only way to reconcile this inconsistency is to demand anisotropic turbulence so that the scaling of $W$ with $v_{\text{turb}}^2$ no longer holds.

To test the sensitivity of our model on the choice of turbulent speed, we now explore a wide range of values for $\zeta$. The precise location of the critical $\dot{M}$ for stability depends sensitively on the value of $\zeta$ (see the spread of solutions in Figs. 3.15 and 3.16). For sufficiently large $\zeta$ (not much larger than our canonical value of 0.1), the turbulent disc becomes stable for all accretion rates. Due to sensitivity to model parameters, this prediction should be treated with caution and must be interpreted more as a proof of concept. Our main result is that turbulent mixing can provide a much stronger
### Table 3.1.: Simulation Derived $\zeta = \nu_{turb}/c_s$ Values

<table>
<thead>
<tr>
<th>Reference</th>
<th>$\alpha$</th>
<th>$\zeta$</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jiang et al. (2013)</td>
<td>0.01-0.02</td>
<td>0.036</td>
<td>Butterfly$^1$</td>
</tr>
<tr>
<td>Guan &amp; Gammie (2011)</td>
<td>0.01-0.02</td>
<td>0.045</td>
<td>Butterfly$^1$</td>
</tr>
<tr>
<td>Blaes et al. (2011)</td>
<td>0.01-0.02</td>
<td>0.021</td>
<td>Direct$^2$</td>
</tr>
<tr>
<td>Penna et al. (2010)</td>
<td>0.05-0.1</td>
<td>0.1-0.2</td>
<td>Direct$^2$</td>
</tr>
<tr>
<td>Davis et al. (2010)</td>
<td>$\sim$0.01</td>
<td>0.026</td>
<td>Butterfly$^1$</td>
</tr>
<tr>
<td>Suzuki &amp; Inutsuka (2009)</td>
<td>$\sim$0.01</td>
<td>0.025</td>
<td>Butterfly$^1$</td>
</tr>
</tbody>
</table>

1 Vertical advection speed is measured from the slope of the simulation butterfly diagram (which shows the characteristic vertical motion of the turbulent dynamo). This slope yields $\nu_{turb}$ in terms of $H\Omega = c_s$. In cases with non-constant slope, we use the slope measured at $z/H = 1$.

2 We compare direct measurements of $\nu_{turb} = v^z$ and $c_s$ at $z/H = 1$ to get $\zeta$. 
stabilizing force than that provided by convection alone. This can be understood on the basis that MRI-induced turbulence is likely much stronger than convective turbulence, and hence should have a stronger impact on the disc physics.

**Comparison with shearing box simulations**

The state-of-the-art in accretion disc modeling are the detailed vertically stratified Radiation-MHD shearing box simulations of Hirose, Blaes & Krolik (2009), Hirose, Krolik & Blaes (2009) and Blaes et al. (2011). By comparing and analyzing a sequence of these simulations with different initial conditions, Hirose, Blaes & Krolik (2009) were able to piece together a simulation derived S-curve. Their resultant S-curve was very similar to the prediction from standard purely radiative disc theory, with one major difference; their radiation pressure dominated branch of solutions was apparently found to be thermally stable. They attribute this stability to non-synchronous evolution of stress and pressure in the simulations (Hirose, Krolik & Blaes 2009). Stress fluctuations were found to precede pressure fluctuations, thereby breaking the usual argument for thermal instability (since the dissipative heating rate $Q^+$ due to stress is no longer set by the pressure, which removes the steep temperature dependence of $Q^+$ in the radiation limit). These results are in contrast to our turbulent disc model, where we find the S-curve to be significantly modified by the action of turbulent mixing.
Figure 3.14: GRMHD derived turbulent vertical velocities for a non-spinning black hole from run A0HR07 of Penna et al. (2010), in which $\alpha = 0.05 - 0.1$. The time-averaged root-mean-squared value of the vertical velocity $v^z$ as measured at the disc midplane (blue solid) and at a scale height (green dotted) are shown and compared to the simulation sound speed (red dashed). The simulation gives $v^z/c_s \sim 0.1$ far from the plunging region (for $r_s > 10$), corresponding to $\zeta \approx 0.1 \approx \alpha$. 

\[ v^z (\text{mid}) \]
\[ v^z (\text{scale}) \]
\[ c_s \]
Figure 3.15: S-curve for various choices of the turbulent mixing parameter $\zeta$. Model parameters: $\alpha = 0.1$, $a_* = 0$, $r_* = 12$, $\Lambda/H = 1$. The red solid track ($\zeta = 0$) represents a convective disc solution (i.e. no turbulent flux). For small values of $\zeta \sim 0.01$, turbulence has negligible impact on the disc solutions. However, for $\zeta \gtrsim 0.05$, the critical $\dot{M}$ is increased substantially, and for $\zeta > 0.1$ the disc is stable for all values of $\dot{M}$. 
Figure 3.16: S-curve for various choices of $\zeta$ in an $\alpha = 0.01$ disc (other parameters identical to those in Fig. 3.15). Qualitatively, the behaviour of the solution tracks is similar to the $\alpha = 0.1$ case shown in Fig. 3.15.

We reconcile this apparent contradiction with the fact that in the shearing box simulations, turbulent mixing is weak. Looking at the Blaes et al. 2011 entry in Table 3.1, we see their simulation yields a low $\zeta \approx 0.02$. If we adopt a small enough value for $\zeta$, our model too gives an S-curve that is almost unmodified from the standard-disc S-curve.
(compare how similar the “No mixing”, \(\zeta = 0\), and \(\zeta = 0.01\) tracks are in Figs. 3.15 and 3.16). One qualitative difference remains between our model and shearing box discs; our model has a thermally unstable upper radiative branch whereas in the simulations of Hirose, Blaes & Krolik (2009), this branch appears to be thermally stable. However, recent work by Jiang et al. (2013) using a code based on Athena (Stone et al. 2008; Jiang et al. 2012), indicates that the radiation branch is in fact unstable. The reason for the difference between the two studies is not understood.

The weak turbulent mixing seen in shearing box simulations may simply be a consequence of their small value of \(\alpha \sim 0.01\). Since discs with larger values of \(\alpha\) ought to produce larger turbulent velocities, we conjecture that once \(\alpha\) becomes sufficiently large (say close to the values observed in real discs, \(\alpha \rightarrow 0.1\) ) turbulent mixing becomes strong, enabling large changes to occur in the S-curve. Thus, although shearing box simulations (small \(\alpha\) ) predict no modification to the standard S-curve, it is possible that nature (large \(\alpha\) ) admits S-curves that are significantly modified due to the presence of stronger turbulent motions.

One final distinction between our model and the shearing-box simulations is that Blaes et al. (2011) find their solutions to be convectively stable everywhere. They do see vertical advection of energy, but it is entirely due to magnetic buoyancy. In contrast, our model for turbulent mixing requires a convectively unstable entropy gradient to achieve a net outwards flux (i.e. for positive \(F_{\text{turb}}\), we require positive \(\Delta T\) in Eq. (3.11e) which can only occur when \(\nabla_e > \nabla\) ). There is no way to reconcile this difference as our model assumes active convection whenever there is outwards turbulent flux. It is perhaps too demanding to ask our highly simplified 1-zone disc model to exactly match the results of detailed 3D MHD simulations.
3.4.5 \( \zeta \) from observations

Spectral state transitions in stellar mass black hole systems may offer a clue in what nature chooses for \( \zeta \). These transitions are brought about by changes in mass accretion rate of the system (McClintock & Remillard 2006). Of particular interest is the transition from the thermally-dominant high/soft state to the steep-power-law very-high state. The high/soft state, believed to be thermally stable, exhibits little variability and spans the luminosity range 2 - 50 per cent of Eddington (Gierliński & Done 2004; Gierliński & Newton 2006). The very-high state, occurring at yet higher accretion rates, is usually accompanied by significant variability including the emergence of high frequency quasi periodic oscillations.

The interpretation from our disc models is that this state transition occurs as the disk migrates from the thermally stable lower branch to the thermally unstable upper branch of solutions. This critical accretion rate is intimately linked to the value of \( \zeta \) (c.f. Figures 3.15, 3.16). Based on observations of the high/soft to very-high state transition, we infer the critical accretion rate to be \( \dot{M}_{\text{crit}}/\dot{M}_{\text{Edd}} \sim 0.5 \) (McClintock & Remillard 2006). According to our models, this value for \( \dot{M}_{\text{crit}} \) suggests \( \zeta \sim 0.1 \) or larger in our fiducial disc model. We are unable to make a very precise prediction regarding \( \zeta \) since the value of \( \dot{M}_{\text{crit}} \) in our model is also affected by the other disc parameters.

3.4.6 Radiative Outer Zone

The vertical transport of energy at the disc surface must be dominated by radiative flux (e.g. Blaes et al. 2011 found that the radiative flux dominates over advective flux above a few disc scale heights in shearing-box simulations). In our one-zone model,
we assume that this radiative outer zone is thin and can be neglected. However, this assumption produces a systematic bias in the disc solutions; the presence of a secondary purely radiative zone acts to increase the interior temperature of the disc. This can be understood on the basis of the radiative diffusion equation. A purely radiative zone has a larger $F_{\text{rad}}$ than an equivalent model with convective and turbulent mixing. According to Eq. (3.9a), a consequence of this larger $F_{\text{rad}}$ is a larger interior temperature.

The exact location of the boundary separating the convective and turbulent interior from the purely radiative surface layer depends crucially on the details of the disc’s vertical structure. Since we only have a one-zone model to work with, we make the following crude and extreme assumption: we assume that the density scale height is the demarcation point between a convective and turbulent interior and a purely radiative exterior (Fig. 3.17).

![Figure 3.17](image)

**Figure 3.17:** A schematic of the vertical structure in our two-zone model (compare to the one-zone model in Fig. 3.1). We assume that the region between the surface and the density scale height is purely radiative. The interior is identical to that of our previous one-zone models.

This jump to two-zones primarily changes the value of $T_{\text{scale}}$. Since in the radiative zone $F_{\text{rad}} = F_{\text{tot}} = \sigma T_{\text{eff}}^4$, the previous radiative diffusion equation (Eq. 3.9a) now
CHAPTER 3. CONVECTION AND TURBULENT STABILIZATION OF DISCS

becomes

\[
T(\tau) = T_{\text{eff}} \cdot \left[ \frac{3}{4} \left( \frac{2}{3} + \tau \right) \right]^{1/4}.
\]  \hspace{1cm} (3.19)

For simplicity, we do not modify the other disc equations so the system can be solved in the same way our one-zone models. Fig. 3.18 compares the one-zone and two-zone disc solutions. As expected, the inclusion of a radiative outer zone pushes the convective and turbulent solutions towards that of a purely radiative standard disc (i.e. the two-zone solution lies in between the purely radiative disc and the homogeneously mixed one-zone disc). The radiative outer zone ultimately pushes down the critical \( \dot{M} \) of the two-zone model towards lower values. However, our main result is unaffected; even in two-zone discs, turbulently mixed discs are still significantly more stable than their purely convective counterparts (compare critical \( \dot{M} \) of left and right panels in 3.18).
Figure 3.18: S-curves of the two-zone model (dotted) compared to the corresponding one-zone model (dashed) and purely radiative standard disc (solid). Note that the critical \( \dot{M} \) for turbulent discs (right panel) are a factor \( \sim 4 - 5 \) larger than in purely convective discs (left panel). Model parameters: \( a_\ast = 0, r_\ast = 12, \alpha = 0.1, \Lambda/H = 1, \) and \( \zeta = \alpha \)
3.4.7 Complete stabilization from turbulence

For certain choices of model parameters (i.e. small $\Sigma_{\text{scale}}$ and/or large $\zeta$), we find that our turbulent and convective disc models admit solutions that are thermally stable at all accretion rates. Solutions of this form exhibit a linear track in the $\dot{M} - \Sigma$ plane with positive slope at all $\dot{M}$ (see the rightmost track in Figs. 3.12, 3.15, and 3.16). To understand what qualitative differences exist between solutions that have a critical transition point and those that are completely stable everywhere, we examine the heating and cooling curves for these two scenarios (compare Fig. 3.19 which exhibits both a stable and unstable disc solution, and Fig. 3.20 where only a single stable solution exists). In these heating/cooling curves, we hold fixed the total vertical column density and solve our usual set of disc equations for various choices of disc midplane temperature.
Figure 3.19: Comparison of cooling flux $Q^-$ (blue, thick solid) with heating flux $Q^+$ (red, thick dashed) for turbulent disc solutions with $\zeta = 0.1$, $\alpha = 0.1$, $r_* = 12$, $a_* = 0$, $\Lambda/H = 1$, and $\log \Sigma = 3.8$. We further break down the cooling flux into its various components: $F_{\text{turb}}$ (thin dotted), $F_{\text{rad}}$ (thin dashed), and $F_{\text{conv}}$ (thin solid). The two black dots denote the two stationary disc solutions where cooling matches heating (i.e. the two points on the S-curve corresponding to the selected column density). The lower dot represents a stable solution and the upper dot an unstable solution.
CHAPTER 3. CONVECTION AND TURBULENT STABILIZATION OF DISCS

Figure 3.20: Same as Fig. 3.19, but for a model with stronger turbulent mixing ($\zeta = 0.15$). Here, the disc solution is completely thermally stable, so there is only one stationary disc solution (black dot, which is thermally stable). The stability is due to the strong turbulent cooling flux which ensures that $Q^- > Q^+$ at large $T_{\text{mid}}$.

We find for our turbulence model that the cooling scales as $F_{\text{turb}} \sim T^8$ – comparable to the scaling for heating in the radiation dominated regime $Q^+ \sim T^8$. If the action of turbulence becomes sufficiently strong (represented in our model by taking a large
value for $\zeta$, then at high temperatures cooling always overtakes heating (see Fig. 3.20). In these strongly turbulent cases, any positive temperature perturbations above the equilibrium solution eventually cool back down to equilibrium. This eliminates the upper unstable branch of solutions, leading to complete stability.

### 3.5 Summary

In this work, we have developed a simple one-zone model for black hole accretion discs with both convective and turbulent vertical energy transport. We find that the action of mixing from convection and turbulence provides a stabilizing effect on the disc, pushing the threshold for thermal instability up towards higher accretion rates (compare the critical $\dot{M}$ for different classes of discs in Fig. 3.4). For stellar mass black holes, convection by itself provides only a modest boost to the thermal instability threshold, only pushing the critical $\dot{M}$ to 5 per cent Eddington in the most favourable cases. On the other hand, models that include additional mixing through MRI-induced turbulence are much more stable. In some cases, we find that turbulent mixing pushes the threshold for instability far above 10 per cent Eddington – even inducing complete stability in the most extreme cases. A similarly strong effect is seen for supermassive black holes. However, since the critical $\dot{M}$ for stability is lower by a factor of $\sim 10$, even with the effect of convection and turbulence, stable solutions are found only up to a few per cent of Eddington.

Previous studies have shown that convective mixing in discs tends to provide a stabilizing effect (Cannizzo 1992; Milsom, Chen & Taam 1994), which raises the critical
accretion rate marking the onset of thermal instability. Since MRI-induced turbulence is much more vigorous than convective turbulence, it is not surprising that turbulently mixed discs experience a much stronger version of this stabilizing effect. Thus, we believe that thermal stabilization from turbulent mixing is a promising mechanism for explaining the apparent lack of instability in luminous (up to 50% of Eddington) black hole X-ray binaries (Gierliński & Done 2004). However, due to the mass scaling for the critical accretion rate, we do not expect supermassive black holes to be stable at such high accretion rates - in our models they become unstable at around a few per cent of Eddington.

Finally, we stress that the precise value for the critical $\dot{M}$ found in our models should not be taken too seriously. We are working in the framework of a one-zone model (a rather severe approximation), and thus the quantitative details of our model are quite sensitive to the choice of model parameters. Our main result is a qualitative one – MRI-induced turbulent discs are much more stable than equivalent convective discs.

3.6 Acknowledgments

The authors would like to thank Robert Penna, Aleksander Sądowski, and Dmitrios Psaltis for insightful discussions about disc convection. We thank the anonymous referee for providing illuminating comments and helping us craft a much clearer paper. YZ also thanks Tanmoy Laskar for helpful suggestions for improving the presentation of the manuscript. YZ was supported by the Smithsonian Institution Endowment Funds. This work was supported in part by NASA grant NNX11AE16G.
3.7 Solving for 8 unknowns

Due to the highly non-linear nature of Eqs. (3.5)-(3.12), we solve the system of equations numerically. The technique is as follows:

1. Assume an exploratory trial value for $P_{\text{mid}}$ (i.e. we guess at its value, and check at the end if it produces a result that is consistent with all 8 equations).

2. From Eqs. (3.5) and (3.6), we obtain a relation for $\Sigma$ in terms of $P_{\text{mid}}$:

   $$\Sigma = \left( \frac{3\Omega k R F \alpha}{\sigma T_{\text{eff}}^4 Q} \right) \cdot P_{\text{mid}}^2.$$  (3.20)

3. Again from Eqs. (3.5) and (3.6), we also obtain $H$ by plugging in our value of $\Sigma$:

   $$H = \sqrt[4]{\frac{4\sigma T_{\text{eff}}^4}{3\Sigma \Omega k QRF \alpha}}.$$  (3.21)

4. From Eq. (3.7) and given the values of $\Sigma$, $P_{\text{mid}}$, and $H$, we can solve for $T_{\text{mid}}$ (numerically, by Newton’s method).

5. Plugging $\Sigma$ into Eq. (3.9a), we immediately get $T_{\text{scale}}$.

6. Using $T_{\text{scale}}$, $\Sigma$, $H$ in Eq. (3.8) yields $P_{\text{scale}}$.

7. Given all these quantities, using the two Eqs. (3.11) and (3.12) allows us to solve for $(\nabla - \nabla_e)$. In particular, we rearrange Eq. (3.11a) to first yield:

   $$A(\nabla - \nabla_e)^{3/2} + Z(\nabla - \nabla_e) = (\nabla_r - \nabla),$$  (3.22)
using the following definitions for the gradients $\nabla$, $\nabla_r$:

$$F_{\text{tot}} \equiv \left( \frac{4acQHT_{\text{mid}}^4}{3\kappa P_{\text{mid}}} \right) \cdot \nabla_r,$$

$$F_{\text{rad}} \equiv \left( \frac{4acQHT_{\text{mid}}^4}{3\kappa P_{\text{mid}}} \right) \cdot \nabla,$$

and the prefactors:

$$A = \frac{3\kappa \Sigma P_{\text{mid}} c_p}{16\sqrt{2}ac\sqrt{QHT_{\text{mid}}^3}} \left( \frac{\Lambda}{H} \right)^2,$$

$$Z = \frac{3\Sigma c_p \bar{v}_{\text{turb}} \kappa P_{\text{mid}}}{8acQH^2T_{\text{mid}}^3} \left( \frac{\Lambda}{H} \right).$$

We eliminate $\nabla$ from the RHS of (3.22) by adding $(\nabla - \nabla_e) + (\nabla_e - \nabla_a)$ to both sides:

$$A(\nabla - \nabla_e)^{3/2} + (Z + 1)(\nabla - \nabla_e) + (\nabla_e - \nabla_a) = (\nabla_r - \nabla_a).$$

Now, Eq. (3.12b) allows us to convert the $(\nabla_e - \nabla_a)$ term into a pure function of $(\nabla - \nabla_e)$, giving the quadratic:

$$\frac{(\nabla - \nabla_e)^2}{(\nabla_e - \nabla_a)^2} - \left( \frac{2\eta \bar{v}_{\text{turb}}}{\nabla_e - \nabla_a} + \frac{\eta^2 \Lambda^2 Q}{8} \right) \cdot (\nabla - \nabla_e) + \eta^2 \bar{v}_{\text{turb}}^2 = 0,$$

where:

$$\eta = \left( \frac{\Sigma c_p}{16\sigma T_{\text{mid}}^3 H} \right) \cdot \left( \frac{1 + \tau_m^2/2}{\tau_m} \right).$$

We take the positive root for $(\nabla_e - \nabla_a)$ in Eq.(3.28), and plug it into Eq.(3.27). This produces a polynomial expression for $(\nabla - \nabla_e)$, which can be solved numerically.

In the limiting case where $\bar{v}_{\text{turb}} \rightarrow 0$ (case of pure convection without additional turbulent mixing), we find that Eq. (3.28) simplifies to:

$$(\nabla_e - \nabla_a) = B(\nabla - \nabla_e)^{1/2}$$
CHAPTER 3. CONVECTION AND TURBULENT STABILIZATION OF DISCS

where

\[ B = \left( \frac{32\sqrt{2} \sigma T_{\text{mid}}^3}{\Sigma c_p \sqrt{Q}} \right) \left( \frac{\Lambda}{H} \right)^{-1} \left( \frac{\tau_m}{1 + \tau_m^2/2} \right). \tag{3.31} \]

Furthermore, \( \bar{v}_\text{turb} \to 0 \) also implies \( Z \to 0 \) and Eq. (3.27) becomes the following cubic, which can be easily solved for \((\nabla - \nabla_e)\):

\[ A(\nabla - \nabla_e)^{3/2} + (\nabla - \nabla_e) + B(\nabla - \nabla_e)^{1/2} = (\nabla - \nabla_e). \tag{3.32} \]

8. Plugging \((\nabla - \nabla_e)\) back into Eq. (3.12) allows us to solve for both \(\nabla\) and \(\nabla_e\) individually.

Now, armed with values for all 8 unknowns, we check to see if we made the correct guess for \(P_{\text{mid}}\) by checking for consistency in \(\nabla\). A correct guess for \(P_{\text{mid}}\) will cause \(\nabla\) computed from the two probe points in Eq. (3.10) to be consistent with \(\nabla\) computed from step (viii). Empirically, we find that this relation for \(\nabla\) is monotonic with \(P_{\text{mid}}\), and hence we are able to solve for the correct value of \(P_{\text{mid}}\) via bisection.
Chapter 4

HERO - A 3D General Relativistic Radiative Postprocessor for Accretion Discs around Black Holes

This thesis chapter originally appeared in the literature as


Abstract

HERO (Hybrid Evaluator for Radiative Objects) is a 3D general relativistic radiative transfer code which has been tailored to the problem of analyzing radiation from simulations of relativistic accretion discs around black holes. HERO is designed to be used as a postprocessor. Given some fixed fluid structure for the disc (i.e. density and
velocity as a function of position from a hydrodynamics or magnetohydrodynamics
simulation), the code obtains a self-consistent solution for the radiation field and for
the gas temperatures using the condition of radiative equilibrium. The novel aspect of
HERO is that it combines two techniques: 1) a short characteristics (SC) solver that
quickly converges to a self consistent disc temperature and radiation field, with 2) a long
characteristics (LC) solver that provides a more accurate solution for the radiation near
the photosphere and in the optically thin regions. By combining these two techniques, we
gain both the computational speed of SC and the high accuracy of LC. We present tests
of HERO on a range of 1D, 2D and 3D problems in flat space and show that the results
agree well with both analytical and benchmark solutions. We also test the ability of the
code to handle relativistic problems in curved space. Finally, we discuss the important
topic of ray-defects, a major limitation of the SC method, and describe our strategy for
minimizing the induced error.

4.1 Introduction

Radiative transport (RT) plays a crucial role in astrophysics. It governs our primary
source of information about the cosmos, viz., the luminosities and spectra of cosmic
sources\(^1\). In addition, radiation acts as a channel for energy transport, shaping the
dynamics and impacting the evolution of astrophysical systems on all scales: stellar
evolution, stellar and galactic winds, super-Eddington accretion, epoch of reionization,
etc.

\(^1\)We do receive additional information from cosmic rays, neutrinos and, hopefully soon, gravitational
waves, but these pale next to the enormous volume of information we receive via electromagnetic radiation.
CHAPTER 4. A 3D GENERAL RELATIVISTIC RADIATIVE CODE

Radiation is quite challenging to model since it behaves in highly nonlinear and nonlocal ways. The radiation field is determined by a six-dimensional (6D) system of coupled equations which link spatial positions (3D), ray directions (2D) and frequencies (1D), all as a function of a seventh coordinate, time (in the general time-dependent problem). The high dimensionality of the solution vector, combined with the complex integro-differential nonlocal nature of the problem, results in an extremely taxing computational challenge (both in terms of memory and computational speed).

Because of the intrinsic complexity of the problem, the earliest radiative solvers made use of fairly restrictive approximations. For example, enforcing spatial symmetries allows one to greatly reduce the dimensionality of the problem and led to the first generation of 1D/2D radiative codes (e.g. Feautrier method that exploits symmetric and antisymmetric radiation moments, see Mihalas et al. 1978). Several of the multidimensional RT ideas have also found application in the field of neutrino transport in core collapse supernovae (Burrows et al. 2000; Rampp & Janka 2002; Liebendorfer et al. 2004; Livne et al. 2004; Hanke et al. 2013). In this case, the infalling matter becomes optically thick to the neutrino flux, which propagates analogously to radiative transport in optically thick media.

3D radiative problems are extremely computationally taxing, which motivates the need for highly efficient techniques. Recent applications that make use of 3D RT include models of young stellar objects (Wolf, Fischer & Pau 1998), protostellar to protoplanetary discs (Indebetouw et al. 2006; Niccolini & Alcolea 2006), reflection nebulae (Witt & Gordon 1996), molecular clouds (Steinacker et al. 2005; Pelkonen, Juvela & Padoan 2009), spiral galaxies (Bianchi 2008; Schechtman-Rook, Bershady & Wood 2012), interacting and starburst galaxies (Chakrabarti et al. 2007; Hayward
et al. 2011), AGNs (Schartmann et al. 2008; Stalevski et al. 2012), and cosmological reionization simulations (Ferland et al. 1998; Abel & Wandelt 2002; Finlator et al. 2009). These applications are primarily concerned with point sources illuminating optically thick media.

In more complex optically thick flows where extended diffuse emission is important, the typical approach is to simplify the directional structure of the radiation field. One idea is to decompose the radiation field into its moments (the first three moments being the radiation energy density, radiation flux, and radiation pressure) and to evolve the radiation field locally as a fluid using some closure relation on the moments. This approach is popular due to its local nature and fast speed, and has been implemented in many hydrodynamic codes (Turner & Stone 2001; Bruenn et al. 2006; Hayes et al. 2006; Gonzáles, Audit, & Huynh 2007; Krumholz et al. 2007; Gittings et al. 2008; Ohsuga et al. 2009; Swesty & Myra 2009; Commerçon et al. 2011; Zhang et al. 2011; Kolb et al. 2013). Flux limited diffusion (FLD, Levermore & Pomraning 1981), which is based on the first two moments, is the simplest and perhaps most popular moment-based radiative method. M1 closure is a generalization of the FLD method which uses the first three moments. It has recently gained traction as a fast solver for diffuse emission (Levermore 1984; Dubroca & Feugeas 1999; Stone et al. 1992; Gonzáles, Audit, & Huynh 2007; Sadowski et al. 2013).

The field of protoplanetary disc dust modelling is one area that has been active in developing RT techniques. In this domain, the monte-carlo approach has been the dominant paradigm owing to its ease of implementation and ability to handle anisotropic scattering kernels (Lopez, Mekarnia, & Lefèvre 1995; Niccolini, Woitke, & Lopez 2003; Wolf, Henning & Stecklum 1999; Bjorkman & Wood 2001; Pinte, Duchene, & Bastien
2006). Monte-carlo RT has also been applied to the problem accretion discs, specifically to handle the problem of modeling Compton scattering in disc coronas (Dolence et al. 2009; Schnittman & Krolik 2013; Ghosh 2013) as well as dust scattering of emission from the Galactic centre (Odaka et al. 2011). The primary downside of this technique is the computational expense and noise associated with photon statistics. The method is also limited to modest optical depths.

The limitations of the various methods mentioned above led to the development of more deterministic discretized finite-differencing methods (Stenholm, Stoeerzer, & Wehrse 1991; Steinacker, Bacmann, & Henning 2002) and discrete ordinates characteristic methods (Dullemond & Turolla 2000; Steinacker, Bacmann, & Henning 2002; Hayes & Norman 2003; Vögler et al. 1982; Heinemann et al. 2006; Woitke, Kamp & Thi 2009; Hayek et al. 2010; Davis et al. 2012; Jiang et al. 2014), which have a number of widely discussed advantages. However, all of these methods suffer from the phenomenon of “ray-defects” (see appendix 4.6 for a detailed discussion). Another twist to the discretization strategy involves expanding and evolving the radiation field as a combination of spherical harmonics (Szu-cheng & Kuo-Nan 1982; Evans 1997; McClarren, Holloway, & Brunner 2008). Spherical harmonics respect rotational symmetry and therefore do not suffer from the linear ray defect patterns that plague other discrete ordinate methods. The tradeoff is that spherical methods typically suffer from artificial sidelobe patterns due to the finite order cutoff used in taking the series expansion of the radiation field.

Raytracing methods are also used for predicting the observed flux from astrophysical objects. These codes typically do not solve for the global radiation field within an object; instead the focus is on calculating the apparent intensities that reach a distant observer. In these codes, the typical assumption is to ignore scattering (i.e. to ignore
nonlocal coupling of the radiation field) since this allows one to immediately compute the evolution of ray intensity by simply integrating the local emissivities along a photon geodesic (e.g. Cunningham & Bardeen 1973; Özel & Di Matteo 2001; Huang et al. 2007; Dexter & Agol 2009; Shcherbakov & Lei 2011; Vincent et al. 2011; Chan et al. 2013; Bohn et al. 2014).

Despite the multitude of radiative solvers available, none are currently able to solve the problem of optically thick emission from accretion discs around black holes (BHs). The main difficulty here is that a general relativistic 3D framework is needed to properly account for both light bending effects and doppler/gravitational redshifts. The work reported here is a first attempt at tackling the radiation problem in full glory around black holes. The code we describe here called HERO is a postprocessor – given some fixed gas structure (density, velocity, energy injection rate), as determined by a separate BH accretion disc simulation, our code solves for the radiation field along with its self-consistent temperature solution. We explain here how our code works and show verification tests to gauge its performance under various conditions.

The organization of the paper is as follows. In §4.2, we first describe the method by which HERO operates. We give a brief overview of the radiative transfer problem that HERO solves, and how it generalizes in curved space. We also detail how the two radiative solvers (short/long characteristics) operate, and the numerical methods involved (i.e. acceleration schemes, interpolation, discretization strategy). Next, in §4.3 we check the code by comparing to analytic and benchmark results for 1D, 2D, and 3D test configurations. Finally, in appendix §4.6 we end with a discussion of ray-defects, a systematic problem that plagues short-characteristic radiative solvers.
CHAPTER 4. A 3D GENERAL RELATIVISTIC RADIATIVE CODE

4.2 Radiative Solver

Our radiative code consists of three primary components:

1. A short characteristic solver for quickly obtaining an approximate solution to the radiation field

2. A long characteristic solver for more accurate modeling of the radiation field

3. An optional raytracer for generating mock observations from some distant observing plane

We employ a hybrid approach to solve the radiation field, hence the name of the code: HERO (Hybrid Evaluator for Radiative Objects). The code is hybrid in the sense that it uses both (i) “short characteristics” (SC, Mihalas et al. 1978) and (ii) “long characteristics” (LC, Feautrier 1964). Here “short” and “long” refer to the length of the light rays that are considered in each iteration of the solver. The short characteristics method is only concerned with propagation of radiation from a given cell to its immediate neighbours, whereas the long characteristic method traces rays all the way to the edge of the computational grid.

The motivation for developing a hybrid approach is to allow for the accurate modelling of radiation in optically thin regions via long characteristics, while retaining the computational speed offered by short characteristics in optically thick regions (see §4.2.1 & §4.2.3 for a detailed comparison of the methods). For any given problem, we first apply short characteristics to quickly solve for the local radiation/temperature. We then feed this solution as the initial guess for a more detailed long characteristics calculation.
Using such a hybrid approach allows us to combine the “best of both worlds.”

In addition to solving for the radiation field, in both SC and LC, we also solve for a self-consistent gas temperature. The general idea is to loop back and forth between solving for the radiation field given a fixed temperature structure, and solving for the equilibrium temperature distribution given a fixed radiation field. This procedure is iterated until convergence, leaving us with a self-consistent solution for both the radiation and the temperature.

Our goal in developing HERO is to model/investigate the observed properties of relativistic accretion discs around black holes, and to compare the radiative properties of simulated discs with data collected by earthbound telescopes. Therefore, after completing the SC and LC steps described above, the final step is to input the radiative and temperature structure as computed from LC and to generate synthetic observations of the disc as seen by a distant observer. For this stage we use the standard raytracing approach (see §4.2.5 for details).

Due to the high temperatures present in accretion discs, Comptonization plays a crucial role in determining the observed radiation from relativistic disc photospheres. Because of the complex nature of the Compton scattering kernel, we defer discussion of Comptonization to a follow-up paper, which will focus exclusively on explaining and testing our relativistic scattering module.
4.2.1 Short Characteristics

![Figure 4.1: Schematic of the short characteristics method. A curved ray (null geodesic) is shot “upstream” to determine the intensity at F. This ray is terminated at the neighbouring cell boundaries, and the intensity $I_0$ at the boundary is computed by interpolation of neighbouring grid points (in this case points H and I). The source function $S(\tau')$ in Eq. 4.22 is also obtained by interpolation on neighbouring points (points E, F, H, I), and the intensity at F is thereby calculated. The procedure is repeated for a number of rays in different directions to obtain an estimate of the radiation field at F.]

Short characteristics is a popular algorithm for tackling multidimensional radiative transfer problems (Mihalas et al. 1978; Olson & Kunasz 1987; Kunasz & Auer 1988). It is a local method and hence very fast. The basic idea is to solve for the radiation field at a given point (reference cell) using only the information provided by neighbouring cells (see schematic in Figure 4.1).
CHAPTER 4. A 3D GENERAL RELATIVISTIC RADIATIVE CODE

Ordinary Radiative Transfer Equation

To set the stage, we first discuss the standard case of flat space, where the radiative transfer equation takes the form:

\[
\frac{dI_\nu}{ds} = -(\kappa_\nu + \sigma_\nu)I_\nu + j_\nu + \sigma_\nu \int \phi_\nu(\Omega, \Omega')I_\nu(\Omega')d\Omega'.
\]  

(4.1)

Here \(I_\nu\) is the intensity of a ray travelling along path \(s\) directed towards \(\Omega\); \(\kappa_\nu\) and \(\sigma_\nu\) are the absorption and scattering coefficients; \(\phi_\nu\) is the scattering shape function, normalized such that \(\int \phi_\nu(\Omega, \Omega')d\Omega' = 1\); and \(j_\nu\) is the emission coefficient. Typically, Eq.4.1 is simplified to the form

\[
\frac{dI_\nu}{d\tau_\nu} = -I_\nu + S_\nu,
\]  

(4.2)

where the optical depth \(\tau_\nu\) is given by

\[
d\tau_\nu = (\kappa_\nu + \sigma_\nu)ds,
\]  

(4.3)

and the source function \(S_\nu\) is defined as

\[
S_\nu = \frac{j_\nu + \sigma_\nu \int \phi_\nu(\Omega, \Omega')I_\nu(\Omega')d\Omega'}{\kappa_\nu + \sigma_\nu}.
\]  

(4.4)

In the case of isotropic scattering, \(\phi_\nu = 1/4\pi\), and the scattering term simplifies to

\[
\int \phi_\nu(\Omega, \Omega')I_\nu(\Omega')d\Omega' = \frac{1}{4\pi} \int I_\nu(\Omega')d\Omega' \equiv J_\nu,
\]  

(4.5)

where \(J_\nu\) is the angle-averaged mean intensity. We assume isotropic scattering in all of the work reported here. We further assume a thermal source, for which the emissivity is given by \(j_\nu = \kappa_\nu B_\nu(T)\), where \(B_\nu(T)\) is the Planck function corresponding to the local temperature \(T\) of the medium. With these two simplifications, the source function can be compactly rewritten as:

\[
S_\nu = \epsilon_\nu B_\nu + (1 - \epsilon_\nu)J_\nu,
\]  

(4.6)
where $\epsilon_\nu$ is the photon interaction destruction probability, given by the ratio of the absorption and total opacity:

$$\epsilon_\nu \equiv \frac{\kappa_\nu}{\kappa_\nu + \sigma_\nu}. \quad (4.7)$$

To avoid confusion with summation indices, in the remainder of this paper we will no longer use a subscript $\nu$ to note the frequency dependence of radiative quantities. Unless otherwise stated, this will apply to all subsequent intensities, opacities, source functions, and emissivities.

Covariant Formulation of the Radiative Transfer Equation

Because HERO has been developed for relativistic problems, we use a generalization of the radiative transfer equation that accounts for relativistic effects such as light bending and doppler boosting. The generalization of the radiative transfer equation to curved space has been discussed in the literature (Mihalas & Mihalas 1984; Lindquist 1966), but we give a brief primer here for completeness. Starting from the time dependent version of Eq. 4.1:

$$\frac{1}{c} \frac{dI}{dt} + (\mathbf{n} \cdot \nabla) I = G, \quad (4.8)$$

where $G$ on the righthand side contains all scattering and absorption source terms that interact with the radiation field, and $\mathbf{n}$ is a normalized 3-vector denoting the propagation direction of photons. We can rewrite Eq. 4.8 in a more revealing form by introducing the direction 4-vector $k^\alpha$ whose orthonormal form is (note that $k \cdot k = 0$):

$$k^\alpha = (1, \hat{n}). \quad (4.9)$$

This lets us rewrite the RT equation (setting $c = 1$) as:

$$k^\alpha \frac{d}{dx^\alpha} I = G \equiv -(\kappa + \sigma) I + j + \sigma J. \quad (4.10)$$
CHAPTER 4. A 3D GENERAL RELATIVISTIC RADIATIVE CODE

By considering the Lorentz transformation properties of the various quantities, Eq. 4.10 can be recast in an invariant form (see Mihalas & Mihalas 1984, §7.1):

\[ p^\alpha \frac{\partial}{\partial x^\alpha} I = G, \]  

(4.11)

where the photon 4-momentum \( p^\alpha = h \nu k^\alpha \) takes on the role of the propagation 4-vector. Instead of the usual definitions of intensity \( I \) and source term \( G \), this equation considers their corresponding relativistically invariant versions,

\[ I = \frac{I}{\nu^3}, \quad G = \frac{hG}{\nu^2}, \]  

(4.12)

where the scaling of \( I \) is determined by photon number conservation in phase space. The \((h\nu)^{-2}\) scaling for \( G \) arises after transforming Eq. 4.10 to use \( I \) and allows us to construct the invariant scalings for the absorption and scattering coefficients as:

\[ k_\nu = h\nu \kappa_\nu, \quad s_\nu = h\nu \sigma_\nu, \]  

(4.13)

and for the emissivity and source function,

\[ j_\nu = hj_\nu/\nu^2, \quad S = S/\nu^3. \]  

(4.14)

In terms of these new quantities, the invariant source \( G \) becomes

\[ G = -(t + s)(I - S). \]  

(4.15)

For convenience, since we usually work with comoving quantities, \( \kappa_c, \sigma_c, S_c \), we can instead write \( G \) as:

\[ G = -h\nu_c(\kappa_c + \sigma_c) \left[ -I + \frac{S_c}{\nu_c^3} \right] = p^\alpha u_\alpha (\kappa_c + \sigma_c) \left[ -I + \frac{S_c}{\nu_c^3} \right], \]  

(4.16)
where $u^\mu$ is the four-velocity of the fluid. Eq. 4.11 is in manifestly covariant form and the generalization to curved space is simply accounted for by replacing the directional derivative $\partial/\partial x^\alpha$ with the covariant derivative:

$$\frac{D}{Dx^\alpha} = \frac{\partial}{\partial x^\alpha} - \Gamma^\gamma_{\alpha\beta} D^\beta \partial/\partial p^\gamma,$$

(4.17)
evaluated along a null geodesic (Lindquist 1966). The general covariant form of the RT equation then takes the form:

$$p^\alpha \frac{D}{Dx^\alpha} I = G$$

(4.18)

We can further simplify the above expression by using the photon geodesic equation,

$$\frac{dp^\gamma}{d\lambda} = -\Gamma^\gamma_{\alpha\beta} p^\alpha p^\beta,$$

(4.19)
where $\lambda$ is an affine parameter defined by $p^\alpha = dx^\alpha/d\lambda$. Using Eq. 4.19 to eliminate the Christoffel symbols in Eq. 4.17, the RT equation becomes:

$$p^\alpha \frac{\partial I}{\partial x^\alpha} + \frac{dp^\alpha}{d\lambda} \frac{\partial I}{\partial p^\alpha} = G.$$

(4.20)

Finally, using the chain rule for partial differentiation, we obtain the very simple equation

$$\frac{dI}{d\lambda} = G.$$

(4.21)

This is the form of the covariant RT equation that we employ in HERO. The only point that requires some care is choosing the correct frequency when evaluating the source term $G$. For instance, if we wish to evaluate $I$ at frequency $\nu$ in the lab frame, but we compute $G$ in the fluid comoving frame (the natural frame since the opacity is defined there), then $\nu_c$ in Eq. 4.16 must be chosen such that it corresponds to the lab frame $\nu$ used to label $I$. 

182
4.2.2 Implementation of Short Characteristics

Given a ray in the reference cell F, we compute the upstream location where it intersects the boundary of that cell (see Fig. 4.1) and use this location to determine the relative weights contributed by each of the four nearest neighbours on that boundary. In flat space, it is possible to write down an analytical formula for the intersection point (e.g. Dullemond & Turolla 2000), but there is no generalization in terms of simple functions for curved space. HERO does the calculation numerically.

Although this is computationally somewhat expensive, it needs to be done only once as part of the initialization steps. The information from this initial computation is saved and used repeatedly during the SC iterations. Given the intersection of a ray with the nearest neighbour cell boundary, we compute the intensity of this ray in the reference cell F using the radiative transfer equation (Eq. 4.21):

\[
I_F = I_0 \exp(-\tau_0) + \int_0^{\tau_0} S(\tau') \exp(-\tau') d\tau',
\]  

(4.22)

where \(I_0\) is the incoming intensity at the boundary, and \(\tau_0\) is the total optical depth of the photon path leading from the boundary to point F (computed using the average of the central and boundary opacities). The intensity \(I_0\) is computed by interpolating from the four boundary points. Our current implementation of short characteristics uses linear interpolations at various stages (this is easily improved in the future). In this spirit, the integral evaluation in Eq. 4.22 is computed as

\[
\int_0^{\tau} S(\tau') \exp(-\tau') d\tau' = \frac{e^{-\tau}(S_0 - S_{\tau}(1 + \tau)) + S_0(\tau - 1) + S_{\tau}}{\tau},
\]

(4.23)

which is the analytic solution when the source function \(S\) varies linearly with \(\tau\) along the
photon trajectory. Higher order interpolation schemes have been explored in other codes
(e.g. Kunasz & Auer 1988); however, special care is needed to ensure non-negativity of
the source function.

Our implementation of SC in HERO retains some memory of the radiative quantities
from the past iteration. This helps with stability. Thus, in computing the intensity at
iteration $n$, we use the linear combination

\begin{equation}
I^n = (1 - m)I_{FS}^{n-1} + mI^{n-1}. \tag{4.24}
\end{equation}

where $I_{FS}^{n-1}$ is the result from the RT formal solution (i.e. evolving the radiation field
through SC). Typically, we set $m = 0.5$. HERO with SC converges within a reasonable
number of iterations $\sim 10 - 1000s$, depending on the degree of scattering/coupling in the
problem.

The above discussion assumed that the gas temperatures are given. More often, one
does not know the temperature but has to solve for it based on boundary conditions and
internal heat sources. In this case, between every two of the above iterations for the
intensity, the code carries out a round of iterations (typically $\sim 10s$) to solve for the
temperature using the condition of radiative equilibrium, i.e., the requirement that the
heating and cooling rates of the gas should balance. Radiative equilibrium requires at
each position $x$

\begin{equation}
Q_{cool}(x) = Q_{heat}(x) + Q_{inj}(x), \tag{4.25}
\end{equation}

where

\begin{align*}
Q_{cool}(x) &= \int \kappa_\nu(x)B_\nu[T](x)d\nu, \\
Q_{heat}(x) &= \int \kappa_\nu(x)J_\nu(x)d\nu, \tag{4.26}
\end{align*}
CHAPTER 4. A 3D GENERAL RELATIVISTIC RADIATIVE CODE

and $Q_{\text{inject}}(x)$ refers to any additional injection of energy into the fluid, e.g., through viscous dissipation in an accretion disc.

4.2.3 Long Characteristics

Figure 4.2: Schematic of long characteristics RT solver. We calculate the local mean intensity at an observing cell (large blue circle) by integrating the source function along photon geodesics (light blue path). The RT integral is evaluated by direct summation of contributions from linear piece-wise segments (small blue dots).

Similar to the short characteristics method, our long characteristics solver obtains the radiation field by evolving ray intensities according to the radiative transfer equation (c.f. Eq. 4.21). Figure 4.2 shows a schematic of our approach to LC – rays are shot upstream, and we evaluate the intensity at the observing cell according to

$$I_0 = I_b \exp(-\tau_b) + \int_0^{\tau_b} S(\tau') \exp(-\tau') d\tau'.$$ (4.27)
The above expression is identical to our SC calculation (Eq. 4.22), except that $I_b$ now represents the intensity distribution at the grid boundaries and the ray integral continues past the immediate neighbouring cells to traverse the entire spatial domain. The integral in Eq. 4.27 is evaluated in discrete steps ($\tau' \rightarrow \tau' + \Delta \tau$), where each source-function piece is evaluated using the same linear interpolation scheme as used in SC (Eq. 4.23).

One major difference between our LC and SC solvers is the choice of angular grid. In SC we used a fixed angular resolution ($N_A = 80$, see §4.2.7) for all cells. This fairly low resolution choice is adequate (though only barely) in SC since one only needs to resolve the neighbouring 27 grid cells. In LC, since the rays traverse the entire spatial domain, much higher angular resolution is needed to resolve the emission from distant source cells. This becomes particularly problematic with spherical coordinates where the grid spans a few decades in radius. Cells located at larger radii have a difficult time “seeing” the inner regions unless one takes care in choosing the ray directions.

Our solution is to dynamically set the angular resolution used by the LC solver such that there is sufficient coverage to resolve all the cells within the domain. Consider, as a typical example, a 3D grid in spherical coordinates which extends from an inner radius $r_{in}$ to an outer radius $r_{out} \gg r_{in}$. The radial dynamic range may be up to three decades, as in some of the examples described later. Let us suppose that our spherical spatial grid uses uniform angular spacing in $\theta$, with $N_\theta$ cells, and logarithmic spacing in radius with $N_r \lesssim N_\theta$ cells per decade. Consider a cell at some radius $r$. For this cell, the apparent density of other cells as viewed in its local ”sky” is highly anisotropic. For rays traveling from the direction of the coordinate centre, this cell needs to sample the local radiation field with an angular resolution of order $\Delta \theta_{res} \sim r_{in}/r N_\theta$ in order to adequately sample the contribution of all the cells at the centre. In the opposite direction, however, an
angular resolution of order $\sim 1/N_\theta$ is sufficient.

The above requirement can be achieved by designing a suitable angular tiling of the sky. One simple prescription is to vary the angular resolution $\Delta \theta_{\text{res}}$ as a function of direction $\chi$ relative to the local radius vector as follows:

$$\Delta \theta_{\text{res}} = \frac{f}{N_\theta} \left( \frac{r_{\text{in}}}{r} + \chi \right),$$  \hspace{1cm} (4.28)

where $f$ controls the factor by which we oversample to reduce noise. Usually $f = 1$ is adequate, however even with this relatively low resolution, the number of tiles varies from $\sim N_\theta^2$ for the innermost radial grid cells to several orders of magnitude larger for outer cells. The increase in resolution at the outer cells is needed in order to resolve the interior. The average number of rays per cell is thus a factor of 100 or more larger than the $N_A = 80$ rays that we consider with short characteristics. In addition, each ray in LC traverses the entire grid, compared to just one cell with SC. For both reasons, one iteration of LC often takes $10^4$ times more computation time than one iteration of SC! On the other hand, one iteration of LC moves the solution much closer to convergence since information is propagated across the entire grid.

The main weakness of LC is its inability to handle optically thick regions in scattering-dominated problems. This is because of the huge number of iterations needed to model properly the diffusion process. Since LC cannot be feasibly run for more than a few dozen iterations, it cannot be used to update radiative information deep inside optically thick objects. Luckily, it is precisely in this optically thick regime where short characteristics is free from the ray-defect problem and can be trusted to produce unbiased results. This motivates our hybrid scheme: first we run many iterations of short characteristics to resolve the complete diffusive process and to correctly obtain the
CHAPTER 4. A 3D GENERAL RELATIVISTIC RADIATIVE CODE

radiation field in the optically thick parts of the problem; second we run a few passes of LC to ensure that the optically thin regions of the radiation field are handled accurately and are devoid of ray-defects.

4.2.4 Acceleration Schemes

The radiative quantities $S$ and $J$ are inextricably linked. For future reference, we quantify the dependence of $J$ on $S$ by means of a Λ operator, defined such that

$$J_i = \Lambda^i_j S_j,$$  \hspace{1cm} (4.29)

where the sub/superscripts denote grid cells. Each iteration of the radiative solver acts as Λ, describing how $J$ is related to $S$. However, since $S$ itself has a scattering term, it needs to be updated along with $J$. Writing this explicitly (c.f. Eq. 4.6),

$$S^{n+1} = (1 - \epsilon)J^n + \epsilon B$$

$$= (1 - \epsilon)\Lambda[S^n] + \epsilon B.$$  \hspace{1cm} (4.30)

By switching between solving for $J$ as a function of $S$ (Eq. 4.29), and for $S$ as a function of $J$ (Eq. 4.30), we seek to find a stationary self-consistent solution for all our radiative quantities. Unfortunately, simply applying these two transformations one after the other and iterating has well known convergence issues whenever scattering strongly dominates (say with $\epsilon < 10^{-2}$).

To avoid slow convergence, acceleration schemes are often employed. Accelerated lambda iteration (ALI) is one such technique. To understand the idea behind ALI,

\[\text{2}\]In both SC and LC, the mean intensity term $J$ is evaluated by averaging $I$ via Eq.4.5
CHAPTER 4. A 3D GENERAL RELATIVISTIC RADIATIVE CODE

rewrite Eq. 4.30 so as to isolate $S$:

$$S = [1 - (1 - \epsilon)\Lambda]^{-1}[\epsilon B].$$ (4.31)

Formally, one can obtain the complete solution by evaluating the inverse matrix on the right hand side of Eq. 4.31 and calculating $S$ from $B$. Then evaluating Eq. 4.29 gives $J$. However, in 3D, $\Lambda$ is an enormous matrix and it is impractical to carry out full matrix inversion. Instead, ALI considers an approximate lambda operator, typically chosen to be the diagonal component $\Lambda_{ii}$ (i.e. the contribution to the local $J$ from solely the local $S$ – usually the largest component). Denoting the shift in our old formal solution without acceleration as $\Delta S^{FS} = S_{i}^{n+1} - S_{i}^{n}$, solving the system taking into account $\Lambda$ yields (see Hubeny 2003 for details):

$$\Delta S_{i}^{ALI} = \frac{\Delta S^{FS}}{1 - (1 - \epsilon_{i})\Lambda_{ii}}.$$ (4.32)

To obtain the diagonal of the $\Lambda$ operator, we consider:

$$\Lambda_{ii} = \frac{J_{ii}}{S_{i}},$$ (4.33)

where $J_{ii}$ represents the local contribution to $J$ from the source function within the reference grid cell. This self-illumination contribution to $J$ is the following sum over the $N_{A}$ ray angles considered:

$$J_{ii} = \frac{1}{N_{A}} \sum_{i=1}^{N_{A}} I_{i}^{\text{local}},$$ (4.34)

where $I$ is evaluated by summing up all local sources within the cell:

$$I_{i}^{\text{local}} = \int_{0}^{\tau_{0}} S(\tau') \exp(-\tau') d\tau'.$$ (4.35)

ALI is crucial for scattering-dominated problems whenever $\epsilon < 0.01$ – see convergence tests in §4.3.2. Even with ALI, we often need to run hundreds of iterations before the
CHAPTER 4. A 3D GENERAL RELATIVISTIC RADIATIVE CODE

solution settles down and converges. This is feasible with SC, which is quite fast, but not practical with the much slower LC. The convergence criterion that we use is:

\[
\max \left( \frac{\Delta S_i}{S_i} \right) \leq \delta_c, \tag{4.36}
\]

where the convergence level \( \delta_c \) is preset to a desired accuracy level (typically \( \delta_c = 10^{-4} \)).

4.2.5 Raytracing

At the end of the day, we wish to calculate images and compute synthetic spectra from our radiating objects. This is accomplished by means of raytracing from some distant observation plane. We consider a large number of rays arriving at the observer and trace each backwards in time to determine its trajectory and point of origin. Then, using the radiative transfer equation and the source function as calculated via the LC method, we compute the observed intensity of each ray. Finally, we combine the rays to construct the observed image of the disc as well as the spectrum of the source. These synthetic results, which are easily calculated for different viewing angles, can be directly compared to observations of accretion discs.

Photon trajectories are handled by numerically evaluating the null geodesic equation Eq. 4.19 (a second order PDE in \( x \) and \( p \)):

\[
\frac{d^2 x^\alpha}{d\lambda^2} = -\Gamma^\alpha_{\beta\gamma} p^\beta p^\gamma
\]

with \( p^\alpha = \frac{dx^\alpha}{d\lambda} \). \tag{4.37}

The emission seen at the observer plane is obtained by summing up the emissivity along the ray paths (same exercise as Eq. 4.27). This yields a grid of ray intensities ("image"
of the system) as projected on the observation plane. Integrating all the rays generates the final observed spectrum via

\[ F_\nu = \int I_\nu \cos \theta d\Omega, \]

where \( \theta \approx 0 \) is the normal incident angle of the ray at the observer plane, \( I_\nu \) is the intensity at the observer plane, and \( d\Omega \) is the angular size of a single pixel at the observer plane.

### 4.2.6 Frequency Discretization

HERO is set up to handle frequency dependent opacities. Both the computational cost and memory requirement scale linearly with \( n_F \) the number of frequency bins. All radiative quantities (opacities, intensities, source functions) are calculated for the same discrete set of frequencies. The code could be set up to handle group mean opacities and emissivities with their appropriate quadrature weights to handle more complex line emission problems. In problems where there is frequency coupling (i.e. gravitational/doppler shifting) the redistribution of photons is handled by linear interpolation. Our typical choice for the frequency grid is a total range of 6 decades with a resolution of 10 points per decade (60 frequency bins).

### 4.2.7 Angular Discretization

The angular setup of our code differs based on the dimensionality of the problem that is being solved.
CHAPTER 4. A 3D GENERAL RELATIVISTIC RADIATIVE CODE

1D Angles

In 1D, we solve plane parallel slab problems which are actually 3D problems that can be collapsed down to 1D by invoking translational and azimuthal symmetry. We subdivide the $4\pi$ steradian sphere into equal solid angle slices in $\theta$; thus, we set our angle spacing $d\theta$ such that $\sin \theta d\theta = 2/N_A$.

2D Angles

In all the 2D test problems described in this paper, the rays are considered to live in a flat 2D space, so we do not consider the 3D solid angle at all. We choose our angle grid so as to cover uniformly the full $2\pi$ of the 2D plane. We use simple linear interpolation on 2D angles to handle the mixing of angles when rays travel from one cell to another.

3D Angles

The goal is to subdivide the $4\pi$ sphere of solid angle into $N_A$ solid angle wedges. A few common approaches include bisecting octants, following a "spiral" that winds around the sphere from the two poles, or a special grid of latitudes/longitudes. All these methods approximately achieve the required goal, but they suffer from undesirable patterns of symmetry lines or "seams" on the surface of the sphere. To sidestep this issue, we obtain our angular grid via a numerical approach.

The strategy that we employ is to deposit $N_A$ fictitious charged particles constrained to move on the surface of a sphere, with initial positions set by simple heuristics (equal proper length spacings in $\theta, \phi$ coordinates). Then, we allow the particles to evolve
CHAPTER 4. A 3D GENERAL RELATIVISTIC RADIATIVE CODE

over time under their mutual inter-particle electrostatic forces until they settle on an equilibrium configuration. This naturally produces a set of positions that are nearly equidistant from one another. We use these positions to specify the angular grid (see Fig 4.3 for an example with $N_A = 80$).

For relativistic problems where light bending introduces mixing between different angle bins, or for curvilinear coordinate grid setups that invoke angle grids fixed along the locally rotated unit direction vectors, we handle angular interpolation via linear combinations. For a ray with a given unit direction vector $\hat{\psi}$, there will always be a set of 3 angular points $(\psi_1, \psi_2, \psi_3)$ on the $4\pi$ sphere which defines a triangle enclosing $\hat{\psi}$. We find the angles $(\psi_1, \psi_2, \psi_3)$ by simply locating the 3 angles with the largest dot product to $\hat{\psi}$ in the angle grid.

Once the three vertices of the triangle have been identified, we decompose $\hat{\psi}$ as a linear combination of $\psi_1$, $\psi_2$, $\psi_3$:

$$\hat{\psi} = c_1 \psi_1 + c_2 \psi_2 + c_3 \psi_3. \quad (4.39)$$

The coefficients $c_1, c_2, c_3$ (after renormalizing to satisfy $c_1 + c_2 + c_3 = 1$) are then used as the interpolation weights. In rare cases (i.e. when $\psi$ is slightly outside the boundary of the triangle formed by $\psi_1, \psi_2, \psi_3$), the interpolation weights may be slightly negative. For stability purposes, we clip any negative weights to zero.

In Fig. 4.3, we show how our linear interpolation scheme handles an example intensity pattern. Notice that in the slowly-varying equatorial regions, the interpolated result provides a much better match to the exact intensity distribution than simply using a discretized version of the intensity field. The linear reconstruction does poorly when the radiation field fluctuates on angular scales smaller than a single beamwidth of the
CHAPTER 4. A 3D GENERAL RELATIVISTIC RADIATIVE CODE

$N_A = 80$ grid (e.g. see polar regions of Fig. 4.3).

**Figure 4.3:** A plot of how our interpolation scheme performs compared to the exact solution. Top: Exact $I = \cos(4\theta)\cos(4\phi)$ beam pattern, Middle: discretized version of top panel using our $N_A = 80$ predefined grid, Bottom: $I$ generated from linear interpolation of the $N_A = 80$ grid.

### 4.3 Numerical Tests

We wish to verify the following properties of HERO: 1) that it correctly solves for the radiation field using both short and long characteristics, 2) that it calculates a self consistent gas temperature taking into account the condition of radiative equilibrium, and 3) that it correctly treats the curved space-time of GR.

In all the following examples, we run HERO decoupled from any hydrodynamics. HERO is a postprocessor to compute the radiation field $I_\nu$ and, if required, a self-consistent temperature $T_{\text{gas}}$, given the optical properties (opacities, emission, absorption) of the medium. Other properties of the background fluid, specifically density and
velocity, are assumed to be given and fixed. It is also assumed that the system is
time-independent, so we effectively solve for the steady state solution for the radiation.

We are particularly interested in problems where scattering dominates the opacity
(as is often true with relativistic accretion flows), leading to diluted-blackbody radiation
fields. Due to the complicated nature of scattering, many iterations are needed for the
radiative solver to converge. This makes scattering problems ideally suited to test the
convergence properties of the code. We describe below a series of 1D, 2D and 3D tests.

4.3.1 1D Plane-parallel Grey Atmosphere

Scattering problems are in general very difficult to solve due to the integro-differential
nature of the scattering kernel in the radiative transfer equation (see righthand side of
Eq. 4.1). As such, there are very few examples that have closed form analytic solutions.

The 1D plane-parallel isothermal slab with grey opacity turns out to be one
particularly simple case with a well known closed form solution in the two-stream limit
(c.f. §1 Rybicki & Lightman 1979). There is also a semi analytic solution in the limit of
infinite angles. For this problem, all quantities, \( T, \kappa, \sigma \), are uniform within the medium,
leading to a solution that is only a function of the optical depth: \( d\tau = (\kappa + \sigma)dz \). Under
the two-stream approximation, and using the Eddington approximation to close the
radiative moment equations, the analytic solution for the mean intensity is given by

\[
J(\tau) = B \left[ 1 - \frac{\exp\left(-\sqrt{3}\epsilon\tau\right)}{1 + \sqrt{\epsilon}} \right],
\]

where \( B \) is the thermal blackbody source function, \( \tau \) is the total optical depth from the
surface, and \( \epsilon \) is the photon interaction destruction probability given by Eq.4.7.
CHAPTER 4. A 3D GENERAL RELATIVISTIC RADIATIVE CODE

Note the exponential transition in $\tau$ that separates a thermal interior from a dilute-blackbody surface layer. In Figure 4.4, we show solutions computed by HERO compared to Eq. 4.40. The $\tau = 0$ surface outer boundary condition was set to have zero ingoing flux, and the innermost $\tau = 10^5$ boundary was set to $J = B$ for some fixed $B$. The agreement between the numerical and analytical solutions is quite good ($\sim 1\%$ errors) with a systematic trend of larger errors in low $\epsilon$ systems. These low $\epsilon$ atmospheres are highly scattering dominated and the induced systematic error is simply a consequence of the nonlocal nature of scattered radiation.

![Figure 4.4](image)

**Figure 4.4:** $J$ computed from HERO (dashed lines) compared with the analytic solution (solid lines) for a plane parallel scattering atmosphere described under the 2-stream approximation. Four values of $\epsilon$ (see Eq. 4.7) are shown.
4.3.2 Convergence Tests

The previous plane-parallel atmosphere calculations for the high scattering cases
($\epsilon < 10^{-3}$) required ALI to converge. In Figure 4.5, we show the convergence properties
of the code for a few example problems to emphasize the importance of acceleration in
obtaining the correct solution within a reasonable number of iterations.

In the absence of acceleration, radiation information can only propagate a distance
$\delta \tau \sim 1$ per iteration. Thus, the number of iterations needed for the radiation to pass
from one end to the other is $\sim \tau_{tot}^2$ iterations (corresponding to random walk diffusion
with stepsize $\delta \tau = 1$).

The power of ALI rests in the fact that it boosts the single iteration range of
influence from $\delta \tau = 1$ to $\delta \tau = \delta \tau_{cell}$. To see why this occurs, consider a tiny perturbation
of the source function $\Delta S$ for some cell. The effect of $\Delta S$ on a neighbouring cell is
attenuated by an exponential optical depth factor $\exp(-\Delta \tau)$, where $\Delta \tau$ represents
the inter-cell optical depth. The ALI boost factor (c.f. Eq. 4.32 when $\epsilon \to 1$) in the
scattering dominated limit is $\Delta S_{\text{ALI}} = \Delta S \exp(\Delta \tau)$, which exactly compensates for the
$\exp(-\Delta \tau)$ attenuation. This allows the $\Delta S$ information to propagate to the next cell
in just a single iteration, hence requiring $N_{\text{cell}}$ iterations for information to traverse the
medium from end to end. For most scattering problems, this leads to a huge speedup
(since $N_{\text{cells}} \ll \tau_{tot}^2$).
Figure 4.5: Convergence rates for highly scattering plane parallel atmospheres. Note that the standard Lambda Iteration (LI) procedure saturates at a fixed error threshold for small values of $\epsilon$ (when $\epsilon \ll 10^{-2}$). Accelerated Lambda Iteration (ALI) using Eq. 4.32 converges for all values of $\epsilon$, even extremely small values.

4.3.3 Multiray Temperature Solution

Here we describe a test of HERO’s ability to calculate equilibrium temperatures. In Figure 4.6, we compute a constant flux nonscattering atmosphere and compare the numerical temperature profile with the analytic result (from a table of the Hopf function, see Chandrasekhar 1950). With sufficient angular resolution ($N_A > 16$), the temperature profile computed with HERO is correct to within 1%.
Figure 4.6: Comparison of analytic solution (solid line) to numerical results (dashed lines) for $\epsilon = 1$ constant flux atmospheres modeled with different numbers of angles. The bottom panel is a zoomed in version of the top panel. The temperature solution in the interior ($\tau \gg 1$) matches perfectly to the analytic solution. The agreement at the surface improves as the number of rays is increased.
4.3.4 Test of Spectral Hardening

As a test of HERO’s multifrequency capabilities, we now examine a frequency dependent 1D atmosphere. This problem is simple enough to admit an exact analytic solution for the spectrum at all depths within the atmosphere. Appendix 4.7 derives the depth dependent spectrum.

We are particularly interested in examining how a strongly scattering atmosphere modifies the thermal emission from deep within a plane parallel atmosphere. This phenomenon is known as spectral hardening, and acts to shift the apparent colour of the photosphere emission. The effect is particularly important in the context of modelling black hole accretion disc spectra (Shimura & Takahara 1995; Davis & Hubeny 2006), where the high degree of scattering leads to a shift in the thermal emission peak by a factor of $\sim 1.5 - 2$ towards higher energies.

Figure 4.7 demonstrates an extreme version of this spectral hardening effect for an $\epsilon = 10^{-6}$ plane-parallel medium. HERO produces spectra that agree to within 5% of the analytic solution for all energies and optical depths. Note that above the photosphere ($\tau < \tau_{\text{eff}} = 1/\sqrt{3\epsilon} \approx 10^3$), the radiation spectrum differs quite substantially from the local thermal emissivity/temperature (dotted curves).

For simplicity, the example shown here is with the two-stream approximation for grey (i.e. frequency independent) homogeneous opacities. We impose a constant flux boundary condition at $\tau = 10^5$ and solve numerically for the equilibrium temperature and frequency-dependent radiation intensity as a function of depth. There is excellent agreement between the code output and the analytic temperature and radiation profiles (Eq. 4.61 in Appendix B).
Figure 4.7: Comparison of depth dependent spectra calculated using HERO (solid lines) and the analytic solution (dashed lines) for a plane parallel thermal atmosphere. Note: this is a highly scattering test problem with $\epsilon = 10^{-6}$, which results in significant spectral hardening at the $\tau = 0$ surface (compare dotted blue blackbody curve with solid blue line). At all depths and frequencies, HERO is able to calculate the spectrum to within 5% of the true solution. The spatial resolution chosen for this calculation was 25 points per decade in $\tau$, and 20 points per decade in $\nu$. The two-stream approximation was used for handling angles ($N_A = 2$).

4.3.5 Effect of a Heating Source

As a final accretion disc-like 1D test that utilizes most of the features of the code, we discuss a problem that includes a heating source. We consider a slab with a total optical depth of $2 \times 10^5$ between two free surfaces. We assume grey opacities with $\epsilon = 10^{-4}$ (a
strongly scattering-dominated disc). The material in the slab is heated at a steady rate of $10^{-5}$ (arbitrary units) per unit optical depth, so that the steady state flux from each surface is unity. We use 20 angles and 61 frequencies.

We start the calculation with some arbitrary temperature profile ($B = 10^4$ at all $\tau$ in this particular example) and we run HERO until it converges. After 3000 iterations (this is many more than needed, but we wished to converge as much as possible for this test), the radiative transfer equation has a maximum error of $10^{-6}$ and the radiative equilibrium equation a maximum error of $10^{-7}$. The results are shown in Fig. 4.8. As expected, the temperature profile is quadratic in $\tau$ and symmetric with respect to the mid-plane (upper panel) and the radiation flux is linear in $\tau$ and antisymmetric (middle panel). The lower panel shows spectra at various depths. In the deep interior of the disc, the radiation is nearly blackbody, but as we approach the surface ($\tau_{\text{eff}} = \tau \sqrt{3} \epsilon < 1$) there are signatures of spectral hardening. The radiation that escapes from the surface is distinctly hardened.
Figure 4.8: Upper Left: Self-consistent solution with HERO for the temperature in the interior of a uniformly heated 1D slab. Upper Right: Numerical solution for the frequency integrated radiative flux $H$ compared with the analytic solution. Bottom: Radiation spectrum $J_\nu$ compared with the blackbody spectrum at the local temperature $B_\nu$ for optical depths (from below) 0, 1, 10, $10^2$, $10^3$, $10^4$ from the surface.
4.3.6 2D Solutions and Ray Defects

A well known limitation of radiative solvers with discretized angular zones is the appearance of “ray defects”. These are sharp linear features that arise because interpolation on the angular grid is imperfect. The presence of ray defects is an important motivator for us to use a hybrid scheme. Our long characteristics method operates independently of an angular grid since it adaptively chooses an angular grid and hence does not suffer from ray-defects. In the following sections, we compare our two solvers, SC and LC, for a few test 2D problems. Appendix 4.6 provides a more extensive discussion of the ray defects.

Opaque Wall Test

Ray-defects are particularly severe when there is a compact source of radiation (see Fig. 4.18 in Appendix 4.6). But there are noticeable effects even in the case of smooth extended sources. As an illustration, we analyze a simple 2D example where it is easy to compute the exact solution. In the top panel of Figure 4.9, we show the radiation field for an empty (zero opacity) box illuminated by a hot spot in the centre of the upper wall. The boundary condition at this wall is an opaque ($\tau = \infty$) isotropic source function with a gaussian profile:

$$\frac{S(x)}{S(x_c)} = \exp \left[ - \frac{(x - x_c)^2}{w} \right],$$

where $x_c$ corresponds to the centre of the box, and the width $w$ is equal to 1/5 of the box size. Since the interior of the box has no opacity, the radiation field at any point can be easily found by tracing rays backwards and finding the source function corresponding to the ray’s intersection point with the upper wall. We can thus calculate the mean
intensity $J$ at every point inside the box.

In Figure 4.9, we compare the exact solution to the result from our two RT solvers. For the SC solver (middle panel), a clear wave pattern appears along each of the angle grid directions (we used $N_A = 20$ in this test). These are the ray defects mentioned earlier. The severity of the defects grows as we approach the upper wall, because of the sharp break in the intensity distribution there. On the other hand, LC perfectly reproduces the radiation pattern, even in the far field limit (compare top and bottom panels of 4.9).
Figure 4.9: Comparison of radiation fields calculated using short and long characteristics with an exact analytic solution for an opaque gaussian wall emitting into vacuum (colour log J). From top to bottom: a) exact solution, b) SC solution ($N_A = 20$), c) LC solution. Note the appearance of systematic banding/waves in the SC case. These are ray defects.
CHAPTER 4. A 3D GENERAL RELATIVISTIC RADIATIVE CODE

Shadowing Test

Shadowing is another classic test that is useful for spotting systematic biases in a radiative solver. Simple moment closure schemes such as FLD (Levermore & Pomraning 1981) or M1 (Levermore 1984) typically have problems resolving the correct shadow structure. For instance, FLD has trouble maintaining the coherency of shadows over long distances due to its diffusive nature, and M1 has issues dealing with multiple light sources.

In Figure 4.10, we consider the shadow structure produced by an optically thick square box illuminated from above by an isotropically emitting wall. The top panel shows the true solution, with a clear umbra and penumbra. The middle panel shows the result using HERO SC with a crude angular grid $N_A = 16$. Note that the beam resolution of $20^\circ$ is our typical angular resolution in 3D problems (80 rays in 3D), so this is a realistic example of how SC would perform in 3D. Increasing the angular resolution makes the ray defects less pronounced – the top “exact” solution was produced with $N_A = 1000$ using our SC solver.

The bottom panel shows the result with the LC solver. Notice the high accuracy of LC in handling the radiation shadow pattern. This is a common theme in all of our tests. LC is always much superior to SC.
Figure 4.10: Comparison of shadow patterns calculated using short and long characteristics for an isotropically emitting top wall shining on a central opaque box. From top to bottom: a) “exact” solution ($N_A = 1000$), b) SC solution ($N_A = 16$), c) LC solution. Note the discrete levels (ray defects) in the SC shadow pattern.
4.3.7 3D Solutions

We now shift our attention to 3D test problems, and begin by discussing ray-defects. All the ray defects discussed in 2D (previous subsection and also Appendix 4.6) are present in 3D as well, especially in the case of cartesian grids. A spherical polar grid (the most natural choice for accretion problems) eliminates some problems by introducing ray mixing via the curvilinear nature of the coordinate system. However, this is at the expense of introducing a particularly serious defect for radial rays moving out from a central source. Specifically, if one applies the SC method blindly on a spherical grid, one will obtain a constant radiation energy density and flux at large radius instead of the inverse square law fall-off one expects.

Dullemond & Turolla (2000) discuss a way to “fix” the radial beam problem. They slightly modify the radiative transfer equation along the radial direction such that the expected inverse-square falloff is recovered. We build on their suggestion, except that, instead of modifying only one ray (the radial one), we treat all rays equally and apply an artificial diffusion that, when coupled with a logarithmic radial grid, naturally produces an inverse square falloff in the flux. Details of our diffusion method are explained in Appendix §4.6.2. The following tests as well as those in §4.3.8 employ this ray diffusion scheme.

Ring Benchmark

One particularly simple test problem is an axisymmetric opaque emitting ring in vacuum, for which it is straightforward to calculate the radiation field at any point in space analytically. For an infinitesimally thin ring emitting at the equatorial plane, the total
light reaching any position $\vec{r}$ is given by a 1-dimensional integral.

$$J(\vec{r}) = \int \frac{C}{|\vec{r} - \vec{r}_{\text{ring}}(\phi')|^2} d\phi',$$

where $C$ is a constant specifying the emission per unit length of the ring. In spherical coordinates, we have

$$J(r, \theta, \phi) = \int \frac{C}{r^2 + r'^2 - 2rr' \sin \theta \sin \theta' \cos(\phi - \phi')} d\phi',$$

where $r' = 0.5$, $\theta' = \pi/2$, $\phi' \in [0, 2\pi]$ are the coordinates of the ring in the setup described here.

In Figure 4.11, we compare the results of SC and LC to the analytic result; the constant $C$ has been appropriately normalized to account for the finite emitting area used in the LC/SC calculations. The LC calculation captures the radiation field perfectly ($\delta J/J \approx 10^{-3}$), whereas the SC result shows strong systematic ray-defect patterns. Note that to enforce a $1/r^2$ falloff of the radiation field, the SC calculation employs our ray diffusion scheme (otherwise, the result would be significantly worse).


**Figure 4.11**: Comparison of true solution (left) with the results obtained with SC (middle) and LC (right) for an axisymmetric emitting ring (colours represent log \( J \)). Note the effect of ray defects producing a “spider” pattern for SC. The LC calculation matches the exact analytic answer to within 0.1%.

**Dusty Torus Benchmark**

Previously, in §4.3.1, we demonstrated that HERO correctly computes both the radiation field and the gas temperature in 1D. Unfortunately, there are no analytic solutions available for nontrivial 3D problems. Therefore, we turn to a standard benchmark problem that has been widely discussed in the literature and use this problem to numerically compare HERO with other radiative codes. The model in question consists
of an axisymmetric dusty torus Pascucci et al. (2004) with density structure given by:

$$\rho(r, z) = \rho_0 \cdot f_1(r) \cdot f_2(z)$$

$$f_1(r) = \left(\frac{r}{r_d}\right)^{-1}$$

$$f_2(z) = \exp\left\{-\frac{\pi}{4} \left[\frac{z}{h(r)}\right]^2\right\}$$

$$h(r) = z_d \left(\frac{r}{r_d}\right)^{9/8}$$

The opacities are tabulated and correspond to 0.12\(\mu\)m silicate grains (Draine & Lee 1984). Scattering is assumed to be isotropic, dominating in the wavelength range 0.2 – 1.0\(\mu\)m. A stellar point source is located at the centre of the disc. This point source shines on the disc and dictates its energetics and temperature structure. In HERO, the radiation emanating from the central point source is set to the exact stellar solution at the innermost radial cells of the grid. The propagation of this radiation outwards is then handled by the short (or long) characteristics solver. To avoid being killed by severe ray-defects, we include the diffusive term in the short characteristics solver as described earlier (see §4.6.2). We apply free outflowing radiation boundary conditions at the outer radius of the grid.

Given the above boundary conditions, HERO solves for the radiation field everywhere, both inside and outside the disc, as well as the self consistent disc temperature, i.e. the temperature that satisfies Eq. 4.26. To benchmark the code, we consider the most difficult example presented in Pascucci et al. 2004, the case corresponding to \(\tau = 100\). The upper panels of Figure 4.12 compare the disc midplane temperatures computed by HERO with the benchmark models. Overall, the agreement is reasonable – the slight differences are likely due to ray defects. We recover the temperature structure to within 10%, with the worst cells being located in the optically
The equilibrium temperature profile is highly sensitive to the amount of scattered light (which dominates over the direct stellar illumination near the disc surface by an order of magnitude). The good agreement between HERO and Pascucci et al. (2004) indicates that: 1) HERO correctly handles/redistributes the scattered light; 2) the temperature solver is robust.

As part of this test, we also show in Fig. 4.12 how the short and long characteristics version of HERO perform in determining the self-consistent disc temperature. Short characteristics does a reasonable job everywhere within the optically thick parts of the disc, but systematically underestimates the radiation field and temperature in the optically thin regions. Generally, SC has difficulties propagating radiation towards the coordinate poles. Panel 2 of Figure 4.12 shows the resultant underestimated radiation temperatures (∼ 10% error) and that the bias increases with increasing θ resolution.

The long characteristics solver does not suffer from this systematic error and recovers the correct temperature profile within tens of iterations throughout the optically thin region. This confirms that it is generally a good idea to run a few iterations of long characteristics after the short characteristics solver has been run to convergence. We make a special point to emphasize the importance of the LC pass. One might be tempted to instead run a high resolution SC pass to pin down a more accurate radiative solution. However we find that high resolution does not reduce the systematic biases that plague the SC method. Despite the computational expense incurred by the LC method, it is the only way to obtain an accurate solution to the radiation field.

Finally, we also show in the lower two panels of Fig. 4.12 integrated disc spectra.
Agreement between HERO and the other benchmark codes is within 20%, which is comparable to the spread amongst the four independent codes discussed in Pascucci et al. (2004). The HERO spectra were computed by raytracing from a distant observing plane, making use of the complete radiative solution (i.e. $S_\nu$) obtained from our SC+LC hybrid solver.
**Figure 4.12:** Comparison of HERO with the most difficult $\tau = 100$ dusty torus benchmark test of Pascucci et al. 2004. The top two panels show slices of the self-consistent temperature solution (top left: radial slices, top right: poloidal slices). We also show the SC solution evaluated for two different choices of grid resolution: $N_\theta = (53, 106)$. We find that the SC solution systematically underestimates the radiation/temperature field near the polar regions, with the effect becoming enhanced at high resolutions. This is a consequence of the angular diffusion scheme used in SC (see Appendix 4.6.2 for more discussion). The bottom two panels compare spectra at different inclination angles as computed from LC raytracing.
4.3.8 GR Solutions

Light Bending

HERO is designed to solve for the radiation field in general relativistic curved spacetimes. One important effect is light bending which is demonstrated in the test problem shown in Figure 4.13. $J$ is computed with both the SC and LC in the 3 panels for a beamed light source located just outside the photon orbit ($r = 3M$) propagating in Schwarzschild space-time. We see that the regions with the highest intensity follow the expected curved trajectory. However, the beam is broadened substantially. This is a consequence of the finite angular grid (number of angles $N_A = 80$) used in the computation.

The lower two panels of Figure 4.13 show LC solutions to the same problem. The LC beam remains narrow and coherent, agreeing very well with the expected behaviour for free-streaming radiation at the photon orbit. The middle panel shows the LC result using an emitting source whose beam has been artificially broadened to match the angular resolution of the SC $N_A = 80$ angular grid. The bottom panel shows the true resolving power of LC, where the emitting source corresponds to a $\delta$-function in angle.
Figure 4.13: Light bending test for our two radiative solvers using a narrow laser beam injected tangent to the $r = 3$ photon orbit. Upper left: SC, Upper right: LC with an artificially broadened beam to match the SC angle grid size, Bottom: LC pure. The solid black line shows the analytic result corresponding to the null geodesic.
CHAPTER 4. A 3D GENERAL RELATIVISTIC RADIATIVE CODE

Disc Spectra

To check that our handling of Doppler and gravitational redshifting is correct, we consider the problem of black hole accretion disc spectra. This problem has been tackled numerous times over the years using many independent codes (e.g. Li et al. 2005; Davis & Hubeny 2006; Kulkarni et al. 2011; Zhu et al. 2012) and is a simple but useful benchmark test.

We place an optically thick, geometrically thin disc around a Schwarzschild \( (a = 0) \) black hole radiating as per the idealized Novikov & Thorne (1973) disc model. In this problem, the disk emission is treated as isotropic thermal radiation with flux given by the NT model. For simplicity, we ignore any spectral hardening effects since we treat the emission in HERO as emanating from a single equatorial grid cell (i.e. for this test, we do not resolve the photon diffusion process that gives rise to spectral hardening). The inner edge of the disk is fixed at the ISCO \( r_{\text{ISCO}} = 6 \) and the outer edge is located at \( r = 1000 \).

We solve for the radiation field in the disc exterior using HERO in full GR. In HERO, the calculation of the disc spectrum is typically handled in two stages – 1) we first solve for the 3D radiation field above the disc using the short/long characteristics solver (this is not needed in the present example because we specify the disc emission profile and assume vacuum outside the disc), and 2) we trace rays backwards (via Eq. 4.22) from a distant observer plane located to create a synthetic image of the disc.

The HERO code solves for the full three-dimensional source function and radiation field within the \( 2M < r < 1000M \) spatial domain of the grid. This information is then fed into a separate raytracing subroutine. In the raytracing stage, parallel rays are shot...
towards the disk distributed according to a squeezed logarithmic polar grid (the same setup as Kulkarni et al. 2011) from an observer plane located at $r = 100,000M$. The final spectrum is generated by integrating the flux across the observer plane.

The top panel of Figure 4.14 shows the computed image for the particular example problem. Integrating over this image yields the observed disc spectrum, which we show in the lower panel. The disc spectrum computed with HERO agrees very well with a previous calculation by Kulkarni et al. (2011), lending confidence that HERO correctly handles GR effects.
Figure 4.14: Upper panel: Raytraced image of a razor thin disc as viewed by an observer located with inclination angle $i = 60$. Colors indicate log of the frequency integrated intensity. The asymmetry is due to the doppler effect combined with gravitational redshifting. The central arc feature corresponds to a secondary image of the accretion disk that arises from strong gravitational lensing about the black hole. Bottom left: Integrated disc spectrum as computed by HERO (red points), compared to the result with the Kulkarni et al. 2011 code (blue line). Bottom right: fractional errors between HERO and Kulkarni et al. 2011.
CHAPTER 4. A 3D GENERAL RELATIVISTIC RADIATIVE CODE

Iron Line Spectra

In a similar vein as for the previous test, which was based on the NT continuum spectrum of the disk, we also benchmark our code via a calculation of the emission profile due to a monoenergetic line (e.g. Fe-Kα as seen in many Seyfert galaxies and some microquasars). For a system that emits only monoenergetic δ-function lines, the final integrated line profile depends only on the geometrical properties of the system (i.e. redshifting from doppler/gravitational curvature and lensing effects from the Kerr metric).

We consider a Keplerian accretion disk in the Kerr metric whose line emission is modulated by a power law emissivity profile $F(r) \propto r^{-3}$. We also set the domain of the disk to span from $r_{in} = r_{ISCO}$ out to $r_{out} = 400$. Figure 4.15 shows the resultant spectra computed for two different choices of BH spin and two different choices of viewing angle. We find good agreement between HERO and the benchmark code RELLINE (Dauser et al. 2010, 2013), with the largest discrepancies occurring in the low energy red tail of the line profiles.

The setup in HERO is identical to that of the previous NT disc tests, except that a monoenergetic 6.4 keV line was used as the local emissivity instead of a thermal continuum source. In addition, we use much higher resolution for the raytracing grid since the low energy red tail of the line profile depends sensitively on how well the inner edge of the disc is resolved.
Figure 4.15: A comparison of the Fe-Kα line profile as computed by HERO and RELLINE (Dauser et al. 2013). Line profiles are computed for different viewing inclinations: top panel shows $i = 30$, whereas bottom panel shows $i = 60$. 
CHAPTER 4. A 3D GENERAL RELATIVISTIC RADIATIVE CODE

Returning Radiation

Another classic accretion disk problem is that of computing the returning radiation due to relativistic light bending around the black hole. For this test, we setup a razor thin accretion disk that emits according to the standard thin disc luminosity profile (Page & Thorne 1974) and measure the amount of returning radiation incident on the disk. The goal is to benchmark both our SC and LC radiative solvers against the solutions of Cunningham (1976), who tackled the same problem by means of relativistic transfer functions.

In HERO, we model the returning radiation problem with a grey calculation (1 bolometric frequency bin) on an axisymmetric polar grid with \((n_r, n_\theta) = (60,30)\) restricted to the upper half plane. The spatial grid was set as uniformly spaced in angle and \(\log(r)\). Boundary conditions invoked are: reflecting for the polar axis (to account for light that passes through the pole), constant flux injection at the equatorial disk plane according to (Page & Thorne 1974), and zero incident radiation at the inner and outer radial boundaries.

In figure 4.16, we plot the incoming radiation flux at various locations above the disk and compare to the solution of Cunningham (1976) for a moderately spinning \(a_\ast = 0.9\) black hole. We find that SC systematically overestimates the amount of returning flux at large radii, presumably caused by the angular interpolation bleeding some of the outbound radiation into inbound rays. LC on the other does not experience any strong systematic biases, but suffers from a lack of resolving power at large radii since at these large distances, the LC ray grid can miss the inner photon ring that is responsible for most of the returning radiation.
Figure 4.16: A calculation of the returning radiation for a razor thin accretion disc around a moderately spinning $a_*=0.9$ black hole. The standard Page & Thorne (1974) luminosity profile is used to set the outgoing flux/rays from the disk plane (blue). We compare the returning flux as calculated by the two radiative solvers in HERO (red, green) with the result from Cunningham (1976) obtained via relativistic transfer functions (black). The black dashed line is an extrapolation of the values from Cunningham’s data table.

Vacuum Test

As a final test of the general relativistic capabilities of HERO, we turn to the problem of light propagation in vacuum from an opaque spherical shell. In Schwarzschild geometry, it is a simple exercise to solve for the apparent angular size of a constant radius shell
as viewed by an observer at some other (larger) radius. Furthermore, if the surface of
the shell radiates isotropically like a blackbody with a constant surface temperature, one
can calculate the radiation quantities (i.e. $J, F, L$) by simply integrating the constant
intensity across the apparent solid angle subtended by the shell. For instance, the radial
profile of luminosity has a particularly simple form, scaling directly with gravitational
redshift as

$$L_{\text{local}} = L_\infty (1 + z)^2$$

(4.45)

where $L_{\text{local}} = F_{\text{local}} 4\pi r^2$ is the total luminosity as measured by an observer at radius
$r$ and $L_\infty$ is the luminosity at infinity. In Figure 4.17, we show the results for the
luminosity as calculated by our general-relativistic short and long characteristics solvers.
We normalize the luminosity by the redshift factors so the analytic solution simply
 corresponds to a flat horizontal line (i.e., constant luminosity as measured at infinity).

In general, we find that our short characteristics solver systematically underestimates
the luminosity profile at large distances. These tests were carried out with the diffusion
prescription described earlier, therefore if the shell radiated purely in the radial direction,
SC would by construction give the correct answer. Here the surface radiates isotropically,
and there is a deviation from the true answer because of angle interpolation and diffusion.
The various SC curves in Figure 4.17 correspond to different choices for the radius of
the inner emitting shell. The high degree of similarity for all choices of inner radius
(i.e. ranging from from highly relativistic $r_0 = 3$ in units of $GM/c^2$ to nonrelativistic
$r_0 = 10^5$) implies that the SC bias is independent of relativistic effects. It arises purely
from the discretization of the spatial and angular grid. While the bias is not negligible
– the luminosity is reduced by a factor of 2.5 at large radius – note that the luminosity
would be a factor $\sim 10^6$ too large if we did not include diffusion. As with the other test problems, LC gives an essentially perfect answer.

Panel two of Figure 4.17 shows another interesting phenomenon, viz., there is a strong dependence between the SC bias on the choice of spatial grid resolution. As we increase $\theta$-resolution (keeping the $r$-resolution fixed), the luminosity profile of our isotropically emitting shell exhibits a stronger bias to lower values. This effect is ultimately caused by the angle diffusion scheme that we employ (c.f. Appendix 4.6.2). The diffusion is tuned such that the radial cell-to-cell attenuation of light recovers the inverse-square law. If the cell aspect ratio in $r - \theta$ is too rectangular, then the diffusion has a strong directional preference. Light rays traveling along the short dimension of the cell are overattenuated since they hit cell boundaries more often than rays traveling in the long-dimension (the degree of diffusion is directly proportional to how often the light ray traverses cells within a given spatial distance). Based on our tests, grid cells should ideally have a square aspect ratio to minimize the error, and any ratio in excess of 2:1 should be avoided.
Figure 4.17: Variation in radial luminosity profiles as a function of a) relativistic effects (lensing, redshifting) and b) grid-θ resolution. Note that the luminosity plotted is the corrected luminosity as measured by an observer at infinity. The analytic solution is therefore a constant $L/(1+z)^2$ for all radii.
CHAPTER 4. A 3D GENERAL RELATIVISTIC RADIATIVE CODE

4.4 Summary

We have described in this paper HERO, a new general relativistic radiative transfer code. The primary aim of this code is to model the radiation field in accretion flows around black holes. The unique features of HERO are: 1) a hybrid short/long characteristics radiative solver that enables accurate and fast modelling of complex anisotropic radiation fields; 2) implementation in a general relativistic framework taking into account the effects of light bending, doppler beaming, and gravitational redshifting.

HERO is written as a post-processing code decoupled from the hydrodynamic evolution of the fluid. It computes the time-independent radiation field assuming a given fixed background fluid structure. Strictly speaking, this approach is valid only for problems where the fluid velocities are small compared to the speed of light (i.e. nonrelativistic flows). Alternatively, and this is the primary application we have in mind, it could also be applied to time-steady relativistic problems.

We provide a detailed explanation of the long/short characteristics method used to solve for the radiation field and our approach for solving the self-consistent gas temperature. To verify that HERO produces physically correct answers, we have performed a comprehensive set of tests designed to examine the code’s convergence properties, accuracy, and capability to handle multidimensional relativistic problems. We confirm the well known result that 2D and 3D problems with compact sources suffer from significant ray defects in the far field when analyzed with the short characteristics method. We present an approximate fix which mitigates the effects in the case of a 3D spherical grid. However, for accurate results, it is necessary to switch to a long characteristics solver which is unaffected by ray-defects.
CHAPTER 4. A 3D GENERAL RELATIVISTIC RADIATIVE CODE

As the subject of a follow-up paper, we intend to apply HERO to radiative MHD simulations of accretion discs – particularly simulations undergoing super-Eddington accretion, where radiation feedback strongly dominates the dynamics of the flow. Using HERO, we will investigate the integrated spectra to see the role that self-shadowing and irradiation plays in these systems. This application requires a Comptonization module which will be described in our next follow-up paper.

4.5 Acknowledgements

The authors would like to thank Nathan Roth, Jack Steiner, Jonathan McKinney, James Guillonchon, Jeff McClintock, Yan-Fei Jiang, Javier Garcia, Eric Keto, and Jiachen Jiang for their excellent insights on radiative transfer and comments/suggestions regarding HERO. RN and YZ aknowledge support from NSF grant AST1312651 and NASA grant NNX 14AB47G. AS acknowledges support for this work by NASA through Einstein Post-doctoral Fellowships PF4-150126, awarded by the Chandra X-ray Center, which is operated by the Smithsonian Astrophysical Observatory for NASA under contract NAS8-03060. Finally, we are grateful for support from NSF XSEDE grant TG-AST080026N, the NASA HEC Program, and the Harvard Odyssey cluster for providing computing resources used in developing this code.
4.6 Ray Defects

The method of short characteristics constitutes the primary workhorse of our radiative solver and is used to generate a good first approximation to the radiation field. However, “ray-defects” are a well known limitation of SC, which is why we need to follow up SC with the more accurate LC method. Here we discuss some of the properties and explore the cause of ray defects. Figure 4.18 shows a few examples of ray defects in 2D arising from point source emitters. The defects manifest as unphysical beam-like patterns far from the emitting source.

Point sources generate the strongest defect pattern so we use them in the following discussion to illustrate the main issues. Point-source-like emission does appear in the problem of black-hole accretion discs, e.g. the hottest innermost region of the disc shines extremely brightly and due to its compact spatial scale, acts like a point source at large distances from the centre.

For a fixed angular grid, the point source ray defects form a series of radial beams that reflect the underlying structure of the angular grid (see Figures 4.18 for a few examples in different coordinate systems). Beam collimation is enhanced for rays travelling in directions where neighbouring grid cells cover a smaller angular size. This “grid-lattice” effect is best seen in panel c) (sheared box) in Figure 4.18, where the thinnest beams are those travelling to the upper-right/lower-left sectors (i.e. the directions where the neighbour points as defined in Figure 4.1 span the smallest angular extent). The same, but less pronounced result is seen in panel d) of Figure 4.18 (polar coordinate system). Here, the rays pointed radially inward suffer less dispersion than their radially outward counterparts, again for the same reason as in the sheared box case.
(tighter angular packing occurs for the neighbour cells at smaller radius).

Finally, ray defects are particularly enhanced for rays that directly intersect neighbouring cell centres (as an example, note that the 45° rays exhibit overwhelmingly strong defect patterns in our cartesian setup in the top two panels of Fig 4.18).

Ideally, we would like to represent the discretized radiation pattern as a smooth field instead of a superposition of laser beams. After much trial and error, we have arrived at an angular diffusion based solution for spherical log polar grids. To motivate our final solution, we first examine the root cause of the defects.
**Figure 4.18**: Comparison of the radiation solution for a point source computed by short characteristics for different grid choices ($N_A = 20$). Colour indicates log($J$). Top left: cartesian grid (100x100); Top right: same as top left, but using quadratic interpolation; Middle: sheared cartesian (100x100); Bottom: Polar grid (60x100). In all cases, we place a delta function source function that emits isotropically. The pencil beams that appear are the ray-defects – physically, the far-field radiation pattern should be spherical and smooth but the limitations of treating radiation only locally in the SC method results in the thin beams (see discussion in §4.6.1).
4.6.1 Mathematical Origin of Ray Defects

Ray defects ultimately arise due to the local nature of the short characteristics solver. Methods that operate using just local propagation of light develop a “Pascal’s Triangle” characteristic beam pattern with increasing distance. If the radiation from a source cell is split with some linear combination to its neighbouring cells and the same method is applied uniformly for all cells, then the propagation of light has the following characteristic shape – consider the case where the mixing coefficients are [1/2, 1/2]:

\[
\begin{align*}
1 \\
1/2^1 & \quad 1/2^1 \\
1/2^2 & \quad 2/2^2 & \quad 1/2^2 \\
1/2^3 & \quad 3/2^3 & \quad 3/2^3 & \quad 1/2^3 \\
1/2^4 & \quad 4/2^4 & \quad 6/2^4 & \quad 4/2^4 & \quad 1/2^4 \\
\vdots & \\
\end{align*}
\]

The final far field pattern simply corresponds to the weights of a discrete random walk. This shape is a Gaussian, with a characteristic width of \( w \propto \sqrt{d} \) where \( d \) is the total propagation distance from the source. The beam far from the source will therefore have angular size \( w/d \propto 1/\sqrt{d} \rightarrow 0 \) as \( d \rightarrow \infty \). Thus, at a sufficiently large distance, the propagation of light via short characteristics will inevitably result in a series of laser
beams.

The defect problem can be further exacerbated if a ray happens to pass exactly through a neighbouring cell centre. In this case, the mixing/interpolation coefficients are $[1, 0]$ and the propagation diagram reduces to:

$$
\begin{array}{ccc}
1 & & 1 \\
\downarrow & \downarrow & \downarrow \\
1 & 0 & 0 \\
\vdots & & \vdots
\end{array}
$$

Notice that for this limiting case the beam pattern does not spread at all, resulting in a zero width beam of constant intensity at all distances. This kind of defect is particularly disastrous in the case of spherical coordinates since it affects all radial rays. Thus, any compact light source near the centre of the spherical grid will develop serious ray-defects at larger radii. The effect is quite devastating if there is a strong point source at the centre. The radial ray defect acts to force a constant intensity on all radial rays, resulting in constant radiation energy density independent of distance rather than the expected inverse square falloff. The natural way to correct non-spreading beams is to introduce some degree of artificial broadening (i.e. force the mixing coefficients to have some minimum floor value). We describe our approach in detail in the next section.
4.6.2 Ray Defect Correction Schemes

The most obvious approach to mitigate the impact of ray defects is simply increasing the angular resolution. However this approach is not feasible for complex problems with large spatial domains due to the unfavorable computational cost scaling with resolution.

One idea to combat ray defects is to use higher order interpolation schemes in treating the propagation of radiation. Davis et al. (2012) found that quadratic monotonic interpolation works well and suggest that it be used over standard linear interpolation. Unfortunately, this does not address the fundamental problem of beam spreading – higher order schemes are actually counterproductive in that they produce even thinner beam patterns and amplify the significance of ray defects (Compare the top two panels of Figure 4.18 for linear vs. quadratic interpolation).

Ultimately, the problem (as illustrated in §4.6.1) has to do with the lack of spreading for a single beam. This motivates us to implement a diffusive scheme to bring back the necessary amount of spreading and to set the intensity values in such a way that an inverse square falloff of radiation density and flux is recovered.

Our approach is the following: every time we read out intensity values from our angular grid (i.e. the interpolation scheme described in §4.2.7), we apply a floor to our interpolation coefficients so that there is always some degree of mixing between nearby angles. That is, we modify the coefficients $c_1, c_2, c_3$ in Eq. 4.39 to new values $c'_1, c'_2, c'_3$ given by:

$$c'_i = \frac{c_i + w}{\left(\sum_i c_i\right) + 3w} \quad (4.46)$$

where $w$ is the diffusion coefficient that controls the amount of angle mixing. A more
physical way to understand this diffusion coefficient is to convert it into an attenuation factor for nearby cells. For a monodirectional beam (i.e. $I = 0$ for all but a single angle bin), the beam will attenuate by a factor $a$ after propagating across a single cell, i.e.,

$$I^{n+1} = aI^n,$$  \hspace{1cm} (4.47)

where from Eq. 4.46, we find:

$$a = \frac{(1 + w)}{(1 + 3w)}.$$  \hspace{1cm} (4.48)

This motivates a method for choosing the appropriate value of $w$ for the case of a log-polar grid (where the radial spacing is uniform in log$r$). Since we desire an inverse square law falloff of the radiation field, we simply solve for $a$ in the equation $a^{N_{\text{dec}}} = 0.01$ where $N_{\text{dec}}$ is the number of cells per radial decade. For example if $N_{\text{dec}} = 20$ points per decade (our canonical choice), we have have $a = 0.794$ and $w = 0.149$.

In Figure 4.19, we show radial profiles of $F$ calculated with the SC for a point source central star. Note that, in the absence of diffusion, the effect of radial ray defects is to produce an unphysical solution where $F$ is constant with radius.

In Figure 4.20, we show the results from this diffusion scheme compared to the exact solution for an off-centre equatorial ring source (e.g. the same setup as described in §4.3.7) for different choices of the spatial grid resolution. We find that our diffusive scheme suffers from systematic overattenuation in the polar regions for spatial grids that too rectangular. We recommend keeping the grid cell aspect ratio as close to square (1:1) as possible in order to minimize this effect.

The scheme described here is a generalization of that proposed by Dullemond & Turolla 2000. They applied the same factor $a$ derived above except that they did it
solely for the radial ray. For a central point source, which produces only radial rays, their method and ours produce the same correct inverse-law behaviour of flux. However, when the source is extended, e.g., the spherical shell test problem described in §4.3.8, their method causes the intensity to fall too steeply. Our method also suffers a similar systematic bias, but it is less severe (see Fig. 4.17). In the limit of an isotropic radiation field, e.g., the interior of a blackbody enclosure, their method would still cause intensity to decline with increasing radius. Our diffusive method, on the other hand, would give the correct result, viz., no change in intensity. These differences are relatively minor. The main feature of both approaches is that they recover the inverse square law at least approximately. However, a major limitation of our diffusive approach is that it only works for the case of uniform log-radial grids.

**Figure 4.19:** Radial profiles of $F$ for various choices of diffusion coefficient as calculated by short characteristics code on a spherical polar grid $(n_r, n_\theta) = (60, 100)$. Given these grid dimensions, the diffusion coefficient must be set to $w = 0.149$ in order to reproduce an inverse square falloff for flux.
Figure 4.20: Ray defect pattern from an isotropically emitting ring for 3 different choices of grid resolution. Here we use a spherical spatial grid with locally defined ray angles. Panels: a) \( n_\theta = 26 \) square aspect ratio; b) \( n_\theta = 52 \) 1:2 aspect ratio; c) \( n_\theta = 104 \) 1:4 aspect ratio. Notice the systematic reduction of radiation near the polar regions when \( n_\theta \) increases and also the characteristic “spider” pattern arising from our choice of spatial grid interpolation.

### 4.7 Analytic 1D Atmosphere Spectrum

In this problem, we set up a constant flux 1D atmosphere subject to fixed grey opacities \( \kappa_\nu = \text{const} \) with an absorption fraction set to \( \epsilon = 10^{-6} \) (i.e. a highly scattering-dominated atmosphere). The local source function is a combination of thermal emission and reflected light (c.f. Eq 4.6).

The analytic solution to this problem can be easily obtained by considering the
moments of the radiative transfer equation. We define the first few moments of the intensity field as

\[
J_\nu = \frac{1}{2} \int_{-1}^{1} I_\nu d\mu, \\
H_\nu = \frac{1}{2} \int_{-1}^{1} \mu I_\nu d\mu, \\
K_\nu = \frac{1}{2} \int_{-1}^{1} \mu^2 I_\nu d\mu,
\]

(4.49)

where \( \mu \equiv \cos(\theta) \) is the angle cosine with respect to the plane normal. Using these quantities allows us to write the moments of the radiative transfer equation as:

\[
\frac{dH_\nu}{d\tau_\nu} = \epsilon (J_\nu - B_\nu) \quad (4.50)
\]

\[
\frac{dK_\nu}{d\tau_\nu} = H_\nu \quad (4.51)
\]

Combining Eqs. 4.50 + 4.51 and invoking the Eddington approximation \( K_\nu = J_\nu/3 \) yields

\[
\frac{d^2 J_\nu}{d\tau^2_\nu} = 3\epsilon (J_\nu - B_\nu),
\]

(4.52)

which is simply a constant coefficient second-order inhomogeneous differential equation in \( J_\nu \). This allows us to construct an exact solution using standard methods. The solutions to the homogeneous system are simply

\[
\phi_1(\tau) = \exp(\sqrt{3\epsilon \tau}),
\]

\[
\phi_2(\tau) = \exp(-\sqrt{3\epsilon \tau}).
\]

(4.53)
The particular solution $J_p$ that satisfies the inhomogeneous system is given by

$$J_p(\tau) = \phi_1(\tau) \int_{\tau}^{\infty} \frac{\phi_2 g}{W} d\tau' + \phi_2(\tau) \int_{0}^{\tau} \frac{\phi_1 g}{W} d\tau',$$

(4.54)

where $g = 3\epsilon B$ is the inhomogeneous function, and $W$ denotes the Wronskian, defined by

$$W \equiv \frac{d\phi_1}{d\tau} \phi_2 - \frac{d\phi_2}{d\tau} \phi_1$$

(4.55)

$$W = 2\sqrt{3}\epsilon$$

(4.56)

Putting everything together, the solution takes the form

$$J = J_p + c_1 \phi_1 + c_2 \phi_2,$$

(4.57)

where the undetermined constants are set by the boundary conditions of the problem.

At the $\tau \to \infty$ inner boundary we expect $J \to B$, so we must eliminate the exponentially growing mode by setting $c_1 = 0$. To set the surface boundary condition, we make use of the two-stream approximation and evaluate

$$H(\tau) = \frac{1}{\sqrt{3}} \frac{dJ}{d\tau}$$

(4.58)

and enforce a surface boundary condition that is consistent with the Eddington approximation:

$$\frac{H(0)}{J(0)} = \frac{1}{\sqrt{3}},$$

(4.59)

which sets

$$c_2 = -\left(\frac{1 - \sqrt{3}}{1 + \sqrt{3}}\right) \phi_2 \int_{0}^{\infty} \frac{\phi_2 g}{W} d\tau'.$$

(4.60)

The final step is to determine the thermal source ($B(\tau)$) that is consistent with our radiation solution from Eq. 4.57. This can be calculated using our two radiative transfer

---

3 The particular solution is constructed via the “variation of parameters” method.
moment equations. Since we have a constant flux atmosphere, $dH/d\tau = 0$, which implies $J(\tau) = B(\tau)$ from Eq. 4.50. We combine this with integrating the pressure equation (Eq. 4.51) to yield the full solution

$$J(\tau) = 3H \tau + J(0)$$

$$\rightarrow B(\tau) = 3H \left( \tau + \frac{1}{\sqrt{3}} \right)$$ (4.61)
Chapter 5

HEROIC - A Comptonization Module for the HERO radiative code

*This thesis chapter will be submitted to MNRAS*


**Abstract**

We describe HEROIC, an upgrade of the relativistic radiative code HERO which now includes Comptonization. To model the Comptonization process, we employ a quadratic Kompaneets operator approach to handle the photon energy redistribution process.

We benchmark and test our code with simple 1D and 3D scattering problems that admit analytic solutions. We then apply HEROIC to the astrophysical problem of a scattering-dominated hot accretion disk. We find that our 3D calculation yields slightly hotter spectra than previous 1D-based estimates. We also find a high energy power-law...
CHAPTER 5. 3D COMPTONIZED ACCRETION DISKS

tail that is sourced by the hot thermal emission from the plunging region of the disk.

5.1 Introduction

Comptonization plays a crucial role in determining the high-energy emission properties of a huge variety of astrophysical objects, such as the X-ray radiation from the nuclei of galaxies/quasars (Veledina, Vurm, & Poutanen 2011), governing the decay rates of neutron star bursters (Joss 1977; Lamb et al. 1977), the light passing through plasma clouds of the intra-cluster medium (Prokhorov et al. 2010), and nonthermal spectral properties of galactic microquasars (McClintock & Remillard 2006). The ubiquity of hot ionized gas in astrophysical settings motivates the need for accurate treatment and modelling of the Compton scattering process.

In the case of black hole accretion disks, cold seed photons emitted from a thermal disk get upscattered by hot coronal electrons (Sunyaev & Titarchuk 1980), producing a whole host of spectral shapes in X-rays such as power-laws and Compton humps. Many systems, especially those in the low-hard spectral state, are observed with a dominant Compton component (Zdziarski et al. 1995; Gierlinski et al. 1997), where interpretation of the data requires accurate modeling of the power-law tail. This is especially true in the case of reflection line modeling for black hole systems (Fabian et al. 1989; Tanaka et al. 1995; Miller 2007), where slight errors in the continuum can lead to large systematic biases in the derived black hole parameters (Weaver, Krolik & Pier 1998; Miller et al. 2006; Kolehmainen & Done 2010; Kolehmainen, Done & Diaz Trigo 2014).

Accurate analytic models of the Comptonized spectrum have been worked out
several decades ago for various regimes such as for optically thick, homogeneous 1D and 3D media (Sunyaev & Titarchuk 1980), the limit of low electron energies (Titarchuk & Hua 1995), or the relativistic limit for an optically thin medium (Coppi 1991). However, the complex nature of accretion flows around black holes precludes the use of analytic models, motivating the development of numerical Comptonization schemes.

Monte-Carlo based methods are by far the most popular approach to the problem (Pozdnyakov, Sobol, Sunyaev 1983; Gorecki & Wilczewski 1984; Stern et al. 1995; Malzac & Jourdain 2000; Schnittman & Krolik 2013) owing to the ease with which the technique can handle relativistic geometries and the complex angle and frequency dependent scattering process. The biggest drawback of the Monte-Carlo (MC) approach is that photon statistics limit the accuracy of the computations, posing difficulties in generating spectral models for data fitting since the results differ each time the code is run. This problem is worst in the limit of extreme photon energies, where there is a dearth of photons and hence poor photon statistics. MC methods also suffer in the limit of high optical depths – here, the full photon diffusion process is incredibly taxing computationally, which in practice restricts MC based codes to problems with only moderate optical depths (i.e. $\tau \lesssim 10$).

Another approach is to discretize the problem and numerically solve the radiative transfer problem for some preset fixed geometry (e.g. compPS: Poutanen & Svensson 1996 for moderate optical depths; COMPTT: Titarchuk & Lyubarskij 1995 for accretion disks; TLUSTY: for 1D atmospheres Hubeny et al. 2001). The main advantage of this

\footnote{The use of an energy-weighted scattering kernel is one work-around for the photon-starvation problem in MC methods.}
CHAPTER 5. 3D COMPTONIZED ACCRETION DISKS

approach is that it can easily handle optically thick problems via a Kompaneets operator approach (Kompaneets 1957), which uses a diffusion approximation to handle the nonrelativistic Comptonization problem. This approach is particularly amenable to the short characteristics fixed-grid framework of HERO, our 3D GR radiation postprocessor code (see Chapter 4). In this work, we implement Comptonization in our code and update the name to HEROIC (HERO Including Comptonization). The primary advance in our work is that we introduce a fully self-consistent relativistic radiation module for the 3D Comptonization problem of hot accretion flows around black holes.

Regardless of the approach taken for solving the Compton problem, a final raytracing calculation is needed to connect the result to the actual spectral observations of astrophysical systems. Typically, geodesic paths are traced backwards from a distant observation plane until they hit the accretion flow (e.g. see Rauch & Blandford 1994; Broderick & Blandford 2003; Dovciak et al. 2004; Dexter & Agol 2009; Johannsen & Psaltis 2010; Kulkarni et al. 2011; Zhu et al. 2012). This yields a transfer function that allows one to map the local Comptonized disk emission to the spectrum as measured by the distant observer. In cases of high scattering optical depths, it is crucial for raytracing methods to resolve the complete nonlocal structure of the scattered radiation field (Schnittman & Krolik 2013). For MC-based methods, this translates to a more computationally expensive “emitter-to-observer” paradigm since this is how the photon diffusion process works in nature (see Laor et al. 1990; Kojima 1991; Dolence et al. 2009; Schnittman & Krolik 2013 for a few recent codes that follow this philosophy). Grid based methods instead require a fully 3D treatment of the radiative problem accounting for all the nonlocal scattering terms in the emissivity profile. In the case of HEROIC, this is achieved by solving for the complete 3D scattered radiation field everywhere around and
inside the disk before the raytracing process is initiated.

The organization of this paper is as follows: in §5.2 we describe how HEROIC handles Comptonization; specifically, we explain how the ray evolution equation works in accordance to the Kompaneets based approach that we have taken. §5.3 follows up with a series of 1D and 3D benchmark tests to verify the correct operation of our code. Finally, we present in §5.4 an application of HEROIC to a couple of general relativistic magnetohydrodynamic (GRMHD) disk simulations. We find that the GRMHD based 3D spectra are slightly hotter than previous 1D estimates.

### 5.2 Radiative Transfer Solution

The approach HEROIC uses for solving radiative transfer is the method of characteristics (see Chapter 4 for an extensive discussion). The intensity field is evolved along a set of discrete rays according to the radiative transfer equation:

\[
\frac{dI}{d\tau} = -I + S, \tag{5.1}
\]

where \( I \) is the radiation intensity along the ray, \( \tau \) is the optical depth, and \( S \) is the source function that governs the rate at which energy is reintroduced back into the beam, accounting for both intrinsic thermal emission and scattering. Specifically, we use

\[
S = \frac{\alpha B + \sigma J}{\alpha + \sigma}, \tag{5.2}
\]

where \( \alpha \) and \( \sigma \) are the absorption and scattering opacities respectively, \( B \) is the thermal Planck function, and \( J \) is the radiation mean intensity averaged over \( 4\pi \) solid angle.

Notice the implicit coupling between the source function and the radiative transfer equation. Ray intensities are updated according to the local spatially varying source
function; however, the source function itself depends on the radiation field through its scattering term. Solving this highly coupled nonlinear system of equations is a difficult task – in practice iterative techniques are used to find the solution.

Lambda iteration is perhaps the most common approach for solving radiative problems. Formally solving the radiative transfer equation (Eq. 5.1) yields the following relationship between the local mean angle-averaged intensity field $J$ given some fixed spatially dependent source function, i.e.

$$J = \Lambda(S),$$

(5.3)

where $\Lambda$ represents the coupling operator that links the locally defined source function as a function of position to the locally measured $J$. Then, one can update $S$ with the new scattering term reflecting the change in $J$. This two-cycle process is then iterated to convergence. One major hurdle is the iterative nature of this approach – for some problems, e.g., whenever scattering dominates the opacity, the system converges at an impractically slow rate, requiring $\sim \tau^2$ iterations to resolve the full random walk diffusion process.

The solution to slow convergence is an approach called “Accelerated Lambda Iteration” (ALI), which seeks to boost the source function update accounting for the iterative coupling between $J$ and $S$. Formally, the acceleration is accomplished by inverting the $\Lambda$ operator and solving for the resultant update to the source function (Hubeny 2003). In practice, the $\Lambda$ matrix is too large and difficult to invert directly, and instead an approximate operator $\Lambda^*$ is used, typically taken as the diagonal component of the $\Lambda$ operator, motivated by the fact that the diagonal terms usually dominate the system since they represent the self illumination contribution to the source term.
Folding everything together, the result from ALI is to boost the source function update as follows:

\[
\Delta S_{ALI} = \frac{\Delta S_{FS}}{1 - (1 - \epsilon)\Lambda^r},
\]

where \(\Delta S_{FS}\) represents the unaccelerated update to \(S\) from the ordinary lambda iteration approach, \(\Delta S_{ALI}\) is the accelerated update according to the ALI scheme, and \(\epsilon = \alpha/(\alpha + \sigma)\) represents the fraction of the absorptive opacity to the total opacity. In general, acceleration is necessary for problems where scattering is dominant, particularly whenever \(\epsilon \ll 10^{-2}\). In the case of Comptonized rays, the acceleration scheme becomes more complex and is described in the next section (§5.2.1).

### 5.2.1 Kompaneets-Ray

The inclusion of Comptonization is handled through a Kompaneets operator approach whereby we modify the source function to have a Compton term. This approach has been successfully tested and applied in other codes (Pomraning 1973; Madej 1989; Hubeny et al. 2001; Peraiah et al. 2010), with some implementations being more sophisticated than ours due to differences in the choice of interpolation scheme. We employ the most basic representation of Compton scattering, viz., we approximate it as an isotropic redistribution of photons in the comoving frame of the fluid.

Specifically, we modify the definition of the source function in Eq. 5.2, replacing the scattered radiation field \(J\) with a newly defined \(J^{\text{compt}}\) that accounts for the effect of Comptonization such that the source function now becomes:

\[
S = \epsilon B + (1 - \epsilon)J^{\text{compt}}.
\]

Here, \(J^{\text{compt}}\) represents the local spectrum of the radiation field after a single scattering
event. To compute this, we turn to the Kompaneets equation which describes the evolution of the mean intensity field $J$ as a function of the number of scatterings. Specifically, we evolve the associated photon number density $n_\nu \sim J_\nu/\nu^3$ via the Kompaneets diffusion equation:

$$\frac{dn}{dt_{\text{scat}}} = x^2 \frac{d}{dx} \left( \frac{1}{x^2} \left[ n^2 + n + \frac{dn}{dx} \right] \right), \quad (5.6)$$

where $t_{\text{scat}}$ measures the characteristic time of the system in units of the number of scattering events, and $x = h\nu/kT_{\text{gas}}$ is the dimensionless frequency of the radiation field. In our case, we are interested in the radiation field after a single scattering event. This corresponds to solving Eq. 5.6 for the photon distribution $n_\nu$ for $\delta t_{\text{scat}} = 1$.

Due to the stiff nature of Eq. 5.6, we solve the system using an implicit scheme following the approach of Chang & Cooper (1970). The system is discretized along a logarithmic frequency grid, and the photon fluxes in neighboring frequency bins are chosen such that the expected quasiequilibrium state for the given gas temperature and total photon number remains stationary. This guarantees convergence towards the expected thermal photon statistics, and guards against instabilities that can arise from the stiffness of the PDE. Additionally, since in the Chang & Cooper (1970) approach we are solving a diffusion equation, the coupling is only between neighboring frequency bins, which results in a simple tridiagonal system that is easy to invert using standard methods. The boundary conditions are simply zero photon flux at the lower and upper frequency boundaries to ensure conservation of photon number, i.e.

$$0 = F|_{x_{\text{BC}}} = \frac{d}{dx} \left( n^2 + n + \frac{dn}{dx} \right)|_{x_{\text{BC}}}. \quad (5.7)$$

One important detail is that we separate the Kompaneets update to $J$ and hence $S$ from the ALI update to $S$. Comptonization is designed to handle frequency redistribution
CHAPTER 5. 3D COMPTONIZED ACCRETION DISKS

of photons, which is a completely separate process from the random walk diffusive nature of light propagating in a scattering medium. By design ALI does not account for the frequency dependent nature of the Compton redistribution kernel and so it cannot be expected to provide any boost on the Compton update. In fact, it can even lead to overshooting and instability in some problems. To avoid this issue altogether, we have carefully separated out the ALI and Compton kernel steps as described in the procedure outlined below:

- For fixed sources, solve the radiative transfer equation for each ray
- Update the local source functions according to the ALI scheme (to compensate for scattering)
- Apply the Kompaneets kernel and update $J$ and also $S$
- Repeat until convergence

5.2.2 Quadratic Variation of the Source Function

In practice, we find that the approach described in §5.2.1 only works well for optically thin regions, where the light ray experiences at most a single scattering event across a cell. Optically thick regions often involve multiple scattering events across each grid cell, necessitating the use of higher order schemes to accurately capture the correspondingly more complex ray evolution.

Based on the results of several numerical tests (see later sections for details), we find that a quadratic based method works well for capturing the Comptonization process in
the limit of optically thick cells. When evaluating the formal solution to the radiative
transfer equation, i.e. the solution to

\[ I(\tau) = I_0 e^{-\tau} + \int_0^\tau S(\tau') \exp(-\tau')d\tau', \]  

(5.8)

we choose to approximate \( S(\tau) \) with a quadratic Taylor-series decomposition:

\[ S(\tau) = S(\tau_0) + \frac{dS}{d\tau}(\tau - \tau_0) + \frac{1}{2} \left( \frac{d^2S}{d\tau^2} \right) (\tau - \tau_0)^2, \]  

(5.9)

where we estimate the first and second derivative terms, \( dS/d\tau \) and \( d^2S/d\tau^2 \), as explained
below.

In HERO’s short characteristic radiative transfer solver (see Ch. 4 for more details),
rays are evolved from cell boundaries to cell centers. The only information accessed are
the intensity value at the neighboring cell boundaries \( I_0 \) plus the boundary and local
source functions \( S_0 \), and \( S_\tau \) respectively. Essentially, we use \( S_0, S_\tau \) to determine the
zeroth and first order terms in Eq.5.9. The second order term is determined purely from
the Comptonization properties of the medium.

Consider a plane-parallel problem and let us work with the first two angular
moments of the radiative transfer equation,

\[ \mu \frac{\partial I}{\partial \tau} = I - S, \]  

(5.10)

where \( \mu = \cos \theta \). Multiplying Eq.5.10 by factors of \( \mu \) and \( \mu^2 \) and integrating yields the
usual radiative transfer equations:

\[ \frac{\partial H}{\partial \tau} = J - S, \]  

(5.11)

\[ \frac{\partial K}{\partial \tau} = H, \]  

(5.12)
where $H$ and $K$ are the radiative flux and pressure respectively. Making use of the Eddington approximation $K = J/3$ (which is equivalent to taking the two-stream limit, cf, Rybicki & Lightman 1977), and combining Eqs. 5.11 and 5.12, we obtain

$$\frac{\partial^2 J}{\partial \tau^2} = 3(J - S).$$ \hspace{1cm} (5.13)

Now we can relate $J$ to $S$ by considering our definition of the source function

$$S_\nu = \epsilon_\nu B_\nu + (1 - \epsilon_\nu)J_\nu^{\text{compt}} = \epsilon_\nu B_\nu + (1 - \epsilon_\nu)(1 + \eta_\nu)J_\nu$$ \hspace{1cm} (5.14)

where $\eta_\nu \equiv (J_\nu^{\text{compt}} - J_\nu) \sim (4kT_{\text{gas}} - h\nu)/(m_e c^2)$ represents the nonrelativistic radiation boost factor resulting from the Comptonization process. Hence from Eqs. 5.14 and 5.13, we get the following second-order dependence of the source function with optical depth:

$$\frac{d^2 S_\nu}{d\tau^2} = \epsilon_\nu \frac{d^2 B_\nu}{d\tau^2} + (1 - \epsilon_\nu)(1 + \eta_\nu) \frac{d^2 J_\nu}{d\tau^2}$$

$$= \epsilon_\nu \frac{d^2 B_\nu}{d\tau^2} + 3(1 - \epsilon_\nu)(1 + \eta_\nu)(J_\nu - S_\nu)$$

$$= \epsilon_\nu \frac{d^2 B_\nu}{d\tau^2} + 3 \left\{ \epsilon_\nu - \eta_\nu (1 - \epsilon_\nu) \right\} S_\nu - \epsilon_\nu B_\nu.$$ \hspace{1cm} (5.15)

Note that the definition of $\tau$ here corresponds to the vertical optical depth of the plane-parallel slab. For oblique rays, $d\tau_{\text{ray}} = d\tau/\mu$, and hence the second derivative as defined for the ray must be suppressed by an additional $\mu^2$ factor relative to Eq. 5.15.

Equation 5.15 allows us to estimate $d^2 S_\nu/d\tau^2$ for each ray. For the pure scattering case ($\epsilon_\nu = 0$) in the absence of Comptonization ($\eta_\nu = 0$), the result for $d^2 S_\nu/d\tau^2 = 0$, implying that the variation of the source function is typically linear with optical depth. However when Comptonization acts, this second derivative term is nonzero, and it is necessary to use a quadratic based solver to account for the radiation curvature.

Combining everything together, we can explicitly evaluate Eq. 5.8 by plugging in Eq. 5.9. This yields the following expression for updating the rays in the short
CHAPTER 5. 3D COMPTONIZED ACCRETION DISKS

classical method:

\[
I(\tau) = [S_{\tau} + S' + S''] - \left\{ S_{\tau} + S'(1 + \tau) + \frac{1}{2} S'' [2. + \tau(2 + \tau)] - I_0 \right\} e^{-\tau}; \quad (5.16)
\]

where \( S'' \) is given by Eq. 5.15 suppressed by the \( \mu^2 \) factor as discussed above, and \( S' \) is set by the boundary value constraints for \( S \) in Eq. 5.9, i.e.:

\[
S' = \frac{S_{\tau} - S_0}{\Delta \tau} - \frac{1}{2} \left( \frac{d^2 S}{d\tau^2} \right) (\Delta \tau). \quad (5.17)
\]

5.3 Numerical Tests

We validate our Compton module in HEROIC by comparing with a suite of particularly simple test problems that admit analytic solutions. For all the test problems, we consider the case of pure scattering with no absorption.

5.3.1 Kompaneets

First, we would like to benchmark our Chang & Cooper (1970) based Kompaneets solver since it is the workhorse that we rely on to handle all Comptonization problems in HEROIC. The problem that we consider is that of photon evolution within a closed box of hot thermal electrons. The simplest problem of this sort is the “Green’s function” problem, whereby a monoenergetic beam of photons (delta-function in \( \nu \)) is injected into the box and allowed to evolve with time.
Figure 5.1: A comparison between the numerical Kompaneets solver employed by HERO (see §5.2.1 for details) and the exact analytic solution (see Eq. 5.18) for the diffusion in energy of an initial delta-source peaked at $x = h\nu/kT = 1$. The different colors correspond to different choices of the Compton $y$-parameter: 0.1 (blue), 1 (green), 10 (red), 100 (black). The slight differences between the exact and our numerical distributions are due to inaccuracies in numerically representing and diffusing a delta-function. Initially the delta source is just a single nonzero bin, and thus the finite bin size slightly fattens the numerical Kompaneets result.

If we ignore the nonlinear $n^2$ term in the Kompaneets equation, the problem is simple enough that an analytic solution for the diffusion process is available. We compare the numerical photon redistribution from our code with the analytic result involving the
complex integrals of Whittaker functions $W_{k,m}(x)$ as worked out by Becker (2003), i.e.

$$f_G(x, y) = \frac{32}{\pi} e^{-9y/4} \frac{x^2}{x^2 - 2} e^{(x_0 - x)/2} \int_0^\infty e^{-u^2 y} \frac{u \sinh(\pi u)}{(1 + 4u^2)(9 + 4u^2)} W_{2, iu}(x_0) W_{2, iu}(x) du$$

$$+ e^{-x} + \frac{e^{-x - 2y}}{2} \frac{(2 - x)(2 - x_0)}{x_0 x},$$

where $x = h\nu/kT$ is the dimensionless frequency for the problem, $f_G(x)$ represents the Green’s function spectral response of the system to a delta source injected at frequency $x_0$ after having evolved over a timescale corresponding to a given Compton $y = (4kT/mc^2)n_{scat}$ parameter. To compare with our numerical code, we set up the diffusion problem such that $x_0 = 1$ (i.e. we initialize HEROIC with a delta function photon distribution peaked at $x = 1$), and we evolve the radiation field according to the scheme as described in §5.2.1. In general, we find excellent agreement between HEROIC and the analytic solution, with the biggest discrepancies occurring when there are a moderate number of scatterings (i.e. when the Compton y-parameter is $0.1 - 1$).

These minor discrepancies are primarily induced by inaccuracies in numerically representing and diffusing a delta-function. Initially the delta source is just a single nonzero bin, and thus the finite bin size slightly fattens the numerical Kompaneets result. Figure 5.2 shows the effect of frequency bin size on the limiting photon distribution slopes. Having insufficient frequency resolutions results in more numerical diffusion, which causes extra broadening of the photon distribution (compare widths of the dotted, dashed, and solid curves in Fig 5.2).
Figure 5.2: Here we compare the numerical result from our Kompaneets solver under various frequency resolutions. The dotted, dashed, and solid lines correspond to 10, 40, 160 bins per frequency decade respectively. The thick solid line denotes the analytical solution as computed from Eq. 5.18.

5.3.2 Escape Time Distributions

We would like to examine and benchmark HEROIC with the more difficult problem of photon diffusion and Comptonization within a scattering atmosphere. The simplest setup is that of a plane parallel atmosphere. Our goal is to compute the spectrum of photons that escape from the upper/lower surface of the atmosphere, given that there is a constant flux of photons injected at the midplane.

Since we have a robust numerical Kompaneets solver that calculates the spectrum
for any given initial photon distribution and Compton-y parameter, we simply need to compute the photon escape time distribution. A convolution of the photon escape time distribution with the Kompaneets derived spectrum for each given escape time then yields the emergent spectrum for the problem. Figure 5.3 shows an example of the spectral evolution of a $T_{rad} = 10^4 K$ initial radiation field evolving in a $T_{gas} = 10^6 K$ gas. This set of spectra constitutes the “Compton kernel” which we will use for calculating the emergent spectrum from a Comptonizing atmosphere.

**Figure 5.3:** An example Compton kernel as computed by our Kompaneets solver. The problem considered is the Comptonization of an initial $T = 10^4 K$ radiation field that is upscatted by hot gas with $T = 10^6 K$. The different curves show Comptonized spectra after various number of scattering events $n_s$. 
CHAPTER 5. 3D COMPTONIZED ACCRETION DISKS

Plane Parallel Scattering Atmosphere

To get the escape time distribution for a 1D plane-parallel atmosphere, we simply solve the diffusion equation, i.e.

\[
\frac{dn}{dt} = D \frac{d^2n}{dx^2},
\]

(5.19)

where \( D = \frac{1}{3} \) is the diffusion constant corresponding to a three-dimensional diffusion process\(^2\) in a homogeneous medium, \( x \) is the distance as measured in units of the scattering mean free path, and \( t \) is measured in units of the number of scattering times.

The problem we seek to model is the process of photon diffusion from a midplane injection point to an outer boundary surface. Specifically, we wish to determine the probability distribution of the photon escape times. This quantity can be measured from the diffusion equation as the escaping photon number flux at the surface. Instead of taking the approach of Sunyaev & Titarchuk (1980) who tackle the problem analytically via series expansions, we opt for a numerical solution.

We employ the same tri-diagonal implicit diffusion approach as we used previously for numerically solving the Kompaneets problem. We only consider the upper half portion of the slab, and make use of the symmetry of the problem through a reflecting boundary condition at the midplane. The system is seeded with a spatial delta function photon source at the midplane and we solve for the time/space evolution of photons. The surface boundary is handled by setting purely outgoing radiation flux, which we take from the two-stream framework of the radiative transfer equation:

\[
\frac{dJ}{dt} \bigg|_{\text{surf}} + \sqrt{3} J_{\text{surf}} = 0.
\]

(5.20)

\(^2\)The coefficient \(1/3\) is also compatible with the two-stream approximation for radiative transfer.
CHAPTER 5. 3D COMPTONIZED ACCRETION DISKS

Figure 5.4 shows the resultant distribution of escape times. Note that in the optically thin limit, the escape time roughly scales as $\tau$, whereas in the optically thick limit the escape time scales as $\tau^2$, reflecting the random walk nature of the diffusion process.

![Figure 5.4](image)

**Figure 5.4:** The escape time probability distribution function for various choices of plane parallel slab thickness. Here, the escape time is measured in units of the characteristic scattering time. Note that the scaling of diffusion time with optical depth changes from $t \sim \tau$ in the optically thin limit to $t \sim \tau^2$ in the optically thick limit.

Finally, we arrive at the emergent spectrum by convolving the escape probability distribution (Fig. 5.4) with the Kompaneets spectral data cube (Fig. 5.3). In Fig. 5.5 we compare this convolution-based benchmark solution (solid), which we consider to be the true solution, with the result from our HEROIC code, both in the 2-stream limit (dotted) and for a multiangle calculation involving 20 rays uniformly distributed in
cosine (dashed). The agreement is quite good (deviations at the 10% level), with the largest errors arising at moderate optical depths ($n_{\text{scat}} \sim 30$).

For the above problem, we set up HEROIC with a purely scattering plane-parallel atmosphere with 100 vertical depth points spaced equidistant in logarithm, with $\tau_{\text{min}} = 10^{-3}$. The code was run with a frequency resolution of 20 points per decade. The midplane was treated as a reflecting boundary with an additional steady source of blackbody radiation with effective temperature $10^4 K$ coupled with a reflecting boundary. At the upper surface of the atmosphere, the boundary condition was that there is no incident radiation flux.

\footnote{The spectral shape for the injected flux is determined by the Planck function for $T = 10^4 K$}
Figure 5.5: Comparison of emergent surface spectra for the plane parallel pure Compton problem. Colors indicate the total optical depth to the midplane of the slab: $\tau = 1$ black, $\tau = 10$ blue, $\tau = 30$ green, $\tau = 100$ red. For this problem, seed photons are injected at the slab midplane with a $T = 10^4 K$ flux and color temperature, which then diffuse outwards through a homogeneous purely scattering atmosphere at electron temperature $T = 10^6 K$. The solid lines show the 1D HEROIC result using the 2-stream approximation and the dashed curves correspond to a multiray 1D calculation.

3D Spherical Scattering Atmosphere

We now repeat the same exercise, except that we invoke 3D geometry. Consider the problem of a spherical gas cloud where photons are injected at the center and allowed to slowly diffuse outwards. This problem is identical to the plane parallel one, except that
CHAPTER 5. 3D COMPTONIZED ACCRETION DISKS

it now occurs in spherical geometry. The modified diffusion equation is:

\[
\frac{dn}{dt} = 1 \frac{d}{r^2} \frac{d}{dr} \left( Dr^2 \frac{dn}{dr} \right).
\] (5.21)

Since we are still modeling a 3D diffusion process, we have \( D = 1/3 \), the same as in 1D. For numerical stability, we make use of logarithmic radial coordinates in evaluation of the above diffusion equation. We inject photons at \( r_{\text{min}} = 10^{-4} \) and measure their escape times at an outer boundary \( r \gg 1 \), which simulates the process of central injection of photons. The escape time probability distribution is measured by the outgoing flux at the outer surface. The central boundary condition is still set as reflection due to the symmetry of the problem, while the outer boundary condition becomes modified to (see Sunyaev & Titarchuk 1980 – Appendix A):

\[
\frac{dJ}{d\tau} \bigg|_{\text{surf}} + \frac{3}{2} J_{\text{surf}} = 0.
\] (5.22)

The dotted and dashed lines in Figure 5.6 show the Comptonized spectra computed from a 3D-axisymmetric HEROIC calculation. For Figure 5.6, the HERO solution curves have been renormalized to match the flux of the analytic solution since the short-characteristic derived flux in spherical geometry is biased due to the presence of ray defects (see discussion in Chapter 4). Generally, we find that 3D spherical diffusion has on average shorter escape times compared to an equivalent 1D plane parallel problem with the same scattering depth. This is a simple consequence of the geometry. Imagine a spherical shell inscribed within an equivalent thickness plane parallel slab. Then for a photon injected at the center of the sphere, the distance to the surface of the slab is \( x_{\text{ray,slab}} = r_{\text{shell}}/\cos \theta \) and hence \( x_{\text{ray,slab}} > r_{\text{shell}} \) for all ray directions. The shorter escape time probabilities in the spherical case translate to less Comptonized spectra as seen by comparing the solid lines in Figure 5.6 with the corresponding curves in Figure 5.5.
CHAPTER 5. 3D COMPTONIZED ACCRETION DISKS

Figure 5.6: Same as figure 5.5, but for the 3D spherical diffusion problem. Color indicates the optical depth to the center of the scattering sphere: \( \tau = 1 \) black, \( \tau = 10 \) blue, \( \tau = 30 \) green, \( \tau = 100 \) red. Note that these spectra are systematically less Comptonized than their plane parallel counterparts shown in 5.3.

The numerical setup for the 3D problem was a \((n_r, n_\theta) = (60, 30)\) axisymmetric grid uniformly spaced in \(\log r\) and \(\theta\), with frequency resolution set to 10 points per decade, and using an angular grid with \(n_a = 80\) rays. Boundary conditions were simply reflecting along the polar axis (to account for axisymmetry), constant flux thermal injection with reflection on ingoing rays at the lower radial boundary (i.e. the same setup as used in 1D), and zero incident flux at the outer boundary where the final spectrum is measured. To be consistent with the choice of constant diffusion coefficient in Eq. 5.21, the scattering sphere was taken as a constant density, constant opacity, homogeneous
CHAPTER 5. 3D COMPTONIZED ACCRETION DISKS

object.

5.4 Application – Accretion Disk

Most of the information that we have for black holes comes from observations of their surrounding accretion disks. Through modelling the X-ray continuum of black hole binary sources, it is possible to deduce the structure of the accretion disk, which can then serve as an indirect probe of the physical properties of the black hole. Unfortunately, this spectral modeling task is complicated by the fact that the accretion disk is scattering dominated and hot. Scattering and Comptonization in the disk interior causes the emerging radiation from the surface of the disk to have a “diluted” blackbody spectrum with a color temperature given by

\[ T_{\text{col}} = f T_{\text{eff}}, \]  \hspace{1cm} (5.23)

where \( f \) is the color correction factor and \( T_{\text{eff}} \) is the effective temperature related to the local disk flux. A key issue then is the determination of \( f \) since it determines the shape of the resultant disk spectrum.

The earliest models typically assumed constant color correction factors \( f \sim 1.5 \) (Mitsuda et al. 1984; Zhang et al. 1997a,b). More recently, much effort has gone towards pinning down \( f \) in the case of X-ray binaries (Shimura & Takahara 1995; Merloni et al. 2000; Davis & Hubeny 2006) via sophisticated radiative transfer calculations. However all prior work on determining \( f \) relied upon simplified 1D treatments of the problem. Here, we present the first attempt at a 3D solution of the Compton disk problem that self-consistently accounts for all relativistic and returning-radiation effects.
CHAPTER 5. 3D COMPTONIZED ACCRETION DISKS

5.4.1 Numerical Disk Setup

For our analysis, we make use of the GRMHD thin disk simulations of Penna et al. (2010) and Sadowski et al. (2015a) to set the gas density, velocity and viscous heating rate in the accretion disk. HEROIC solves for the time-stationary radiation field, therefore all simulation quantities extracted correspond to their time and azimuthally averaged values. We dimensionalize all quantities from the GRMHD data by considering the case of a $10 M_\odot$ black hole accreting at $\sim 60\%$ Eddington. This value for $L/L_{\text{Edd}}$ was set such that the simulation disk thickness $h/r \sim 0.1$ matches the theoretical expectation in the Novikov & Thorne disk model. We also extract an integrated luminosity profile from the GRMHD simulations using the same technique as described in Zhu et al. (2012). Further details of the various simulations analyzed are listed in table 5.1.

We then feed into HEROIC the 3D simulation density, velocity, and energy injection information for the radiative transfer calculation. This data is further rebinned from the original GRMHD data to a new grid with $(n_r, n_\theta) = (30, 64)$ that is uniformly spaced in $\log(r)$ but preserving the same $\theta$ structure as in the original GRMHD grid. Both the inner and outer radial grid boundaries are treated with outflow conditions, whereas the polar boundaries are given reflecting boundary conditions to account for axisymmetry. We set the frequency resolution to be 10 points per decade with $\nu_{\text{min}} = 10^{15}\text{Hz}$ and $\nu_{\text{max}} = 10^{21}\text{Hz}$, and we make use of $n_A = 80$ rays in angular space. The opacities are taken to be free-free bremsstrahlung for absorption,

$$\left(\frac{\alpha_{\nu}^{ff}}{\text{cm}^{-3}}\right) = 1.34 \times 10^{56} \left(\frac{T}{\text{K}}\right)^{-1/2} \left(\frac{\rho}{\text{g/cm}^3}\right)^2 \left(\frac{\nu}{\text{Hz}}\right)^{-3} (1 - e^{-h\nu/kT}), \quad (5.24)$$
and Thomson for scattering:

\[ \left( \frac{\sigma_{es}}{\text{cm}^{-1}} \right) = 0.4 \left( \frac{\rho}{\text{g/cm}^3} \right) . \]  (5.25)

In the accretion disk environment where typically we have \( T \sim 10^7 K \) and \( \rho \sim 10^{-4} \text{g/cm}^3 \), scattering dominates over absorption opacity by a factor of \( \sim 100 \).

This extremely high degree of scattering within the accretion disk requires a more sophisticated temperature solver than the simple minded local radiative equilibrium approach described in Ch. 4. We instead turn to an Unsold-Lucy inspired scheme for solving the disk temperatures. We start with the three-dimensional radiative moment equations

\[ \nabla \cdot H = -\frac{Q_{\text{inject}}}{4\pi} , \]

\[ \nabla K = \chi H , \]  (5.26)

where \( Q_{\text{inject}} \) is the local energy injection per unit volume, and \( \chi = \alpha + \sigma \) is the total absorption and scattering opacity. Combining Eqs. 5.26 with the Eddington approximation \( K = J/3 \) yields the second order Poisson’s equation for the radiation field

\[ \nabla^2 J = -\frac{3\chi Q_{\text{inject}}}{4\pi} . \]  (5.27)

Next, we convert the second order equation for radiation into one for temperature by making use of the radiative equilibrium condition

\[ 4\pi \int \alpha_{\nu}(J_{\nu} - B_{\nu})d\nu + Q_{\text{inject}} = 0 . \]  (5.28)

Hence, \( \nabla^2 J \approx \nabla^2 B - Q_{\text{inject}}/(4\pi\alpha_J) \) where \( \alpha_J = \int \alpha_{\nu}J_{\nu}d\nu/\int J_{\nu}d\nu \). This leads to a simple Poisson equation for the thermal source \( B(T) \):

\[ \nabla^2 B = -\frac{3\chi Q_{\text{inject}}}{4\pi} + \frac{\nabla^2 Q_{\text{inject}}}{4\pi\alpha_J} . \]  (5.29)
which can be solved iteratively for the temperature. The resulting solver is robust and is fairly insensitive to the behaviour of the radiation field. The advantage of the Poisson approach is that given some fixed distribution of $Q_{\text{inject}}$ on the RHS of Eq. 5.29, we can iteratively solve for $\nabla B$ using a simple implicit Poisson solver (we are essentially just solving the heat equation in the presence of sources/sinks of energy). The radiation coupling only modifies the $Q_{\text{inject}}$ term in Eq. 5.29 via the Compton heating/cooling process.

For numerical stability, we operate HEROIC by cycling between our temperature and radiation solvers instead of solving all quantities simultaneously. Typically, we run 20 iterations of ALI + Comptonization for pinning down the radiation field followed by 20 iterations of the temperature solver. We also allow the code to fall back on more robust radiative and Compton solvers if any error is detected such as negative light intensities. Problematic rays are handled by switching to ordinary lambda iteration as well as defaulting to a more basic version of the radiative evaluator that does not account for the any second derivative curvature in the source function.

5.4.2 Results

In total, we have run HEROIC with 3 different disk setups, as described in table 5.1. For simplicity, we only consider Schwarzchild black holes ($a_s = 0$), and we investigate how the temperature and spectral properties of the disk depend on the luminosity profile chosen (the analytic NT model, vs the dissipation profile from simulations). Convergence in the radiative and temperature solutions were typically achieved after $\sim 3000$ iterations for all runs. The results presented below correspond to the solution after 5000 iterations.
Table 5.1: HEROIC disk parameters

<table>
<thead>
<tr>
<th>Model</th>
<th>Luminosity $L/L_{Edd}$</th>
<th>Energy Injection†</th>
<th>Basis of Model‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.7</td>
<td>$\rho^3$</td>
<td>NT</td>
</tr>
<tr>
<td>B</td>
<td>0.7</td>
<td>$\rho^3$</td>
<td>GRMHD</td>
</tr>
<tr>
<td>C</td>
<td>0.6</td>
<td>sim</td>
<td>GRMHD+Rad</td>
</tr>
</tbody>
</table>

† The prescription used to translate the total luminosity into a 3D energy injection distribution. For models A and B, we adopt a centrally concentrated injection distribution that loads the energy according to $\rho^3$ in each annulus. Model C simply takes the radiation-gas net heating rate directly measured from the simulation.

‡ NT – Novikov Thorne analytic energy injection profile; GRMHD – based on numerical disk simulations of Penna et al. (2010); GRMHD+Rad – based on numerical thin disk simulations of Sądowski et al. (2015a)

Figure 5.7 shows the temperature distributions in the poloidal plane for the three runs. We find in all cases that a hot optically-thin spherical corona forms around the disk, and that the corona is hottest close to the black hole. In model B, we also find that the final equilibrium temperature of the corona depends sensitively on the amount of
energy injection present in the corona. The base GRMHD disk model (Penna et al. 2010) lacks any treatment of radiation, and so the derived temperature for the optically thin coronal region is likely incorrect since it depends sensitively on gas heating distribution. The temperatures of Model C on the other hand come directly from another GRMHD code that includes a simplified treatment of radiation (Sadowski et al. 2015a). The extreme temperatures seen along the polar axes of model C are a consequence of strong radiative jets that are present in this disk simulation.

In Figure 5.8, we compare the disk temperature profile as computed by HEROIC with the 1D radiative transfer calculation of TLUSTY. In both cases, the temperature profile within the optically thick disk takes on a characteristic \( T \propto \tau^{1/4} \) form, where \( \tau \) is the optical depth from the disk surface. The primary difference between TLUSTY and HEROIC occurs above the disk’s scattering effective photosphere – the temperature profile in HEROIC tends towards isothermality at all locations near the surface. This is partly due to the 3D propagation of radiation, specifically the effect of returning radiation from the disk. This incoming flux can penetrate downwards through the cool photospheric surface to isothermalize via Compton heating of the local gas to match the temperature of the incoming radiation.

Some differences between the 3D GRMHD result and the 1D TLUSTY calculation are also due to differences in the mass distribution within the disk. TLUSTY recomputes

---

4In model B, we find that the coronal temperatures can vary by up to factors of 3, depending on the exact energy injection prescription used (e.g. whether \( Q_{\text{inj}} \propto \rho \) or \( Q_{\text{inj}} \propto \rho^3 \))

5Note: TLUSTY and HEROIC have slightly different setups for the vertical structure of energy injection – TLUSTY injects energy within the vertical column according to a \( \rho^3 \) prescription, whereas in our computations with HEROIC, we used \( \rho^3 \).
CHAPTER 5. 3D COMPTONIZED ACCRETION DISKS

the vertical structure of the disk accounting only for the gas and radiation pressure. However the action of magnetic pressure support within the GRMHD simulations tends to puff up the disk, producing structures with more mass loading at large scale heights.

![Figure 5.7: Comparison of 3D temperature profiles for the 3 runs listed in Table 5.1. From left to right, the three panels correspond to models A, B, and C respectively.](image)

We show the integrated disk spectra for the 3 runs in Figure 5.9. These spectra are computed via long characteristic raytracing from a distant observation plane located at $r = 100,000M$, and we only consider emission from the inner $r < 20M$ region of the accretion disk. In general, we find that the 3D radiative transfer calculation results in somewhat hotter spectra with larger hardening factors when compared to 1D models (see Figure 5.11 for a comparison). This effect can be largely attributed to 1) hot radiation from the corona irradiating the disk and modifying the photospheric temperatures to be

270
CHAPTER 5. 3D COMPTONIZED ACCRETION DISKS

hotter, and 2) disk returning radiation (primarily, the hot lensed radiation emitted close to the black hole) that scatters off the disk photosphere. In our hottest spectra (model C), the hardening factor as compared to pure blackbody annuli is $f_{col} \sim 1.79^6$.

![Comparison of TLUSTY and HEROIC vertical temperature profiles](image)

**Figure 5.8:** Comparison of TLUSTY and HEROIC vertical temperature profiles for various radii (4,8,16,32) in model B. The differences at low column mass are due to disk irradiation isothermalizing the surface layers.

---

6This hardening factor is estimated by the ratio of spectral centroids as defined by $\nu_c = \int \nu I_\nu d\nu / \int \nu d\nu$. Specifically, we take $f = \nu_c^{\text{sim}} / \nu_c^{\text{BB}}$, where $\nu_c^{\text{sim}}$ and $\nu_c^{\text{BB}}$ are the simulation and blackbody spectral centroids respectively.
Figure 5.9: Comparison of integrated disk spectra for the inner disk ($r < 20$) as computed by HEROIC for the three accretion disk models. The top panel shows spectra for a face-on observer ($i = 0$), whereas the right panel shows spectra for an inclined ($i = 60^o$) observer.
Figure 5.10: A plot of the individual annuli spectra for model C. The annuli spectra shown (from hottest to coldest) correspond to radii \( r = 2.8, 3.6, 4, 5.5, 8, 13, 20 \). To better show the shape of the locally emitted spectra, we have ignored relativistic doppler and gravitational broadening effects in this plot.

Interestingly, for the hottest disk (model C), we find that the high energy region of the spectrum exhibits a power law tail. Figure 5.10 decomposes the spectrum into contributions from individual radial annuli. Consistent with Zhu et al. (2012); Dexter & Blaes (2014), we find that the power law tail is due to emission from a range of radii inside the plunging region. One difference between this work and that of Zhu et al. (2012) is that the steep power law is present even in a nonspinning \( (a_* = 0) \) disk model whereas previously it was seen only for rapidly spinning black holes. The difference is due to the strong nonlocal heating action of the corona on the inner plunging region gas.
CHAPTER 5. 3D COMPTONIZED ACCRETION DISKS

At high luminosities, radiation-dominated disks are expected to become effectively optically thin. When coupled to the fact that the inner radii still produce a significant amount of emission (e.g. Noble et al. 2009), we find that this leads to a rapid increase in gas temperature in the inner radii resulting in a high-energy powerlaw spectral tail from the superposition of nearly saturated Wien spectra.

We note that this thermal superposition mechanism for producing high energy power-laws is different from another earlier idea known as bulk Comptonization (Blandford & Payne 1981; Titarchuk, Mastichiadis & Kylafis 1997; Titarchuk & Zannias 1998), which also predicts the onset of power laws. For bulk Comptonization, relativistic boosting from bulk gas motions within a converging flow can produce power law tails by upscattering the incident disk light. The bulk Comptonization mechanism is extremely sensitive to the nature of the irradiating thermal light (Psaltis & Lamb 1999), and also has a strong viewing inclination dependence (Niedźwiecki 2005). In our calculations, we do not find any change in the power law behaviour at different viewing inclinations (i.e. compare how similar the power law slope and strengths are in Fig. 5.9 for two different inclinations), so it is unlikely that the HEROIC power law region is sourced by bulk Comptonization.

5.5 Discussion

Recently, there has been great interest in fitting X-ray spectral observations of black holes for the purpose of measuring black hole spin (Zhang et al. 1997a; Shafee et al. 2006; Steiner et al. 2009a; McClintock et al. 2011). The ‘continuum fitting method’ relies upon accurate modeling the soft X-ray thermal continuum emission. In particular, the
CHAPTER 5. 3D COMPTONIZED ACCRETION DISKS

color correction factor (c.f. Eq. 5.23) is needed to translate models of the accretion disk into models of the observed disk spectra. For decades, this color factor was estimated either simply through ad-hoc observational arguments (Mitsuda et al. 1984; Zhang et al. 1997a,b), or from sophisticated 1D radiative transfer calculations (Davis & Hubeny 2006). For reference in Fig. 5.11, we show the 1D-based results for the hardening factor as a function of black hole spin and luminosity.

Figure 5.11: A plot of the overall disk hardening factor from 1D models (using the $\alpha = 0.1$ results from BHSPEC – Davis & Hubeny 2006) as a function of black hole spin and mass accretion rate. We indicate with a dotted horizontal line the hardening factor ($f = 1.79$) as measured from model C (a nonspinning 60% Eddington accretion disk). The upper 60% Eddington benchmark curve (dashed) is computed by extrapolating the hardening factors measured from the highest accretion rate disk models available within the BHSPEC grid.
CHAPTER 5. 3D COMPTONIZED ACCRETION DISKS

The calculations in §5.4.2 of this work allow for perhaps the best estimation of the hardening factor to date (i.e. a hardening factor based on numerical simulations of black holes with 3D radiation transport). Our 3D result (Fig. 5.9) yields a slightly higher overall hardening factor at $f = 1.79$ for a nonspinning black hole accreting at 60% Eddington. We see in Figure 5.11 that this new calculation is not too different from the previous 1D estimates.

Of particular interest is the impact that the new spectral hardening factor has on the recovered spins of black hole systems. Unfortunately, due to the high accretion rates of our simulated disks ($L/L_{Edd} \sim 0.6$), the disk spectrum is outside the parameter ranges of the currently used X-ray disk models. Rather than directly fitting our simulated spectrum with the current spectral models, we instead rely on the trend seen in Fig. 5.11 between the color correction factor and spin at a fixed $L/L_{Edd}$. We observe a minor discrepancy in $\Delta f \sim 0.05$ between 1D and 3D models, which when translated in terms of a spin bias corresponds to measuring $a_* = 0.3$ for a $a_* = 0$ system. If the black hole were highly rotating, then the same $\Delta f \sim 0.05$ would translate to a spin discrepancy of $(a_* = 0.88) \rightarrow (a_* = 0.83)$. This estimate for the spin bias is also consistent with the experiences of observers (Steiner 2015, private communication) when fitting disk models using fixed spectral hardening factors.

As mentioned earlier, it is the disk returning radiation that causes the additional hardening we see in 3D models. In figure 5.12, we compare both the emergent and returning fluxes as a function of radius. Typically, our HEROIC calculation yields larger

---

7Steiner (2015, private communication) typically finds that $\Delta f \sim 0.05$ translates to spin biases at the level of $(a_* = 0.1) \rightarrow (a_* = 0)$ or $(a_* = 0.87) \rightarrow (a_* = 0.83)$. 

276
returning fluxes compared to previous work with analytic disks (Cunningham 1976; Li et al. 2005). For instance HEROIC predicts the returning flux to be at the $\sim 10\%$ level, whereas figure 2 of Li et al. (2005) suggests it is only at the few percent level.

Disk flaring was ignored in the work of Li et al. (2005) and is likely the cause for HEROIC having such large returning radiation fluxes (see rightmost panel of Figure 5.7, model C, for the most obvious example of flaring in our simulated disks). For a flared disk, the additional flux scales as simply

$$F = \xi \frac{L_i}{4\pi r^2},$$

(5.30)

where $L_i$ is the characteristic luminosity of the interior irradiating source (i.e. the hot innermost annuli that shine on the rest), and $\xi = rd(h/r)/dr$ is a geometric scaling constant related to the flaring angle within the disk. If $L_i$ is large, then even tiny values of $\xi$ (i.e. small flaring angles) can still result in substantial irradiated flux at the flared surface.
CHAPTER 5. 3D COMPTONIZED ACCRETION DISKS

Figure 5.12: A plot of the vertical net ingoing and outgoing fluxes measured at the disk photosphere from the upper (red) and lower (blue) half planes of the disk in model C. Note that the returning inward flux scales roughly as $r^{-3}$.

A consequence of this large incident flux is a modification of surface temperatures in the atmosphere. In figure 5.8, we saw that HEROIC has a tendency to isothermalize the disk above the effective photosphere. Recent work by Kubota et al. (2010) supports this notion of isothermality in the photosphere with observations of black hole disk spectra. They find that the shape of the soft X-rays for real black holes never exhibits the absorption features associated with an inverted temperature atmosphere like those found in 1D calculations (c.f. TLUSTY solutions in Fig 5.8). Figure 5.13 shows that the HERO isothermalized temperature profile also produces spectra that are a better match to the observational data of Kubota et al. (2010). In other words, the lack of absorption
features indicates that the surfaces of real accretion disks are close to isothermal – in line with the HEROIC temperature solutions. The mechanism for isothermalization is likely Compton heating of the atmosphere where due to the irradiation of hot photons, the gas couples with the mean energy of the photon field, i.e. $\bar{E} \sim 4kT_{\text{gas}}$. This also explains why the isothermalization depth is the effective optical depth – the irradiating photons can only diffuse to the effective photosphere before being absorbed, losing all information about the external incident radiation field.

![Figure 5.13: A comparison of surface spectra for different temperature profiles corresponding to an annulus at radius $r = 32M$. The spectra were both calculated in 1D via TLUSTY, making use of the annuli boundary conditions. For the blue spectrum, we allow TLUSTY to compute temperatures completely self-consistently, which leads to an inverted temperature profile as seen in Fig. 5.8. Green corresponds to fixing the atmospheric temperature structure to match HEROIC (i.e. an atmosphere with isothermal surface temperatures).](image-url)
CHAPTER 5. 3D COMPTONIZED ACCRETION DISKS

5.6 Summary

In this work, we develop a new 3D relativistic Comptonization module within the framework of the radiative transfer code HERO. In this enhanced code HEROIC, Comptonization is handled through a quadratic solver that evolves light rays by means of a modified Comptonized source function. We test our Comptonization module by comparing with exact solutions of light diffusing through purely scattering media. These exact solutions are generated by consideration of the photon escape time probability distribution. We find excellent agreement between HEROIC and the exact solutions for both 1D and 3D test problems.

Next, we apply HEROIC to the astrophysical problem of radiation emerging from an accretion disk. We calculate the self consistent temperature and radiation fields throughout the accretion disk taking into account relativistic effects, returning disk radiation, and Compton cooling. The complete solution results in a two-phase disk structure, with a hot spherical corona surrounding a cooler thermal disk. Performing raytracing on the solution yields integrated disk spectra with slightly higher spectral hardening factors compared to previous 1D based estimates (i.e. we find $f = 1.79$ for our hottest disk models, compared to $f \sim 1.75$ in a 1D model).

We also verify the Zhu et al. (2012) result of a high-energy power law tail feature emerging in the black hole accretion disk spectrum. We find the power law arises from the superposition of saturated Wien spectra from the inner plunging region. Another feature of the 3D radiation calculation is the impact of returning radiation – namely, we find that the irradiation flux from the hot inner annuli tends to isothermalyze the surface layers of the disk, and is also responsible for the extra degree of spectral hardening.
present in our 3D models.

We stress that the results presented here are preliminary. In the future, we will extend our analysis to the case of spinning black holes, as well as more exotic states of black hole accretion such as the supereddington regime. A current limitation of HEROIC is its inability to treat outflows and jets due to the lack of a proper advection operator in its temperature solver. Relativistically hot systems are also presently inaccessible to HEROIC due to the soft Comptonization assumption inherent with the Kompaneets operator approach. Future releases of the code will include a switch to toggle between Kompaneets and a new relativistic photon redistribution kernel to handle the case of extremely hot Comptonization.
Chapter 6

Summary and Future Directions

6.1 Summary

In this thesis, we investigate the radiation from accretion disks around black holes. A major motivating factor for studying radiation from accretion disks is the recent observational effort to measure black hole spin via spectral fitting of the X-ray thermal continuum. The task of continuum fitting is complicated by several factors: 1) the high degree of scattering present within the accretion disk leads to complex nonlocal coupling of the radiation field; 2) the true structure of the disk may deviate from the standard model of Novikov & Thorne (1973); and 3) the presence of a hot comptonizing corona can complicate the determination of the thermal disk component of the spectra. To combat these issues, a firmer theoretical understanding of the underlying radiative processes is needed.

This thesis tries to address some of the limitations lurking in the spectral modeling
business. Our exploration starts in Chapter 2, where we compute the spectral signatures of recent GRMHD global simulations of turbulent accretion. We calculate synthetic disk spectra by combining 1D-radiative transfer calculations with relativistic raytracing. We find that our numerical GRMHD turbulent accretion disks are nearly indistinguishable from prevailing analytic disk models in their spectra. The only surprise from this exercise was for the spinning black hole case. We observed a weak power law tail at high energies that was being generated by the hot comptonized emission of the plunging region. However due to the 1D treatment of problem becoming increasingly invalid inside the plunging region, doubts were raised on the veracity of this apparent power law. This motivated the need for 3D radiative transfer calculations to confirm the result hinted at by the 1D calculation.

Chapter 3 builds upon the 1D disk modeling theme to answer a longstanding question: how can radiatively dominated disks remain thermally stable? The prediction from the standard disk model is that beyond moderate accretion rates ($L > 0.05L_{Edd}$), the disk should enter a radiation dominated regime that is susceptible to a runaway heating/cooling process (the thermal instability). Yet there are many observations of well-behaved accretion disks accreting at $L \sim 0.3L_{Edd}$ that are stable for long periods of time. Our 1D disk model reconciles this tension between theory and observations by noting that the introduction of turbulent mixing leads to a suppression of the radiative heat flux, which restores stability up to $L \sim 0.5L_{Edd}$.

We take a departure from the 1D radiative modelling approach starting in chapter 4. Here we develop HERO, a new 3D relativistic radiative transfer code. The bulk of this chapter is simply devoted to explaining how HERO operates. We also present a series of 1D, 2D, 3D, and 4D(relativistic) benchmark tests to verify that the code is producing
physically reasonable results.

Chapter 5 focuses on extending the capabilities of HERO by adding a Compton module to the code. The new upgraded code is named HEROIC – HERO Including Comptonization. Again, we benchmark HEROIC against a suite of 1D and 3D analytic solutions before turning to the astrophysical application of a Comptonized thin accretion disk. We find that our disk spectra from the 3D radiative calculations match fairly closely to the earlier 1D estimates. The slight differences that arise are due to a combination of: 1) returning radiation providing additional heating in the disk photosphere, and 2) differences in the energy injection/mass profile for 1D vs 3D.

6.2 Future Directions

HEROIC represents a first attempt at a robust 3D general-relativistic radiative-transfer code. It is designed to work for all manners of optically thick Compton problems with $T_{\text{gas}} < 10^8$K (i.e. the nonrelativistic Comptonization regime). This covers a wide range of problems, and several of these potential applications are listed below.

6.2.1 Astrophysical Applications

Perhaps the most immediately obvious application of HEROIC is revisiting the spin fitting bias calculation of (Zhu et al. 2012). It would be useful to see how the 3D spectral analysis differs from the 1D analysis to check if the GRMHD disk deviations for BH spin fitting are still negligible. Based on the preliminary results of Chapter 5, it is likely that the updated 3D HEROIC calculation would have larger deviations than was previously
Suppressed accretion disks are another interesting application for HEROIC. This relates to the spectral modeling of ultraluminous X-ray (ULX) sources, which have luminosities that greatly exceed (by factors of up to 1000 – Farrell et al. 2009) the Eddington limit for a typical stellar mass black hole ($M \sim 10M_\odot$). This large violation of the Eddington limit suggests that these objects probably correspond to intermediate mass black holes with $M \gg 10M_\odot$. An interesting project would be for someone to calculate the spectral properties of numerically simulated ULXs via HEROIC. This would test if there is truly an upper limit to the apparent luminosity as seen by distant observers to help settle the debate on the existence of IMBHs.

Computing the radiation from BH jets would be another interesting future application of HERO. Jet-structures, their formation, acceleration, and their relationship to the accretion phenomena is still an area of open research (Tchekhovskoy et al. 2011; Tchekhovskoy & McKinney 2012). Luckily, there is a wealth of observational data available from a variety of astrophysical systems (e.g. AGN, microquasars). Due to the importance of relativity and 3D effects on jet production, HERO would be an ideal code to analyze the observable consequences of jets produced by numerical simulations.

### 6.2.2 Future Plans for HERO

HERO as a code is still in its infancy – so far, it has only been applied to the problem of modeling thermal emission from BH accretion disks. Below we discuss a partial list of features that we intend to explore in future releases.

Perhaps the feature with the most immediate astrophysical payoff would be the
inclusion of line opacities and emissivities. A proof of concept has already been demonstrated from investigations in chapter 4 (i.e. the iron line reflection problem), which involved handling line emission for a single atomic transition. Iron line fitting is an active area of research since it offers another way of measuring black hole spin. Unfortunately, there are several major uncertainties underpinning the whole business of fluorescence line modeling. For instance, the exact nature of the disk illumination from a hot corona is not well understood. The hope is that HERO will be able to self-consistently calculate both the coronal component and reflection components of the radiation field. If successful, this exercise would be extremely useful for constraining illumination models for the iron line profile.

Adding time dependence to HERO is another alluring next step. The variability properties of black hole systems are very diverse and depend on the spectral state of the BH. Particularly interesting is the phenomenon of high-frequency Quasi-Periodic Oscillations. These QPOs are seen as narrow peaks in the power spectrum that sometime arrive in packs with fixed integer frequency ratios. These are thought to be caused by resonant coupling of epicyclic frequencies in the disc although there is still much debate on their true origin. It would be interesting to see if the time dependent emission as computed by HERO for a realistic accretion flow leads to the phenomenon of QPOs. Other ideas related to time dependence includes the phenomenon of X-ray reverberation in accretion disks. It would be very interesting to see if we can reproduce the same time lags in a simulation as the lags observed in nature. So far, we have made some minor progress towards developing a time-dependent HERO package in collaboration with Olek Sadowski by interfacing with his hydrodynamic KORAL code via a Variable Eddington Tensor approach. We have only completed a few preliminary test problems
that demonstrate proof of concept (i.e. shadow tests, crossing beam tests, relativistic light bending tests, etc...).

Polarization offers yet another dimension for gathering information about astrophysical sources. One of the major goals of the Event Horizon Telescope collaboration is the measure the linear polarization of our Galaxy’s radio emission at 10\(\mu\)as scales. A scintillating project would therefore be to add polarization capabilities to HERO and calculate the expected polarized field from to a realistic simulated accretion flow. The main advantage HERO has is that it correctly computes the self-consistent scattered radiation field in all parts of the flow (including optically thick parts).

Other desirable features for HERO would be the ability to handle relativistic Comptonization. This would require an overhaul of the current Compton solver from the current Kompaneets diffusion approach to a more detailed frequency by frequency redistribution analysis. Another future improvement would be the ability to handle angle dependent scattering kernels instead of the isotropic approximation that is currently in place. This becomes especially important in the limit of relativistic scattering. For studying emission from say nonthermal shocks in jets, it would be nice if the Comptonization module also had support for nonthermal electron temperature distributions.
References


Balbus S.A., Hawley J.F., 1998, Rev. Mod. Phys., 70, 1


Beckwith K., Hawley J. F., Krolik J. H., 2008a MNRAS, 390, 21

REFERENCES


Bolton C. T., 1972, Nat, 240, 124


REFERENCES


Chang J.S., Cooper G., 1970, J. Comp. Phys, 6, 1


Coppi P.S., 1991, PhD, 2


REFERENCES

Dubroca B., & Feugeas J. L. 1999, CRAS, 329, 915
Feautrier P., 1964, Comptes Rendus Academie des Sciences (serie non speciee), 258, 3189
REFERENCES


Gittings M., Weaver R., Clover M., Betlach T., et al., 2008, CS&D, 1a, 5005


Gorecki A., Wilczewski W., 1984, AcA, 34, 141

REFERENCES

Gou L., McClintock J. E., Steiner J. F., Narayan R., Cantrell A. G., Bailyn C. D.,


Hayek, W., Asplund, M., Carlsson, M., Trampedach, R., Collet, R., Gudiksen, B. V.,

Hayes J. C., Norman M. L., Fiedler R. A., Bordner J. O., Li P. S., Clark S. E.,


Hayward C. C., Keres D., Jonsson P., Narayanan D. Cox, T. J., Hernquist L.,


293
REFERENCES

Hubeny, I., 2003, ASPC, 288, 17
Kawaguchi T., Shimura T., Mineshige S., 2000, NewAR, 44, 443
REFERENCES

Kompaneets A.S., 1957, Sov. Phys. JETP, 4, 730


REFERENCES

McClintock J. E. et al., 2011, Classical and Quantum Gravity, 28, 114009
REFERENCES


Mitsuda K. et al., 1984, PASJ, 36, 741


Narayan R., McClintock J.E., 2008, NewAR, 51, 733


297
REFERENCES

Ohsuga K., Mineshige S., Mori M., Kato Y., 2009, PASJ, 61, 7O
Orosz J. A. et al., 2007, Nat, 449, 872
Paczyński B., 2000, astro-ph/0004129
REFERENCES


Pringle J.E., 1981, ARAA, 19, 137


Psaltis D., Lamb F. K., 1999, ASPC, 161, 410


Remillard R. A., McClintock, J. E., 2006b, ARA&A. 44, 49


299
REFERENCES


Schwarzschild K., 1916, Über das Gravitationsfeld eines Massenpunktes nach der Einsteinschen Theorie, Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften, 7, 189


REFERENCES

Steinacker J., Bacmann A., Henning T., 2002, JQSRT, 75, 765
Stenholm L. G., Stoerzer H., Wehrse R., 1991, JQSRT, 45, 47
REFERENCES

Szu-Cheng S., Kuo-Nan L., 1982, JQSRT, 28, 271
Vincent F. H., Paumard T., Gourgoulhon E., Perrin G., 2011, CQGra, 28, 5011
Webster L., Murdin P. 1972, Nat, 235, 37

302
REFERENCES


Wu X.-B., et al., 2015, Nature, 518, 512


Zhang S. N., Cui W., Chen W., 1997a ApJ, 482, L155


