

THE WEATHER AROUND BLACK HOLES:  
ACCRETION DISK AND JET EVOLUTION IN  
GRS 1915+105 AND OTHER SYSTEMS

A Dissertation

Presented to the Faculty of the Graduate School  
of Cornell University

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

by

David Mark Rothstein

January 2007

© 2007 David Mark Rothstein

ALL RIGHTS RESERVED

THE WEATHER AROUND BLACK HOLES: ACCRETION DISK AND JET  
EVOLUTION IN GRS 1915+105 AND OTHER SYSTEMS

David Mark Rothstein, Ph.D.

Cornell University 2007

A fundamental problem in astrophysics involves the origin of jets and the dynamics of accretion disks. How does a collimated outflow arise from hot material that is spiraling into a central object (be it a black hole, neutron star, white dwarf or young star)? Why do jets appear to turn “on” and “off” as the accretion disk changes between different states?

In this thesis, we attempt to shed some light on this question through observational, theoretical and computational studies. We present high time resolution observations of the Galactic black hole candidate GRS 1915+105 and show how subtle differences in the accretion disk evolution in this object during different episodes of activity are related to different types of jet ejections. We then develop several new theoretical results about accretion disks that allow us to parametrize our uncertainty about the jet and other complex physics in a way that can be studied in one- or two-dimensional numerical simulations. Following this, we present FRIENDLY, a new code for numerical integration of arbitrary partial differential equations to arbitrary orders of accuracy that was developed during the course of this thesis. Finally, we present a preliminary application of FRIENDLY, in which we attempt to simulate an accretion disk that experiences a sudden increase in the strength of turbulence, as might be expected if an ordered magnetic field were

ejected from the system in the form of a jet. We compare the simulations to the observations of GRS 1915+105 and conclude that further study of this mechanism may lead to an explanation for the behavior of this enigmatic object.

## BIOGRAPHICAL SKETCH

David Mark Rothstein, who does not usually write about himself in the third person, is the only child of Frederick and Suzanne Rothstein of Trenton, New Jersey. He was born on July 4, 1978. (He eventually concluded that the fireworks and parades aren't in his honor, although he still suspects they might be.)

He began his educational career in 1980, enrolling in the "Mommy and Me" program at the local JCC. He did well, although he struggled with the part of the curriculum where you had to leave your mommy and go play with the other kids. Nonetheless, he progressed to nursery school, where he cleaned up academically despite the high-pressure environment; by this time, he had moved on to the Abrams Hebrew Academy in Yardley, Pennsylvania, where he would remain until his 8th grade graduation in 1992.

Bright things were expected of the young Rothstein, but his academic interests while at Abrams took a turn towards the obscure; his first published paper, on the topic of "cooties" and which girls do and do not have them, was not well-received (it was published on looseleaf paper and was never read by anyone besides himself). Also during this time period, he had a fish named Fishy who could do tricks and come back from the dead, and another fish who sadly died before being given a name.

In 1992, he moved on to Princeton Day School, where he would graduate with a high school degree in 1996. By this time, he had acquired the nickname "Ice Ice Rothstein" (for reasons that remain obscure except in the mind of one Rex Reyes, who bestowed the name upon him), and he therefore will be referred to as "Ice" for the remainder of this biographical sketch.

While in high school, Ice developed an interest in journalism and physics. He also participated in athletics, winning the Most Improved Player award for the Junior Varsity basketball team (which basically meant that he improved from “extremely bad” to “just plain bad”) but was shockingly cut during Varsity tryouts his senior year. Although Michael Jordan *also* was allegedly cut from his high school basketball team, Ice eventually concluded that he was not going to be selected out of high school in the N.B.A. draft, so he proceeded instead to Haverford College in Pennsylvania, where he matriculated in the fall of 1996.

Haverford affected Ice profoundly. The Quakers that founded and still influence the college believed that the light of God is within every human being, and it can shine at any time. Though Ice is proudly Jewish, he remains fascinated by this central tenet of Quakerism, if not always able to believe or live up to it; the mix of stubborn individualism, quiet integrity, and tempered optimism that it implies was a good fit for Ice (in that it matched the values his parents had already tried to instill in him), and it helped him mature to adulthood while serving as an anchor for his later life.

Haverford’s hypocrisies and flaws are the same as anywhere else, perhaps more pronounced because of its high ideals, and Ice fought against them mightily, especially in his role with the campus newspaper, which had become his primary extracurricular pursuit. Thanks to some wonderfully-dedicated professors, his academic interests, meanwhile, moved through physics and into astronomy. He eventually graduated with a B.A. in Physics and Astronomy in the year 2000.

In May of 2000, on the advice of his longtime friend Ben Seelig, he ran for election as God (a.k.a., Supreme Ruler of the Universe), but was unfortunately trounced at the polls by the incumbent. He therefore decided to pursue a career in

astrophysics (not realizing that science was already a very popular career choice for people who think they are God). He entered the graduate program in Astronomy at Cornell University in the fall of 2000, where he has been ever since.

Cornell does not wear its ethos on its sleeve quite so prominently as Haverford (or perhaps it does wear it on its sleeve, but the sleeve is covered by a very heavy winter coat six months out of the year). Nonetheless, the idea that Cornell is a place where “any person can find instruction in any study” inspires Cornell to invest in the surrounding community, and Ice found solace in many outreach activities during his time as a graduate student, including the Cornell Science Inquiry Partnerships (CSIP) program, the Graduate Student School Outreach Project (GSSOP), Focus for Teens, and the Ask an Astronomer website. The second part of the creed (“any study”) he occasionally tried to live up to singlehandedly, taking courses in science journalism and entrepreneurship in addition to physics and astronomy, writing a thesis that involves observational, theoretical and numerical studies, and participating in numerous other activities, including Ultimate Frisbee, Karatedo Doshinkan, Darfur activism, Habitat for Humanity, Birthright, and even an infamous production of *Romeo and Juliet* (now out on DVD!).

During six years in Ithaca, he fell in love with its gorges, its hills, and its diversity, and eventually with a wonderful woman, Marielle Macher, who he is lucky to have met. He will stay at Cornell after completing his Ph.D., as the recipient of a National Science Foundation Astronomy and Astrophysics Postdoctoral Fellowship.

This thesis is dedicated to my thumbs,  
without whom it would just be one long word.

This thesis is also dedicated to Marielle,  
for helping me come up with the above dedication.

Finally, this thesis is dedicated to my parents,  
without whom I wouldn't have any thumbs.

## ACKNOWLEDGEMENTS

First, I would like to thank my advisors during the course of writing this thesis: Steve Eikenberry for Chapter 2, and Richard Lovelace for the rest.

I want to thank Richard for encouraging me to follow my own path and for believing that I could do a thesis involving theory and observations when no one else seemed to. It was an honor to have an advisor who, in addition to being helpful, is also one of the nicest and most genuine people I have ever met.

Without Steve, meanwhile, I suspect I might have drifted out of grad school not so long after I entered it. He introduced me to the wonders of accretion disks and jets, and his enthusiasm and energy for science (and life in general) are infectious. Working with Steve taught me to be bold and not to be afraid to question conventional scientific wisdom, and I have greatly enjoyed remaining his colleague and continuing to collaborate with him after he left Cornell.

I would also like to thank my funding sources over the years. In particular, much of the material in this thesis is based upon work supported under a National Science Foundation Graduate Research Fellowship, which I held for three years and which gave me the freedom to follow my research wherever it led me. Additional support came from Steve Eikenberry's NSF CAREER award (NSF-9983830), the Cornell Science Inquiry Partnerships (CSIP) program at Cornell (a National Science Foundation Graduate Teaching Fellows in K-12 Education program), the NASA Space Grant program, and the Stewardship Sciences Academic Alliances program of the National Nuclear Security Administration under US Department of Energy Cooperative agreement DE-FC03-02NA00057. This thesis has also made use of NASA's Astrophysics Data System, as well as the High Energy Astrophysics Science Archive Research Center (HEASARC) provided by NASA's

Goddard Space Flight Center.

I would like to thank everyone who collaborated with me on the research presented in this thesis or otherwise helped out with it. In particular, the observations in Chapter 2 wouldn't have happened without the staff at Palomar Observatory and the members of the *Rossi X-Ray Timing Explorer* team. They also wouldn't have happened without the incomparable Keith Matthews, who knows the instruments at Palomar better than most people know their own children, and who is a fun guy to observe with as well. I would also like to thank Don Barry for the many times he gave me computer assistance, as well as many people for discussions about my research, including Marina Romanova. My Special Committee members, Jim Cordes and Dong Lai, have always been helpful, kind during exams and, miraculously, in their offices whenever I needed a signature. Finally, I would like to thank Valerie Mikles, a collaborator, friend, and kindred soul in this weird world of astronomy, to whom I dedicate my next webpage update (whenever it may be).

I would like to thank all the teachers who I've worked with over the years, both as a student and as a colleague. From elementary school through high school through college, grad school and then CSIP, I've always been privileged to learn with and work alongside amazing people who truly care about their students and the subjects they teach. I hesitate to start naming names, because I wouldn't know where to stop, but I will mention one particular group: the professors I worked with most closely at Haverford (including Bruce Partridge, Jerry Gollub, Steve Boughn and the rest of the physics department), who prepared me for grad school and, just as important, gave me a vision of what it is like to have an honorable career in science that one can go home at the end of the day and be proud of. I

would also like to thank all the students I've had, especially in the past year, who made it worthwhile for me to teach.

I would like to thank everyone at Cornell who participated in extracurricular activities with me and thereby helped make my time here more enjoyable. To name a few (in no particular order): the Ask an Astronomer team, Nancy Trautmann and all the amazing grad students I met through CSIP, the cast, crew and video producers of *Romeo and Juliet*, and everyone involved with GSSOP, Karatedo Doshinkan, Birthright, GAP and Hillel.

I would like to thank all the people who silently make the Space Sciences building run, and I would also like to thank all my not-so-silent officemates and labmates over the years at Cornell. From Joe and Dae-Sik in the early days all the way up through Ryan Yamada, Ann, Suniti, and Ryan Yamada (er, Shannon) today, I've been fortunate to share my workspace with interesting people who make the office so much more than a place to go and stare at a computer screen. Even in the last few months, when you all probably thought I wanted you to stop talking and let me get back to my thesis, trust me—I was happy for the diversions!

I would also like to thank my housemates over the years in Ithaca, including Matija Cuk and Tim McConnochie (with whom I shared the most unlikely “party house” ever), Kate Becker (who will hopefully remember me when she achieves fame and fortune with her adjustable brownie pan, or perhaps one of the many inventions dreamed up together in our living room), and Shannon Guiles (who had the misfortune of being my housemate during one of the busiest years of my life, but who I wish I had more time to hang out with in our humble abode). I would also like to thank all my other friends at Cornell, especially the Shabbat crew for making me feel like I had a home on Friday evenings. And I would be remiss

not to thank the Old Order Mennonites of Yates County, who aren't actually my friends, but I wish they were. I also want to thank my friends from college, high school and before, who I don't always keep in touch with as much as I should, but whose presence in my thoughts keeps me grounded and reminds me of who I am and where I came from.

If I'm forgetting anyone else (which I'm sure I am), blame it on fatigue after the time spent writing the next couple hundred pages.

The final thanks are the most important ones, however. First, Marielle. You put up with me more than anyone else while I was writing this thesis, and you never once complained. Your wisdom keeps me sane, your quirkiness keeps me happy, and your kindness keeps me hopeful. I wish you much ketchup in the future, and I hope I can continue to share it with you.

And finally, my parents. You have encouraged me in everything I have ever done, and you've taught me to think for myself and to always remember that the truly important things in life are health, peace, happiness and caring for others. I hope I can take that lesson with me as I move on from graduate school to the rest of my life.

## TABLE OF CONTENTS

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	The Energy Released by Black Holes . . . . .	1
1.2	Observations of Black Holes . . . . .	2
1.3	Overview of the Thesis . . . . .	4
<b>2</b>	<b>Observations of Rapid Disk-Jet Interaction in the Microquasar GRS 1915+105</b>	<b>6</b>
2.1	Summary . . . . .	6
2.2	Introduction . . . . .	7
2.3	Observations . . . . .	10
2.4	Light Curves . . . . .	12
2.4.1	Multiwavelength Features of the Light Curves . . . . .	14
2.4.2	Origin of the Infrared Flares . . . . .	14
2.5	Infrared/X-Ray Correlation . . . . .	15
2.6	X-Ray Spectral Evolution . . . . .	22
2.6.1	1997 and 2002 State Changes . . . . .	23
2.6.2	1997 Trigger Spike . . . . .	27
2.6.2.1	Nonthermal Models for the Spike . . . . .	29
2.6.2.2	Thermal Accretion Disk Models for the Spike . . . . .	31
2.6.2.3	The Spike in the Context of Other X-ray Binaries . . . . .	34
2.6.3	1997 and 2002 Oscillations . . . . .	37
2.7	Origin of the “Class C” Subflares . . . . .	39
2.8	Discussion . . . . .	43
2.8.1	A Continuum of Behavior . . . . .	43
2.8.2	When Does the Ejection Occur? . . . . .	45
2.8.3	Infrared and Radio Flares . . . . .	47
2.9	Conclusions . . . . .	51
<b>3</b>	<b>The Equations Governing the Evolution of an Accretion Disk: A Study of the Vertically-Integrated Structure</b>	<b>54</b>
3.1	Summary . . . . .	54
3.2	The Equations of Radiative Magnetohydrodynamics . . . . .	57
3.2.1	MHD Assumptions . . . . .	57
3.2.2	General Principles of Conservation . . . . .	62
3.2.3	Conservation of Mass . . . . .	63
3.2.4	Conservation of Momentum . . . . .	64
3.2.4.1	Surface Forces . . . . .	65
3.2.4.2	Volume Forces (Gravity) . . . . .	75
3.2.4.3	The Final Momentum Equation . . . . .	82
3.2.5	Conservation of Energy . . . . .	84
3.2.6	Final Equations for Our Accretion Disk Model . . . . .	91

3.3	Simulations in Fewer Than Three Dimensions . . . . .	94
3.3.1	Removing the Dependence on Azimuthal Angle ( $\phi$ ) . . . . .	95
3.3.2	Removing the Dependence on Height ( $z$ ) . . . . .	96
3.3.2.1	The Dimensionless Vertical Parameter $\{u\}_z$ . . . . .	97
3.3.2.2	Typical Values of the Vertical Parameter $\{u\}_z$ . . . . .	101
3.3.2.3	The Vertical Scale Height . . . . .	118
3.3.2.4	The Surface Flux and Luminosity . . . . .	123
3.4	Steady State Disk Models . . . . .	128
3.4.1	Mass and Angular Momentum Conservation . . . . .	130
3.4.2	Local Energy Balance . . . . .	132
3.4.3	Evolution Timescales . . . . .	142
3.4.4	Steady State Solution . . . . .	145
3.5	Conclusions . . . . .	147
<b>4</b>	<b>FRIENDLY: A New Code for Integrating Time-Dependent Equations to Arbitrary Orders of Accuracy</b>	<b>149</b>
4.1	Summary . . . . .	149
4.2	Motivation . . . . .	150
4.3	Separating Space and Time Derivatives with the Method of Lines . . . . .	152
4.4	Choosing an Order of Accuracy for the Space and Time Derivatives . . . . .	154
4.5	Second Order Method of Lines in Comparison to More Specialized Alternatives . . . . .	158
4.5.1	The Two-Step Lax-Wendroff Scheme . . . . .	160
4.5.2	The Staggered Leapfrog Scheme . . . . .	161
4.6	Calculating Spatial Derivatives on Irregular Grids to Arbitrary Orders of Accuracy . . . . .	163
4.7	Conservative Integration Schemes for Irregular Grids . . . . .	165
4.7.1	Nth Order Conservative Integration Schemes . . . . .	167
4.7.2	Applications to User-Specified Boundary Conditions . . . . .	171
4.7.3	Conservation on Non-Uniform Grids . . . . .	173
4.8	Boundary Conditions . . . . .	174
4.9	Conclusions . . . . .	175
<b>5</b>	<b>Simulations of Accretion Disk Evolution in GRS 1915+105</b>	<b>178</b>
5.1	Summary . . . . .	178
5.2	Introduction . . . . .	178
5.3	Simulations . . . . .	182
5.4	Results . . . . .	184
5.5	Conclusions and Future Work . . . . .	198
<b>A</b>	<b>Appendix: The Divergence of a Tensor in Cylindrical Coordinates</b>	<b>201</b>
A.1	Evaluation of the Divergence . . . . .	202
A.2	Specialization to Cylindrical Coordinates . . . . .	204

## LIST OF FIGURES

2.1	A typical $2.2\mu\text{m}$ infrared image of the field of GRS 1915+105 at 1-second time resolution, from our observations on 2002 July 27. . . . .	11
2.2	Simultaneous infrared and X-ray light curves of GRS 1915+105 during July 2002. . . . .	13
2.3	Simultaneous infrared and X-ray light curves of GRS 1915+105 during August 1997, from data originally presented in Eikenberry et al. (1998a). . . . .	16
2.4	The number of X-ray oscillations in each $\sim 30$ minute cycle plotted against the integrated flux density of the infrared flare accompanying that cycle. . . . .	19
2.5	Simultaneous infrared and X-ray light curves of GRS 1915+105 at the time of the sixth infrared flare observed on 2002 July 27. . . . .	22
2.6	Comparison of one cycle of jet formation in GRS 1915+105 from observations in 1997 and 2002. . . . .	24
2.7	Comparison of the X-ray spectral evolution of GRS 1915+105 during 1997 and 2002. . . . .	26
2.8	X-ray spectra of GRS 1915+105 during 1997 and 2002. . . . .	28
2.9	Simultaneous infrared and X-ray light curves of GRS 1915+105 at the time of the infrared “subflare” observed on 2002 July 27. . . . .	40
3.1	The dependence of the vertical parameter $\{u\}_z$ on the scale height, for Gaussian and exponential vertical distributions. . . . .	103
3.2	The turbulent heating rate per unit mass as a function of height $z$ in the accretion disk, for three different polytropes. . . . .	110
3.3	Same as Figure 3.2, but for the turbulent heating rate per unit volume. . . . .	111
3.4	The “smoothness” of the vertical profile of the turbulent heating rate per unit mass, as a function of the equatorial plane gas pressure fraction $\beta_0$ and the polytropic index $N$ . . . . .	112
3.5	Same as Figure 3.4, but this time measuring the compactness of the vertical profile of the turbulent heating rate per unit volume. . . . .	113
3.6	The dependence of the vertical parameter $\{u\}_z$ on the exponent $s$ , for polytropic vertical equations of state. . . . .	116
3.7	The effect of the jet ram pressure on the local energy balance curve of an accretion disk. . . . .	139
3.8	The effect of the jet energy transport on the local energy balance curve of an accretion disk. . . . .	141
3.9	The effect of the jet vertical structure on the local energy balance curve of an accretion disk. . . . .	143
4.1	Four snapshots of the numerical evolution of a sound wave in a one-dimensional adiabatic gas, made using FRIENDLY. . . . .	157

4.2	Twelve snapshots of the numerical evolution of a sound wave in a two-dimensional adiabatic gas, made using FRIENDLY. . . . .	159
4.3	A comparison of the evolution of two identical sound waves, one of which begins on the spatial grid and the other of which is initially imposed on the grid boundary, made using FRIENDLY. . . . .	176
5.1	The surface mass density evolution for a run in which the value of $\alpha$ is raised from 0.01 to 0.1 in the inner region of the disk. . . . .	186
5.2	The profile of the surface mass density at five times from Figure 5.1. . . . .	187
5.3	The evolution on the local energy balance curve at two radii, when $\alpha$ is raised from 0.01 to 0.1 in the inner region of the disk. . . . .	188
5.4	The outburst light curves for the run shown in Figures 5.1 through 5.3, compared to the light curve for a run in which the value of $\alpha$ is raised in a smaller region of the disk. . . . .	189
5.5	The surface mass density evolution for a run in which the value of $\alpha$ is raised from 0.01 to 0.1 in the middle part of the disk. . . . .	193
5.6	Same as Figure 5.5, but the outer boundary of the region where $\alpha$ changes is now $100r_g$ closer to the black hole. . . . .	194
5.7	The outburst light curves for the runs shown in Figures 5.5 and 5.6. . . . .	195
5.8	The evolution on the local energy balance curve at a radius $r = 110r_g$ , for the runs shown in Figures 5.5 and 5.6. . . . .	196
5.9	The evolution of the luminosity per unit radius for the run shown in Figure 5.5. . . . .	197
5.10	A rough measure of the degree to which various assumptions made for our one-dimensional disk model break down during the inner disk outburst shown in Figure 5.1, at a radius of $r = 22r_g$ . . . . .	199
5.11	Same as Figure 5.10, but for the middle disk outburst shown in Figure 5.5, at a radius of $r = 110r_g$ . . . . .	200

# Chapter 1

## Introduction

Imagine an object with the mass of the Sun, condensed into the size of a small city.

This object is a black hole, and its extremely high mass density produces, in the region of spacetime around it, the strongest gravity in the universe. The black hole’s gravity is so strong that neither the city nor the Earth itself could hold it up; the black hole would crush through these objects, and its gravitational tides would shred them apart.

This thesis concerns the effects of a black hole’s gravity on objects—in particular, on concentrations of ionized plasma—that are found near actual black holes in the universe.

### 1.1 The Energy Released by Black Holes

The technical definition of a black hole is that it is an object whose gravity is so strong that nothing, including light, can escape from within it. A black hole, in fact, does not have a hard surface—its gravity is so strong that it pulls any matter and light within it in to the center, where the forces of quantum gravity deal with it in some unknown way. The “sphere” that defines the black hole is really just a region of space whose invisible boundary (known as the event horizon) marks the region within which light cannot escape. Outside the black hole, one looks inside and sees only darkness.

Despite the fact that no light is emitted from a black hole, observing black holes is one of the main topics of this thesis. It is a curious fact that the inability of light to escape from a black hole is the very property that not only *allows* these tiny, city-sized objects to be detected all the way across the Galaxy (or all the way across the visible Universe for the largest black holes, which are still only about as big as our solar system), but that also virtually *guarantees* that black holes are among the most important contributors of light and energy to the universe that surrounds them, and therefore worthy of study.

Specifically, if we consider an object of mass  $M$  and require only that it has an escape velocity equal to the speed of light  $c$  at its “surface” (i.e., that it meets the fundamental definition of a black hole), then the radius of the black hole must be  $R \sim 2GM/c^2 = 2.95 (M/M_\odot)$  km from simple Newtonian physics ( $G$  is Newton’s gravitational constant and  $M_\odot$  is the mass of the Sun), and the amount of energy liberated by an object of mass  $m$  falling from a large distance onto the black hole, which is  $\sim GMm/R$ , must therefore become  $\sim mc^2$  upon substitution of the expression for  $R$ ; i.e., it must roughly equal the object’s rest mass energy.

In other words, it is a fundamental property of a black hole (or, indeed, of any object whose escape velocity approaches the speed of light) that it can efficiently convert a large fraction of the rest mass of accreting material into other forms of energy *before* the accreting material enters the black hole.

## 1.2 Observations of Black Holes

What happens to the energy that is liberated by a black hole each time an object falls in? There are several possibilities: the energy can be liberated in the form of radiation, making the region around the black hole shine bright; it can be liberated

by being transferred to other particles that do not enter the black hole but instead speed away from it in the form of a jet; or it can be carried directly into the black hole itself in the form of kinetic energy. It is one of the goals of this thesis to help understand what makes astrophysical black holes switch between different modes where one or another of the above energy release mechanisms is dominant.

In fact, observations have already been quite successful in determining the various properties that black holes have when they are in one or another of the above states. As might be expected from energy conservation arguments, the general trend is that black holes which emit a large amount of energy in the form of thermal radiation from an accretion disk show less evidence for energy dissipation in a jet, and vice versa. For good reviews of the current state of knowledge and the detailed properties of the different black hole states, see Fender et al. (2004), McClintock & Remillard (2006) and Remillard & McClintock (2006).

Much less is known about the manner in which black holes *transition* between the different states, however. The basic picture appears to be that if one plots a black hole's evolution on a "hardness-intensity" diagram (essentially the X-ray astronomer's version of a color-magnitude diagram, although with a slightly different physical interpretation of the observed color, due to the different spectral components that are active in the X-ray band), one finds that black holes undergo outbursts in which they start off "blue" and faint, then become gradually brighter and "redder" as thermal emission from the accretion disk becomes visible; at some point, the steady jet that dominates the energy release in the faint state disappears in a brief burst of fast, transient jet ejection that is observed in the form of flares in radio and infrared light curves and, occasionally, discrete moving ejecta in high resolution radio images. After the ejection, the emission from the region around

the black hole remains red and bright for some period of time, before eventually getting fainter and bluer again.<sup>1</sup> However, the physical parameters that trigger black holes to enter an outburst and control the speed and manner with which they cycle through various parts of the outburst remain uncertain.

Perhaps the most fascinating black hole in our galaxy is an object known as GRS 1915+105. This object is interesting because it essentially has spent all of its time since its discovery (nearly fifteen years ago) in an “active” state, cycling right around the region of the hardness-intensity diagram where transient jets form (e.g., Fender et al. 2004). What makes GRS 1915+105 unique is likely its enormous accretion disk, which is much bigger than that of any other similar system and therefore effectively gives the black hole’s strong gravity a lot of mass that it can continually “play with” (Done et al. 2004; Fender & Belloni 2004; Remillard & McClintock 2006). Thus, GRS 1915+105 is a perfect source in which to study the behavior of plasma in the strong gravitational field of a black hole and the way in which this material changes between different states—in other words, the weather around a black hole.

### **1.3 Overview of the Thesis**

This thesis investigates GRS 1915+105 and black hole accretion disks and jets in general from an observational, theoretical and numerical perspective, with an eye towards shedding some light on the reasons black hole “weather” behaves as it does.

Chapter 2 presents high time resolution observations of GRS 1915+105 in the

---

<sup>1</sup>These observations are primarily based on stellar mass black holes with companion stars that feed the black hole via Roche lobe overflow; other categories of black holes can and do behave differently.

infrared and X-rays during periods of jet ejection activity and identifies subtle features in the accretion disk evolution that are related to different types of jet ejections. This chapter is published in Rothstein, Eikenberry, & Matthews (2005).

Chapter 3 rederives a large portion of accretion disk theory with an eye towards incorporating ad-hoc models of a jet or other changes in the disk's vertical structure in one-dimensional accretion disk models. Several new quantities are introduced that can be used to parametrize our uncertainty about a jet and other complex physics, and the effects of these quantities on the energy balance and evolution timescales of an accretion disk are explored.

Chapter 4 presents FRIENDLY, a new code for numerically integrating partial differential equations that can be used to study accretion disks, as well as for a wide variety of other astrophysical and general scientific applications.

Finally, Chapter 5 presents simulations of an accretion disk using FRIENDLY in which the strength of turbulence increases suddenly, in a manner that is roughly consistent with observational data as well as theoretical models for jet ejection. The results of the simulations are compared to observations of GRS 1915+105, and further directions for study are suggested.

## Chapter 2

# Observations of Rapid Disk-Jet

# Interaction in the Microquasar

# GRS 1915+105\*

## 2.1 Summary

In this chapter, we present evidence that  $\sim 30$  minute episodes of jet formation in the Galactic microquasar GRS 1915+105 may sometimes entirely be a superposition of smaller, faster phenomena. We base this conclusion on simultaneous X-ray and infrared observations in July 2002, using the *Rossi X-ray Timing Explorer* and the Palomar 5 meter telescope. On two nights, we observed quasi-periodic infrared flares from GRS 1915+105, each accompanied by a set of fast oscillations in the X-ray light curve (indicating an interaction between the jet and accretion disk). In contrast to similar observations in 1997, we find that the duration of each X-ray cycle matches the duration of its accompanying infrared flare, and we observed one instance in which an isolated X-ray oscillation occurred at the same time as a faint infrared “subflare” (of duration  $\sim 150$  seconds) superimposed on one of the main flares. From these data, we are able to conclude that *each* X-ray oscil-

---

\* This chapter is based on the published paper by Rothstein et al. 2005 [Rothstein, D. M., Eikenberry, S. S., & Matthews, K. 2005, *The Astrophysical Journal*, 626, 991; © 2005. The American Astronomical Society. All rights reserved.] It is reprinted here with minor changes, based on rights retained by the author.

lation had an associated faint infrared flare and that these flares blend together to form, and entirely comprise, the  $\sim 30$  minute events we observed. Part of the infrared emission in 1997 also appears to be due to superimposed small flares, but it was overshadowed by infrared-bright ejections associated with the appearance of a sharp “trigger” spike in each X-ray cycle that were not present in 2002. We also study the evolution of the X-ray spectrum and find significant differences in the high energy power law component, which was strongly variable in 1997 but not in 2002. Taken together, these observations reveal the diversity of ways in which the accretion disk and jet in black hole systems are capable of interacting and solidify the importance of the trigger spike for large ejections to occur on  $\sim 30$  minute timescales in GRS 1915+105.

## 2.2 Introduction

GRS 1915+105 is one of the most fascinating objects in astrophysics today. It was one of the first systems in the Galaxy to be identified as a “microquasar”—an X-ray binary with relativistic jets which mimics some of the behavior of quasars on a smaller and closer scale. (For a review of these objects, see Mirabel & Rodríguez 1999 and Fender 2006; for a review focusing on GRS 1915+105, see Fender & Belloni 2004.) Because timescales in microquasars are expected to be a factor of  $\sim 10^8$  shorter than in quasars (owing to the much smaller mass of the compact object which powers the system), microquasars are excellent laboratories for investigating accretion disk evolution and jet formation in black hole systems.

GRS 1915+105 was first discovered as a transient X-ray source (Castro-Tirado et al. 1994), and its radio and infrared counterparts have since been identified (Mirabel et al. 1994). The system contains a  $\sim 14 M_{\odot}$  black hole fed by Roche

lobe overflow from a K-M giant companion (Greiner et al. 2001a,b). GRS 1915+105 displays extreme variability in many different wavebands and on many different timescales, making it unique among the X-ray binaries. Its uniqueness is likely due to its extremely high accretion rate, which allows it to regularly reach luminosities unseen in other Galactic X-ray sources (Done et al. 2004; Fender & Belloni 2004).

Perhaps the most spectacular variability observed from GRS 1915+105 is seen in high resolution radio maps, which occasionally reveal the presence of resolved, bipolar ejections with flux densities up to  $\sim 600$  mJy that form on timescales of weeks and move away from the system at relativistic speeds ( $> 0.9c$ ; Mirabel & Rodríguez 1994; Rodríguez & Mirabel 1999; Fender et al. 1999b). These events are referred to as “class A” ejections by Eikenberry et al. (2000) to distinguish them from smaller “class B” radio and infrared flares on  $\sim 30$  minute timescales which are also thought to correspond to jet ejection. In addition to the radio and infrared flares, GRS 1915+105 displays a broad range of variability in the X-rays, where it switches between many different states (reviewed extensively by Belloni et al. 2000a) and where quasi-periodic oscillations (QPOs) on timescales as fast as  $\sim 168$  Hz have been observed (Remillard et al. 2002; McClintock & Remillard 2006).

Multiwavelength observations of “class B” events in GRS 1915+105 by Pooley & Fender (1997), Eikenberry et al. (1998a,b), and Mirabel et al. (1998) were the first to reveal the intimate link between accretion disk evolution and relativistic jet formation on short timescales in any black hole system. Eikenberry et al. (1998a) observed a one-to-one correspondence between repeating X-ray variability cycles and infrared flares ( $\sim 100$  mJy dereddened) on timescales of  $\sim 30$  minutes. The observations are consistent with a picture in which emptying and refilling

of the X-ray emitting inner disk (e.g. Belloni et al. 1997a,b) coincides with the ejection of material into a jet, which radiates through synchrotron emission to produce a flare (Fender et al. 1997; Pooley & Fender 1997). The infrared and X-ray light curves decouple as the ejecta becomes causally separated from the inner disk (Eikenberry et al. 1998a), and as the plasma moves out of the accretion disk plane it can radiatively pump emission lines originating within the disk (Eikenberry et al. 1998b). Each infrared flare appears to be accompanied by a delayed radio flare, perhaps indicating adiabatic expansion of the ejected cloud (Mirabel et al. 1998) or motion along a conical jet (Fender & Pooley 1998), while the jet itself is resolved by the Very Long Baseline Array as an optically thick synchrotron source of length  $\sim 20$  AU that is variable on similar timescales as the X-rays and infrared (Dhawan et al. 2000).

Though subsequent observations have revealed smaller, more complex infrared flaring behavior in this source (e.g. Eikenberry et al. 2000), the above represents the basic picture for the ejection of “class B” jets in GRS 1915+105. In this chapter, we present multiwavelength observations of GRS 1915+105 on four nights during July 2002, when the source was undergoing another period of “class B” ejection. The similarities and differences between these observations and those obtained previously allow us to begin to map out how the many different X-ray states of GRS 1915+105 affect jet production in this unusual source. We conclude that the observed ejections may sometimes entirely be a superposition of smaller, more complex phenomena, and that the appearance of a “trigger” spike in the X-ray light curve (seen in the observations of Eikenberry et al. 1998a, but not in our July 2002 observations) is a key ingredient for large, infrared-bright ejections to occur on  $\sim 30$  minute timescales.

## 2.3 Observations

We obtained infrared observations of GRS 1915+105 on the nights of 2002 July 27–29 UT in the K ( $2.2\mu\text{m}$ ) band, using the D-78 camera at the Cassegrain focus of the Palomar Observatory 5 meter Hale telescope. X-ray observations (with coverage between  $\sim 2\text{--}100\text{ keV}$ ) were obtained on July 27–28 and July 30 using the Proportional Counter Array (PCA) on the *Rossi X-Ray Timing Explorer* (RXTE); further details regarding this instrument can be found in Greiner et al. (1996) and references therein.

We configured the infrared camera to take  $128 \times 128$  pixel ( $16'' \times 16''$ ) images at a rate of one frame per second, with absolute timing provided by a WWVB radio signal from the National Institute of Standards and Technology ( $\sim 1\text{ ms}$  accuracy). We observed GRS 1915+105 in this mode for  $\sim 8$  hours on each of the first two nights and for  $\sim 1$  hour on July 29. We processed each image by subtracting an averaged sky frame, dividing by a flat field, interpolating over bad pixels and filtering it in the Fourier domain to remove electronic pattern noise that corrupted many of the images. On July 27, we also observed the faint HST/NICMOS standard GSPC P182-E (star no. 9177 from Persson et al. 1998) for absolute flux calibration.

A typical processed image is shown in Figure 2.1. The field of view is large enough to capture GRS 1915+105 and several nearby stars. We performed differential photometry on GRS 1915+105 and the surrounding field stars, measuring their fluxes within a  $1''$  radius software aperture. We used the two brightest field stars as a reference, averaging their light curves together and dividing the normalized result into the GRS 1915+105 light curve to correct for atmospheric variability and changes in the point spread function. This procedure was also applied to “Star

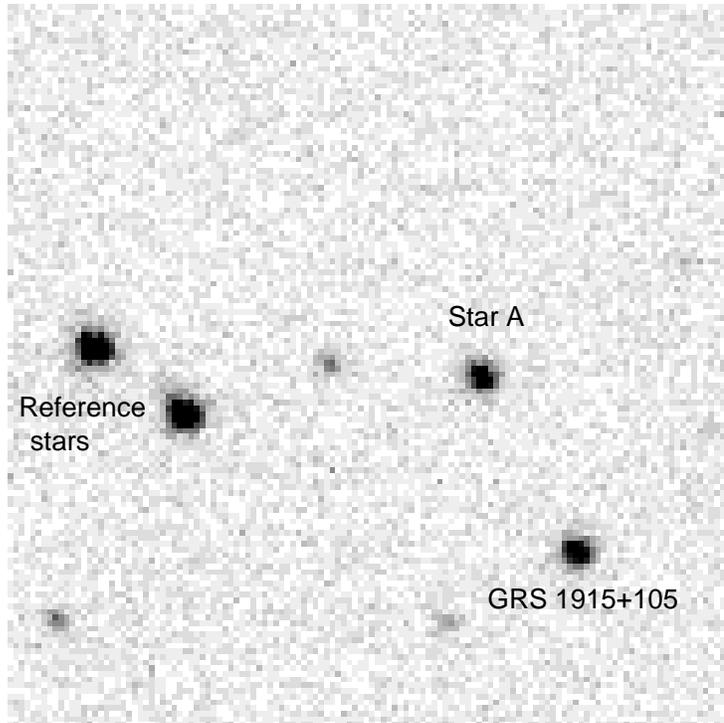


Figure 2.1: A typical  $2.2\mu\text{m}$  infrared image of the field of GRS 1915+105 at 1-second time resolution, from our observations on 2002 July 27. North is up, east is to the left, and the field of view is  $16'' \times 16''$ . The two reference stars were used to correct the GRS 1915+105 and Star A light curves for atmospheric variability and changes in the point spread function.

A'' in Figure 2.1; its flux density remained steady at a value of 3.1 mJy (consistent with the reported value in Fender et al. 1997) and was used to determine the flux density of GRS 1915+105. The GRS 1915+105 light curve was then dereddened by  $A_K = 3.3$  magnitudes to correct for absorption in the Galactic plane (Fender et al. 1997). We chose  $A_K = 3.3$  to be consistent with the previous literature, although more recent work suggests that  $A_K = 2.2$  is a better estimate for this source (Fuchs et al. 2003; Chapuis & Corbel 2004). If  $A_K = 2.2$  is adopted, then all the infrared flux densities in this chapter should be reduced by a factor of  $\sim 3$ .

For the X-ray analysis, we extracted light curves (at 1-second time resolution) from PCA Standard-1 data using FTOOLS v5.2. We also extracted X-ray spectra

in the  $\sim 3\text{--}25$  keV range, using data from PCA binned mode B\_8ms\_16A\_0\_35\_H\_4P and event mode E\_16us\_16B\_36\_1s. We used standard procedures for response matrix generation, background estimation and subtraction, and correction for PCA deadtime. We then used XSPEC v11.2 to fit each spectrum with a standard model for black hole candidates consisting of a “soft” component (which peaks in the low energy X-rays) modeled as a multitemperature disk blackbody (e.g. Mitsuda et al. 1984) and a “hard” component (which extends to the higher energy X-rays) modeled as a power law, both modified by hydrogen absorption fixed to a column density of  $6 \times 10^{22}$  atoms  $\text{cm}^{-2}$  (Muno et al. 1999). A systematic error of 1% was added to each spectrum before the fit was performed.

## 2.4 Light Curves

The resulting light curves from July 27–28 are shown in Figure 2.2. On each night, we detected quasi-periodic infrared flares with dereddened amplitudes of  $\sim 30$  mJy. The flares have durations of  $\sim 15\text{--}30$  minutes and repeat on timescales of  $\sim 30\text{--}60$  minutes. The accompanying X-ray light curves show GRS 1915+105 undergoing a series of long “dips” and fast oscillations. The particular pattern of X-ray variability seen here has been observed many times in GRS 1915+105; it is the “class  $\alpha$ ” state defined by Belloni et al. (2000a), characterized by oscillations which grow longer and fainter with time before the source enters an X-ray dip of duration  $\sim 1000$  seconds.

Our observations on July 29 were limited to  $\sim 1$  hour but showed two infrared flares similar to those seen on July 27–28. This suggests that the state we observed lasted for more than two days. The light curves from July 29 are not shown here because there were no accompanying X-ray observations, but they are available

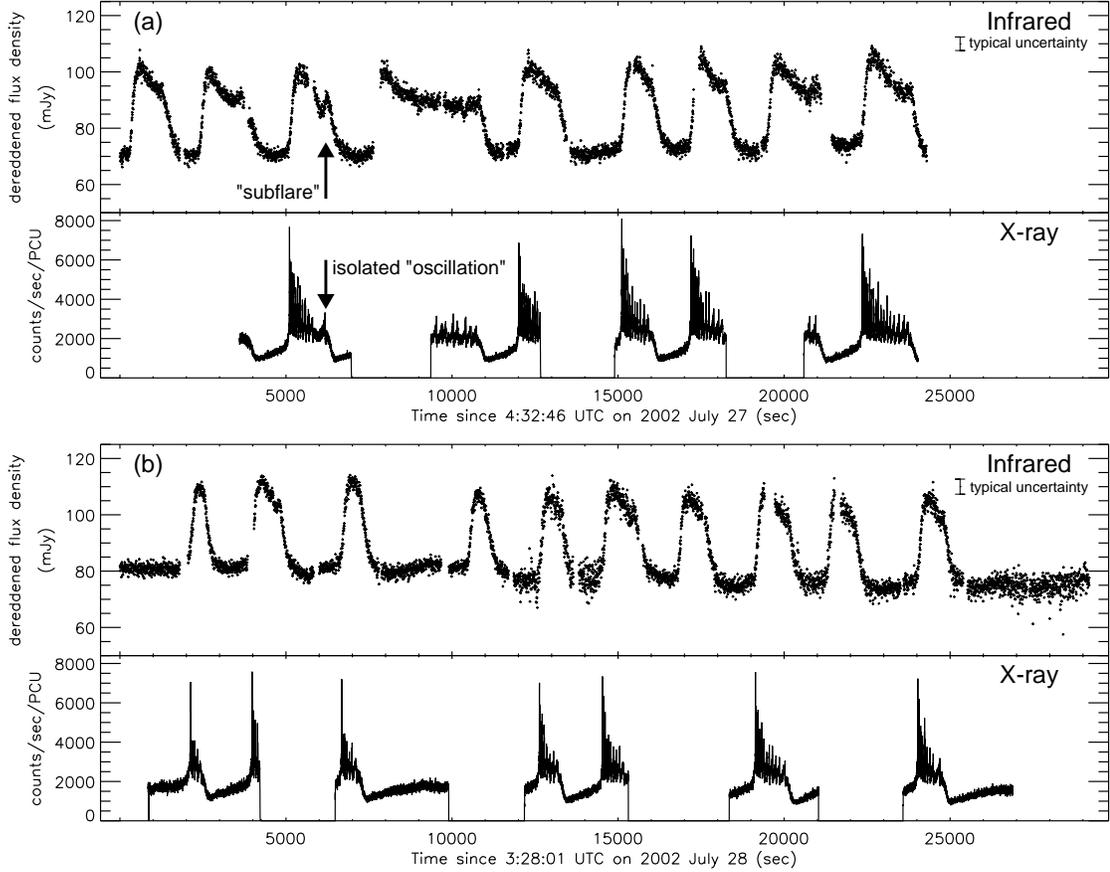


Figure 2.2: Simultaneous infrared and X-ray light curves of GRS 1915+105 on (a) 2002 July 27 and (b) 2002 July 28. Both light curves are at 1-second resolution, but the infrared data have been smoothed to 5-second resolution in this plot. Typical 1-second  $\pm 1\sigma$  uncertainties are shown for the infrared; the X-ray uncertainties are assumed to be Poissonian and are too small to be seen on this scale. The infrared data were taken in the K ( $2.2\mu\text{m}$ ) band and have been dereddened by 3.3 magnitudes, while the X-ray data are in the  $\sim 2\text{--}100$  keV energy range and are plotted in units of counts per second per Proportional Counter Unit (PCU) of the RXTE PCA instrument. Gaps in the X-ray light curves represent times when RXTE was not observing the source. The “subflare” and isolated “oscillation” in (a) are discussed in §2.5.

upon request. X-ray observations on July 30 indicated that GRS 1915+105 was no longer undergoing oscillations and had entered a period of hard, steady X-ray emission (the “class  $\chi$ ” state of Belloni et al. 2000a).

### **2.4.1 Multiwavelength Features of the Light Curves**

As can be seen from Figure 2.2, each time the X-ray light curve of GRS 1915+105 switches from a dip into a period of oscillation, it is accompanied by an infrared flare. Furthermore, the flare appears to be triggered by a sharp X-ray “spike” of duration several seconds which is present during each transition.

These features are broadly consistent with previous multiwavelength observations of GRS 1915+105 during “class B” ejections (e.g. Eikenberry et al. 1998a; Mirabel et al. 1998); in fact, many of the X-ray states classified by Belloni et al. (2000a) in which the source transitions between a long dip and a period of oscillations are now known to trigger radio (and therefore presumably infrared) flares (Klein-Wolt et al. 2002).

What makes the current observations interesting is their high time resolution and broad coverage of many individual episodes of jet production, matched only by the similar observations of Eikenberry et al. (1998a). By looking at how the subtle differences in X-ray evolution between these two sets of observations lead to differences in the infrared behavior, we can begin to unravel the details of the complicated evolution of GRS 1915+105.

### **2.4.2 Origin of the Infrared Flares**

Infrared flares similar to those presented here have been observed many times from GRS 1915+105 and have been consistently interpreted as synchrotron emission

from a jet (Fender et al. 1997; Fender & Pooley 1998; Eikenberry et al. 1998a; Mirabel et al. 1998; Ueda et al. 2002). We propose that the flares in Figure 2.2 have the same origin.

This interpretation is supported by 15 GHz observations at the Ryle Telescope which were obtained several hours before our July 28 observations (G. Pooley 2002, private communication). Radio flares with similar timescales as the infrared flares and amplitudes of  $\sim 10$ – $20$  mJy (above a baseline flux density of  $\sim 10$  mJy) were observed. If the infrared flares we observed had similar radio counterparts, then the relatively flat infrared-to-radio spectrum (as well as the  $\sim 10^9$  to  $10^{10}$  K brightness temperatures of the radio flares) indicates a nonthermal origin.

## 2.5 Infrared/X-Ray Correlation

The most interesting result from these observations may be the strong correlation between the duration of each infrared flare and the duration of its accompanying X-ray oscillation period. All the flares have similar rise times ( $\sim 200$ – $300$  seconds), but each flare appears to “wait” to return to its quiescent level until the X-rays stop oscillating. This gives rise to asymmetric profiles for some of the longer flares, particularly those on July 27.

How does this behavior compare to that seen in previous observations of GRS 1915+105? The flares observed in August 1997 by Eikenberry et al. (1998a) were a few times stronger than those presented here, with peak amplitudes as high as  $\sim 200$  mJy dereddened. A correlation between the infrared flare duration and the X-ray oscillation period was not directly evident in 1997, and a few flares appear to have decayed almost completely while the X-rays were still undergoing strong oscillations. An example of this behavior is shown in Figure 2.3a.

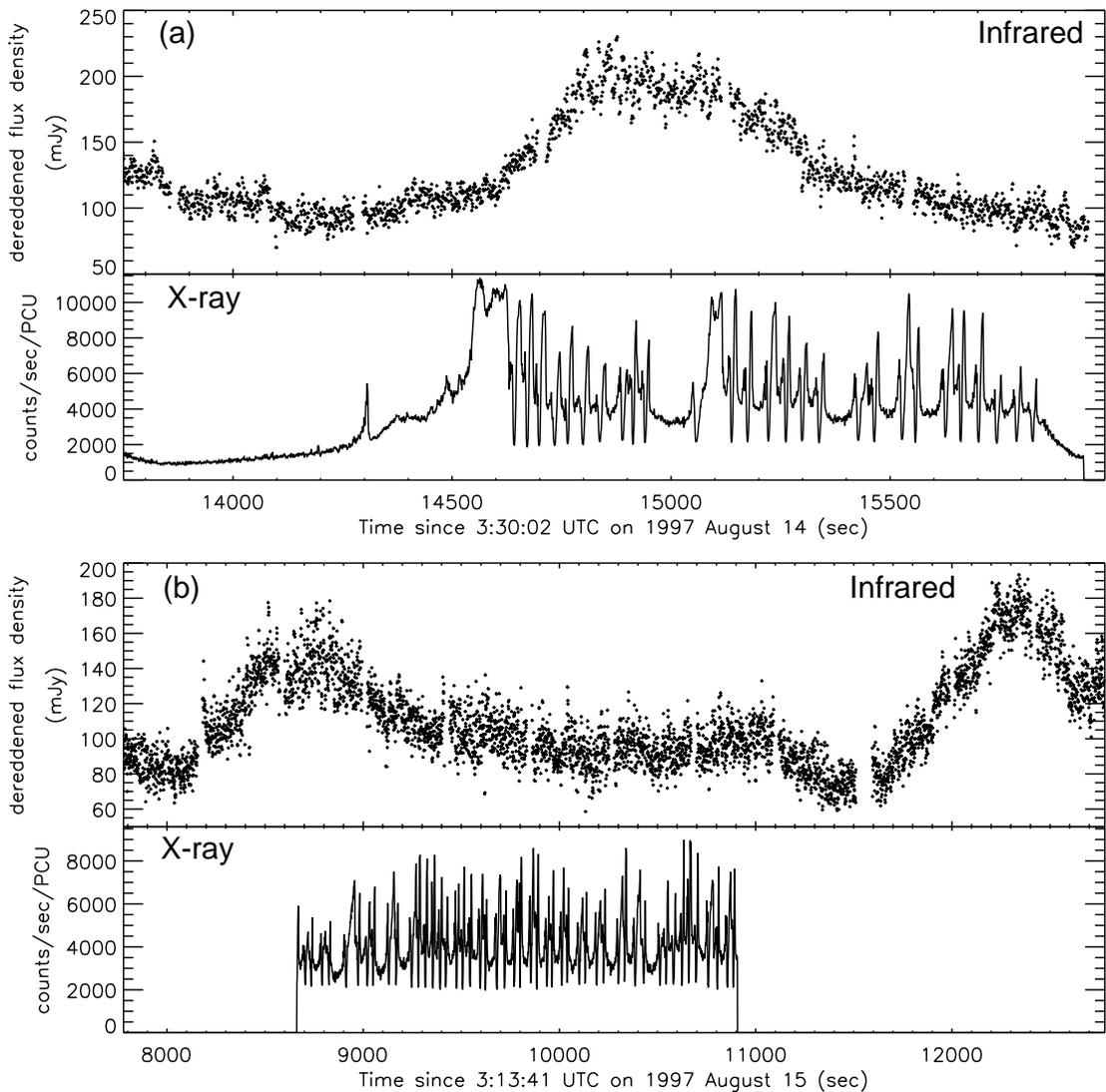


Figure 2.3: Simultaneous infrared and X-ray light curves of GRS 1915+105 at 1-second resolution during August 1997, from data originally presented in Eikenberry et al. (1998a). The infrared data have been dereddened by 3.3 magnitudes before plotting. In (a), the infrared flare appears to end and the flux density decays back to its quiescent level while the X-rays are still undergoing oscillations. In (b), there is an infrared flare beginning at  $\sim 8,000$  seconds which appears to decouple from the X-rays (though there is no X-ray coverage for the initial flare rise), but there is also a  $\sim 20$  mJy “infrared excess” that continues until  $\sim 11,000$  seconds and that is associated with an unusually long series of X-ray oscillations.

However, there is some evidence for a smaller contribution to the infrared light curve (on the  $\sim 20$  mJy level) associated with the 1997 X-ray oscillations. In particular, a long series of oscillations occurred on August 15 at the same time as a period of “infrared excess” following a flare, in which the GRS 1915+105 flux density remained steady at a level above its quiescent value (see Figure 2.3b). Eikenberry et al. (2000) showed that if each X-ray oscillation in this series had a  $\sim 5$ –10 mJy infrared flare associated with it, then the superposition of these flares could reproduce the observed infrared excess. Their conclusion was based on the spectral similarity of the 1997 oscillations to isolated X-ray oscillations in July 1998, each of which had an accompanying, time-resolved  $\sim 5$ –10 mJy infrared flare (the “class C” flares).

The observations presented in our current work strengthen the argument made by Eikenberry et al. (2000). The key piece of evidence comes from the third infrared flare on July 27 (Figure 2.2a). This flare contains clear evidence for a “subflare”—a secondary peak that occurs after the flare has begun to decay. Furthermore, the subflare is associated with an isolated oscillation in the X-ray light curve that occurs after an unusually long delay between it and the previous set of oscillations. It is therefore plausible that each X-ray oscillation has a faint infrared subflare associated with it, but in general the oscillations are so closely spaced in time that the subflares blend together to create the illusion of one continuous flare.<sup>1</sup>

Unlike in the 1997 data, where these faint flares simply contribute an “infrared excess” on top of a larger episode of flaring behavior, we propose that the  $\sim 30$  minute flares in 2002 are composed *entirely* of these superimposed events. This

---

<sup>1</sup>Note that there are a few other infrared flares in Figure 2.2 that appear to have very faint subflares, but none of these events have good enough X-ray coverage to determine whether they are associated with isolated X-ray oscillations.

would immediately explain the correlation between the infrared flare duration and the X-ray oscillation period seen in 2002; the flare is only sustained by the continual production of subflares associated with the X-ray oscillations.

To test whether this idea is feasible, we show in Figure 2.4 the number of X-ray oscillations in each  $\sim 30$  minute cycle plotted against the integrated flux density of the infrared flare accompanying that cycle. (The integrated flux densities were calculated by smoothing the light curve to 30 second resolution, determining the start and end times of each flare by eye, subtracting out the baseline and integrating the result numerically.) It is clear from Figure 2.4 that X-ray cycles with more oscillations produce more infrared emission, consistent with our suggestion. Figure 2.4 also shows linear fits to the data; the best fit line has a slope of  $\sim 1.9$  Jy sec oscillation $^{-1}$  and a vertical intercept of  $\sim 3.6$  Jy sec, while the best fit line constrained to pass through the origin has a slope of  $\sim 2.2$  Jy sec oscillation $^{-1}$ . Both lines appear to be reasonable fits, and if we assume a plausible value of  $\sim 1.5$  Jy sec for our measurement errors in determining the integrated flux densities, then both fits are found to be statistically acceptable, with reduced chi-squared values of  $\chi^2_{\nu} \sim 0.4$  and 1.1, respectively.

We also tried fitting a power law to the data in Figure 2.4 and found a slightly nonlinear relationship (power law index  $\sim 0.8$ ), but the quality of the fit was not significantly better than that of the linear fit ( $\chi^2_{\nu} \sim 0.34$ ). Therefore, the simplest conclusion we can draw from our observations is that there is a roughly linear relationship between the number of X-ray oscillations in each  $\sim 30$  minute cycle and the strength of the infrared flare accompanying that cycle. In addition, we tested whether the infrared flare strength might actually be related to the total amount of X-ray emission during the oscillations (rather than the number of

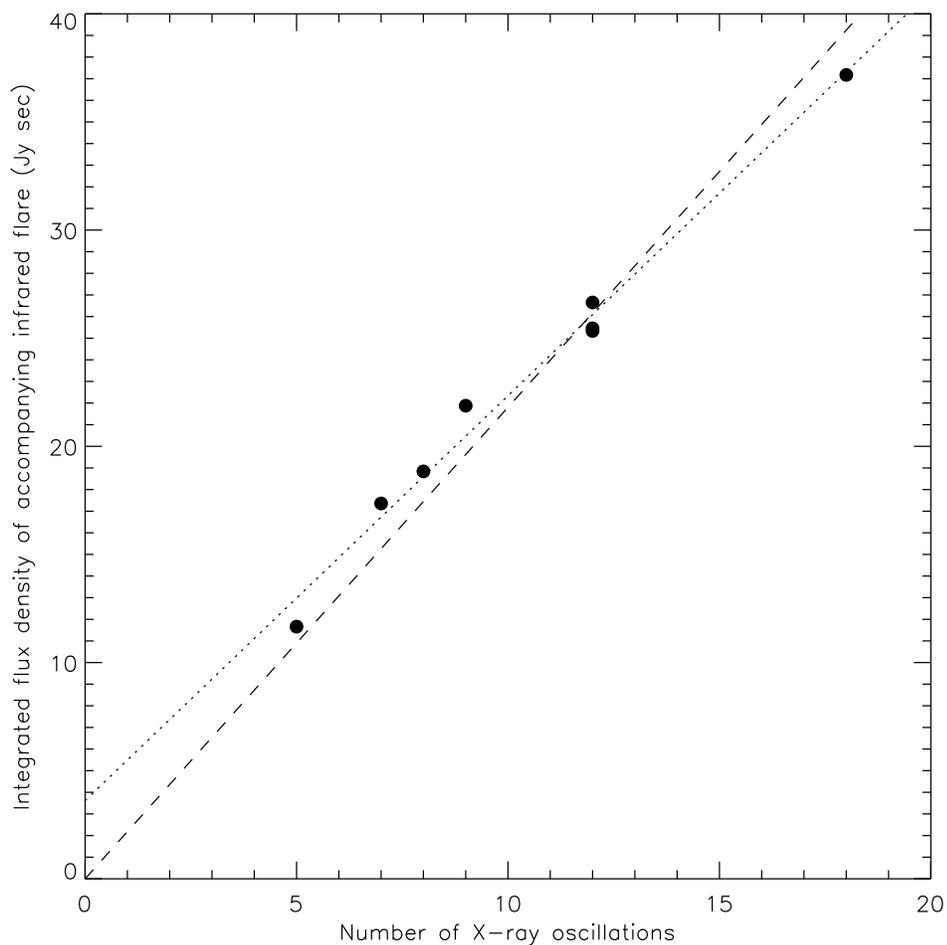


Figure 2.4: The number of X-ray oscillations in each  $\sim 30$  minute cycle plotted against the integrated flux density of the infrared flare accompanying that cycle, from our observations on 2002 July 27–28. Data from all eight infrared flares that had uninterrupted X-ray coverage are shown in this figure, plotted as points. (Note that two of the points with 12 X-ray oscillations are difficult to discern because they almost completely overlap.) The dotted line shows the best linear fit to the data, and the dashed line shows the best linear fit constrained to pass through the origin.

oscillations) and found no evidence to support this; if the infrared flare strength is plotted against the integrated X-ray flux in each cycle (above a baseline level of 1800 counts/sec/PCU), the correlation has a larger scatter than that shown in Figure 2.4, and both linear and power law fits to the data are poor ( $\chi^2_\nu > 2$ ).

The simplest interpretation of the linear relationship in Figure 2.4 is that each X-ray oscillation contributes a constant amount of infrared emission to the overall light curve, in the form of a faint flare. In addition, the fact that the vertical intercept of our fit is small (equivalent to an infrared excess of only a few mJy over the typical duration of a flare) and possibly consistent with zero supports our suggestion that the infrared events are composed *entirely* of these superimposed faint flares. Furthermore, the slope that we measure from our fit indicates that each X-ray oscillation should be responsible for  $\sim 2$  Jy sec of infrared emission. This number is slightly larger than, but still consistent with, the integrated flux density in the subflare we observed (between  $\sim 0.5$  and 2 Jy sec, depending on how we choose to distinguish the “subflare” from the underlying main flare). It may also be reasonable to assume that an average X-ray oscillation triggers a slightly larger subflare than the one we observed, since this subflare was associated with one of the weakest oscillations in its series.

To further test our idea, we attempted to simulate the infrared light curve as a superposition of faint subflares. We began our simulation with a constant infrared background level of  $\sim 70$  mJy and added one faint flare to this background for each oscillation in the X-ray data. We modeled each flare as a Gaussian with fixed amplitude and full width at half maximum (FWHM). Based on the observed properties of the July 27 subflare, we chose the position of each Gaussian so that its peak occurred 60 seconds after the peak of its accompanying X-ray oscillation.

We performed several simulations with different, fixed values of the amplitude and FWHM, since these quantities were hard to determine from the observed subflare.

In general, these simulations did not fit the data well; the superimposed flares we produced reached peaks which were initially too high or dropped to values which were too low by the end of the  $\sim 30$  minute cycle. We therefore decided to allow the FWHM of the flares to vary based on the duration of each X-ray oscillation. We chose a simple linear scaling, in which the FWHM of each flare was taken to be proportional to the rise time of its accompanying X-ray oscillation. Results of this simulation for the sixth infrared flare on July 27 are shown in Figure 2.5; in this case, we fixed the amplitude of each Gaussian to be 8 mJy and set each Gaussian's FWHM to be a factor of 5 greater than the rise time of its accompanying X-ray oscillation. This choice guaranteed that a typical oscillation in each X-ray series, with a rise time of  $\sim 50$  seconds, had an associated infrared flare with a FWHM of 250 seconds and an integrated flux density of  $\sim 2$  Jy sec, consistent with the slope of the fit in Figure 2.4.<sup>2</sup>

Simulations such as the one in Figure 2.5 do a good job of reproducing the observations, especially given the gross approximations we are making. Even so, some difficulties are encountered. The July 27 subflare is not reproduced using the above procedure; its accompanying X-ray oscillation has an extremely long rise time, and therefore the simulation tries to model it with an unrealistically large FWHM. In general, the simulations work better at the beginning of each  $\sim 30$  minute cycle than at the end, suggesting that there is residual infrared emission

---

<sup>2</sup>Note that by allowing the FWHM to vary, we are also varying the amount of infrared emission associated with each X-ray oscillation. However, this does not contradict the linear relationship found in Figure 2.4; it turns out that each X-ray cycle has about an equal number of oscillations with rise times above and below the typical value of  $\sim 50$  seconds, so the stronger infrared flares may “cancel out” the weaker ones and still lead to a roughly linear relationship.

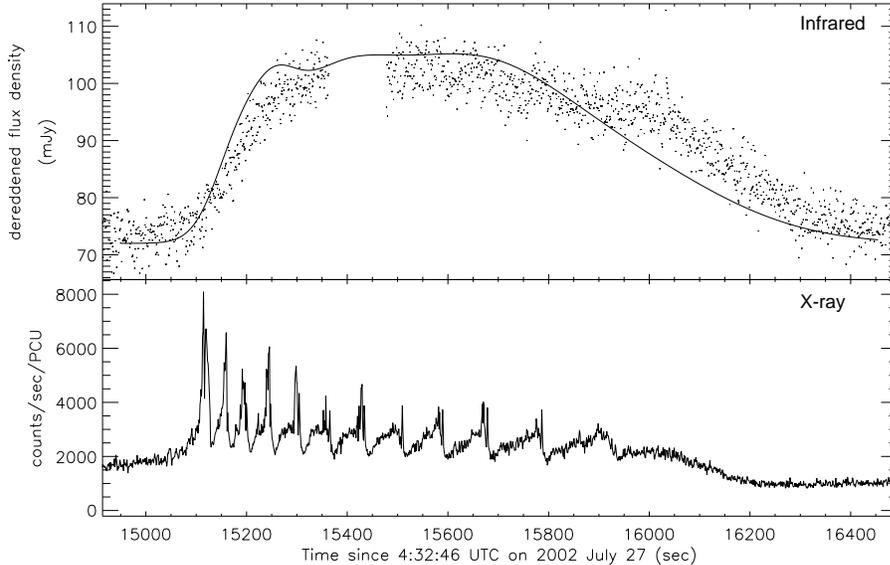


Figure 2.5: Simultaneous infrared and X-ray light curves of GRS 1915+105 at the time of the sixth infrared flare observed on 2002 July 27 (Figure 2.2a). The infrared data (dereddened by 3.3 magnitudes) are plotted as points at 1-second resolution; the curve in the top panel shows the results of our simulation (discussed in §2.5), in which the infrared flare is modeled as a superposition of small flares, one for each X-ray oscillation.

associated with the end of each X-ray cycle that is not captured in our model. Nonetheless, the results shown in Figure 2.5 are fairly typical, and we can be confident that the 2002 flares consist primarily of superimposed faint flares associated with each X-ray oscillation.

## 2.6 X-Ray Spectral Evolution

We can compare our X-ray light curves and spectra to those obtained in August 1997 by Eikenberry et al. (1998a). Though both sets of observations show transitions between X-ray dips and oscillations, the details of the behavior are different; in the classification of Belloni et al. (2000a), GRS 1915+105 was in the “class  $\beta$ ” state in 1997 and the “class  $\alpha$ ” state in 2002.

In Figure 2.6, we present one cycle of jet formation from each of the 1997 and

2002 observations. Several parameters obtained from our spectral fitting are plotted at 4-second resolution: the unabsorbed 2–25 keV flux from the multitemperature disk blackbody and power law components of the spectrum, the temperature at the inner edge of the accretion disk  $T_{in}$ , the radius of the inner edge  $R_{in}$ , and the power law index  $\gamma$  (where the photon flux density due to the power law, at photon energy  $E$ , is  $\propto E^{-\gamma}$ ). The fits shown in Figure 2.6 have typical reduced chi-squared values of  $\chi^2_{\nu} \sim 1.2$ , and all fits with  $\chi^2_{\nu} > 2$ , as well as those for which XSPEC was unable to return uncertainties on the fit parameters, were excluded from the plot. Note that  $R_{in}$  is calculated from the disk blackbody normalization ( $R_{in} \propto \sqrt{N}$ , where  $N$  is the normalization) assuming a distance to GRS 1915+105 of 11 kpc and an accretion disk inclined  $66^\circ$  to the line of sight (Fender et al. 1999b), but the values derived from this procedure should not be viewed as accurate estimates of the physical size of the disk (see §2.6.2.2 and Merloni et al. 2000).

### 2.6.1 1997 and 2002 State Changes

Figure 2.6 shows that in both 1997 and 2002, the X-ray spike which appears near the beginning of the infrared flare (at  $\sim 2,200$  seconds in 1997 and  $\sim 24,000$  seconds in 2002) represents a “state change” within the accretion disk, in the sense that it initiates changes that persist after the spike is finished. During the X-ray dip, the source is in a spectrally hard state in which the disk blackbody component is nearly absent, but the increase in  $T_{in}$  and in the disk emission which occur at the spike indicate that the disk becomes hot and visible at this point and remains that way throughout the X-ray oscillations.

In 1997, there is a prolonged period of a few hundred seconds after the X-ray spike when the inner accretion disk has “turned on” but before the period of fast

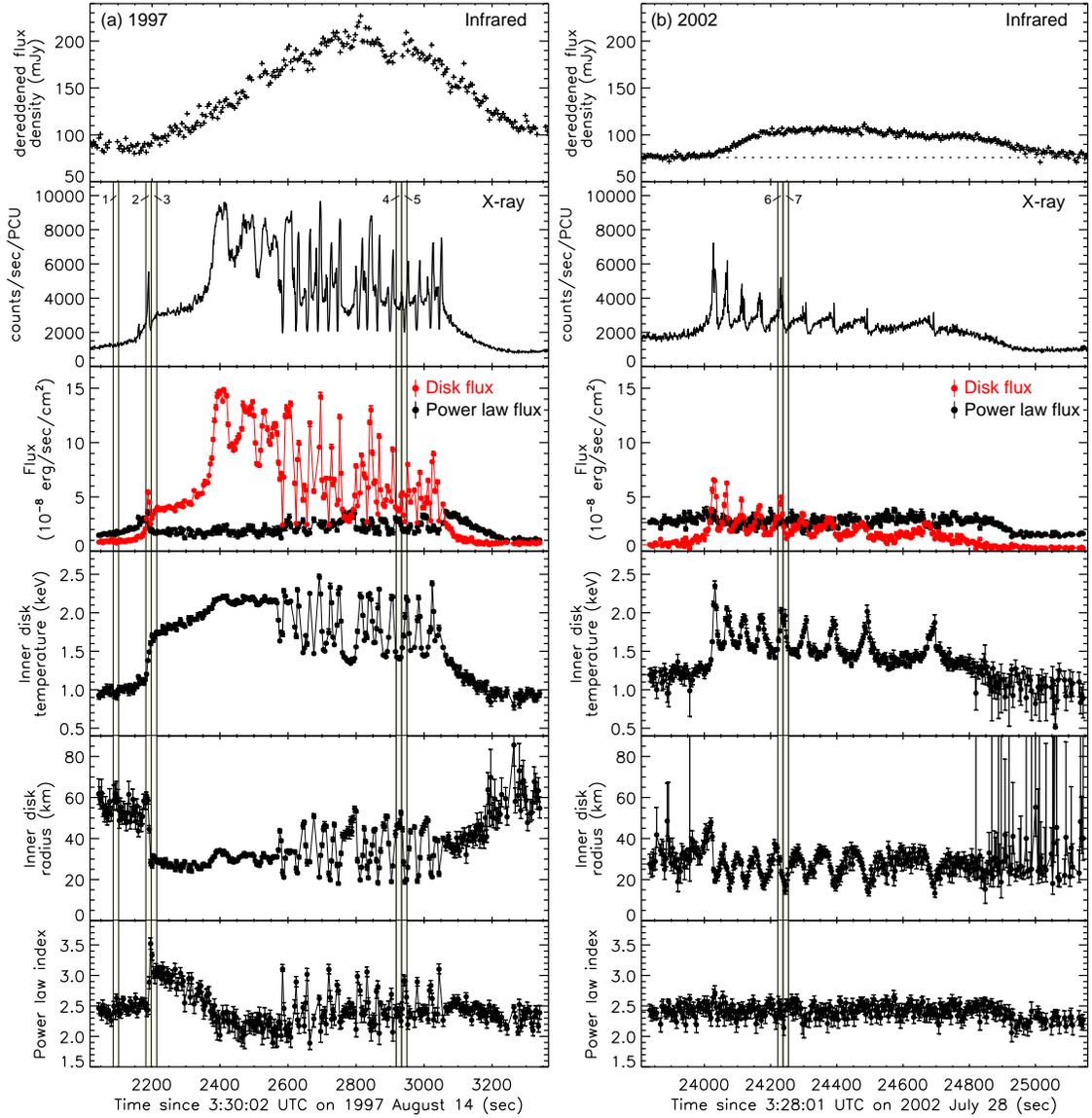


Figure 2.6: Comparison of one cycle of jet formation in GRS 1915+105 from observations in (a) 1997 and (b) 2002. The top two panels in each column show the infrared (dereddened by 3.3 magnitudes) and X-ray light curves, respectively (at 1-second resolution, but the infrared data have been smoothed to 5-second resolution in this plot); the dotted horizontal line in (b) shows the baseline infrared level. The bottom four panels show the results of our X-ray spectral fitting at 4-second resolution, plotted along with  $\pm 1\sigma$  uncertainties; these include the unabsorbed 2–25 keV flux from the multitemperature disk blackbody and power law components of the spectrum, the temperature at the inner edge of the accretion disk, the radius of the inner edge of the accretion disk, and the power law index. The vertical shaded regions numbered 1 through 7 show the parts of the light curve from which the spectra plotted in Figure 2.8 were extracted.

oscillations. It is clear that the infrared flare begins well before the oscillations, so although we may be able to ascribe some of the infrared emission to blended together “subflares” associated with each oscillation (as discussed in §2.5), there must be some other mechanism at work which leads to the initial rise of the infrared light curve right around the moment of the spike and allows it to reach a peak amplitude of  $\sim 100$  mJy.

In 2002, however, the spike seems to represent the first oscillation in a series; it is spectrally similar to the oscillations which immediately follow it, and differs mainly in that it is the strongest and fastest oscillation in the group. This is consistent with the idea that the flares observed in 2002 consist entirely of superimposed  $\sim 5$ – $10$  mJy subflares associated with each oscillation, and that no abnormally large ejection occurs at the moment of the first spike.

The differences between the 1997 and 2002 behavior can be more clearly seen in Figure 2.7, where we have plotted the results of X-ray spectral fitting performed at high time resolution (1 second). Again, all fits with  $\chi^2_\nu > 2$ , as well as those for which XSPEC was unable to return uncertainties on the fit parameters, were excluded from the plot. Figure 2.7a shows the initial X-ray spike in 1997, Figure 2.7b shows the initial X-ray spike in 2002 and the oscillations which immediately follow it, and Figure 2.7c shows a typical set of oscillations in 1997.

It is clear from Figure 2.7 that although the details of the 1997 and 2002 oscillations are different, there are also many similarities: both consist primarily of periodic changes in the disk emission and  $T_{in}$ . The initial oscillation in 2002, which signals the beginning of the disk activity and coincides with the rise of the infrared flare, is no different in this regard. Figure 2.7a, however, shows that the initial spike in 1997 (which we refer to as the “trigger” spike) is an entirely

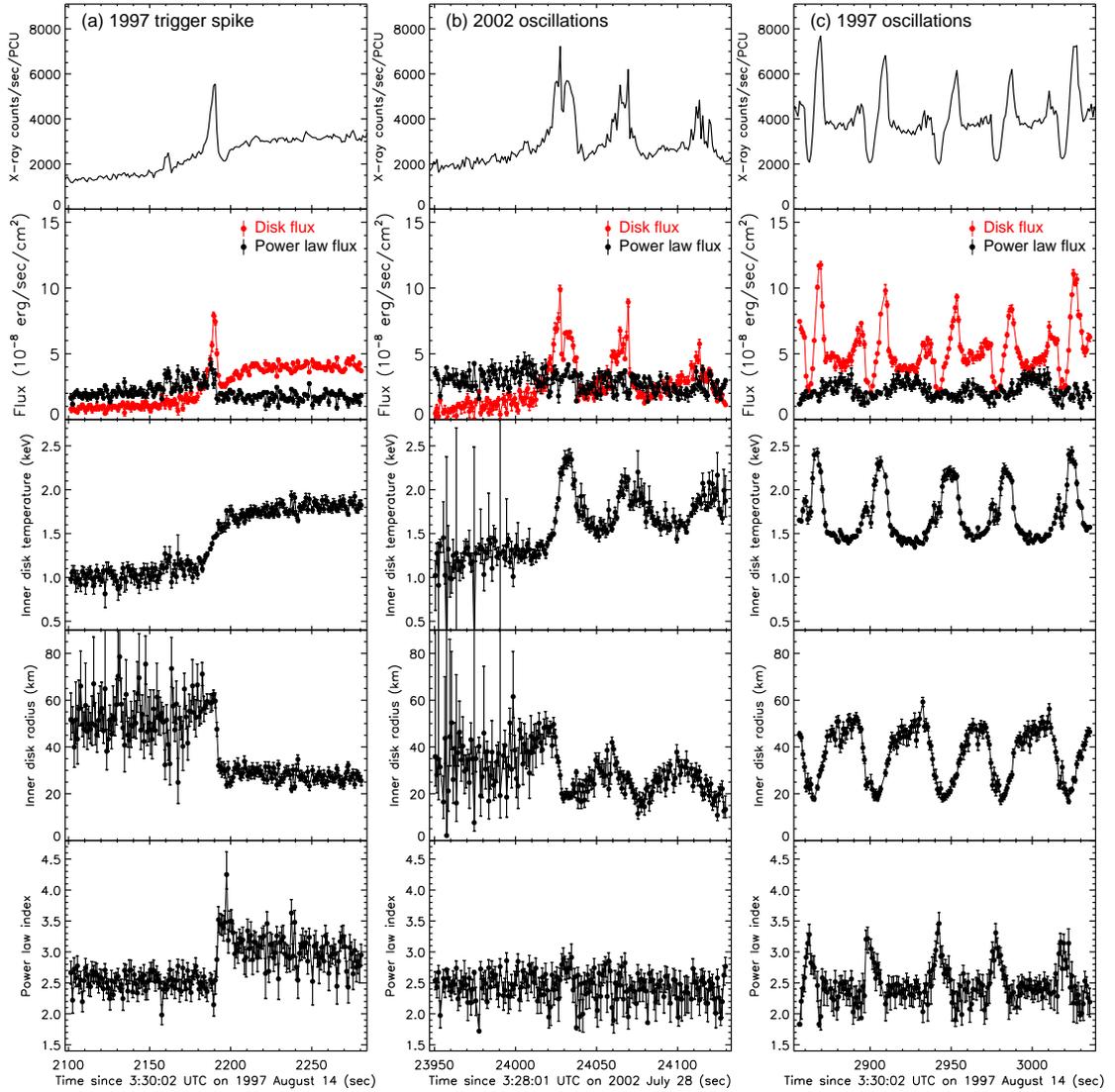


Figure 2.7: Comparison of the X-ray spectral evolution of GRS 1915+105 during (a) the 1997 trigger spike (from Figure 2.6a), (b) the initial spike in 2002 and the oscillations which immediately follow it (Figure 2.6b), and (c) a typical set of oscillations in 1997 (Figure 2.6a). The top panel in each column shows the X-ray light curve (at 1-second resolution), and the bottom four panels show the same X-ray spectral fitting parameters plotted in Figure 2.6, but the fits shown here were performed at 1-second resolution. It is clear from this figure that the trigger spike in (a) is spectrally different from the other oscillations in 1997 and 2002.

different phenomenon. During this spike,  $T_{in}$  rises monotonically to  $\sim 1.8$  keV, where it remains, while the power law component of the spectrum does not vary noticeably during the spike but changes sharply as the spike ends. Similar behavior was observed by Migliari & Belloni (2003), but our spectral fits are performed at higher time resolution and therefore confirm that the power law does not change significantly even during the last few seconds of the spike’s rise.

### 2.6.2 1997 Trigger Spike

Given the importance of the 1997 trigger spike and its possible connection to large ejections in GRS 1915+105, it is useful to examine its evolution in a model-independent way. This is shown in Figure 2.8a, where we plot three X-ray spectra associated with the trigger spike (the parts of the X-ray light curve from which these spectra have been extracted are indicated by the vertical shaded regions numbered 1 through 3 in Figure 2.6a). We extracted the spectra at a time resolution of 16 seconds using PCA Standard-2 data, which have higher energy resolution than the binned and event mode data used for the fitting procedure.

As can be seen from Figure 2.8a, between regions 1 and 2 (as the spike begins) there is virtually no change in the hard ( $\gtrsim 20$  keV) X-ray emission, but the soft X-rays brighten dramatically. After the drop at the end of the spike, however (region 3), the spectrum steepens sharply, and the hard X-rays become nearly undetectable. The soft X-rays in region 3 retain a similar spectral shape as in region 2, but decrease in intensity.

In this model-independent description, the trigger spike consists of:

1. A factor of a few increase in the soft X-ray emission on a timescale of  $\sim 20$  seconds or longer.

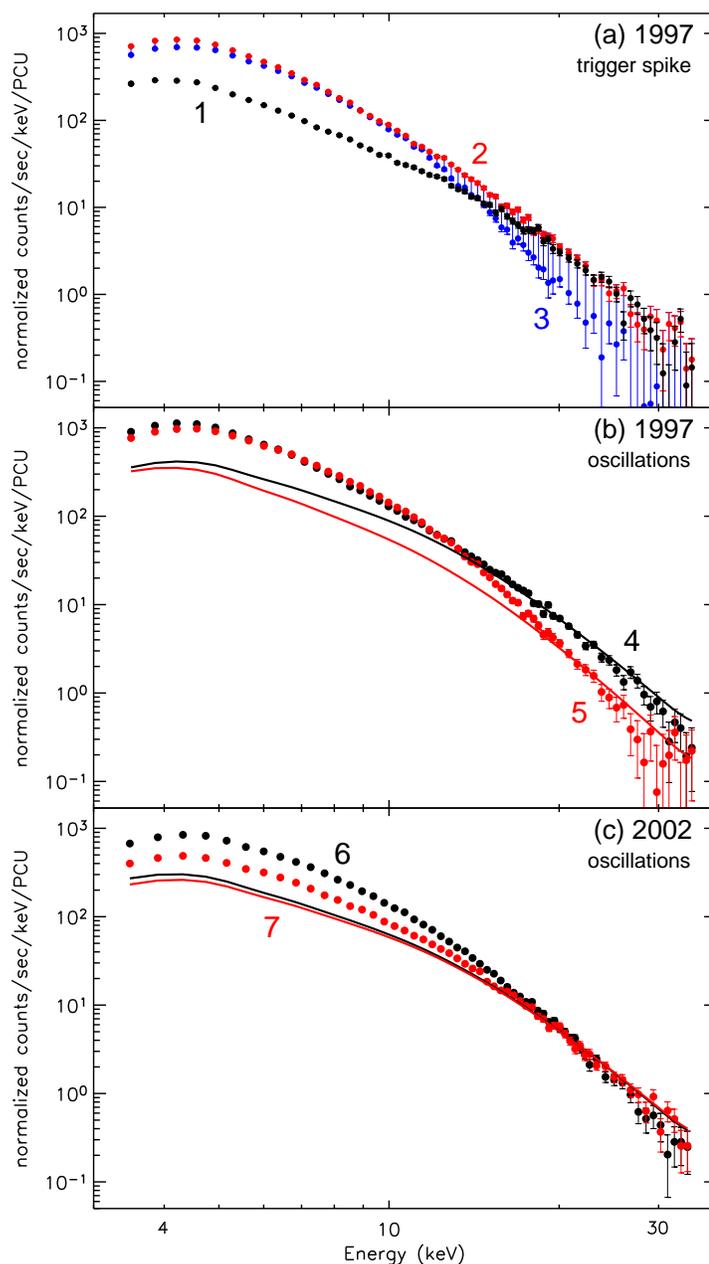


Figure 2.8: X-ray spectra of GRS 1915+105 during (a) the 1997 trigger spike in Figure 2.6a, (b) the 1997 oscillations in Figure 2.6a, and (c) the 2002 oscillations in Figure 2.6b. The spectra are numbered 1 through 7, corresponding to the regions of the X-ray light curve from which they were extracted (see Figure 2.6). The spectra were obtained at 16-second resolution and are plotted as points with  $\pm 1\sigma$  uncertainties; the curves in (b) and (c) show the power law component of the fit to each spectrum, modified by hydrogen absorption (this component accounts for all of the hard X-ray emission and a fraction of the soft X-ray emission).

2. A dramatic, factor of a few decrease in the hard X-ray emission and less dramatic decrease in the soft X-ray emission, on a timescale of a few seconds.

How do these changes in the soft and hard X-ray emission during the trigger spike relate to each other physically? There are several possible ways to understand this, depending on how much we believe the accuracy of the multitemperature disk blackbody model that we have employed above.

### 2.6.2.1 Nonthermal Models for the Spike

Although our spectral fitting indicates that the soft X-ray excess during the trigger spike consists of thermal emission from the accretion disk, this is not an ironclad conclusion. The energy resolution of our data is quite poor, and it is possible that other models could reproduce this brief burst of soft X-ray emission. Before the spike, the accretion disk component of the spectrum is extremely weak, and in fact, we are able to obtain acceptable fits at 1-second resolution ( $\chi^2_\nu \sim 1.4$ ) for a model that only includes a power law and no accretion disk. Thus, it is possible to conclude that the spike is simply a “blip” on an otherwise smooth rise of the disk flux that corresponds (along with the smooth rise in  $T_{in}$ ) to the accretion disk “turning on.” In this picture, when the 1997 X-ray light curve transitions from a dip into a period of oscillations, the thermal emission from the disk *gradually* becomes visible (as it does in 2002). At some point during this transition, however, a  $\sim 10$  second nonthermal burst of soft X-rays is released, creating the spike. This interpretation certainly seems plausible in light of the second panel of Figure 2.7a, which shows the disk flux rising smoothly before the spike and finishing its smooth rise afterwards; the spike can easily be imagined as a sharp feature superimposed on an otherwise smooth event.

If this is the case, then when the nonthermal soft X-ray emission disappears at the end of the spike, the nonthermal hard X-ray emission (from the power law component of the spectrum) is also observed to drop (Figure 2.8a). The origin of the power law in X-ray binaries is a subject of some debate; it is generally assumed to represent either inverse Compton emission from a corona of relativistic electrons (Poutanen 1998) which may constitute the base of a jet (Fender et al. 1999a), or the high frequency tail of synchrotron emission from the jet itself (Markoff et al. 2001, 2003). In either case, it comes from a region close to the accretion disk, so the decrease in the power law emission at the end of the spike indicates that changes are occurring immediately outside of the accretion disk at this time. Meanwhile, the infrared flare also begins, as energy is dissipated in the jet further downstream from the disk.

To test the nonthermal scenario, we tried fitting the soft X-ray emission during the spike as a power law (that is, we fit two power laws to the spectrum, the original one which extended to the hard X-rays, and a stronger, steeper one which covered the soft X-ray excess) but were not able to obtain acceptable fits at 1-second resolution. This implies the need for a “curved” spectral component to explain the emission during the spike; however, there are many possible sources for such a component besides thermal emission from the disk.

The above suggestions are tentative, and much more data would be required to confirm them. Nevertheless, a logical picture can be constructed in which the spike does not originate from thermal emission in the accretion disk. This picture would be consistent with several recent models, including the “magnetic bomb” (Eikenberry & van Putten 2003) and “magnetic flood” (Tagger et al. 2004) models for GRS 1915+105 and the more general model proposed by Livio et al. (2003).

These models all evoke the existence of large scale magnetic fields during the X-ray dip and associate the state transition at the end of the dip with the destruction of these magnetic fields, perhaps through a reconnection event that produces a  $\sim 10$  second burst of nonthermal emission along with an ejection of material.

It is worth pointing out that our 2002 observations strengthen some of the above models, since they show that the large ejections ( $\sim 100$  mJy infrared flares which decouple from the X-ray light curve) only seem to occur in the presence of the trigger spike. However, they also show that the  $\sim 30$  minute state transitions can occur *without* an initial large infrared event. Why is this the case? If the spike and ejection represent the destruction of the magnetic field which triggers the state transition, then why do state transitions sometimes seem to require a trigger spike and ejection  $\sim 300$  seconds before the fast oscillations, but other times occur without them and jump immediately into the oscillations? Ideally, a model for the GRS 1915+105 behavior should answer these questions.

### 2.6.2.2 Thermal Accretion Disk Models for the Spike

If we instead elect to believe that the soft X-ray excess during the trigger spike is due to thermal emission from the accretion disk, we need to explain how the disk interacts with the corona or jet (which is responsible for the power law component of the spectrum) in order to produce the observed emission.

Yadav (2001) suggested a general picture for the spike in which the soft X-ray luminosity increases until it reaches a critical value, thus triggering the ejection and subsequent changes in the light curve and spectrum. In this picture, all of the “action” takes place at the end of the spike; at this time, the X-ray power law and infrared light curve change in a way that is consistent with jet ejection (just

as they did in the nonthermal spike scenario of §2.6.2.1), but this change is now accompanied by a sharp drop in the accretion disk luminosity. This scenario has some advantages for observations in which the trigger spike does not have a well-defined beginning but rather seems to rise gradually to its maximum throughout the preceding X-ray dip (see, for example, the observation presented in Migliari & Belloni 2003).

What could cause the accretion disk luminosity to decrease at the end of the spike, at the same time as the power law component changes? One possibility is that the drop in disk luminosity indicates a sharp increase in the fraction of the accretion power which is being dissipated in the GRS 1915+105 jet. Although detailed modeling of the disk spectrum would be required to confirm this suggestion, it is consistent with the changes observed in the disk parameters at the end of the spike (as discussed below), and it is also consistent with the observed changes in the X-ray power law and infrared light curve, which show that material may be ejected into the jet at the end of the spike.

To study in detail the changes in the disk parameters at the end of the spike, it is useful to examine Figure 2.7a. As can be seen from this figure, it is only the flux of radiation coming from the accretion disk which changes dramatically at the end of the spike, not the basic shape of the disk spectrum; the end of the spike occurs because of a sharp drop in  $R_{in}$  over a few seconds, while  $T_{in}$  increases over a slower timescale ( $\sim 20$  seconds) throughout the spike.

This behavior is puzzling. In general, changes in the disk parameters occur on similar timescales, in which case an increase in  $T_{in}$  coupled with a decrease in  $R_{in}$  can be understood by a model in which the physical conditions in the outer part of the disk remain similar but the inner part of the disk has “turned

on,” either by refilling with matter or by beginning to emit detectable radiation (Belloni et al. 1997a,b). This scenario has been successfully simulated by Watarai & Mineshige (2003); the decrease in  $R_{in}$  is due to the fact that at high accretion rates (where the “slim disk” model of Abramowicz et al. 1988 applies), significant thermal radiation can emerge from inside the last stable circular orbit around the black hole (Watarai et al. 2000). However, the fact that in our observations the drop in  $R_{in}$  occurs *rapidly* at the end of the spike, without any corresponding rapid changes in  $T_{in}$ , is more difficult to understand, and it is therefore necessary to look for other interpretations for what is happening to the radiation from the accretion disk at the end of the spike.

Rodriguez et al. (2002) point out that when the normalization of the multitemperature disk blackbody is observed to drop, the simplest interpretation is that the effective area of the emitting region of the disk has decreased; the calculation of  $R_{in}$  from the normalization is model-dependent and not necessarily reliable. A drop in the fitted value of  $R_{in}$  can therefore be interpreted as the appearance of a hot spot or spiral shock pattern in the disk (Rodriguez et al. 2002), or more generally as a situation where some representative portion of the disk stops emitting radiation which we receive in the soft X-ray band (see also Munro et al. 2001). In this vein, Merloni et al. (2000) showed that if a more realistic model of the disk flow is used to simulate the X-ray spectrum, then changes in several parameters, including the fraction of the accretion power dissipated in regions outside of the disk (for example, in a corona or jet), could lead to apparent changes of  $R_{in}$  by a factor of  $\sim 2$  when the multitemperature disk blackbody model is used to fit the spectrum.

It is therefore plausible that the end of the spike coincides with a transfer of accretion power from the disk (where it is dissipated thermally) to the jet (where it

initiates changes in the X-ray power law and infrared flare). During the  $\sim 300$  second break between the spike and the subsequent oscillations, the disk then settles into the standard “slim disk” state, where the oscillations can occur. However, as in §2.6.2.1, it is still not clear from this scenario what physical parameter triggers the sharp transition and jet ejection in our 1997 observations but not in 2002.

### 2.6.2.3 The Spike in the Context of Other X-ray Binaries

Fender et al. (2004) have proposed a unified model for jet formation in X-ray binary systems. The model is “unified” in the sense that the trigger spike in GRS 1915+105 is shown to have counterparts in several other sources, where a radio flare is produced around the time of a peak in the X-ray light curve that occurs during the transition from a hard to soft spectral state. Fender et al. (2004) proposed a physical model in which a steady jet exists during the spectrally hard X-ray state preceding the spike and the speed of the jet increases sharply as the X-ray spectrum becomes softer; the jet is then quenched completely once the softness reaches a critical value. Thus, synchrotron flares are produced via internal shocks in the preexisting jet around the time of the X-ray spike, when the jet velocity is increasing the fastest.

Our observations are relevant to this model in several ways. First, we have shown that the sharp X-ray trigger spike observed in 1997 is associated with the appearance of large infrared flares that do not occur in 2002, when the spike is absent. This would seem to be evidence in favor of the Fender et al. (2004) model. On the other hand, a hard-to-soft state transition clearly occurs in 2002 as well, and the first oscillation in each 2002 cycle might therefore meet the definition of a trigger spike in this sense (see also §2.8.1). If so, the model would need to explain

why the 1997 trigger spike is associated with a  $\sim 100$  mJy infrared flare, while the 2002 spike is only associated with a  $\sim 5\text{--}10$  mJy flare that is rivaled, at least in the infrared, by the flares that immediately follow it.

There are several reasons why the strength of the flare might vary. As suggested by Vadawale et al. (2003), the flare strength may be related to the total amount of material in the path of the faster jet—that is, if more material was ejected during the spectrally hard X-ray dip that preceded the spike, the resulting flare might be stronger. However, our observations do not show any particular relationship between the duration of each X-ray dip and the strength of the subsequent infrared flare. Therefore, there is no evidence to suggest that this effect is responsible for the stronger flares in 1997 as compared to 2002.

Another mechanism for varying the strength of the flare in the Fender et al. (2004) model is to change the speed of the jet. Fender & Belloni (2004) suggested that the proposed increase in jet speed as the X-ray spectrum softens may be due to the inner edge of the accretion disk moving closer to the black hole, where the escape velocity increases rapidly. If we take the results of our X-ray spectral fits in Figure 2.7 at face value, we can attempt to test this suggestion, by seeing whether there are significant differences in the evolution of  $R_{in}$  between 1997 and 2002.

It can be seen from Figure 2.7 that although  $R_{in}$  decreases during the spike, it does so by a similar order of magnitude in 1997 as in 2002, suggesting that there may not be any large differences in the jet velocity. On the other hand, the value of  $R_{in}$  just before the spike appears to be slightly larger in 1997 than it is in 2002. Therefore, if the argument of Fender & Belloni (2004) is correct, the velocity of the steady jet is slower in 1997, leading to a greater velocity differential and stronger internal shock when the highly relativistic jet is released during the

spike. In addition, the decrease in  $R_{in}$  happens on a faster timescale in 1997 than in 2002, which could mean that the jet velocity increases faster, leading to an internal shock in 1997 which occurs closer to the disk and therefore in the peak infrared-emitting region of the jet.

However, this interpretation carries with it many caveats since, as we have pointed out in §2.6.2.2, the observed change in  $R_{in}$  may not correspond to an actual change in the inner radius of the disk, and even if it does, it only corresponds to the radius of the *thermally emitting* part of the disk (i.e. the region that is optically thick), which is not necessarily the same as the region from which the jet is ejected. Furthermore, there is a considerable range of X-ray behavior in both the 1997 and 2002 observations, so any conclusions that apply to the data shown in Figure 2.7 do not necessarily apply as strongly to the “class  $\alpha$ ” and “class  $\beta$ ” states as a whole.

Another way in which our observations are relevant to the Fender et al. (2004) model concerns the infrared subflare and our suggestion that each X-ray oscillation in 2002 is associated with a similar amount of infrared emission (see §2.5). If these faint infrared flares are interpreted as emission from the jet, then the jet in GRS 1915+105 may be active during the time of the soft X-ray oscillations, in contrast to the suggestion of Fender et al. (2004). On the other hand, we cannot conclusively prove that the subflares represent jet emission (see §2.7), and even if they do, it is not clear exactly where the “dividing line” between jet and non-jet states in the Fender et al. (2004) model actually exists; it is possible that the low points of the 2002 X-ray oscillations (where  $T_{in} \lesssim 1.5$  keV) are spectrally hard enough to produce a steady jet.

In general, our observations do not contradict the Fender et al. (2004) model at this stage of its development; they will, however, likely be important in testing

a more quantitative model based on the ideas that Fender et al. (2004) proposed. Furthermore, any successful model will need to explain what parameter triggers the sharp X-ray state transitions and infrared-bright ejections that are observed in 1997 but not in 2002.

### 2.6.3 1997 and 2002 Oscillations

Perhaps the most interesting difference between the 1997 and 2002 X-ray spectra is in the evolution of the power law component. In 1997, the power law index undergoes much more dramatic changes than in 2002, especially during the period of oscillations (see Figures 2.6 and 2.7). If this difference is real, it would suggest that significant changes are occurring in the corona or jet during the 1997 oscillations that do not take place during 2002, which is somewhat puzzling since both sets of oscillations produce similar “class C” infrared flares (see §2.5).

We investigated the possibility that the observed variations in the power law index were artificial. For example, if the multitemperature disk blackbody model provided a poor fit to the soft X-ray data, this could force the fitting routine to alter the power law component in order to fix the problems with the disk model on the soft X-ray end. Since the power law contributes significantly to both the soft *and* hard X-rays, the model spectrum might then fit the hard X-rays poorly, which would not necessarily be noticed in our fitting routine because the hard X-rays only provide a small percentage of the total flux and contribute little to the chi-squared value of the fit. Therefore, it might be possible to obtain changes in the fitted power law index even if there were no changes in the actual flux of hard X-ray photons, which comprise the bulk of the power law shape.

To get a handle on whether or not this is occurring, we show plots of X-ray

spectra from the 1997 and 2002 oscillations in Figure 2.8b and 2.8c. The parts of the X-ray light curve from which these spectra have been extracted are indicated by the vertical shaded regions numbered 4 through 7 in Figure 2.6. As before, we extracted the spectra at a time resolution of 16 seconds using PCA Standard-2 data, in order to obtain the highest possible energy resolution.

It is clear from Figure 2.8 that the spectrum at energies  $\gtrsim 20$  keV does not change significantly during the 2002 oscillations but does during 1997, where the hard X-ray flux appears to drop by a factor of a few as the light curve enters the low point of its oscillation (the low point of the oscillation is an example of an X-ray “soft dip” discussed in e.g. Eikenberry et al. 2000). Furthermore, the solid lines in Figure 2.8 show the power law fit to the data; at high energies, the fit does seem to match the spectrum relatively well. There is a tentative suggestion that the actual data are steeper than the fitted power law, but it appears to be a minor effect.

We attempted to investigate this issue further by performing power law fits to the hard X-ray data only, thereby circumventing any problems introduced by the multitemperature disk blackbody model. We found that the power law index did appear to systematically increase by a value of  $\sim 0.3$  to 1 when this procedure was used, but we were unable to obtain high enough signal-to-noise to reliably investigate the variability properties. Therefore, while we can conclusively state that the high energy power law component changes much more dramatically in 1997 than 2002, our data showing that this change specifically occurs in the power law index are less robust.

Finally, although the power law is much more variable in 1997 than 2002, there does appear to be a small peak in the fitted power law index during the

first oscillation in 2002 (Figure 2.7b). There is also tentative evidence for this behavior in the first oscillation of other X-ray cycles in our 2002 observations. These detections are very close to our signal-to-noise limit, and thus we cannot determine whether the power law component varies in *all* the 2002 oscillations but is in general too weak for us to detect (since the subsequent oscillations are not as strong as the first one), or whether the first X-ray oscillation is in fact physically distinct, and produces power law variations even though the others do not.

## 2.7 Origin of the “Class C” Subflares

It is interesting to speculate on the origin of the  $\sim 5\text{--}10$  mJy “class C” infrared flares which seem so ubiquitous in GRS 1915+105.

Eikenberry et al. (2000) showed that when these flares were observed in July 1998, each flare *preceded* its associated X-ray oscillation by  $\sim 200\text{--}600$  seconds, including examples in which the infrared light curve peaked and started to decay before the X-ray disruption began. To see if this property applies to the events we observed in 2002, we show in Figure 2.9 a closer view of the isolated “subflare” from our first night of observations. The X-ray oscillation associated with this subflare consists of a smooth rise in the accretion disk flux, followed by a sharp spike. As can be seen from Figure 2.9, the subflare peaks after the X-ray spike, but its initial rise does seem to begin  $\sim 30$  seconds before the spike. This evidence is tenuous, but it is another hint that the  $\sim 5\text{--}10$  mJy flares in GRS 1915+105 might all have a similar cause.

Note that the X-ray event shown in Figure 2.9 is in no way unusual except for the long delay between it and the previous oscillations; a similar spike is seen at the end of most, if not all, of the 2002 oscillations (for example, see the last event

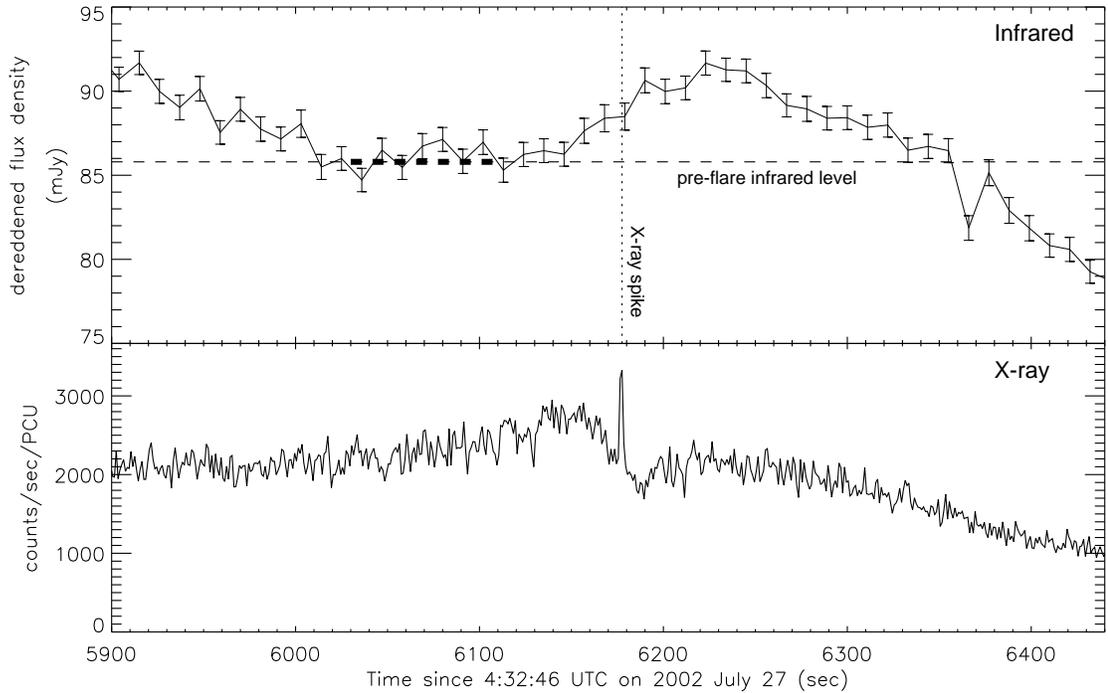


Figure 2.9: Simultaneous infrared and X-ray light curves of GRS 1915+105 at the time of the infrared “subflare” observed on 2002 July 27 (Figure 2.2a). The X-ray data are at 1-second resolution, and the infrared data (dereddened by 3.3 magnitudes) have been smoothed to 11-second resolution to increase the signal-to-noise ratio; they are plotted along with their  $\pm 1\sigma$  uncertainties. The dotted vertical line marks the time of the X-ray spike, and the dashed horizontal line marks the median value of the infrared light curve calculated between 6,030 and 6,110 seconds (the bold region of the dashed line). The subflare appears to rise above this value  $\sim 30$  seconds before the X-ray spike begins.

in Figure 2.7b). Similar oscillations have been discussed by Taam et al. (1997), where the spike is referred to as a “secondary burst.” Spectral fitting of these spikes is inconclusive but indicates that they may consist of a momentary increase in the flux from the accretion disk *and* power law components of the spectrum, which makes them different from any other part of the 1997 or 2002 oscillations. Note that the “class C” infrared flares originally defined by Eikenberry et al. (2000) were associated with “soft dips” in the X-ray light curve that occur in the 1997 oscillations (see §2.6.3), but for the purposes of this chapter, we will refer to any infrared subflare that occurs during the  $\sim 30$  minute events and is associated with an individual oscillation as a “class C” flare.

We cannot directly rule out a thermal origin for the 2002 subflare (since it is an isolated event with no radio coverage), but the near-simultaneous presence of radio flares during our observations (see §2.4.2), combined with the idea that all of the infrared flaring behavior in 2002 was due to superimposed smaller events (see §2.5), strongly suggests that each small infrared flare has a radio counterpart, and therefore a nonthermal origin (see, however, §2.8.3). In addition, Eikenberry et al. (2000) argued that the July 1998 “class C” infrared flares were probably related to synchrotron jet emission, based on observations by Feroci et al. (1999) of a delayed,  $\sim 40$  mJy radio counterpart to an X-ray oscillation spectrally similar to the July 1998 oscillations (again, however, we point out that the 2002 oscillations do not share this similarity).

It is clear that the “class C” flares are a complex phenomenon, with a variety of X-ray counterparts. There is also further fast variability within the “class B” events that has not been discussed here: large (up to  $\sim 70$  mJy) subflares in some of the 1997 data that unfortunately have no X-ray coverage, and smaller ( $\sim 10$ – $20$  mJy)

flares that do *not* occur during the X-ray oscillations but rather during the initial rise of the infrared light curve, right after the trigger spike (Rothstein & Eikenberry 2002). If we accept a synchrotron jet origin for all the flares in GRS 1915+105, then the observations discussed here show that the accretion disk can interact with the jet under a wide variety of timescales and circumstances.

What is the nature of this interaction? The fact that at least some of the flares precede their associated X-ray oscillation argues for an “outside-in” origin for these events, in which the flare occurs somewhere outside of the inner accretion disk and eventually propagates inward to cause a disturbance in the X-rays emitted there (Eikenberry et al. 2000). Simulations of magnetized accretion disks by Lovelace et al. (1994) have shown that an increase in the poloidal magnetic field in the outer part of the disk can lead to an enhanced accretion rate, which triggers a slightly delayed burst of radiation from the inner disk. An alternate idea is that the flare could be produced by a reverse shock originating within a continuous jet outflow, which then propagates back down to the inner disk (Eikenberry et al. 2000).

The possibility that shocks within a preexisting jet could be responsible for the infrared flares from GRS 1915+105 is intriguing in light of the variability that we find within each flare. A model for internal shocks in microquasar jets was presented by Kaiser et al. (2000). It was primarily used to explain the larger “class A” events, but a modified version might be applicable to the  $\sim 30$  minute “class B” events as well (see Türler et al. 2004), especially given that a continuous jet has been observed in GRS 1915+105 and is known to be associated with the  $\sim 30$  minute cycles (Dhawan et al. 2000). In addition, Collins et al. (2003) showed that particularly large ( $\sim 300$  mJy) “class B” events observed by Fender & Pooley (2000) cannot be explained by a simple model in which the flare results from an

increase in the injected particle density at the base of the jet.

## 2.8 Discussion

In this section, we outline our scenario for the  $\sim 30$  minute variability cycles in GRS 1915+105, based on the observations presented in this chapter. The key feature of this scenario is that although GRS 1915+105 always dissipates energy in the jet (via infrared and radio flares) when the accretion disk becomes active, this emission is at least partially due to a superposition of smaller events, one for each oscillation in the X-ray light curve. Some of the events in GRS 1915+105 that would previously have been described as “class B” flares (in the sense that there is one large ejection every  $\sim 30$  minutes) may really just be a superposition of smaller “class C” flares, which blend together and mimic a larger event. Episodes in which a large ejection really *does* occur, leading the infrared flare to dissociate from the X-ray light curve as the ejected material becomes causally separated from the inner disk (Eikenberry et al. 1998a), seem to require a “trigger” spike in the X-ray light curve at the moment of the state transition, when GRS 1915+105 does not immediately begin the fast X-ray oscillations but rather enters a phase of duration  $\sim 300$  seconds in which the accretion disk is hot and visible but changes on a much slower timescale.

### 2.8.1 A Continuum of Behavior

It is important to realize that the  $\sim 30$  minute cycles in GRS 1915+105 do not neatly fall into discrete categories. From perusing the figures in this chapter, one can easily imagine a continuum of behavior, from an extreme example in which the initial ejection completely dominates the infrared flare (Figure 2.3a), to examples

where the initial ejection is weaker and the excess emission from blended together faint flares is easier to see (Figure 2.3b), to examples from our 2002 observations where *most* of the infrared emission is due to blended together faint flares but a small “peak” is still observed at the beginning of the event, perhaps indicating a slightly larger initial ejection (some of the flares in Figure 2.2a), to examples on the other extreme, where there is no peak at the beginning of the flare and the event is consistent with the initial ejection being exactly the same as the subsequent ones (Figure 2.6b).

In this vein, we point out that X-ray observations which contain a strong trigger spike (the “class  $\beta$ ” state defined by Belloni et al. 2000a) have been observed to be accompanied by infrared flares with a wide range of amplitudes. The events observed by Eikenberry et al. (1998a) clustered around  $\sim 100$  mJy but ranged from  $\sim 60$  to 200 mJy, while a single event observed by Mirabel et al. (1998) during this X-ray state only reached a peak amplitude of  $\sim 30$  mJy. Meanwhile, the superimposed faint flares reported in this chapter during the “class  $\alpha$ ” state in July 2002 peak relatively consistently at  $\sim 30$  mJy, but even these less extreme events may have some variation in amplitude; a single “class  $\alpha$ ” event observed by Mirabel et al. (1998) was accompanied by a flare that only appeared to peak at  $\sim 10$  mJy.

The X-ray behavior during the  $\sim 30$  minute cycles may also represent a continuum, between cases where there is a clear X-ray trigger spike and cases where this distinction is not so clear. Although the initial X-ray oscillation in each 2002 series is very similar to the subsequent oscillations, it does share a few subtle characteristics with the 1997 trigger spike. For example, Figure 2.7b shows that the initial 2002 oscillation is “double-peaked,” and at the end of the first peak the accretion

disk flux drops sharply while the inner disk temperature continues its gradual rise. This is qualitatively the same behavior that was observed during the 1997 trigger spike (see §2.6), and thus the first peak in this oscillation may be regarded as a weak analog of the trigger spike in 1997, albeit one which leads *immediately* to the X-ray oscillations rather than requiring a  $\sim 300$  second phase in which the accretion disk is much less variable. Based on these properties, it is not surprising that the initial 2002 oscillation sometimes appears to be accompanied by a slightly larger infrared flare (Figure 2.2a, as discussed above) and slightly more variation in the power law index (see §2.6.3) than the oscillations which follow it; it may really be a weak variant of the 1997 spike which leads to the large ejections.

### 2.8.2 When Does the Ejection Occur?

The observations presented in this chapter contradict some of the conclusions of Klein-Wolt et al. (2002), who analyzed simultaneous X-ray and radio observations of GRS 1915+105 and argued that the radio flares are produced via a continuous ejection of material during the long, spectrally hard dips in the X-ray light curve (“state C” in the classification of Belloni et al. 2000a).

Independent of any of the specific conclusions we have reached in this chapter, our observations pose a serious challenge for this argument. As is easily visible from Figure 2.2, the duration of each infrared flare in our July 2002 observations is correlated with the duration of its accompanying period of X-ray oscillations, not with the previous X-ray dip. This strongly suggests that the flares are associated with the X-ray oscillations, while providing no indication that they are influenced by any properties of the preceding dip, as might be expected in the model of Klein-Wolt et al. (2002), where a longer dip means that more material is being ejected

into the jet.

The differences which we observe between the 1997 and 2002 infrared flares also argue against the model of Klein-Wolt et al. (2002). The X-ray dips in 1997 and 2002 are very similar to each other, and they are separated by X-ray oscillation cycles which have similar durations. The only real difference is that the 2002 dips appear to be slightly longer ( $\sim 1000$  seconds as opposed to  $\sim 600$  seconds) and to have disk blackbody emission that is weaker by a factor of  $\sim 2$  (although it is important to note that the disk blackbody emission is so weak in both cases that it is barely detectable).<sup>3</sup> It is not clear how these slight differences in the X-ray dips could lead to dramatic differences in the infrared flares under the Klein-Wolt et al. (2002) model. Therefore, the fact that there appears to have been a large ejection during each  $\sim 30$  minute cycle in 1997 but not in 2002 argues for a scenario in which the flares are associated with properties of the X-ray state transition (such as the presence or absence of the trigger spike) and subsequent accretion disk oscillations, which do proceed much differently in 1997 than in 2002.

We do not dispute the conclusion of Klein-Wolt et al. (2002) that the *presence* of “state C” dips is required for infrared and radio flares to occur; for example, it is definitely possible that mass which is later available to be ejected steadily builds up during the dips (e.g. Belloni et al. 2000b). It is also possible that matter is being continuously ejected during the dips but that this material does not directly produced the large flares; instead, the flares could be produced via internal shocks which occur in the continuous jet when faster-moving material is ejected at the moment of the state transition (Vadawale et al. 2003; see also Fender et al. 2004).

---

<sup>3</sup>But see Mikles, Eikenberry, & Rothstein (2006) for recent evidence of a relationship between QPO properties during the dip and X-ray and infrared behavior during the subsequent oscillations.

The main conclusion of Klein-Wolt et al. (2002) which we refute is simply that the flares can be *entirely* explained by the evolution of particles which were ejected during the dip, without appealing to some other event that occurs afterwards in the accretion disk.

The above ideas were hinted at by previous multiwavelength observations containing infrared data (Eikenberry et al. 1998a; Mirabel et al. 1998; Eikenberry et al. 2000), but they are confirmed by our present work. Nonetheless, it will be important to test them with future radio observations. The observations analyzed by Klein-Wolt et al. (2002) did not contain any examples of the “class  $\alpha$ ” X-ray state for which there was significant overlap in the X-ray and radio coverage, and in general, there are only a few radio observations in the literature which overlap with the “class  $\alpha$ ” state.

### 2.8.3 Infrared and Radio Flares

One of the few previous multiwavelength observations during the “class  $\alpha$ ” state raises some questions about our current results. Ueda et al. (2002) obtained  $\sim 3$  hours of X-ray and radio (and, separately, X-ray and infrared) observations during this state. They detected at least one radio flare that had a *shorter* duration than its accompanying X-ray oscillation period, with the main part of the flare decaying several hundred seconds before the end of the oscillations. This flare was most clearly seen at 1.3 cm (23 GHz), where it reached an amplitude of  $\sim 15$  mJy (Figure 2 of Ueda et al. 2002).

In most models, radio flares are expected to peak later and last longer than their associated infrared flares, due to optical depth effects in the ejected material (van der Laan 1966; Mirabel et al. 1998; Collins et al. 2003). Therefore, the radio

flare observed by Ueda et al. (2002) does *not* appear to be the counterpart of the superimposed infrared flares we report in our current work, and, in fact, Ueda et al. (2002) used their observations to specifically rule out the possibility that the radio flares are produced via a superposition of smaller flares associated with each X-ray oscillation. In addition, the baseline flux density of the radio emission during these observations was  $\sim 10$  mJy; thus, if there was any “radio excess” associated with the X-ray oscillations after the decay of the flare, it was limited to, at most, this value.

How can we understand these results? If the Ueda et al. (2002) observation is representative of the radio behavior during the “class  $\alpha$ ” state, then the initial ejections at the moment of the state transition may really be distinct, even though they do not appear that way in the infrared light curve. The initial ejection might have a relatively flat infrared-to-radio spectrum (and thus a prominent radio signature), while the repeating events associated with each X-ray oscillation could have little associated radio emission. If this is the case, then the differences between the “class  $\alpha$ ” and “class  $\beta$ ” states are not really as strong as we have indicated in this chapter; both might produce ejections at the moment of the state transition which are substantially different from the subsequent ones (although it is still the case that the “class  $\beta$ ” events produce, on average, much larger infrared and radio flares than the “class  $\alpha$ ” events).

Alternatively, it is possible that the Ueda et al. (2002) event was not representative of the normal behavior during the “class  $\alpha$ ” state. We have analyzed the X-ray light curve and spectrum from this event and found that there is a relatively sharp double peak in the first oscillation in this series, whose first peak has spectral and temporal characteristics particularly reminiscent of a “class  $\beta$ ”

trigger spike (see §2.8.1). In addition, there is an unusual,  $\sim 10$  second spike in the X-ray light curve which occurs  $\sim 30$  seconds *before* the state transition (although it is spectrally different from the trigger spike and appears to simply be a thermal “bump” in the spectrum). This behavior suggests that Ueda et al. (2002) may have observed an anomalous event, in which the initial ejection was much different than those which normally occur during the “class  $\alpha$ ” state.

Furthermore, the only other “class  $\alpha$ ” radio observation we could find in the literature appears to be consistent with the interpretation we have put forth in this chapter. Figure 1 of Mirabel et al. (1998) shows that during the “class  $\alpha$ ” state, the flare duration increases with wavelength as expected, including one case in which the flare was simultaneously observed in the infrared and three different radio wavelengths. The X-ray oscillation period associated with this event was not observed, but it is constrained to have been within a short ( $\sim 30$  minute) gap in the X-ray coverage that overlaps with the infrared flare and is of significantly shorter duration than the radio flares. Thus, it is broadly consistent with our current observations.

The ultimate answer to the questions raised here may rest on how much variation there is in the amplitude of the “class C” flares associated with each X-ray oscillation. If the “radio excess” seen in the Ueda et al. (2002) observations ( $< 10$  mJy if it exists at all) is typical, then it is difficult to see how the repeating, larger radio flares during the “class  $\alpha$ ” state (see §2.4.2, as well as Mirabel et al. 1998), could be explained without resorting to initial ejections that are different from the subsequent events. Also note that the “class  $\rho$ ” state in GRS 1915+105, which consists of an extremely long series of oscillations very similar to those seen during the “class  $\alpha$ ” state (Belloni et al. 2000a), is associated with steady radio

emission of only  $\sim 4$  mJy at 15 GHz (Klein-Wolt et al. 2002).

To further complicate the picture, the infrared observations of Ueda et al. (2002) show at least one instance (and likely more) in which a *delayed*  $\sim 30$  minute infrared flare is seen during the “class  $\alpha$ ” state, beginning at the end of its associated X-ray oscillation period rather than the beginning. (Note that as in our observations, the duration of this flare still matches the duration of the oscillations.) It is unclear what could have caused this unusual delay, which is inconsistent with all the infrared flares we report in this chapter. Ueda et al. (2002) suggested the possibility of an internal shock occurring far out along the jet, but it is a puzzling event under any interpretation.

Clearly, the only way to resolve the issues raised here is to obtain simultaneous X-ray, infrared and radio observations over a long stretch of events. In the absence of such observations, we can only point out that the conclusions we have reached in this chapter apply to all  $\sim 15$  simultaneous episodes that we observed. Overall, it is clear that the 1997 “class  $\beta$ ” events produce qualitatively different flaring behavior than the 2002 “class  $\alpha$ ” events and that these differences are related to the different ways in which the state transition proceeds and in particular to the X-ray trigger spike, which appears most strongly in the “class  $\beta$ ” observations. However, the exact manner in which properties of the trigger spike control the behavior of the initial flare—for example, whether the “sharpness” of the spike or the delay between it and the subsequent oscillations is the most important feature—cannot yet be conclusively determined.

## 2.9 Conclusions

We have presented simultaneous infrared and X-ray observations of the microquasar GRS 1915+105 (at 1-second time resolution) during a period of jet ejection in July 2002. By comparing these observations to those obtained in August 1997 by Eikenberry et al. (1998a), we arrive at a picture in which a large ejection may not always be present during the  $\sim 30$  minute state transitions in GRS 1915+105. The observed infrared flares may sometimes entirely be a superposition of smaller, more complex phenomena, with large, infrared-bright ejections superimposed on them only when there is a “trigger” spike in the X-ray light curve. In particular, we find that:

1. The duration of each infrared flare in 2002 matches the duration of its accompanying X-ray oscillation period.
2. There is one instance in which a single, isolated X-ray oscillation occurs in the light curve, accompanied by a faint infrared “subflare” superimposed on one of the main flares.
3. These observations are consistent with a scenario in which each X-ray oscillation has an associated faint infrared flare and these flares blend together to form, and entirely comprise, the larger events.
4. The main difference between our X-ray observations and the August 1997 observations is the presence of a sharp “trigger” spike in the 1997 X-ray light curve during each  $\sim 30$  minute cycle, which occurs  $\sim 300$  seconds before the accretion disk oscillations and signals a sharp change in the spectral properties of the source.

5. The trigger spike appears to be associated with initial, large infrared flares, which are observed in 1997 but not in 2002.
6. The 1997 X-ray oscillations contain dramatic changes in the high energy power law component of the X-ray spectrum, which are not observed during the 2002 oscillations. This is somewhat puzzling, since both sets of oscillations produce similar infrared subflares.
7. There is tentative evidence that the infrared subflare observed in 2002 leads its associated X-ray oscillation by  $\sim 30$  seconds, possibly suggesting an “outside-in” origin for these events, as in observations by Eikenberry et al. (2000).
8. It is unlikely that any of the 2002 flares can be explained solely by the continuous ejection of material during the spectrally hard X-ray “dip” which precedes each set of X-ray oscillations, as suggested by Klein-Wolt et al. (2002).

Although many authors have speculated on the importance of the trigger spike as a signature of large jet ejections (e.g. Eikenberry et al. 1998a; Mirabel et al. 1998; Yadav 2001; Fender et al. 2004), the observations presented in this chapter confirm it. Our 2002 X-ray observations, which have many similarities to the Eikenberry et al. (1998a) observations except for the absence of a distinctive trigger spike, do not show the same large flares in the infrared. Thus, these data are important in allowing us to pinpoint which aspects of the X-ray behavior are related to which types of events in the jet.

Overall, we find that jet production in GRS 1915+105 is a complex, unsteady phenomenon, down to the fastest timescales that have been observed. The jet

and the accretion disk appear to interact on a regular basis, perhaps every time an oscillation occurs within the inner portion of the disk. These details need to be taken into account in accretion disk models, such as in the simulations of radiation pressure dominated disk evolution by Janiuk et al. (2002), which found that energy dissipation in a corona or via an outflow from the disk may be important in reproducing the GRS 1915+105 X-ray emission.

## Chapter 3

# The Equations Governing the Evolution of an Accretion Disk: A Study of the Vertically-Integrated Structure

### 3.1 Summary

In this chapter, we derive a set of equations that govern the physical evolution of an accretion disk in a more general and comprehensive way than has been previously attempted in the literature. Along the way, we present several new results, particularly with regard to the handling of the time-dependent structure in a vertically-integrated disk model. We introduce a dimensionless number  $\{u\}_z$  that can be defined for any physical quantity  $u$  and that characterizes the vertical distribution of  $u$  within the disk in a model-independent way; this parameter has much in common with the dimensionless  $\alpha$  parameter for turbulent stress defined by Shakura & Sunyaev (1973). Allowing  $\{u\}_z$  to change in space and time is a crude way to model changes in the disk's vertical structure (e.g., due to a jet, or to changes in the vertical distribution of the turbulent heat dissipation). We also derive a steady-state solution, energy balance equation and evolution timescales for an accretion disk that are more general than the well-known Shakura & Sunyaev (1973) and Abramowicz et al. (1988) solutions; in particular, our solution

accommodates the possibility of an arbitrary vertical disk structure and the energy loss and pressure contribution from a steady jet. This solution can be used as the initial condition in (or for comparison with) numerical disk simulations.

Our discussion in this chapter is introductory in nature and is geared towards researchers who are interested in studying and numerically simulating accretion disks in any number of situations; however, we primarily focus on the particular applications that are relevant to the simulations of GRS 1915+105 attempted in this thesis. The chapter contains three main sections, in which we derive the basic equations (§3.2), simplify them to one or two dimensions (§3.3), and derive a steady-state solution (§3.4).

We begin by outlining the basic equations of radiative magnetohydrodynamics that govern the ionized plasma in the inner part of an accretion disk (§3.2), with a focus on expressing these equations in conservative form. Published papers on accretion disks use a wide variety of equations, but the specific assumptions that go into the equations are not often carefully considered. Here, we attempt to show the connections between different disk models and carefully elucidate the various assumptions that each one makes in order to simplify the equations; the validity of these assumptions can thereby be checked in the results of numerical simulations.

Next, we focus on versions of the equations that can be used for numerical simulations in fewer than three dimensions (§3.3). This is particularly important for the simulations of GRS 1915+105 that we undertake in this thesis. The timescales of interest in GRS 1915+105 are  $\sim 30$  minutes, as previously discussed, which is very “slow,” being equal to  $\sim 10^7$  characteristic dynamical timescales of the inner disk. State of the art three-dimensional simulations are extremely computationally intensive and can currently run for only  $\sim 10^4$  characteristic timescales (e.g.,

Krolik et al. 2005), or  $\sim 1$  second for GRS 1915+105.<sup>1</sup> Therefore, it is crucial to have simulations that can run for many more timesteps than the three-dimensional ones, especially if we want to run many different trials and compare them to the diversity of behavior seen in the long term X-ray light curve of this source; one-dimensional simulations, in particular, are the “workhorse” that can accomplish this. To further the goal of making one-dimensional simulations as physically believable as possible, we derive in §3.3 new results for vertically-integrated disks that are designed to improve the accuracy of these simplified disk models and the ability of the models to handle changes in the vertical disk structure, such as that due to a first-order model of a jet.

Finally, we present a steady state solution to our modified accretion disk equations (§3.4) that can be used as a starting point for numerical simulations and for comparing to the results of time-dependent work (e.g., via plotting on an local energy balance diagram to study the stability properties). We find that the timescales on which this steady disk changes have an interesting and previously unknown dependence on the vertical structure of the disk.

There are many previous works that discuss the equations governing an accretion disk; some good examples include Kato et al. (1998), Bisnovatyi-Kogan & Lovelace (2001), and Frank et al. (2002). Portions of this chapter benefited from reading these and other works; however, the full presentation of the equations contained in this chapter has not previously been published in the literature.

---

<sup>1</sup>These simulations, and others like them, also neglect radiation physics. This is fine for the objects that these simulations are specifically designed to investigate (low luminosity flows such as those surrounding Sgr A\*, the black hole at the center of our Galaxy), but clearly not adequate for GRS 1915+105, which often radiates near its Eddington luminosity and for which radiation pressure and cooling are therefore of fundamental importance.

## 3.2 The Equations of Radiative Magnetohydrodynamics

We present in this section a physically-motivated derivation of the magnetohydrodynamic (MHD) equations that govern an ionized plasma containing both gas particles and radiation (in other words, a three-species fluid of ions, electrons and photons), with an eye towards specializing these equations to the case of an accretion disk. We begin by discussing some of the basic assumptions that are inherent in our use of the MHD equations (§3.2.1), and after a brief discussion of the general principle of conservation that guides our decision to express the MHD equations in conservative form (§3.2.2), we go on to discuss the equations expression conservation of mass (§3.2.3), conservation of momentum (§3.2.4) and conservation of energy (§3.2.5), respectively. Finally, we summarize the equations we will use for our accretion disk model (§3.2.6).

### 3.2.1 MHD Assumptions

The equations discussed in this chapter can be derived directly from fundamental particle equations such as the Boltzmann equation. (For example, see Nicholson 1983 for a derivation of the non-radiative MHD equations and Mihalas & Mihalas 1984 for a derivation of the radiative hydrodynamic equations without magnetic fields.<sup>2</sup>) We do not go through these derivations in this chapter because they are mathematically very complex. Rather, we present a more intuitive approach to the equations expressing conservation of mass, conservation of momentum and conservation of energy in the fluid.

Nonetheless, it is important to understand the assumptions that go into deriv-

---

<sup>2</sup>We are not aware of any textbook that rigorously derives the combined radiative MHD equations, however.

ing our equations from the full kinetic theory. There are two main assumptions that are made. First, in order to treat the gas particles or photons as a fluid, we must assume an “equation of state”—a relationship between the rest-frame kinetic energy that a group of particles carries with them (i.e., the internal energy density) and the rate at which these particles transport momentum to their surroundings per unit surface area, due to microscopic motions (i.e., the pressure tensor). Whereas for an arbitrary configuration of particles, one can easily imagine changing the rate of momentum transport without changing the energy (simply change the direction that the particles are moving in), the fluid approximation assumes that such changes will always statistically cancel out. We will primarily consider cases where the distribution of both particles and photons is roughly thermal, so that the energy density is well-characterized by a local temperature for each species (this generally occurs when the ions and electrons are both in local thermodynamic equilibrium so that the primary emission mechanism is thermal radiation, although strictly speaking the photons do not necessarily have to be thermally distributed in this case; Blaes & Socrates 2003). Possible equations of state under these conditions are discussed in §3.2.4.1.

The second assumption is more subtle, and, in the case of an accretion disk, potentially more dangerous. In going from the fluid equations to the MHD equations, we assume that the various species (ions, electrons and photons) are all characterized by the *same* local temperature—i.e., that they are all in local thermodynamic equilibrium with each other. This assumption dramatically simplifies the equations that need to be considered. Instead of employing energy equations for each species of particle separately, and including within these equations terms dealing with the energy transfer between each species and the others, we can simply add together

the equations for the different species and produce a single equation governing the total energy density of the fluid, in which the interaction terms between the various species cancel out (see §3.2.5). The temperature (and therefore the apportioning of the energy density between different species) can be calculated from the total energy density, giving us all the information we need to know about the fluid.

There are several accretion disk models that relax this assumption. The most prominent are the so-called Radiatively Inefficient Accretion Flows (RIAFs) or Advection Dominated Accretion Flows (e.g., Shapiro et al. 1976; Narayan & Yi 1995), which allow the temperature of electrons and ions to differ. These apply in extremely low-density environments, where the interaction of ions and electrons via Coulomb collisions is inefficient; therefore, if angular momentum transfer within the disk occurs via a viscous-like process that primarily heats the ions, the ions can remain much hotter than the electrons and carry their energy into the black hole without allowing the electrons to radiate it away. However, Bisnovatyi-Kogan & Lovelace (1997, 2000) have argued that the dissipation of magnetic fields in the disk can provide a heating source for the electrons that can keep the electron and ion temperatures much closer together than is normally assumed in these models. In any case, RIAFs are unlikely to be important for the disks considered in this thesis, where the accretion rates and densities are high, Coulomb collisions are likely to be abundant, and observations unequivocally show that the disks are radiatively efficient, at times shining near their Eddington luminosities.

Although the presence of a power law in the X-ray spectrum extending to the hard X-rays has been used to suggest that accretion disks must contain at least some nonthermal or two-temperature plasma in the vicinity of the accretion flow (e.g., Shapiro et al. 1976), other work has suggested that this radiation may be due

to synchrotron emission from a jet further downstream from the disk (Markoff et al. 2001, 2003). The exact origin of the power law, and how closely related it is to the dynamics of the disk (i.e., whether it originates in the disk itself, in a corona, in the “base” of a jet, or in the jet itself) is a matter of much debate in the literature and is still unresolved. Here, we note that in GRS 1915+105, thermal emission persists even when the power law is strong, and thermal emission is particularly prominent during the fast oscillations that we are most interested in explaining (Chapter 2). Therefore, we focus on the main body of the disk in this chapter, where the ions and electrons are expected to be in thermal equilibrium, although in §3.3.2 and the following sections, we discuss ways that the nonthermal jet structure could impact the thermal emission from the disk.

A more interesting possibility is that the particles and *photons* do not have the same temperature. This occurs whenever a photon traversing the fluid encounters a gas temperature that changes more quickly than the timescale for the photon to be thermalized by absorption and reemission—a circumstance that is definitely possible in a radiatively efficient accretion disk shining near the Eddington luminosity, where the electron temperatures are high and thus the absorption opacity is small while the gas timescales are fast. The basic condition for thermalization is briefly discussed by Mihalas & Mihalas (1984), but we expand on it here to specifically focus on the high-temperature environment of an accretion disk, where the opacity is expected to be due to a mix of scattering *and* absorption, rather than absorption alone (see equation 3.51), and the condition for thermalization is therefore different.

Consider a disk in which  $H$  and  $t$  are the typical distance and timescale over which the gas temperature changes ( $v \sim H/t$  is therefore the speed at which the

temperature change propagates along  $H$ ), and where the total optical depth across  $H$  is given by  $\tau$ , the optical depth due to absorption only (i.e., if scattering in the disk were completely ignored) is given by  $\tau_{abs}$ , and the effective optical depth is given by  $\tau_{eff} = \sqrt{\tau\tau_{abs}}$  (e.g., Rybicki & Lightman 1979), defined so that a photon traversing  $H$  will be absorbed and reemitted  $\sim \tau_{eff}^2$  times, out of a total of  $\sim \tau^2$  interactions with the gas. Under these circumstances, there are two conditions that must be met in order for a photon to thermalize with the gas. First, we must have

$$\tau_{eff} \gg 1, \quad (3.1)$$

so that the photon undergoes many absorptions and reemissions as it traverses  $H$ ; in other words, the gas temperature scale height must be optically thick. The second condition, however, is that the photon must undergo  $\gg 1$  absorptions during the time  $t$  that it takes the gas temperature to change; otherwise, the photon will not be able to thermalize with the *local* gas, regardless of what happens once it has diffused along  $H$ . This second condition boils down to

$$t \gg \frac{H}{c\tau_{abs}} \quad (\text{or, alternatively, } v/c \ll \tau_{abs}), \quad (3.2)$$

which can be derived by noting that the photon experiences  $\sim ct/\lambda$  interactions during a time  $t$ , of which a fraction  $\sim \tau_{abs}/\tau$  are absorptions ( $\lambda \equiv H/\tau$  is the mean free path). If absorption dominates ( $\tau_{abs} \approx \tau_{eff} \approx \tau$ ), this second condition is trivially satisfied if the first one is, but if scattering is important,  $\tau_{abs}$  can be significantly smaller than  $\tau_{eff}$ , and conditions (3.1) and (3.2) should be checked independently.

Good presentations of the independent gas and radiation equations that must be used when the above conditions are violated can be found in Mihalas & Mihalas

(1984) and Stone et al. (1992). Papers about accretion disks that consider the gas and radiation equations separately and do not assume the same temperature for gas particles and photons include Blaes & Socrates (2001, 2003), Okuda et al. (2004), Turner et al. (2005), and Hirose et al. (2006).

Now that we have explained the assumptions underlying the MHD equations, we discuss the equations themselves and their physically-motivated derivations in the following sections.

### 3.2.2 General Principles of Conservation

The instantaneous rate of change of any physical quantity can, in principle, only have two contributions: the rate at which the quantity flows into (or out of) the boundaries of the region being considered, and the rate at which the quantity is locally produced or depleted. For any physical quantity  $u$ , its time evolution can therefore be expressed mathematically as:

$$\frac{d}{dt} \int_V \rho_u dV = - \oint_A \mathbf{F}_u \cdot d\mathbf{A} + \int_V S_u dV, \quad (3.3)$$

where  $\rho_u$  is the density of  $u$ ,  $\mathbf{F}_u$  is the flux of  $u$  (the amount of  $u$  passing through a unit surface area per unit time), and  $S_u$  is the source term (i.e., the volume production rate, or the rate at which  $u$  is produced per unit volume and unit time). Here, we consider the evolution of the total amount of  $u$  that is initially contained within a volume  $V$  bounded by a surface area  $A$ .

Tassoul (1978) derives the Reynolds theorem, a mathematical result which allows us to write the left hand side of equation (3.3) as

$$\frac{\partial}{\partial t} \int_V \rho_u dV + \oint_A \rho_u \mathbf{v} \cdot d\mathbf{A}, \quad (3.4)$$

where  $\mathbf{v}$  is the fluid velocity. This can be understood intuitively; the first term represents the rate of change of  $u$  within a fixed volume  $V$ , and the second term represents the rate at which  $u$  is transported out of  $V$  by the bulk fluid flow (for more details, see Oswatitsch 1956).

Manipulating the above equations using the divergence theorem, we can derive equation (3.3) in differential form:

$$\frac{\partial}{\partial t}(\rho_u) = -\nabla \cdot (\rho_u \mathbf{v} + \mathbf{F}_u) + S_u. \quad (3.5)$$

Note that mathematically, the time evolution of any quantity can be written this way (provided we have flexibility in choosing  $\mathbf{F}_u$  and  $S_u$ ), but this formulation becomes useful for *conserved quantities*, where  $S_u$  is either 0 or, at the very least, a truly “local” term that only depends on local properties of the fluid (i.e., one that does not contain any spatial derivatives).

Writing all of our equations in conservative form makes the physical behavior of different terms in the equations easy to understand (each is either a flux or a source of the physical quantity being considered), and it is also useful for numerical simulations, since it guarantees that regardless of the underlying physics, the computational routine will always deal with equations of the same structure—whose numerical conservation properties can easily be checked.

### 3.2.3 Conservation of Mass

In the nonrelativistic limit, mass cannot be produced or destroyed either locally or through interactions with neighbors; therefore,  $S_u = 0$  and  $\mathbf{F}_u = 0$  in equation (3.5). Equation (3.5) therefore becomes:

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\rho \mathbf{v}), \quad (3.6)$$

where  $\rho$  is the mass density. This equation can be written in cylindrical coordinates  $(r, \phi, z)$  as:

$$\frac{\partial}{\partial t}(r\rho) = - \left[ \frac{\partial}{\partial r}(r\rho v_r) + \frac{\partial}{\partial \phi}(\rho v_\phi) + \frac{\partial}{\partial z}(r\rho v_z) \right]. \quad (3.7)$$

Cylindrical coordinates are the most natural choice for modeling a thin accretion disk, since they allow the disk's vertical and radial structure to be treated separately.

### 3.2.4 Conservation of Momentum

The momentum density of a fluid is given by  $\rho\mathbf{v}$ ; inserting this into  $\rho_u$  in equation (3.5), we find that the equation has units of force density; in other words, in addition to the transport term (which now becomes  $\rho v_j \mathbf{v}$  for each component  $\rho v_j$  of the momentum density, in Cartesian coordinates), we must take into account on the right hand side of the equation all forces that act on the fluid. Equation (3.5) shows that there are effectively two types of forces:  $\mathbf{F}_u$  represents forces per unit area that only act on the surface of an infinitesimal fluid element and therefore produce a flux of momentum into or out of that surface, while  $S_u$  represents forces per unit volume that act on the *entire* fluid element and therefore can be considered a local “source” of momentum. Formally, these two types of forces are often called “surface forces” and “body forces,” respectively (e.g., Tassoul 1978), although we prefer the term “volume forces” for the latter.

In curvilinear coordinates, we need to be more careful, because the forces in each coordinate direction are now coupled to each other, and thus the above discussion does not apply independently for each direction. The best way to proceed is to derive equation (3.5) for the momentum in a more rigorous way, using Newton's

Second Law:

$$\begin{aligned} \rho \frac{d\mathbf{v}}{dt} &= (\text{surface} + \text{volume forces applied}) \\ \implies \frac{\partial}{\partial t}(\rho\mathbf{v}) + \nabla \cdot (\rho\mathbf{v} \otimes \mathbf{v}) &= (\text{surface} + \text{volume forces applied}), \end{aligned} \quad (3.8)$$

where we have used standard vector manipulations and substitution from equation (3.6) to rewrite the acceleration term  $\rho d\mathbf{v}/dt$ . We can now see that the momentum transport term in equation (3.5) is properly cast as a tensor  $\rho\mathbf{v} \otimes \mathbf{v}$ , whose components  $\rho v_i v_j$  represent the flux of the  $j$ th direction of momentum through a surface perpendicular to the  $i$ th direction, due to the bulk fluid flow.

### 3.2.4.1 Surface Forces

By similar arguments as above, the surface forces should also be defined in terms of a tensor. We define a *stress tensor*  $\mathbf{G}$ , whose components  $G_{ij}$  represent the flux of the  $j$ th direction of momentum through a surface perpendicular to the  $i$ th direction, due to the surface forces. The divergence of this tensor is a vector whose components  $(\nabla \cdot \mathbf{G})_j$  represent the rate at which the  $j$ th component of momentum density is transported out of each infinitesimal fluid element by surface forces. With this definition, we can write all three components of the momentum equation at once simply by substituting  $\mathbf{F}_u \equiv \mathbf{G}$  in equation (3.5).

The stress tensor that we will consider includes contributions from microscopic forces (i.e., pressure and viscosity), which we will denote  $P_{ij}$ , Maxwell stress (i.e., electromagnetic forces), which we will denote  $M_{ij}$ , and, finally, an additional “turbulent” stress component  $t_{ij}$  which is formally equal to zero but whose purpose we will discuss later. It is also convenient to redefine the transport term  $\rho v_i v_j$  as part of the stress tensor (i.e., the “Reynolds stress”); in that case, we label the total

stress tensor as  $\mathbb{T}$ , and equation (3.5) for the momentum can be written as:

$$\frac{\partial}{\partial t}(\rho \mathbf{v}) = -\nabla \cdot \mathbb{T} + \mathbf{f}_{volume}, \quad (3.9)$$

where  $\mathbf{f}_{volume}$  is the volume force and

$$T_{ij} = \underbrace{\rho v_i v_j}_{\text{Reynolds}} + \underbrace{P_{ij}}_{\text{Pressure}} + \underbrace{M_{ij}}_{\text{Maxwell}} + \underbrace{t_{ij}}_{\text{Turbulent}}. \quad (3.10)$$

### ***Pressure Stress ( $P_{ij}$ )***

As noted earlier, we write the pressure stress (i.e., the contribution to the stress from pressure and other microscopic forces, such as viscosity) in terms of the internal energy density via an equation of state. For gas particles and radiation, respectively, the internal energy density under conditions of thermal equilibrium is given by  $E = E_{gas} + E_{rad}$ , where

$$E_{gas} = \left( \frac{1}{\gamma_{gas} - 1} \right) \frac{\rho k_b T}{\mu m_p} \quad (3.11)$$

$$E_{rad} = a_{rad} T^4, \quad (3.12)$$

$\gamma_{gas}$  is the adiabatic index of the gas particles (equal to 5/3 for a nonrelativistic, monatomic gas),  $k_b$  is Boltzmann's constant,  $T$  is the temperature,  $m_p$  is the mass of a proton,  $a_{rad}$  is the radiation constant (equal to  $4\sigma/c$ , where  $\sigma$  is the Stefan-Boltzmann constant), and  $\mu$  is the mean molecular weight of a particle in the accretion disk. For a fully ionized hydrogen plasma we have  $\mu \approx 0.5$ , while  $\mu \approx 0.615$  for a fully ionized ‘‘cosmic mixture’’ with hydrogen mass fraction  $X = 0.70$  and helium mass fraction  $Y = 0.28$ ; see Frank et al. (2002) and Ostlie & Carroll (1996).

Notice that we are assuming equilibrium between gas and radiation by using the same temperature  $T$  in both of the above equations; in that case, it can be

seen that a knowledge of  $\rho$  and  $E$  at any point in time (which can be obtained from numerical integration of the conservation of mass and conservation of energy equations discussed elsewhere in this chapter) is sufficient to determine  $T$  and therefore the apportioning of energy between  $E_{gas}$  and  $E_{rad}$ .

Now that we have the internal energy density, we can write the equation of state. We set

$$P_{ij} = P\delta_{ij} = (P_{gas} + P_{rad})\delta_{ij}, \quad (3.13)$$

where  $P$  is the total thermodynamic pressure and  $P_{gas}$  and  $P_{rad}$  are the pressures produced by gas particles and radiation, respectively, given by:

$$P_{gas} = (\gamma_{gas} - 1) E_{gas} \quad \left[ = \frac{\rho k_b T}{\mu m_p} \right] \quad (3.14)$$

$$P_{rad} = \frac{1}{3} E_{rad} \quad \left[ = \frac{1}{3} a_{rad} T^4 \right]. \quad (3.15)$$

The expression for  $P_{gas}$  is the standard equation of state for an ideal gas, while the expression for  $P_{rad}$  is the standard equation of state for isotropic radiation. A more compact and sometimes more useful way of writing the combined equation of state is:

$$E = \left[ 3(1 - \beta) + \frac{\beta}{\gamma_{gas} - 1} \right] P, \quad (3.16)$$

where  $\beta \equiv P_{gas}/P$ .

This equation of state makes several assumptions. First, it assumes that all microscopic forces besides the pressure that potentially contribute to  $P_{ij}$  can be ignored in an accretion disk; these include the forces due to molecular, radiative and bulk viscosities. The molecular viscosity is ignored because the mean free path of particles that control viscous mixing is typically much smaller than other length scales of interest in an astrophysical setting. We have also ignored the radiative viscosity  $\nu_{rad} \sim P_{rad}/c\kappa\rho^2$ , where  $\kappa$  is the opacity (this expression is roughly valid

for pure absorption or pure scattering; e.g., Masaki 1971; Tassoul 1978), which is generally negligible in an accretion disk.<sup>3</sup> Finally, we ignore the bulk viscosity, as is customary (especially for the monatomic gas expected in an ionized disk).<sup>4</sup>

A second assumption that is made in deriving equation (3.13) is that the radiation pressure has a simple scalar value given by  $P_{rad} = \frac{1}{3}E_{rad}$ . This is the standard Eddington approximation that follows from assuming an isotropic radiation field, which is roughly true in any region where thermal emission or isotropic electron scattering dominates (Rybicki & Lightman 1979). However, Mihalas & Mihalas (1984) and Stone et al. (1992) have pointed out that in any physical region, the *boundary* of that region (and in particular the boundary through which photons escape to an outside observer) is, by definition, not isotropic. Therefore, although it is customary to follow the lead of stellar atmospheric modelers and use the Eddington approximation to evaluate the flux of radiation everywhere in the disk

---

<sup>3</sup>We can justify this statement by calculating the ratio of the radiative viscosity  $\nu_{rad}$  to the turbulent viscosity  $\nu \sim \alpha c_s H$  that dominates the  $T_{r\phi}$  stress in the Shakura & Sunyaev (1973) disk solution that we derive later on in this chapter (here,  $c_s$  is the sound speed and  $\alpha$  is a dimensionless parameter usually expected to be less than 1). In the innermost region of a Shakura & Sunyaev (1973) disk, where the pressure is dominated by  $P_{rad}$ , the ratio of these viscosities reduces to a particularly simple expression:  $\nu_{rad}/\nu \sim (H/r)(v_c/c)/\alpha\tau$  (where  $v_c$  is the velocity of a circular orbit in the equatorial plane), thereby justifying our statement for a geometrically thin ( $H/r \ll 1$ ), optically thick ( $\tau \gg 1$ ) disk. More generally, Mihalas & Mihalas (1984) show that the importance of the radiative viscosity is always of order  $\sim v/c\tau$  for a fluid velocity  $v$ . This is always  $\ll 1$  for a disk where the gas and radiation are in equilibrium, if we assume that all fluid velocities are comparable to or less than the velocity at which changes to the gas temperature propagate (see §3.2.1 and the assumption already made in equation 3.2).

<sup>4</sup>The bulk viscosity can be thought of as a correction to the thermodynamic pressures given in equations (3.14) and (3.15) that takes into account the fact that those values are formally calculated under equilibrium conditions, which may not hold in an accretion disk that evolves in time. However, the bulk viscosity can always be ignored for monatomic gases, and ignoring it is generally a valid approximation in many other situations as well. In the case of radiation, meanwhile, the bulk viscosity is of the same order of magnitude as the radiative viscosity  $\nu_{rad}$ , which we have already ignored. For more details, see Mihalas & Mihalas (1984), Landau & Lifshitz (1987) and Sonin (1994).

(e.g., Kato et al. 1998), including at the boundary, a more accurate method would necessarily involve using a more complicated equation of state for the radiation.

To summarize, the two biggest assumptions we have made in this section are (1) equilibrium between the gas and radiation, and (2) an isotropic radiation field. The first assumption affects the equation for the energy density, and the second affects the equation of state used to determine the pressure tensor.

When the second assumption is violated, very often the first one is as well (since an anisotropic radiation field by definition must occur in a region where the spectrum is not equal to the local thermal value). A variety of choices exist for the radiative equation of state in this more complicated, non-isotropic case. One of the simplest methods is to use flux-limited diffusion (e.g., Levermore & Pomraning 1981; Mihalas & Mihalas 1984; Kley 1989; Turner & Stone 2001), where the radiation pressure is treated as a tensor that is allowed to have a different magnitude in each coordinate direction; the relative values of each “directional pressure” are prescribed in an ad-hoc way based on the gradient of the energy density so as to guarantee that the radiative energy flux in each direction never exceeds  $cE_{rad}$ , and also so that the equation of state reduces to the correct values for the extreme optically thick and extreme optically thin limits. However, because the equation of state is derived in an ad-hoc way, it is not guaranteed to be valid at intermediate optical depths (Mihalas & Mihalas 1984). A more complicated method is to actually *calculate* the anisotropy of the radiation field at any moment in time and use that to get a more accurate estimate of the pressure tensor (Stone et al. 1992); however, this method proves to be extremely computationally expensive in practical situations (Turner & Stone 2001). The fundamental problem with all of these methods is that if the radiation field is highly anisotropic, the fluid assumption

*itself* starts to break down, and one really needs to go back to the original radiative transfer equation that governs the behavior of individual photons. Therefore, we do not consider disks that are highly optically thin and highly anisotropic further in this chapter.

As a final note, we consider how to handle the equation of state when the gas and radiation are *not* in equilibrium but the radiation field is still roughly isotropic (in other words, when the first assumption listed above fails but the second does not). In this case, the radiation spectrum will not necessarily be thermal, although it is still often characterized by a “temperature”  $T_{rad}$  defined so as to make equation (3.12) true (Mihalas & Mihalas 1984). Generally, one must solve separate gas and radiation energy equations in this case, as discussed above. However, if a specific set of assumptions are made, Hubeny (1990) and Artemova et al. (1996a) have shown that it is possible to derive an expression for the radiation energy density that closely relates it to the gas temperature  $T$  in an accretion disk. We will discuss and expand upon this work in §3.3.2.4, but we reproduce the equation for the equatorial plane energy density here as a point of comparison with equation (3.12):

$$E_{rad} = a_{rad} T^4 \left(1 + \frac{C_1}{\tau}\right) \left(1 + \frac{C_1}{\tau} + \frac{C_2}{\tau_{eff}^2}\right)^{-1}, \quad (3.17)$$

where  $C_1$  and  $C_2$  are dimensionless numbers of order 1 and  $\tau$  and  $\tau_{eff}$  are the optical depths measured from the equatorial plane out to the “top” of the disk; see equation (3.91) for a more exact expression. The radiation pressure is then given by the Eddington approximation, so that  $P_{rad} = \frac{1}{3}E_{rad}$  as usual. Artemova et al. (1996a) claim that their procedure is valid for an accretion disk with any optical depth, although this is unlikely to be true because of their use of the Eddington approximation. Instead, we believe that this equation is most likely to be useful

as a point of comparison for our work when the optical depth becomes small ( $\gtrsim 1$ ) but the disk is still effectively optically thick so that the Eddington approximation holds.

### ***Maxwell Stress ( $M_{ij}$ )***

Formally, the electromagnetic force that contributes the Maxwell stress is a volume force rather than a surface force; in fact, the force per unit volume is simply equal to the Lorentz term  $\rho_e \mathbf{E} + (\mathbf{J} \times \mathbf{B}/c)$ , where  $\rho_e$  and  $\mathbf{J}$  are the charge and current densities in the fluid element and  $\mathbf{E}$  and  $\mathbf{B}$  are the local values of the electric and magnetic fields.

However, when Maxwell’s equations are used to substitute for  $\rho_e$  and  $\mathbf{J}$  in terms of the electromagnetic field, derivatives of  $\mathbf{E}$  and  $\mathbf{B}$  are introduced into the equations. Since it is not desirable to have derivatives in the “source term” of equation (3.5)—i.e., in the volume forces—it is appropriate to rearrange the equation to get the electromagnetic contribution inside the divergence so that it can be treated *like* a surface force, and so that the equation once again returns to conservative form.

With the help of Maxwell’s equations and vector identities, it is found that this can be accomplished if a quantity is transferred to the left hand side of the equation such that the momentum density is “redefined” to include an electromagnetic component  $\mathbf{E} \times \mathbf{B}/4\pi c$  (e.g., Jackson 1999). However, if the usual magnetohydrodynamic assumption is made that the displacement current is negligible,<sup>5</sup> and if electric forces are also assumed negligible (which is valid if the fluid is not highly

---

<sup>5</sup>This assumption is valid for low-frequency phenomena, which is all we usually are interested in considering. This is especially true in the case of radiative MHD, since high-frequency changes in the electric and magnetic fields are by definition already dealt with as radiation (Kaufman 1966).

relativistic, since typical values for the electric field in a highly-conducting medium are  $E \sim v \times B/c$ , then it is possible to transform the Lorentz force  $\mathbf{J} \times \mathbf{B}/c$  into the divergence of a tensor *without* redefining the momentum; the resulting “Maxwell stress” is given by:

$$M_{ij} = -\frac{B_i B_j}{4\pi} + \frac{B^2}{8\pi} \delta_{ij}, \quad (3.18)$$

where the expression is valid for any orthonormal coordinate system. (Note that if the above assumptions are not made,  $M_{ij}$  will have addition terms for the electric field analogous to those shown here for the magnetic field).

If the Maxwell stress is retained in the final equations, an additional equation (the induction equation) is required to determine the time evolution of  $\mathbf{B}$ ; this equation can be obtained by combining Maxwell’s equations with Ohm’s law. The induction equation will not be discussed further here.

### ***Turbulent Stress ( $t_{ij}$ )***

In the above discussion, we have included a “turbulent stress”  $t_{ij}$  in the equations and stated that it is formally equal to zero. The reason for keeping this term is to deal with effects of turbulence that are not captured by the other terms.

Specifically, the length scales and velocities inherent in an astrophysical setting will be much greater than the length scales and velocities associated with molecular viscosity (this is why we have ignored the molecular viscosity in the above equations, as previously discussed). These conditions suggest that the fluid will be highly turbulent and that the Reynolds and Maxwell stresses can have strong, chaotic fluctuations on small length scales and fast timescales that serve to transport momentum and energy within the disk and, ultimately, dissipate some of the disk’s kinetic energy into heat; the latter is observed to occur via a “turbulent

cascade,” in which large-scale fluctuations pass their energy down to progressively smaller spatial scales until eventually fluctuations exist on microscopic scales that transfer their energy to the internal energy of the gas. (For further discussion of turbulence, see Blandford & Thorne 2004.)

These fluctuations can contribute significantly to the mixing in an accretion disk, but in a numerical computation, the grid resolution and/or time resolution is unlikely to be fine enough to capture some of these effects in the existing  $\rho v_i v_j$  and  $-B_i B_j / 4\pi$  terms (almost by definition, the grid resolution will not be good enough to capture the lower end of the turbulent cascade). To allow for the possibility of modeling the mixing due to this small-scale turbulence, which can be crucial for allowing the angular momentum transport in an accretion disk, it is customary to include a turbulent stress  $t_{ij}$  in the accretion disk equations and to include some ad-hoc prescription for its value (e.g., Shakura & Sunyaev 1973), keeping in mind that it should be set to zero in a “real” disk.

The prescription for the turbulence is generally of a statistical nature—that is, it attempts to capture the average rate at which the turbulence transports energy and momentum and dissipates kinetic energy into heat. One simple way to do this is to treat  $t_{ij}$  like a viscous stress, which is often an extremely good approximation for the dissipation (since molecular viscosity can be a major contributor to the heating at the smallest scales of the turbulent cascade) and is sometimes a good approximation for the transport as well. Nonetheless, one should not blindly treat turbulence as being exactly like a viscosity, because the length scales and velocities for the turbulent fluctuations are not necessarily the same in all directions (whereas they are for an isotropic molecular viscosity); the turbulence can take advantage of macroscopic asymmetries in the flow in different ways than a linear

Newtonian viscosity (Ogilvie 2001; Armitage 2004), although, to first order, the magnetically-driven turbulence that occurs via the magnetorotational instability (MRI) is expected to act like a viscosity (Balbus & Papaloizou 1999).

For the transport terms (e.g., the term that appears in the conservation of momentum equation 3.9), it is useful to determine under what conditions the turbulent stress  $t_{ij}$  is needed and under what conditions it can be set to zero. In general, because the turbulent stress allows mixing to take place between the boundaries of the grid cells in a numerical simulation, it will become useful whenever the mixing length is smaller than the grid resolution. The Shakura & Sunyaev (1973) prescription discussed later in this chapter (in which the stress is prescribed as  $t_{r\phi} = \alpha P$  for a dimensionless parameter  $\alpha$ ) will track mixing on length scales of  $\sim \alpha H$ , where  $H$  is the pressure scale height of the accretion disk; thus, it automatically probes scales smaller than the grid size  $\Delta r$  as long as  $\alpha H < \Delta r$ . For a radiation pressure-dominated inner accretion disk, the above condition becomes, approximately,  $\alpha \dot{M} < (\Delta r / 0.2 r_s)$ , where  $\dot{M}$  is the dimensionless accretion rate into the inner disk (expressed in units of the Eddington luminosity divided by the speed of light squared) and  $r_s$  is the Schwarzschild radius of the black hole. In other words, unless  $\dot{M} \gg 1$  (highly super-Eddington accretion), turbulent stress with a physically reasonable value ( $\alpha < 1$ ; Shakura & Sunyaev 1973) will allow us to probe turbulent transport on size scales smaller than a moderate-resolution grid, without having to increase the grid resolution and therefore sacrifice computational time.

In simulations where the space and time resolution are good enough to track the magnetic and kinetic structure of the turbulence on its largest, most important scales, the transport terms involving  $t_{ij}$  can be set to zero. The dissipation, how-

ever, is another story; this will be discussed further in §3.2.5 when conservation of energy is considered.

### 3.2.4.2 Volume Forces (Gravity)

Having finished our discussion of the surface forces, we now turn our attention to the volume forces. Besides the electromagnetic force, which we have already handled and found a way to treat as a surface force, the only volume force that we must consider is gravity.<sup>6</sup> The appropriate force per unit volume is given by  $-\rho\mathbf{g}$ , where  $\mathbf{g} \equiv \nabla\Phi_g$  is the gravitational acceleration in the negative coordinate direction, and  $\Phi_g$  is the gravitational potential.

In what follows, we will assume that the self-gravity of the accretion disk is negligible and that the gravitational force is entirely due to the central object of mass  $M$ . This occurs when the Toomre Q parameter (Toomre 1964) is small. In a disk, the Toomre Q parameter is roughly equal to  $\sim (t_{ff}/t_{orb})^2$ , where  $t_{ff}$  is the free fall timescale due to the self-gravity of the disk and  $t_{orb}$  is the orbital timescale. Only for very massive central objects or very large distances is the orbital timescale large enough for self gravity to be important; therefore, we ignore self-gravity when studying the inner disks around stellar mass black holes considered in this thesis.

There are several ways to prescribe the gravitational potential for a black hole of mass  $M$ . The simplest, of course, is to assume Newtonian gravity, in which case  $\Phi_g = -GM/R$ , where  $R = \sqrt{r^2 + z^2}$  is the distance from the central object. This

---

<sup>6</sup>It is particularly appropriate that in the formalism, the gravitational force enters equation (3.5) as a local source of momentum, since from the point of view of general relativity, gravity really *does* arise from the local curvature of spacetime at the position of each fluid element.

leads to the following components for the acceleration in cylindrical coordinates:

$$g_r = \left( \frac{GM}{R^2} \right) \frac{r}{R} \approx \frac{GM}{r^2} \left( 1 - \frac{3}{2} \frac{z^2}{r^2} \right) \quad (3.19)$$

$$g_z = \left( \frac{GM}{R^2} \right) \frac{z}{R} \approx \frac{GMz}{r^3}, \quad (3.20)$$

where the approximate equalities are valid to second order in  $z/r$  (these approximations are sometimes useful when calculating the vertical integration of the gravitational acceleration in equations 3.63 and 3.65; in this context, the importance of including the second order dependence on  $z$  in the equation for  $g_r$  was noticed by Matsumoto et al. 1984).

Naturally, the Newtonian equations will become less and less accurate as one approaches the event horizon of the black hole, where the effects of relativity become more and more important. Artemova et al. (1996b), in fact, have shown that the steady-state accretion disk structure computed with a Newtonian potential can differ up to  $\sim 50\%$  from the true quantities calculated for a non-rotating black hole.

Many methods have been proposed in the literature to prescribe a value of  $\Phi_g$  (and therefore  $\mathbf{g}$ ) that better matches the predictions of general relativity; these “pseudo-potentials” are, in effect, designed to bridge the gap between Newtonian and fully relativistic simulations of the accretion disk. In general, three methods are possible:

1. Prescribe a value of  $\mathbf{g}$  that leads to roughly accurate predictions for physical quantities in a steady-state accretion disk.
2. Prescribe a value of  $\mathbf{g}$  that leads to roughly accurate predictions for the *timescales* on which these physical quantities can change.

3. Prescribe a value of  $\mathbf{g}$  that is based on the *exact* expression that occurs in the theory of relativity, where the acceleration felt by a locally non-rotating observer in the vicinity of a black hole is given by  $c^2 \nabla (\ln \alpha_{grav})$  (MacDonald & Thorne 1982), where  $\alpha_{grav}$  represents the lapse rate (i.e., the ratio of elapsed proper time  $d\tau$  to elapsed “universal time”  $dt$ , where  $dt$  is the time measured by an observer at infinity).

Examples of these methods for both rotating and non-rotating black holes are discussed in the following sections. At first glance, one might assume that the last method is automatically the best because it is taken directly from the theory of relativity; however, this does not turn out to be the case. One cannot simply graft one part of the theory of relativity onto a fundamentally Newtonian treatment and expect everything to work out okay when many other parts of the equations do not take relativity into account at all. (As a simple example, consider the frequency of a circular orbit in the equatorial plane. This is given by  $\Omega_c = \sqrt{g_r/r}$  in Newtonian theory, but this relation does not hold in general relativity; therefore, a Newtonian simulation which uses the “correct” relativistic value of  $\mathbf{g}$  is essentially guaranteed to get  $\Omega_c$  and other disk timescales wrong.) It is therefore usually better to use one of the other prescriptions, which, although they are formally less valid, are specifically designed to get the “important” physical quantities as close to their correct values as possible in a self-consistent way. However, we do argue below that the relativistic method might be competitive in one specific physical situation that is important astrophysically but not often discussed in the context of pseudo-potentials: the case of a thick disk around a rotating black hole.

### *Prescriptions for $g$ Designed to Match Physical Quantities*

One of the first methods devised to approximate the predictions of general relativity was the pseudo-Newtonian method of Paczyński & Wiita (1980), where the gravitational potential is given by  $\phi_g = -GM/(R - r_s)$ , with  $r_s = 2r_g$  the Schwarzschild radius of the black hole and  $r_g \equiv GM/c^2$  the gravitational radius. This method is extremely popular because of its simplicity and the fact that it reproduces many features of general relativity correctly for a non-rotating black hole, including the locations of the innermost stable circle orbit and innermost bound orbit (Paczyński & Wiita 1980) and the behavior of a particle falling at rest from infinity (Becker & Le 2003). In fact, Artemova et al. (1996b) have shown that the steady-state accretion disk structure calculated using this potential is accurate to at least 10–20%, and often better than 10%.

This acceleration for this potential is usually presented in the literature in spherical coordinates; evaluating it in cylindrical coordinates, we obtain:

$$g_r = \left( \frac{GM}{(R - r_s)^2} \right) \frac{r}{R} \approx \frac{GM}{(r - r_s)^2} \left( 1 - C_g \frac{z^2}{r^2} \right) \quad (3.21)$$

$$g_z = g_r \frac{z}{r} \approx \frac{GMz}{r(r - r_s)^2}, \quad (3.22)$$

where

$$C_g = \frac{3}{2} + \frac{r_s}{r - r_s} \quad (3.23)$$

and the approximate equalities are valid to second order in  $z/r$  as long as the factor  $r - r_s$  in the denominator of  $C_g$  is significantly larger than  $z$ . (Note that as expected, these equations reduce to the Newtonian case when  $R \gg r_s$ .)

Artemova et al. (1996b) have derived an extension to this potential for rotating black holes that preserves many of the features of the Paczyński & Wiita (1980) formulation. The extension was discussed by Artemova et al. (1996b) in the context

of orbits in the equatorial plane (i.e., the plane that is perpendicular to the spin axis of the black hole), and its useful properties appear most strongly in that context. Here, however, we present a simple generalization of this potential to three dimensions that will reduce to Paczyński & Wiita (1980) potential at *any* location if the black hole is non-rotating; within the equatorial plane, our prescription has the same values as Artemova et al. (1996b). In this case, we find that the components of the acceleration in cylindrical coordinates are:

$$g_r = \left( \frac{GM}{R^{2-\beta} (R - r_h)^\beta} \right) \frac{r}{R} \approx \frac{GM}{r^{2-\beta} (r - r_h)^\beta} \left( 1 - C_g \frac{z^2}{r^2} \right) \quad (3.24)$$

$$g_z = g_r \frac{z}{r} \approx \frac{GMz}{r^{3-\beta} (r - r_h)^\beta}, \quad (3.25)$$

where

$$C_g = \frac{3}{2} + \left( \frac{\beta}{2} \right) \frac{r_h}{r - r_h}, \quad (3.26)$$

$\beta = r_{ms}/r_h - 1$  as defined by Artemova et al. (1996b),  $r_h = r_g (1 + \sqrt{1 - a^2})$  is the radius of the event horizon of the rotating black hole,  $a$  is the dimensionless angular momentum of the black hole (measured in units of  $GM^2/c$ ), and  $r_{ms}$  is the radius of the innermost marginally stable circular orbit in the equatorial plane (Bardeen et al. 1972), given by

$$r_{ms} = r_g \left[ 3 + Z_2 \pm \sqrt{(3 - Z_1)(3 + Z_1 + 2Z_2)} \right] \quad (3.27)$$

$$Z_1 = 1 + (1 - a^2)^{1/3} \left[ (1 + a)^{1/3} + (1 - a)^{1/3} \right] \quad (3.28)$$

$$Z_2 = (3a^2 + Z_1^2)^{1/2}; \quad (3.29)$$

the minus sign in equation (3.27) is for an accretion disk that corotates with the black hole spin, and the plus sign is for an accretion disk that is counterrotating. As expected, equations (3.24) and (3.25) reduce to the Paczyński & Wiita (1980)

values for a non-rotating black hole ( $\beta = 2$ ), and they reduce to the Newtonian values when  $R \gg r_h$ . Artemova et al. (1996b) have shown that as for the Paczyński & Wiita (1980) case, the prescription given here for a rotating black hole generally gives steady-state accretion disks that are accurate to  $\sim 10$ – $20\%$ .

### *Prescriptions for $g$ Designed to Match Physical Timescales*

Nowak & Wagoner (1991) introduced a prescription for gravity that does a better job approximating physical timescales around a non-rotating black hole than the Paczyński & Wiita (1980) method, and Mukhopadhyay & Misra (2003) have generalized this prescription for rotating black holes. We present here the Mukhopadhyay & Misra (2003) version only, since it reduces to the Nowak & Wagoner (1991) case when  $a = 0$ . As with the Artemova et al. (1996b) prescription, the Mukhopadhyay & Misra (2003) prescription was specifically designed for use on the equatorial plane, but we generalize it here so that we can derive equations for the gravitational force at an arbitrary height  $z$  in the disk. In cylindrical coordinates, the components of the gravitational acceleration are:

$$g_r = \left(\frac{GM}{R^2}\right) \left(1 - \frac{r_{ms}}{R} + \frac{r_{ms}^2}{R^2}\right) \frac{r}{R} \approx \frac{GM}{r^2} \left(1 - \frac{r_{ms}}{r} + \frac{r_{ms}^2}{r^2} - C_g \frac{z^2}{r^2}\right) \quad (3.30)$$

$$g_z = g_r \frac{z}{r} \approx \frac{GMz}{r^3} \left(1 - \frac{r_{ms}}{r} + \frac{r_{ms}^2}{r^2}\right), \quad (3.31)$$

where

$$C_g = \frac{3}{2} - 2\frac{r_{ms}}{r} + \frac{5}{2} \left(\frac{r_{ms}}{r}\right)^2. \quad (3.32)$$

This prescription is expected to give values for the orbital frequency that are accurate to better than 25% (as compared to  $\sim 50\%$  or worse for the inner radii in a Paczyński & Wiita 1980 potential), while still remaining accurate to  $\sim 10\%$  in the specific energy of infalling material (Mukhopadhyay & Misra 2003).

### ***Prescriptions for $g$ from General Relativity***

Finally, we consider the value of  $\mathbf{g}$  calculated directly from general relativity that we have alluded to above. For a non-rotating black hole, the lapse rate is given by  $\alpha_{grav} = \sqrt{1 - r_s/R}$  (MacDonald & Thorne 1982), leading to a particularly simple expression for the acceleration in cylindrical coordinates:

$$g_r = \alpha_{grav}^{-2} \left( \frac{GM}{R^2} \right) \frac{r}{R}, \quad (3.33)$$

with  $g_z = (z/r) g_r$  as usual. This is the acceleration that would be measured and actually felt by local observers; one could also imagine defining the acceleration as the change in velocity per unit *universal* time and therefore multiplying the above expression by  $\alpha_{grav}$ . This would do a poorer job of modeling the gravitational force actually experienced by the disk, but a better job of modeling the disk timescales as measured by faraway observers.

Both of these possibilities were considered (although in a slightly roundabout and less physically intuitive way) by Artemova et al. (1996b), who found that they do not do as good of a job as either the Paczyński & Wiita (1980) or Nowak & Wagoner (1991) potentials at reproducing the steady-state disk structure. Furthermore, it is easy for us to check how well they do at reproducing typical accretion disk timescales, since the actual orbital frequency for a non-rotating black hole in general relativity is equal to the Keplerian value  $\Omega_k$ . We find that the timescales calculated from the above two prescriptions are  $\Omega_k/\alpha_{grav}$  and  $\Omega_k/\alpha_{grav}^{1/2}$ , respectively, which generally give worse agreement (although in some cases comparable agreement) to general relativity than the Nowak & Wagoner (1991) values. Overall, therefore, the relativistic prescription does not appear to be competitive for a non-rotating black hole.

For a rotating black hole, however, the value of  $\mathbf{g}$  calculated from general

relativity may be useful in the case of thick disks, where we are interested in values of  $\mathbf{g}$  far above the equatorial plane. This is because the potentials of Artemova et al. (1996b) and Mukhopadhyay & Misra (2003) were designed to match specific quantities that are only defined on the equatorial plane (in particular, the radius  $r_{ms}$  of the innermost stable circular orbit and the orbital frequency  $\Omega_k$ ). There is no reason to expect these potentials to be accurate at large distances from the equatorial plane, where they effectively treat a point at distance  $R$  (measured in an arbitrary direction) as though it were located on the equatorial plane at a distance  $r = R$ .

The value of  $\alpha_{grav}$  for a rotating black hole and the corresponding value of  $\mathbf{g}$  can be found in MacDonald & Thorne (1982); we do not reproduce them here because of their mathematical complexity. To our knowledge, no one has explored using these values in an otherwise Newtonian simulation; this may be a worthwhile possibility to study further.

### 3.2.4.3 The Final Momentum Equation

Combining all the above information, we can write out the momentum equation as:

$$\frac{\partial}{\partial t}(\rho\mathbf{v}) = -\nabla \cdot \mathbb{T} - \rho\mathbf{g}, \quad (3.34)$$

where the stress tensor  $\mathbb{T}$  is equal to the sum of surface force and transport terms specified previously. Depending on the assumptions being made about the accretion disk, some of these terms may be zero. Here, we do not make any assumptions about which of these terms are zero, but rather write the momentum equation in cylindrical coordinates using  $\mathbb{T}$  only; it is easy to derive equations for a specific set of assumptions by substituting the appropriate values of  $\rho v_i v_j$ ,  $P_{ij}$ ,  $M_{ij}$  and  $t_{ij}$

into these equations.

Using the results of Appendix A to write the divergence of  $\mathbb{T}$  in cylindrical coordinates, we obtain:

$$\frac{\partial}{\partial t}(r\rho v_r) = - \left[ \frac{\partial}{\partial r}(rT_{rr}) + \frac{\partial}{\partial \phi}(T_{r\phi}) + \frac{\partial}{\partial z}(rT_{rz}) \right] + T_{\phi\phi} - r\rho g_r \quad (3.35)$$

$$\frac{\partial}{\partial t}(r^2\rho v_\phi) = - \left[ \frac{\partial}{\partial r}(r^2T_{r\phi}) + \frac{\partial}{\partial \phi}(rT_{\phi\phi}) + \frac{\partial}{\partial z}(r^2T_{\phi z}) \right] \quad (3.36)$$

$$\frac{\partial}{\partial t}(r\rho v_z) = - \left[ \frac{\partial}{\partial r}(rT_{rz}) + \frac{\partial}{\partial \phi}(T_{\phi z}) + \frac{\partial}{\partial z}(rT_{zz}) \right] - r\rho g_z, \quad (3.37)$$

where we have manipulated these equations slightly using the fact that  $\mathbb{T}$  is a symmetric tensor (which is obvious from examining the constituent terms of  $T_{ij}$  listed in §3.2.4.1). Also notice the factor of  $r^2$  within the derivatives in equation (3.36); this comes from applying the chain rule to subsume an extra “source term” involving  $T_{\phi r}$  ( $= T_{r\phi}$ ) that appears in this equation when  $\nabla \cdot \mathbb{T}$  is computed in cylindrical coordinates (see Appendix A). Physically, this extra factor of  $r$  represents the fact that angular momentum density  $r\rho v_\phi$  is conserved in the fluid.

A particularly useful implementation of these equations for our purposes involves ignoring the Maxwell stress and setting all terms in the turbulent stress to zero except for  $t_{r\phi}$  (i.e. the component that takes advantage of the orbital shear in the disk). We write this component as  $t_{r\phi} = \alpha P$  following Shakura & Sunyaev (1973), where  $\alpha$  is a dimensionless measure of the stress. Here we make no assumptions about the value of  $\alpha$  or its possible dependence on space, time, or other parameters; it is simply a mathematical device (in other words, there is no requirement that  $t_{r\phi}$  depends on pressure, because it is possible to have  $\alpha \propto 1/P$ ). Under these simplifications, equations (3.35) through (3.37) become:

$$\begin{aligned} \frac{\partial}{\partial t}(r\rho v_r) &= - \frac{\partial}{\partial r} \left[ r(P + \rho v_r^2) \right] - \frac{\partial}{\partial \phi} (\alpha P + \rho v_r v_\phi) \\ &\quad - \frac{\partial}{\partial z} (r\rho v_r v_z) + P + \rho (v_\phi^2 - r g_r) \end{aligned} \quad (3.38)$$

$$\frac{\partial}{\partial t} (r^2 \rho v_\phi) = -\frac{\partial}{\partial r} \left[ r^2 (\alpha P + \rho v_r v_\phi) \right] - \frac{\partial}{\partial \phi} \left[ r (P + \rho v_\phi^2) \right] - \frac{\partial}{\partial z} (r^2 \rho v_\phi v_z) \quad (3.39)$$

$$\frac{\partial}{\partial t} (r \rho v_z) = -\frac{\partial}{\partial r} (r \rho v_r v_z) - \frac{\partial}{\partial \phi} (\rho v_\phi v_z) - \frac{\partial}{\partial z} \left[ r (P + \rho v_z^2) \right] - r \rho g_z. \quad (3.40)$$

Notice that since the velocity of a circular orbit in the equatorial plane is given by  $v_c = \sqrt{r g_r}$ , the last term in equation (3.38) can be written as  $\approx \rho (v_\phi^2 - v_c^2)$  near the equatorial plane; in other words, it is proportional to the difference in the squares of the azimuthal and orbital velocities. From a practical standpoint, this term will usually be small compared to either of the components  $\rho v_\phi^2$  and  $\rho v_c^2$  that are subtracted in order to form it, although it still can be significant compared to the other terms in equation (3.38). Therefore, numerical simulations that retain a large number of significant figures (e.g., double floating point precision) may be required to accurately measure this term.

### 3.2.5 Conservation of Energy

We can intuitively construct an equation for the evolution of the total energy density  $E_{tot}$  by replacing  $\mathbf{F}_u$  and  $S_u$  in equation (3.5) with the power produced by each of the forces considered in §3.2.4. In other words, we could use  $\mathbf{T} \cdot \mathbf{v}$  for the power produced by surface forces and  $\rho \mathbf{g} \cdot \mathbf{v}$  for the power produced by the gravitational volume force (cf. equation 3.34). However, this ignores some subtle effects, and, perhaps more important, provides no information about the different types of energy density that constitute  $E_{tot}$ . Of course, these can be inserted “by hand” based on other physical knowledge, but their origin becomes directly clear if we follow a different, more mathematical procedure to derive the energy equation.

This more illuminating approach begins by taking the dot product of  $\mathbf{v}$  into equation (3.34). Recalling that the combined time derivative and Reynolds stress

in equation (3.34) can be written  $\rho(d\mathbf{v}/dt)$  as was done in equation (3.8), we obtain:

$$\rho\mathbf{v}\cdot\frac{d\mathbf{v}}{dt} = -\mathbf{v}\cdot(\nabla\cdot\mathbf{G}) - \rho\mathbf{g}\cdot\mathbf{v}, \quad (3.41)$$

where  $\mathbf{G}$  is once again the stress tensor not including the Reynolds stress (i.e., not including the transport term). Standard vector manipulations and use of the conservation of mass equation (3.6) then yields

$$\frac{\partial}{\partial t}\left(\frac{1}{2}\rho v^2\right) = \left[-\nabla\cdot\left(\frac{1}{2}\rho v^2\mathbf{v} + \mathbf{G}\cdot\mathbf{v}\right) - \rho\mathbf{g}\cdot\mathbf{v}\right] + \mathbf{G}\cdot\nabla\mathbf{v}. \quad (3.42)$$

This, of course, is an equation for the kinetic energy density of the fluid, and it is *almost* in conservative form; as expected, the stress tensor  $\mathbf{G}$  produces an energy flux that takes the role of  $\mathbf{F}_u$  in equation (3.5), while the gravitational field is a local source of kinetic energy that takes the role of  $S_u$ . However, we are left with an “extra” source term  $\mathbf{G}\cdot\nabla\mathbf{v}$  that is undesirable because it contains spatial derivatives. This term represents the rate at which other forms of energy are converted to kinetic energy, and in order to evaluate it for each constituent of  $\mathbf{G}\cdot\nabla\mathbf{v}$ , it is necessary to use the individual energy equations that are obtained when gas, radiation, and electromagnetic fields are considered separately.

An instructive example is the equation for gas particles; there, the relevant component of  $\mathbf{G}$  is  $\mathbf{P}_{\text{gas}}$  (we allow for the arbitrary case in which the gas pressure is a tensor rather than a scalar; e.g., if molecular viscosity is important), and the first law of thermodynamics derived from the Boltzmann equation states:

$$Q_{\text{gas}} = \rho \left[ \frac{d}{dt} \left( \frac{E_{\text{gas}}}{\rho} \right) + \mathbf{P}_{\text{gas}} \cdot \frac{d}{dt} (\rho^{-1}) \right], \quad (3.43)$$

where  $Q_{\text{gas}}$  is the volume rate at which the gas is heated (i.e., the rate at which energy density is transferred to the gas particles from other species) and  $E_{\text{gas}}$  is

the internal energy density defined in equation (3.11). Manipulating this equation and again using the conservation of mass equation (3.6) yields

$$\mathbf{P}_{\text{gas}} \cdot \nabla \mathbf{v} = - \left[ \frac{\partial}{\partial t} (E_{\text{gas}}) + \nabla \cdot (E_{\text{gas}} \mathbf{v}) \right] - Q_{\text{gas}}. \quad (3.44)$$

Therefore, replacing the  $\mathbf{P}_{\text{gas}} \cdot \nabla \mathbf{v}$  component of  $\mathbf{G} \cdot \nabla \mathbf{v}$  with its more explicit form is equivalent to adding a new type of energy  $E_{\text{gas}}$  to the kinetic energy in equation (3.42). We also obtain an extra term  $Q_{\text{gas}}$ , but as mentioned in §3.2.1, this term will cancel out with the equivalent terms that are derived for radiation and electromagnetic fields, since the total amount of energy that these species exchange with each other is conserved. Therefore, equation (3.42) is one step closer to being in conservative form.

A similar procedure can be followed for all the constituent terms of  $\mathbf{G}$ . The radiation stress is considerably more complicated (see Mihalas & Mihalas 1984 for details), while the Maxwell stress is best handled by going back to its original formulation as a Lorentz volume force (§3.2.4.1); in that case, one finds that

$$\mathbf{v} \cdot \left( \rho_e \mathbf{E} + \frac{\mathbf{J} \times \mathbf{B}}{c} \right) \approx -\frac{J^2}{\sigma_e} + \mathbf{J} \cdot \mathbf{E}, \quad (3.45)$$

using simple vector manipulations and Ohm's law (here,  $\sigma_e$  is the conductivity) and ignoring relativistic terms of order  $v/c$  as was previously done for the electromagnetic field. We can identify  $-J^2/\sigma_e$  as the Ohmic heating of electrons by the field and therefore treat it similarly as the  $Q_{\text{gas}}$  term above (i.e., it will cancel out in the total energy equation), leaving

$$\mathbf{J} \cdot \mathbf{E} \approx - \left[ \frac{\partial}{\partial t} \left( \frac{B^2}{8\pi} \right) + \nabla \cdot \left( \frac{c}{4\pi} \mathbf{E} \times \mathbf{B} \right) \right], \quad (3.46)$$

where we have used vector manipulations and Maxwell's equations (e.g., Jackson 1999) to rewrite this term, again in the nonrelativistic limit. If we further assume

ideal MHD, where  $\mathbf{E} = -(\mathbf{v} \times \mathbf{B})/c$ , equation (3.46) can be transformed into

$$\mathbf{J} \cdot \mathbf{E} \approx - \left[ \frac{\partial}{\partial t} \left( \frac{B^2}{8\pi} \right) + \nabla \cdot \left( \frac{B^2}{8\pi} \mathbf{v} \right) + \nabla \cdot (\mathbf{M} \cdot \mathbf{v}) \right], \quad (3.47)$$

where  $\mathbf{M}$  is the Maxwell stress tensor defined in equation (3.18). This equation therefore defines the magnetic energy density  $B^2/8\pi$  that can be added to the kinetic energy in a similar manner as  $E_{gas}$  was, and it also recovers the Maxwell stress component of the  $\mathbf{G} \cdot \mathbf{v}$  term in equation (3.42); however, it should be remembered that in the case of non-ideal MHD, the advection of magnetic energy and the divergence of  $\mathbf{M} \cdot \mathbf{v}$  need to be replaced by the more complicated expression involving  $\mathbf{E} \times \mathbf{B}$  in equation (3.46).

Putting all the above information together, the following energy equation is obtained:

$$\frac{\partial}{\partial t} (E_{tot}) = -\nabla \cdot (E_{tot} \mathbf{v} + \mathbf{G} \cdot \mathbf{v} + \mathbf{F}_{rad}) - \rho \mathbf{g} \cdot \mathbf{v}, \quad (3.48)$$

where  $E_{tot}$  is the total energy density, given by

$$E_{tot} = \frac{1}{2} \rho v^2 + E_{gas} + E_{rad} + \frac{B^2}{8\pi}, \quad (3.49)$$

and  $\mathbf{F}_{rad}$  is the flux of radiative energy. The origin of the  $\nabla \cdot \mathbf{F}_{rad}$  term in equation (3.48) is mathematically obscure, although physically intuitive; it represents the energy that photons take with them as they leave a radiating volume element.<sup>7</sup> In the diffusion limit,  $\mathbf{F}_{rad}$  is given by

$$\mathbf{F}_{rad} = -\frac{c}{\rho \kappa} \nabla \cdot \mathbf{P}_{rad}, \quad (3.50)$$

---

<sup>7</sup>The equivalent term that arises from the Boltzmann energy equation for gas particles is thermal conduction, which we implicitly set to zero here because it is generally of the same order of magnitude as the molecular viscosity, which we have already ignored. The radiative energy flux  $\mathbf{F}_{rad}$  is more important than thermal conduction because the photon speed  $c$  in the numerator of equation (3.50) is much larger than the typical gas particle speed, while in situations with reasonable optical depths, the opacity  $\kappa$  in the denominator will be much smaller than the equivalent cross section per unit mass at which gas particles' paths are interrupted (Mihalas & Mihalas 1984).

where  $\kappa$  is the opacity. This simply states that the radiative acceleration experienced by the fluid ( $-\nabla \cdot \mathbf{P}_{\text{rad}}/\rho$ ) equals the incident radiative momentum flux ( $\mathbf{F}_{\text{rad}}/c$ ) multiplied by the cross-sectional area per unit mass for the fluid's interaction with the radiation ( $\kappa$ ). Technically, the Rosseland mean opacity  $\kappa_R$  should appear here instead, but it can be replaced by  $\kappa$  if we assume that the opacity does not depend on the frequency of the radiation (Kato et al. 1998). In this case, we adopt a mix of electron scattering and free-free absorption:

$$\kappa = \kappa_{sc} + \kappa_{abs}, \quad (3.51)$$

where

$$\kappa_{sc} = 0.2 (1 + X) \text{ cm}^2/\text{g} \quad (3.52)$$

$$\kappa_{abs} = 3.68 \times 10^{22} g_{ff} (1 + X) (X + Y) \left( \frac{\rho}{\text{g/cm}^3} \right) \left( \frac{T}{1 \text{ K}} \right)^{-7/2} \text{ cm}^2/\text{g}, \quad (3.53)$$

the latter being the Kramers opacity that approximates the opacity due to free-free absorption for the temperatures of interest in our accretion disk (Ostlie & Carroll 1996). Here,  $X$  and  $Y$  are the hydrogen and helium mass fractions discussed earlier, and  $g_{ff}$  is the Gaunt factor for free-free absorption, which we take to be  $\approx 0.87$  following the (unstated) assumption of Kato et al. (1998) and others, although we note that this number can actually have significant dependence on photon energy in the X-ray regime (e.g., Karzas & Latter 1961).

It is possible to manipulate equation (3.48) further so as to remove the gravitational source term and make the equation fully conservative. As one would expect, this procedure (which can be accomplished through straightforward manipulation of the  $\rho \mathbf{g} \cdot \mathbf{v}$  term and use of the conservation of mass equation) leads to a redefinition of  $E_{\text{tot}}$  to include the gravitational potential energy  $\Phi_g$ . However, what is not often noticed is that this manipulation leaves behind a term  $\rho \partial \Phi_g / \partial t$  on

the right hand side of the equation—i.e., a term that represents the “injection” of gravitational energy into the system due to the change in the local gravitational field. This term is usually negligible in the situations we consider here, where the accretion disk is not self-gravitating and the mass accretion rate  $\dot{M}$  onto the central object of mass  $M$  is slow enough so that  $\int \dot{M} dt \ll M$  during a typical simulation; thus, the energy equation in its “fully conservative” form, including  $\Phi_g$ , can be used (e.g., the simulations of Watarai & Mineshige 2003).

However, the  $\rho \partial \Phi_g / \partial t$  term can be important in certain circumstances, in particular in a self-gravitating disk. To preserve generality, we prefer to leave gravity out of the definition of  $E_{tot}$  and let it remain as a source term in equation (3.48), which is perfectly easy to handle in our numerical simulations; this also has the added benefit of allowing us to use more complicated prescriptions for the gravitational acceleration  $\mathbf{g}$  that do not easily lend themselves to calculation of  $\Phi_g$  (for example, some of the prescriptions listed in §3.2.4.2).<sup>8</sup>

Finally, a word about the appearance of the turbulent stress in the conservation of energy equation. In equation (3.48), we have implicitly included the turbulent energy transport as part of the stress tensor  $\mathbf{G}$ ; i.e., the turbulent energy flux is given by  $\mathbf{t} \cdot \mathbf{v}$ . In ideal MHD, this assumption is perfectly reasonable, because the Reynolds and Maxwell stresses that ultimately make up the turbulent stress enter the equations in the same way. For non-ideal MHD, however, where the disk has a finite conductivity, we can see from equation (3.46) that the Maxwell stress enters the energy equation in a different way; specifically, there is a component of the energy flux due to Maxwell stresses that has *no* dependence on the fluid

---

<sup>8</sup>But see Blandford & Thorne (2004) for discussion of a gravitational stress tensor (analogous to the Maxwell stress tensor) that can be derived to handle gravity in a fully-conservative manner, even in an arbitrary case where the gravitational potential is not assumed to be constant.

velocity  $\mathbf{v}$ . Given the fact that numerical simulations of the magnetorotational instability (MRI) usually show the turbulent Maxwell stress dominating the turbulent Reynolds stress by a factor of a few (e.g., Hawley et al. 1995, 1996; Hawley 2000; Sano et al. 2004), this difference is potentially important for an accretion disk with low to moderate conductivity, since in this case our energy equation effectively treats the turbulence as if it were only due to a Reynolds stress. This issue is often overlooked and could be of importance in outer regions of the disk where the temperature is low, but in the inner regions of the disk that we study here, the assumption of infinite conductivity is likely to be a sufficient one.

A related issue involves the dissipation of turbulent energy into heat, which, in the non-ideal MHD case, can proceed either by Ohmic dissipation of Maxwell stress that primarily heats the electrons, or viscous dissipation of Reynolds stress that primarily heats the ions. Being able to distinguish between these two possibilities undoubtedly involves knowledge of the detailed microphysics of the disk and has important consequences for two-temperature disk models and some of the objections that have been raised to them (e.g., Bisnovaty-Kogan & Lovelace 1997, 2000), but, fortunately, it has absolutely no effect in our simulations. As long as the gas and radiation are assumed to be in equilibrium, the dissipation of turbulent energy into heat must ultimately be divided in a well-defined way between the electrons, ions and photons. This once again illustrates the advantages of using the full conservation of energy equation whenever it is possible to do so; the interaction terms between different species cancel out and therefore make the physics that needs to be modeled considerably simpler. In cases where the gas and radiation are not assumed to be in equilibrium, however, different methods can be found to attempt to model the turbulent dissipation; for example, the simulations

of Hirose et al. (2006) used a non-conservative numerical scheme and then added all the “missing” energy to the internal energy of the gas to roughly track the heating due to turbulence.

Returning to the energy equation (3.48) and writing it under the specific disk assumptions we have previously made (isotropic equations of state and no magnetic fields), we obtain:

$$\frac{\partial}{\partial t} (E_{tot}) = -\nabla \cdot \left[ (E_{tot} + P) \mathbf{v} + \mathbf{t} \cdot \mathbf{v} + \mathbf{F}_{rad} \right] - \rho \mathbf{g} \cdot \mathbf{v}, \quad (3.54)$$

where  $E_{tot} = \frac{1}{2} \rho v^2 + E_{gas} + E_{rad}$ . Specializing this equation to cylindrical coordinates in the case where the turbulent stress is assumed to be entirely due to the  $t_{r\phi} = t_{\phi r} = \alpha P$  component, we finally obtain:

$$\begin{aligned} \frac{\partial}{\partial t} (r E_{tot}) &= -\frac{\partial}{\partial r} \left[ r (E_{tot} + P) v_r + r \alpha P v_\phi + r F_{rad,r} \right] \\ &\quad -\frac{\partial}{\partial \phi} \left[ (E_{tot} + P) v_\phi + \alpha P v_r + F_{rad,\phi} \right] \\ &\quad -\frac{\partial}{\partial z} \left[ r (E_{tot} + P) v_z + r F_{rad,z} \right] - \rho (g_r v_r + g_z v_z). \end{aligned} \quad (3.55)$$

### 3.2.6 Final Equations for Our Accretion Disk Model

In this section, we summarize the equations of conservation of mass, conservation of momentum and conservation of energy that we will consider in our current accretion disk models:

$$\frac{\partial}{\partial t} (r\rho) = - \left[ \frac{\partial}{\partial r} (r\rho v_r) + \frac{\partial}{\partial \phi} (\rho v_\phi) + \frac{\partial}{\partial z} (r\rho v_z) \right] \quad (3.56)$$

$$\begin{aligned} \frac{\partial}{\partial t} (r\rho v_r) &= -\frac{\partial}{\partial r} \left[ r (P + \rho v_r^2) \right] - \frac{\partial}{\partial \phi} (\alpha P + \rho v_r v_\phi) \\ &\quad -\frac{\partial}{\partial z} (r\rho v_r v_z) + P + \rho (v_\phi^2 - r g_r) \end{aligned} \quad (3.57)$$

$$\frac{\partial}{\partial t} (r^2 \rho v_\phi) = -\frac{\partial}{\partial r} \left[ r^2 (\alpha P + \rho v_r v_\phi) \right] - \frac{\partial}{\partial \phi} \left[ r (P + \rho v_\phi^2) \right] - \frac{\partial}{\partial z} (r^2 \rho v_\phi v_z) \quad (3.58)$$

$$\frac{\partial}{\partial t} (r \rho v_z) = -\frac{\partial}{\partial r} (r \rho v_r v_z) - \frac{\partial}{\partial \phi} (\rho v_\phi v_z) - \frac{\partial}{\partial z} \left[ r (P + \rho v_z^2) \right] - r \rho g_z \quad (3.59)$$

$$\begin{aligned} \frac{\partial}{\partial t} (r E_{tot}) = & -\frac{\partial}{\partial r} \left[ r (E_{tot} + P) v_r + r \alpha P v_\phi + r F_{rad,r} \right] \\ & -\frac{\partial}{\partial \phi} \left[ (E_{tot} + P) v_\phi + \alpha P v_r + F_{rad,\phi} \right] \\ & -\frac{\partial}{\partial z} \left[ r (E_{tot} + P) v_z + r F_{rad,z} \right] - \rho (g_r v_r + g_z v_z), \quad (3.60) \end{aligned}$$

where  $E_{tot} = \frac{1}{2} \rho v^2 + E_{gas} + E_{rad}$ .

These equations involve several simplifications, even for the one-temperature disks (thermal equilibrium between gas particles and radiation) that we are already implicitly assuming. The first major simplification is that the equations are not fully relativistic. In fact, it would not be that difficult to convert the gravitational and hydrodynamic terms in these equations to their fully relativistic forms—the primary differences are that the mass and energy equations would need to be combined, the relativistic value  $c^2 \nabla (\ln \alpha_{grav})$  would need to be used for the gravitational acceleration (§3.2.4.2), and several factors of  $\alpha_{grav}$  and  $\Gamma \equiv \sqrt{1 - v^2/c^2}$  would need to appear to take gravitational time dilation and special relativistic effects into account; for rotating black holes, meanwhile, additional terms would need to be added to the equations to handle inertial frame dragging (MacDonald & Thorne 1982). However, the *radiation* equations in the relativistic limit (when terms of order  $v/c$  are not ignored) are considerably more complex (Mihalas & Mihalas 1984), and that is the reason why fully relativistic MHD disk simulations in the literature do not yet include radiation. Given the importance of radiation in the accretion disk of GRS 1915+105, there is very little reason to employ a fully relativistic code to simulate it; even if we correctly probe gravity in the in-

ner regions of the disk (where  $v/c \lesssim 1$ ), our errors are likely to be swamped by uncertainties in the radiation physics operating in these regions. The best we can hope for is that the pseudo-Newtonian methods approximately capture the effect of relativity on the gravitational force—though it must be said that the ubiquity of jets in astrophysics (including in systems such as young stars, where relativity is unimportant) suggests that we can still learn something very important about the process of jet formation in GRS 1915+105 using our methods.

The second major simplification in these equations is that we ignore the presence of large-scale magnetic fields. This is not realistic, since magnetic fields are thought to be an important contributor to jet formation, and it is a long term goal of our simulations to include them. However, the lack of certainty about an appropriate initial condition for the magnetic field structure, as well as our desire to simplify these equations further by reducing them to one- or two-dimensional versions (§3.3), suggests that a good preliminary first step is to ignore large-scale magnetic fields in our equations. However, as discussed in §3.3.2, we do provide ways to alter the vertical structure of the disk and the  $\alpha$  parameter for the stress that are roughly designed to mimic the effect of magnetic fields and jets; this is in contrast to most radiative accretion disk simulations in the literature, which completely ignore the possible effects of the large scale magnetic field.

The final simplification in these equations is that we parametrize the turbulent stress as  $t_{r\phi} = \alpha P$  following Shakura & Sunyaev (1973), with all other components set to zero. The particular parametrization of  $t_{r\phi}$  is not by itself an assumption because  $\alpha$  is just a mathematical device that can have any possible dependence on space, time or other physical parameters (a fact often forgotten in the literature). However, setting all the other components of the turbulent stress to zero

is an assumption, although it is straightforward to insert the appropriate terms into equations (3.56) through (3.60) if this assumption is to be relaxed. Nonetheless, to the extent that the turbulence behaves like a Newtonian viscosity, the vastly stronger shear in the  $r\phi$  plane strongly suggests that this component will be dominant, at least for a thin disk.

Most well-known disk models make further simplifications to the above equations. In addition to assuming axisymmetry and integrating over the disk height (discussed in the next section), the so-called “slim disk” model of Abramowicz et al. (1988) ignores the diffusion of radiative energy in the radial direction, while the original disk model of Shakura & Sunyaev (1973) additionally ignores the radial advection terms in equation (3.60) and the *entire* radial momentum equation (3.57), replacing the latter with the assumption that orbits within the disk are close enough to circular so that  $v_\phi \approx v_c = \sqrt{rg_r}$ . The Abramowicz et al. (1988) model is therefore valid up to higher accretion rates than the Shakura & Sunyaev (1973) model.

### 3.3 Simulations in Fewer Than Three Dimensions

Equations (3.56) through (3.60) correctly govern the evolution of an accretion disk under the assumptions listed. However, it can be computationally expensive to simulate these equations numerically; the presence of spatial derivatives in all three physical directions means that they must be modeled on a three-dimensional grid. It is therefore desirable to find ways to modify these equations to retain as much of the physics as possible while removing the dependence on one or two physical coordinates so that, for example, they only need to be evolved on a one-dimensional grid in  $r$  or a two-dimensional grid in either  $(r, \phi)$  or  $(r, z)$ . These

modifications are the subject of this section.

### 3.3.1 Removing the Dependence on Azimuthal Angle ( $\phi$ )

Removing the dependence on  $\phi$  is relatively simple; to first order, we can assume that the accretion disk is axisymmetric (i.e., cylindrically symmetric about the rotation axis) and therefore set  $\partial/\partial\phi = 0$  in the above equations. This assumption is reasonable, because although physical quantities in an accretion disk are expected to vary significantly with distance  $\sqrt{r^2 + z^2}$  from the central object (due to the importance of gravity) and with distance  $z$  along the disk's rotation axis (due to the importance of angular momentum), there is no obvious reason why they should depend on  $\phi$ . This assumption may break down in the extreme outer disk, however (where the formation of the disk from, e.g., Roche lobe overflow from a binary companion can lead to dependence on  $\phi$ ; e.g., Truss et al. 2000; Truss & Wynn 2004), and it also precludes the formation of spiral waves and other axial instabilities (e.g., Tagger & Pellat 1999), but for the inner accretion disk, it is at least a reasonable assumption to start out with.

Simulations that make this assumption while still retaining all three dimensions of any vector quantities in the equations (e.g.,  $v_\phi$  and other quantities that point in the  $\phi$  direction are *not* set to zero) are often referred to in the literature as “2.5-dimensional” simulations. Despite the name, the reader should be assured that these simulations do not involve any fractal dimensions or other exotic processes. They are more properly (and less confusingly) referred to as “three-dimensional axisymmetric” simulations.

### 3.3.2 Removing the Dependence on Height ( $z$ )

Removing the dependence on  $z$  from the accretion disk equations is considerably more complex, because physical quantities *do* change significantly with  $z$  inside the disk. To first order, most quantities will have a maximum at  $z = 0$  and then drop off with height; this is because the vertical gravitational force tends to concentrate quantities near the equatorial plane.

The general procedure for solving this problem is to integrate the equations over  $z$ ; what makes this idea useful is that in many cases, the scale height of the accretion disk will be small compared to other relevant length scales in the system, and therefore the integration is not likely to introduce errors that are too large; in other words, the timescales for change in the vertical direction are likely to be shorter than other timescales in the system, and therefore the entire “vertical column” of material at a given location in the disk can be assumed to roughly act like a single object on the timescales of interest.

Vertical integration of the above equations is often done in the literature, but not often done carefully; a wide variety of models are used and unstated assumptions are often made. This is often considered to be a minor concern because the vertical integration in pure disk models introduces dimensionless numbers into the equations that are all of order  $\sim 1$ . However, the discovery that jets are a ubiquitous phenomenon associated with accretion disks certainly challenges this conclusion; a jet, by definition, constitutes a major disruption to the vertical structure of the disk from which it emanates. To allow for a first-order treatment of the effect of a jet, we provide in this section a model-independent way of performing the vertical integrations making as few assumptions as possible, and only later introducing assumptions that are necessary for actually making use of the

height-integrated equations.

### 3.3.2.1 The Dimensionless Vertical Parameter $\{u\}_z$

It can be seen that many of the terms in equations (3.56) through (3.60) are multiplied by a factor of  $\rho$  that will need to be integrated over height (this is especially true when we realize that there are “hidden” factors of  $\rho$  in the gas pressure, internal gas energy and kinetic energy components of  $P$  and  $E$ ). It is therefore convenient to define the following function for any physical quantity  $u$ :

$$\{u\}_z = \frac{\int \rho (u/u_0) dz}{\int \rho dz}, \quad (3.61)$$

where  $u_0$  is the value of  $u$  on the equatorial plane. The value of  $\{u\}_z$  is simply a dimensionless number that characterizes the vertical distribution of  $u$  within the disk at a particular instant in time and a particular location within the equatorial plane; we will discuss its properties more fully below.

Note that we have deliberately not defined the bounds of either integral in equation (3.61); the only requirement is that both are taken over the same bounds and that these bounds do not depend on  $u$  (in other words, that the bounds are the same for each term in the accretion disk equations that we are integrating over). The choice of bounds effectively corresponds to choosing the upper and lower faces of the single “pixel” in the vertical direction that will be used when integrating our equations numerically. Normally, of course, we will choose this “pixel” to be centered on the equatorial plane, so that the upper and lower bounds are equal.

If we follow the above rules, then the introduction of  $\{u\}_z$  means that we can transform our equations to their appropriate height-integrated versions simply by doing the following:

- Replace all  $\partial/\partial z$  terms with the difference of their boundary values,

- Replace all occurrences of  $\rho u$  with  $\{u\}_z \Sigma u_0$  (where  $\Sigma \equiv \int \rho dz$  is the surface mass density),
- Replace all occurrences of  $u$  alone (that are not multiplied by the density) with  $2 \{u/\rho\}_z u_0 H$ , where  $H \equiv \frac{1}{2} \int (\rho/\rho_0) dz = \Sigma/2\rho_0$  is a measure of the half thickness of the disk (equal to the density scale height if  $\rho$  has an exponential vertical distribution and if the bounds on the integral are large enough).

If we make these mathematical substitutions, we obtain the following equations (where we have assumed an axisymmetric disk to limit the number of terms that need to be written, but this procedure works equally well for a non-axisymmetric disk that is being simulated in  $r$  and  $\phi$ ):

$$\frac{\partial}{\partial t} (r\Sigma) = -\frac{\partial}{\partial r} \left[ \{v_r\}_z r\Sigma v_r \right] - r \left[ \rho v_z \right]_{bottom}^{top} \quad (3.62)$$

$$\begin{aligned} \frac{\partial}{\partial t} \left[ \{v_r\}_z r\Sigma v_r \right] &= -\frac{\partial}{\partial r} \left[ r \left( 2 \{P/\rho\}_z PH + \{v_r^2\}_z \Sigma v_r^2 \right) \right] \\ &\quad + 2 \{P/\rho\}_z PH + \Sigma \left( \{v_\phi^2\}_z v_\phi^2 - r \{g_r\}_z g_r \right) \\ &\quad - r \left[ T_{rz} \right]_{bottom}^{top} \end{aligned} \quad (3.63)$$

$$\begin{aligned} \frac{\partial}{\partial t} \left[ \{v_\phi\}_z r^2 \Sigma v_\phi \right] &= -\frac{\partial}{\partial r} \left[ r^2 \left( 2 \{\alpha P/\rho\}_z \alpha PH + \{v_r v_\phi\}_z \Sigma v_r v_\phi \right) \right] \\ &\quad - r^2 \left[ T_{\phi z} \right]_{bottom}^{top} \end{aligned} \quad (3.64)$$

$$\begin{aligned} \frac{\partial}{\partial t} \left( 2 \{E_{tot}/\rho\}_z r E_{tot} H \right) &= -\frac{\partial}{\partial r} \left[ r \left( 2 \{v_r E_{tot}/\rho\}_z v_r E_{tot} H + 2 \{v_r P/\rho\}_z v_r PH \right. \right. \\ &\quad \left. \left. + 2 \{v_\phi \alpha P/\rho\}_z v_\phi \alpha PH + 2 \{F_{rad,r}/\rho\}_z F_{rad,r} H \right) \right] \\ &\quad - \{\mathbf{g} \cdot \mathbf{v}\}_z \Sigma g_r v_r \\ &\quad - r \left[ E_{tot} v_z + (\mathbf{G} \cdot \mathbf{v})_z + F_{rad,z} \right]_{bottom}^{top}. \end{aligned} \quad (3.65)$$

In these equations and in the discussion that follows, we have simplified the notation by removing the “0” subscript that indicates a quantity is evaluated on the equatorial plane; in other words, all factors in the form of  $\{u\}_z X$  in the above equations should actually be read as  $\{u\}_z X_0$ . This change in notation should be easy to follow if we remember that we are working with height-integrated equations, so all physical variables in the equations that are not themselves height-integrated must refer to their values at a particular location in the accretion disk (in our case,  $z = 0$ ). We have also used the notation “top” and “bottom” to refer to the upper and lower bounds of the integral over  $z$ , and we have written the vertical stresses on these boundaries as  $T_{rz}$  and  $T_{\phi z}$  and the work done by the stress on these boundaries as  $(\mathbf{G} \cdot \mathbf{v})_z$ ; this is a more general notation that takes into account the possibility of crudely modeling the effect of a jet, in which, for example, the magnetic stresses may play an important role even though we have not modeled their dynamics in the main body of the disk. Under the formal assumptions we have made, however, these terms are  $T_{rz} = \rho v_r v_z$ ,  $T_{\phi z} = \rho v_\phi v_z$  and  $(\mathbf{G} \cdot \mathbf{v})_z = P v_z$ , respectively. In this case, if the disk is in vertical hydrostatic equilibrium (i.e.,  $v_z = 0$  and no jet), then the only vertical flux of energy is due to the radiation term  $F_{rad,z}$ . Alternatively, if  $v_z \neq 0$  but the integral is taken over an infinite range in  $z$  (or, in reality, to a large enough extent in  $z$  such that the value of  $\rho$  at the boundary is negligible), then once again the only vertical flux of energy is due to the radiation term  $F_{rad,z}$ ; thus, vertical motions in the disk that are confined to a region  $|z| \lesssim H \lesssim r$  in general do not influence the disk evolution on the timescales we are interested in.

As previously noted, there are “hidden factors” of  $\rho$  within some of the terms in the above equations (e.g.,  $P$  and  $E_{tot}$ ), which from a practical standpoint means

that many of the factors of  $2H \equiv \Sigma/\rho_0$  can be removed. Specifically, the following relations hold for any physical quantity  $u$ :

$$2 \{uP/\rho\}_z uPH = u \left[ \{uT\}_z \left( \frac{\Sigma k_b T}{\mu m_p} \right) + 2 \{uT^4/\rho\}_z P_{rad} H \right] \quad (3.66)$$

$$2 \{uE_{tot}/\rho\}_z uE_{tot}H = u \left[ \{uv^2\}_z \frac{1}{2} \Sigma v^2 + \{uT\}_z \left( \frac{\Sigma k_b T}{(\gamma_{gas} - 1) \mu m_p} \right) + 2 \{uT^4/\rho\}_z E_{rad} H \right]. \quad (3.67)$$

Thus, the *only* terms in equations (3.62) through (3.65) that have any dependence on an actual vertical length scale  $H$  are those involving the radiation pressure  $P_{rad}$  and radiation energy density  $E_{rad}$ ; in all other cases, the vertical structure of the disk is entirely described by the dimensionless numbers  $\{u\}_z$ , which do not depend on the overall size scale of the system. We will discuss the value of  $H$  further in §3.3.2.3, but for now we focus on the  $\{u\}_z$ , which are the most common “artifact” of the height-integration procedure.

Many previous authors have introduced dimensionless numbers similar to  $\{u\}_z$  to characterize the vertical distribution of physical quantities in an accretion disk and to relate vertically averaged quantities to their values on the equatorial plane (e.g., Hōshi 1977; Paczyński & Bisnovaty-Kogan 1981; Abramowicz et al. 1988; Szuszkiewicz & Miller 1997; Kato et al. 1998; Artemova et al. 2001; Bisnovaty-Kogan & Lovelace 2001; Janiuk et al. 2002). However, these dimensionless numbers are not always defined consistently, nor are their possible variations with space and time often considered. We believe that the new formalism we have presented in equation (3.61), in which all of the dimensionless numbers have a consistent, model-independent definition (which as far as we know has not been previously considered in the literature), allows for a more transparent study of the plausible

values that these numbers can take and the ways in which they can change in a real accretion disk. We present this analysis in the next section.

### 3.3.2.2 Typical Values of the Vertical Parameter $\{u\}_z$

Since most physical quantities  $u$  are expected to decrease with distance  $z$  away from the equatorial plane, we can see from equation (3.61) that  $\{u\}_z \lesssim 1$ . In addition, the faster that  $\rho$  drops off with height in the disk compared to  $u$ , the closer  $\{u\}_z$  will be to 1; this is simply because  $\rho$  acts as a “weight” that limits the contribution of the outermost values of  $u/u_0$  to the numerator in equation (3.61). However, it is *not* valid to say that a thin disk (i.e.,  $H \ll r$ ) will automatically have  $\{u\}_z = 1$ ; even in a thin disk, both  $\rho$  and  $u$  can (and likely will) decrease with height at a roughly similar rate, such that  $\{u\}_z < 1$ .

Note the similarities between the quantity  $\{u\}_z$  we have introduced here and the quantity  $\alpha$  introduced by Shakura & Sunyaev (1973); in both cases, we have defined a dimensionless number that is expected to be  $\lesssim 1$  and whose value characterizes some physical aspect of the accretion disk. In both cases, the quantity can vary with space and time (and we therefore keep it within the space and time derivatives in the equations above), although it is erroneously often assumed *not* to vary with space and time in the literature. In fact, changes to the disk’s vertical structure (e.g., due to a jet) can be crudely modeled by allowing  $\{u\}_z$  to change. For example, as pointed out above,  $\{u\}_z \rightarrow 1$  when the density  $\rho$  drops off with height much faster than a typical physical quantity  $u$ ; therefore, if we imagine trying to model the ejection of a jet that is “clean,” in the sense that it contains a small amount of mass but a large amount of energy at large distances from the accretion disk (for example, a highly relativistic jet, or one that is Poynting-flux dominated), it

might be appropriate to take  $\{u\}_z \approx 1$ , whereas a jet with a high mass content will generally have smaller values of  $\{u\}_z$ .

It is possible to study the values of  $\{u\}_z$  more quantitatively in certain cases, in order to get a better estimate of how they behave. As a simple example, consider the case in which both  $u$  and  $\rho$  follow Gaussian or exponential profiles in the vertical direction, with “scale heights”  $H_u$  and  $H_\rho$ , respectively; i.e.,  $u \propto \exp(-z^2/H_u^2)$  for the Gaussian case, and  $u \propto \exp(-z/H_u)$  for the exponential case, with analogous equations for  $\rho$ . If we define  $h_u \equiv H_u/H_\rho$  to be the ratio of the scale heights, it is easy to derive the following from equation (3.61), in the case where the integrals in the definition of  $\{u\}_z$  are taken over a range much larger than the scale heights:

$$\{u\}_{z,Gauss} = \sqrt{\frac{h_u^2}{1+h_u^2}}. \quad (3.68)$$

$$\{u\}_{z,exp} = \frac{h_u}{1+h_u}. \quad (3.69)$$

Figure (3.1) shows a plot of these functions. Extreme values of  $\{u\}_z \approx 0$  or  $\{u\}_z \approx 1$  require the scale heights of  $u$  and  $\rho$  to differ by a large amount, approximately an order of magnitude. It is especially difficult to get  $\{u\}_z \approx 0$ ; for example, even if the density scale height exceeds the scale height of  $u$  by a factor of 10, we still only have  $\{u\}_z \approx 0.1$ , but  $u$  will now have to decrease from its equatorial value by a factor  $\sim 8,000$  times more than  $\rho$  does by the time they reach the density scale height (this is for the exponential case; the factor becomes  $\sim 10^{43}$  in the Gaussian case!). These arguments suggest that in the absence of any further information about the vertical structure, values of  $\{u\}_z$  corresponding to  $h_u \approx 1$  should be used; i.e.,  $\{u\}_{z,Gauss} \approx 1/\sqrt{2}$  and  $\{u\}_{z,exp} \approx 1/2$ .

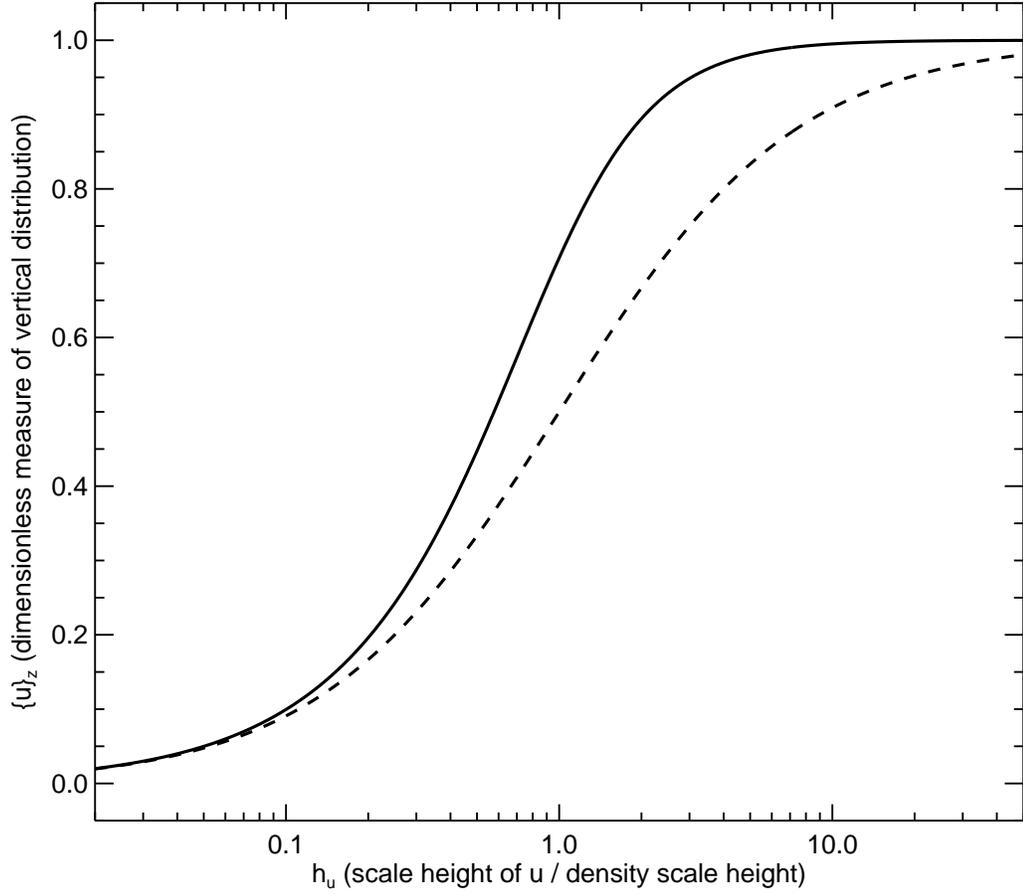


Figure 3.1: The dependence of the vertical parameter  $\{u\}_z$  that we have defined in equation (3.61) on the scale height of the physical quantity  $u$  that it is defined for, in the case where both  $u$  and the mass density have a Gaussian (solid line) or exponential (dashed line) dependence on height from the disk's equatorial plane. In order to have  $\{u\}_z \approx 0$  or  $\{u\}_z \approx 1$ , the scale heights of  $u$  and the mass density must differ by a large amount, approximately an order of magnitude.

### ***Physical Models***

It is possible to appeal to physics to get a more specific estimate of the disk's vertical structure than the rough estimates provided above. One potential avenue, of course, is to look at the results of detailed three-dimensional simulations of a small patch of the disk (e.g., Hirose et al. 2006), compare the scale heights of different physical quantities in these simulations, and use the above equations to estimate the appropriate values of  $\{u\}_z$ .

Another, more general approach is to consider simple physical models. In fact, the vertical structure of an accretion disk has much in common with the atmosphere of a star, since it is ultimately regulated by a competition between gravity pulling in and the pressure from an internal energy source pushing out. However, there are two main differences (Hubeny 1990): first, the gravitational force is due to a central point mass only (and increases roughly linearly with  $z$  at a given location above the accretion disk plane; see §3.2.4.2), and second, the internal energy source is not nuclear reactions concentrated at the center, but rather the heat produced by the dissipation of turbulence and magnetic fields, which can be more uniformly distributed throughout the height of the disk.

The key difference is in many ways the second one; as we will see, our lack of knowledge about turbulence *directly* leads to a lack of knowledge about the vertical structure of the disk. However, the formalism for the vertical structure that we have adopted in this thesis gives us a convenient way to handle this situation; provided we can determine how different models for the turbulence affect the dimensionless numbers  $\{u\}_z$ , we can turn the problem around and *choose* particular values of  $\{u\}_z$  to use in our simulation that will allow us to investigate different kinds of turbulence. Just like the  $\alpha$  parameter of Shakura & Sunyaev (1973), the

dimensionless numbers  $\{u\}_z$  allow us to “hide” our ignorance about the turbulence in a convenient manner.

### ***Polytropes***

In order to accomplish our goal of roughly tracking how different types of turbulence affect  $\{u\}_z$ , we will stick with the simplest models possible. As with stellar atmospheres, a popular simple model for accretion disks involves assuming that at each location in the disk’s equatorial plane, the vertical structure above that location follows a polytropic relation, with  $P \propto \rho^{1+1/N}$  (Hōshi 1977). Hōshi (1977) and Kato et al. (1998) have shown that for a polytrope in vertical hydrostatic equilibrium (which should be true on our timescales of interest if the disk is thin<sup>9</sup>), several physical quantities in the disk will be distributed vertically according to the formula

$$u(z) = u_0 \left( 1 - \frac{z^2}{H_N^2} \right)^s, \quad (3.70)$$

where  $s$  is an exponent that depends on which physical quantity  $u$  is being considered, and  $H_N$  is the height above the equatorial plane at which  $u$  would go to zero if the polytropic assumption is strictly followed (see also Shakura et al. 1978). Following Kato et al. (1998), typical values of  $s$  for some important physical quantities include  $s = N$  for the density and  $s = N + 1$  for the pressure; Kato et al. (1998) also claim that  $s = 1$  for the temperature  $T$ , but this is not true because  $T \propto \beta P / \rho$  by definition (where  $\beta \equiv P_{gas} / P$  as defined above), so that  $T \propto \beta (1 - z^2 / H_N^2)$  for a polytrope. Therefore,  $s = 1$  for the temperature requires  $\beta$  to be constant with height, which, as we will show below, is not true unless the polytrope has  $N = 3$ .

Using the above information, inserting equation (3.70) into equation (3.61)

---

<sup>9</sup>We are implicitly ignoring the possibility of a jet or convective energy transport in the vertical direction, but we will consider these elements below.

and using formulas given in Hōshi (1977) and Kato et al. (1998) to evaluate the appropriate integrals up to a distance  $H_N$ , we obtain:

$$\{u\}_{z,poly} = 2^{2s} \left[ \frac{\Gamma(N+s+1)}{\Gamma(N+1)} \right]^2 \frac{\Gamma[2(N+1)]}{\Gamma[2(N+s+1)]}, \quad (3.71)$$

where  $\Gamma(x)$  is the gamma function. This allows us to calculate  $\{u\}_z$  for different physical parameters provided we have chosen a value of  $N$ .

The polytropic assumption is usually made in the literature for the sake of convenience, and its physical interpretation is not discussed. However, it is possible to gain some rough insight into the physical meaning of a polytrope for the case of a steady disk, and this will allow us to determine reasonable values of  $N$  that we can use to calculate  $\{u\}_z$  with the help of equation (3.71). In fact, it will turn out that different values of  $N$  correspond to different models for the turbulent dissipation.

To begin, we first note that the polytropic condition  $P \propto \rho^{1+1/N}$  can easily be rewritten as  $(1-\beta)/\beta^4 \propto \rho^{(3-N)/N}$  using the equations of state (3.14) and (3.15) and can be further rewritten as

$$\frac{(1-\beta)}{\beta^4} = \frac{(1-\beta_0)}{\beta_0^4} \left( 1 - \frac{z^2}{H_N^2} \right)^{3-N} \quad (3.72)$$

based on our discussion above, which therefore shows that an  $N = 3$  polytrope corresponds to a disk in which  $\beta$  is constant with height; this is exactly analogous to the Eddington standard model for stellar atmospheres (e.g., Hansen & Kawaler 1999). For an accretion disk, however, we can gain more insight into this and other polytropes by writing the equations of energy balance and hydrostatic equilibrium for a steady, thin disk:

$$Q_{heat} \approx -\frac{\partial}{\partial z} \left( \frac{c}{\rho\kappa} \frac{\partial P_{rad}}{\partial z} \right) \quad (3.73)$$

$$\frac{\partial P}{\partial z} \approx -\rho\Omega_c^2 z, \quad (3.74)$$

where  $Q_{heat}$  is the heating rate per unit volume due to turbulence and magnetic fields and  $\Omega_c \equiv \sqrt{g_r/r}$  is the angular velocity of a circular orbit on the equatorial plane. (The assumption of a steady disk dictates that we set  $Q_{heat} = \nabla \cdot \mathbf{F}_{rad}$  everywhere so as to balance heating and cooling, while the assumption of a thin disk allows us to approximate this equation using derivatives with respect to  $z$ , leading to the expression in equation 3.73).

The fact that  $\beta$  is constant with height for an  $N = 3$  polytrope means that  $P_{rad} \propto P$ , and therefore equation (3.74) can be inserted into equation (3.73) to yield  $Q_{heat} \propto \partial(z\kappa^{-1})/\partial z$ . This shows that in regions of the disk that are scattering-dominated (i.e., where  $\kappa$  is constant), choosing an  $N = 3$  polytrope is the same as specifying that the heating rate  $Q_{heat}$  be constant with height. However, since we can write equation (3.53) for the absorption opacity as

$$\kappa_{abs} \propto \left( \frac{\beta}{1-\beta} \right)^{7/6} \rho^{-1/6}, \quad (3.75)$$

or  $\kappa_{abs} \propto (1 - z^2/H_N^2)^{-1/2}$  for an  $N = 3$  polytrope with constant  $\beta$ , we can see that this polytrope must become absorption-dominated at a large enough distance from the equatorial plane, thereby modifying the vertical profile of  $Q_{heat}$ . We will return to the absorption case later on, but for now we point out that an  $N = 3$  polytrope in a scattering-dominated disk corresponds to a *physical assumption* that the heating rate per unit volume due to the turbulence is uniform throughout the disk height. This requires that the heating rate per unit mass  $Q_{heat}/\rho$  increase sharply with distance from the equatorial plane, which seems unlikely (although possible if chaotic magnetic fields play an important role and if the scale height for magnetic stress is larger than the density scale height). The three-dimensional simulations of Hirose et al. (2006) did show one side of the disk where the dissipation rate corresponding to our definition  $Q_{heat}$  was roughly constant with disk height, but

on the other side of the disk this quantity decreased with height. Therefore, the  $N = 3$  polytropes which are very popular in the literature (e.g., Kato et al. 1998) show some promise, but might not be the most accurate model in certain cases.

The solution of equations (3.73) and (3.74) and the determination of  $Q_{heat}(z)$  for an arbitrary value of  $N$  is more complex, but it can be done numerically. To obtain the appropriate equations, we explicitly write out  $\partial P_{rad}/\partial z$  in terms of  $\beta$  and  $P$ , insert our result into equation (3.73) and then, with the help of equation (3.74) as well as other results stated above, obtain:

$$Q_{heat}(z) = \frac{\partial}{\partial z} \left[ \frac{c}{\kappa} \left( (1 - \beta) \Omega_c^2 z + \frac{P_0}{\rho_0} (1 - z^2/H_N^2) \frac{\partial \beta}{\partial z} \right) \right]. \quad (3.76)$$

If we substitute for the opacity using equation (3.75) and for  $H_N$  using the expression  $\Omega_c^2 H_N^2 = 2(N + 1) P_0/\rho_0$  found by Hōshi (1977) and Kato et al. (1998) for a polytrope in hydrostatic equilibrium, this equation can be written in a convenient dimensionless form that depends only on  $\beta$ ,  $z$  and  $N$  (for a given set of physical conditions on the equatorial plane). We can then proceed by solving equation (3.72) numerically to obtain  $\beta(z)$ , which can be used to calculate  $\partial\beta/\partial z$  analytically from the differentiation of equation (3.72), and, finally, the results can be inserted into equation (3.76). The end result of our work is that for a particular  $\beta_0$  and  $(\kappa_{abs}/\kappa_{sc})_0$  (i.e., the gas pressure fraction and ratio of absorption to scattering opacity at a given point on the equatorial plane), we are able to calculate the vertical distribution of the heating rate  $Q_{heat}(z)$  above that point.

We plot  $Q_{heat}(z)/\rho(z)$  (the heating rate per unit mass) in Figure 3.2 and  $Q_{heat}(z)$  (the heating rate per unit volume) in Figure 3.3 for three different polytropes ( $N = 1, 2$  and  $3$ ), using conditions in the equatorial plane that are consistent with a region near the radiation-dominated, inner portion of a Shakura & Sunyaev (1973) disk. We can see from these figures that in this regime, a polytrope with

$N \approx 1$  is appropriate for modeling a disk in which the turbulent heating per unit mass is constant with height, while, as noted previously,  $N \approx 3$  is appropriate for a situation in which the turbulent heating per unit *volume* is constant with height (the turbulent heating per unit volume for the  $N = 1$  and  $N = 2$  profiles more closely tracks the mass density and therefore drops off with distance from the equatorial plane; Figure 3.3). Thus, we find a general trend whereby more centrally-concentrated dissipation profiles correspond to smaller values of  $N$ .

We investigate these ideas more generally in Figures 3.4 and 3.5, for a range of values of  $N$  and  $\beta_0$ . Moving from left to right in these figures (i.e., increasing  $\beta_0$ ) is equivalent to moving from the radiation-dominated inner region to the gas-dominated middle region of a Shakura & Sunyaev (1973) disk. In Figure 3.4, we plot a measure of the “smoothness” of the vertical profile of the heating rate per unit mass and show that in order to get as close as possible to a heating rate that is constant with height (i.e., to a “smooth” vertical distribution), the polytropic index  $N$  used at a particular location in the accretion disk must depend strongly on the gas pressure fraction  $\beta_0$  in the equatorial plane; the value of  $N \approx 1$  that we found above for Figure 3.2 depended strongly on our choice of  $\beta_0 = 0.2$ . By contrast, Figure 3.5 shows that if we are instead interested in obtaining a centrally-concentrated heating rate per unit volume, the lowest values of  $N$  generally do correspond to the most “compact” (or centrally-concentrated) profiles. We have also investigated the effect of increasing  $(\kappa_{abs}/\kappa_{sc})_0$  to allow for the presence of significant absorption opacity (as would be expected in the outer regions of a Shakura & Sunyaev 1973 disk), but we find that this does not have a significant effect on the results in Figures 3.4 and 3.5, in particular for the high values of  $\beta_0$  at which the absorption opacity is expected to become significant.

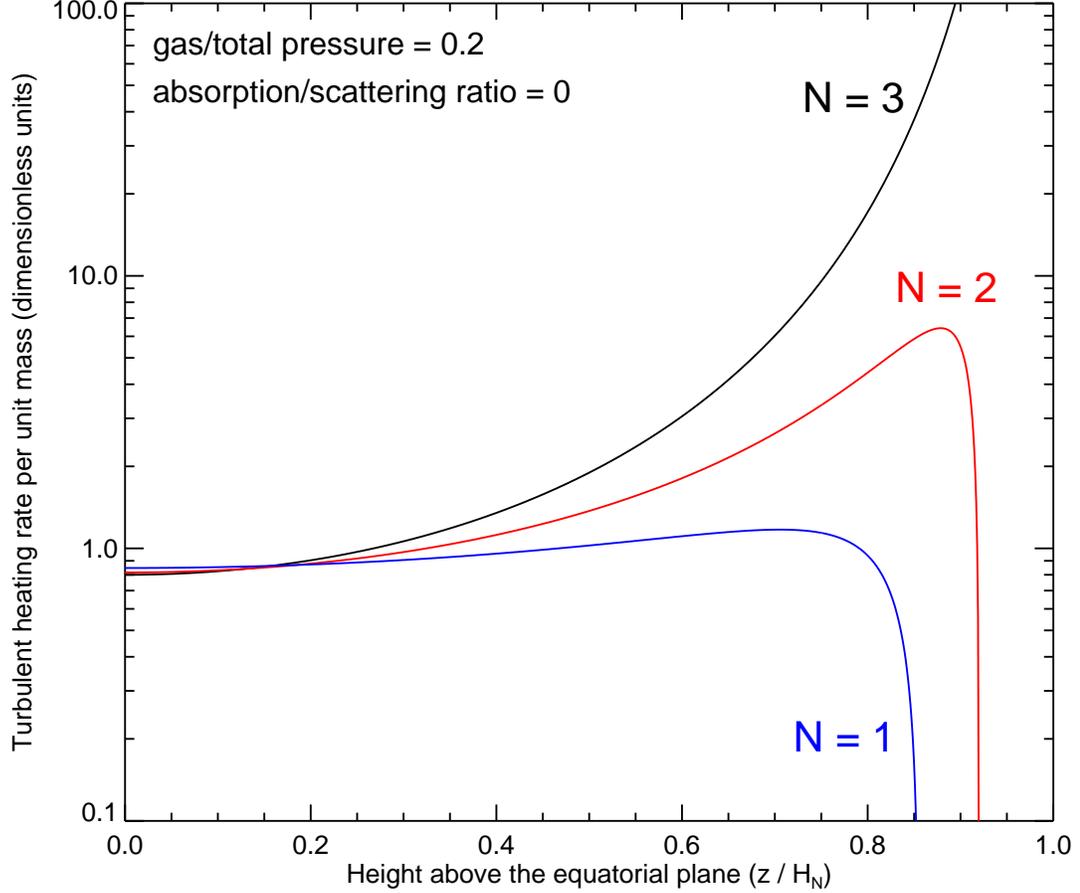


Figure 3.2: The turbulent heating rate per unit mass as a function of height  $z$  in the accretion disk, for three different polytropes ( $N = 1, 2$  and  $3$ ). The conditions on the equatorial plane are  $\beta_0 = 0.2$  and  $(\kappa_{abs}/\kappa_{sc})_0 = 0$ , typical of a region near the radiation-dominated, inner portion of a Shakura & Sunyaev (1973) disk. The heating rate is measured in units of  $c\Omega_c^2/(\kappa_{sc}\rho_0)$ . We can see from this graph that the  $N = 1$  polytrope has a relatively constant heating rate per unit mass throughout most of the disk, while the others do not. (Note that because  $H_N$  depends on  $N$ , a given value of  $z/H_N$  in this graph does *not* correspond to a particular fixed distance above the equatorial plane; the graph should only be used to compare the relative *shapes* of the heating profiles for different polytropes, not their absolute values.)

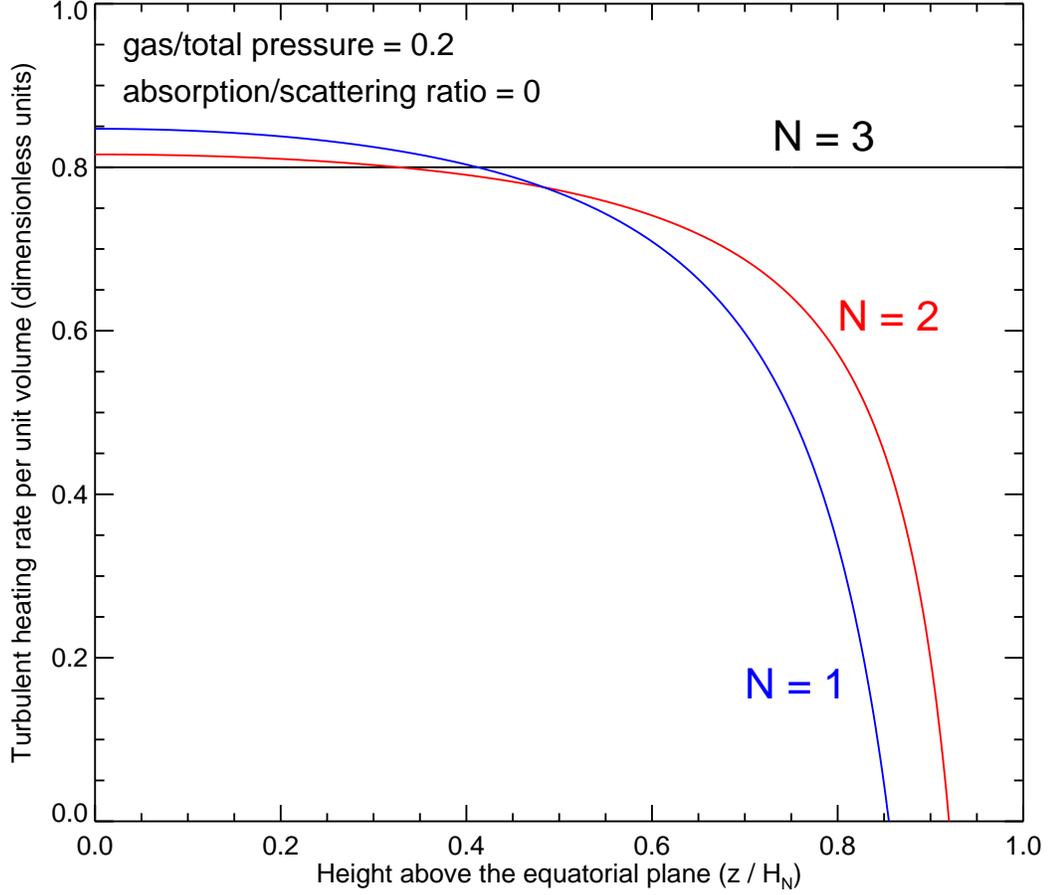


Figure 3.3: Same as Figure 3.2, but for the turbulent heating rate per unit *volume*, measured in units of  $c\Omega_c^2/\kappa_{sc} \approx 3.7 \times 10^{21} f_{grav} (M/M_\odot)^{-2} (r/r_g)^{-3} \text{ erg cm}^{-3} \text{ s}^{-1}$ , where the latter expression assumes a hydrogen mass fraction  $X = 0.7$  and defines  $f_{grav} = g_r / (GM/r^2)$  as the deviation of the gravitational force on the equatorial plane from the Newtonian value. We can see from this graph that the  $N = 3$  polytrope has a constant heating rate per unit volume throughout the disk and that the heating rate becomes progressively more centrally-concentrated as  $N$  is decreased.

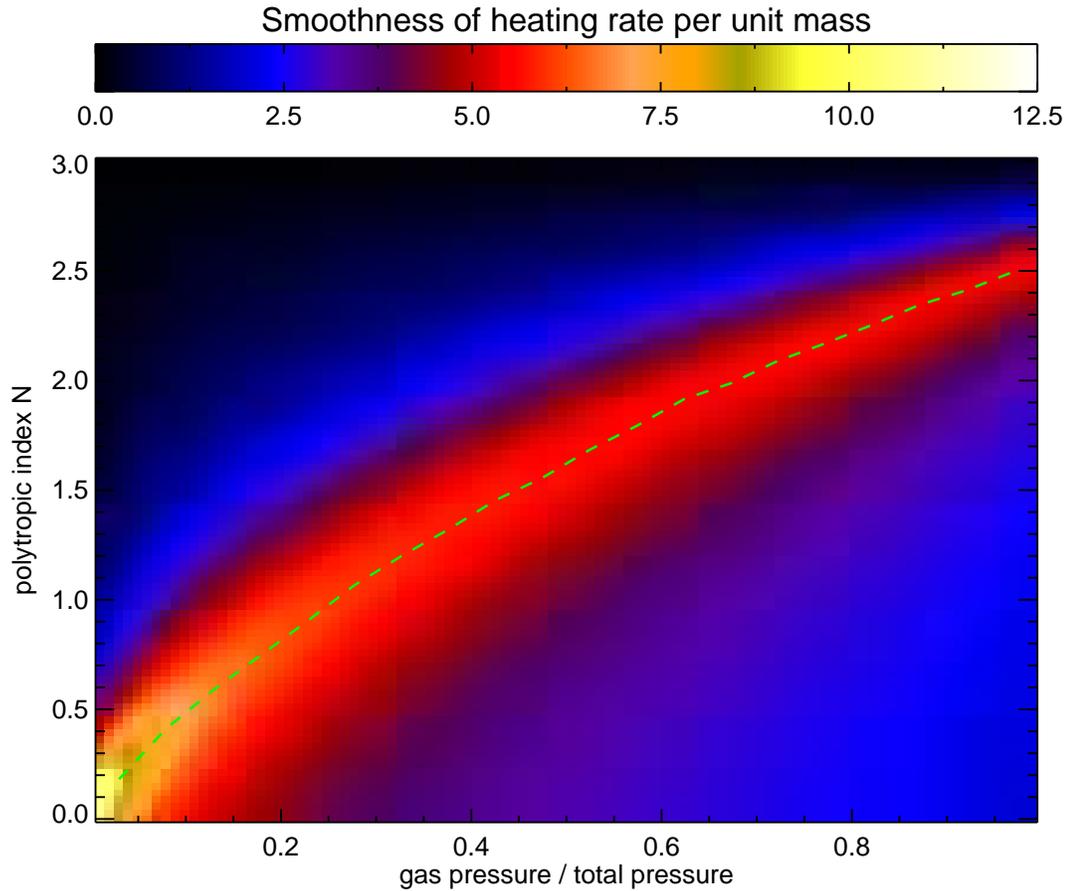


Figure 3.4: The “smoothness” of the vertical profile of the turbulent heating rate per unit mass (cf. Figure 3.2), as a function of the equatorial plane gas pressure fraction  $\beta_0$  and the polytropic index  $N$ . The dashed line roughly traces out the value of  $N$  that gives the smoothest profile for each  $\beta_0$ . The smoothness is estimated by taking the mean value of the turbulent heating profile divided by its standard deviation, over the range in  $z$  where the polytrope approximation is applicable. In all cases, the disk is assumed to be scattering-dominated.

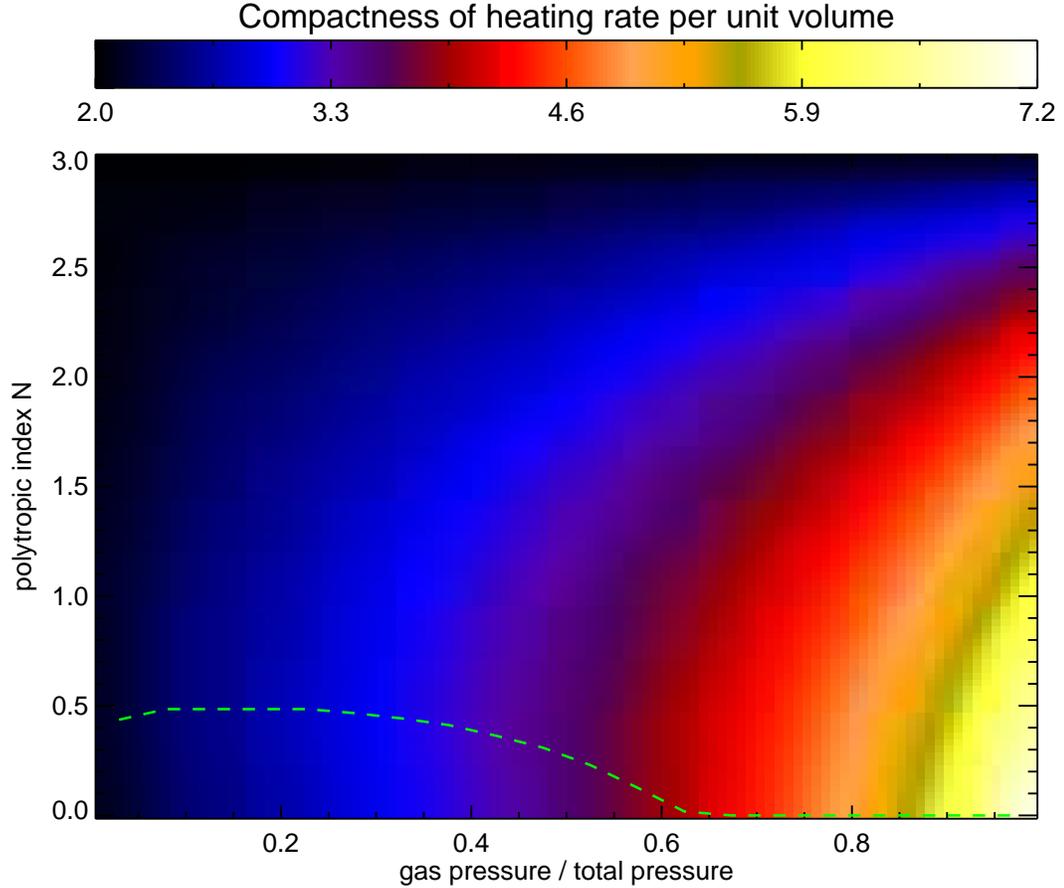


Figure 3.5: Same as Figure 3.4, but this time measuring the “compactness” of the vertical profile of the turbulent heating rate per unit *volume* (cf. Figure 3.3). The compactness is estimated by taking the inverse of the first spatial moment of the turbulent heating profile, over the range in  $z$  where the polytrope approximation is applicable. The dashed line roughly traces out the value of  $N$  that gives the most compact (i.e., most centrally-concentrated) profile for each  $\beta_0$ ; in contrast with Figure 3.4, we find that the most centrally-concentrated profiles always occur for very low values of  $N$  (and, in fact, would always occur for  $N = 0$  if we corrected this figure for the fact that  $H_N$  depends on  $N$ , which we have chosen not to do). In all cases, the disk is assumed to be scattering-dominated.

Our general conclusions are consistent with the model by Shakura et al. (1978), who “worked backwards” compared to the method we have used here; they calculated (analytically) a polytropic index of  $N = 1.17$  based on a specific model for the turbulent heating profile in a radiation-pressure dominated disk. The model they chose had a constant turbulent heating rate per unit mass, and thus it is not surprising that they obtained a small value of  $N$ , based on what we have argued above and in Figure 3.4. Our method is much more general, however, and provides a way to see the general connection between the polytropic index  $N$  and the turbulent heating profile. Shakura et al. (1978) also allowed for the possibility that the disk was unstable to convection in the vertical direction, so that convection as well as radiation played a role in carrying the turbulent heat out of the disk and in maintaining the vertical pressure balance (in fact, many of the profiles we have considered here are likely to be at least slightly unstable to convection, especially in the radiation-dominated regime; Bisnovatyi-Kogan & Blinnikov 1977; Shakura et al. 1978). They found that their polytropic index dropped when convection was included, reaching  $N = 0.85$  in the case where the turbulent convection velocity was set equal to the sound speed. We suggest that the presence of a steady-state jet would also tend to lower the polytropic index, since it enters into the equations in a similar way as convection (with the jet velocity replacing the turbulent convection velocity; i.e., a jet emanating from within the disk can essentially be treated as an “extreme” version of a convective bubble, as discussed further in §3.3.2.3). In this case, the jet can easily be faster than the sound speed, so we may have the freedom to choose even lower values of  $N$  than found by Shakura et al. (1978), though a full quantitative analysis of this problem is beyond the scope of our discussion here.

Having discussed possible values of the polytropic index  $N$ , we can now return to equation (3.71) and calculate  $\{u\}_z$  for different polytropes. We plot  $\{u\}_z$  in Figure 3.6 for three different polytropes: the  $N = 1$  and  $N = 3$  cases we have considered above, as well as a more extreme case of  $N = 0.2$  that might correspond to a jet (or, at least, to a heating profile that is very centrally concentrated). As can be seen from this figure, it is once again difficult to get extremely low values of  $\{u\}_z$  unless  $s$  (the exponent in the vertical distribution of  $u$ ; see equation 3.70) is implausibly large. For values of  $s$  that scale with  $N$ , we find that there is actually very little variation in  $\{u\}_z$  as  $N$  changes, and the fiducial values of  $\{u\}_z$  that we estimated initially (for Gaussian and exponential distributions with similar scale heights for  $u$  and  $\rho$ ) are roughly good estimates regardless of the turbulent heating profile.

On the other hand, if  $s$  is held constant as  $N$  changes, we find that  $\{u\}_z$  decreases when the polytropic index  $N$  decreases (i.e., as the turbulent dissipation per unit volume becomes more and more centrally concentrated). We suggest that terms such as these (which include  $\{P/\rho\}_z$ , which appears prominently in equations 3.62 through 3.65) are also affected more strongly than the others by the presence of a jet, and the magnitude of this effect can be estimated from Figure 3.6; it appears to be on the order of 20–25% in the cases we have considered here.

Even more interesting is the behavior of a term such as  $\{\alpha P/\rho\}_z$ , which also appears quite frequently in equations (3.62) through (3.65). Although a full analysis of this term is beyond the scope of our work here, we generally expect  $\alpha$  to become more centrally concentrated as  $N$  is lowered and the turbulent dissipation becomes more concentrated, because  $\alpha$  effectively represents the efficiency of the turbulent dissipation. (More specifically, we can write  $\alpha \sim t_{r\phi}/P \sim Q_{heat}/(Pr\partial\Omega/\partial r)$ ,

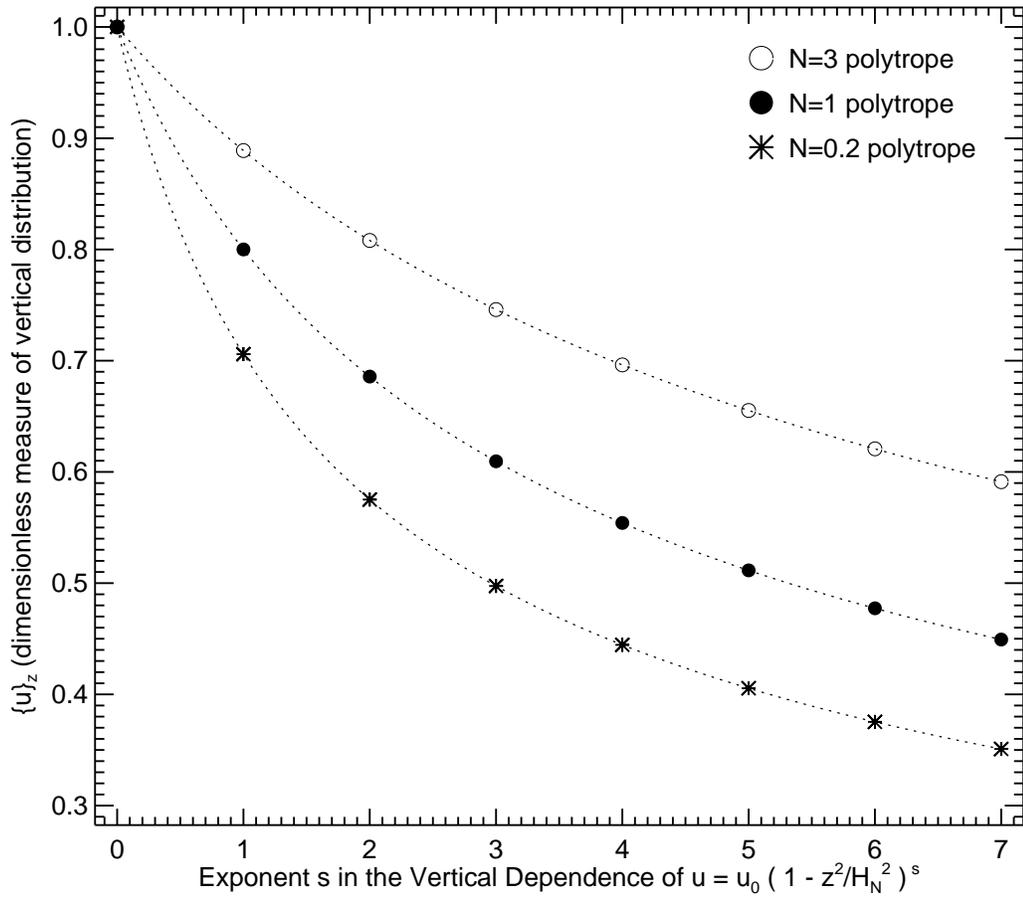


Figure 3.6: The dependence of the vertical parameter  $\{u\}_z$  that we have defined in equation (3.61) on the exponent  $s$ , when the physical quantity  $u$  is vertically distributed according to  $u = u_0 (1 - z^2/H_N^2)^s$ , as occurs in the case of a polytropic vertical equation of state. Plots are shown for three different polytropes ( $N = 0.2, 1$  and  $3$ ).

where the last expression uses results we derive later in equation 3.101, and then notice that  $Q_{heat}$  becomes more centrally concentrated as  $N$  is lowered while  $P$  becomes less so, relative to  $H_N$ .) This behavior of  $\alpha$  suggests that  $\{\alpha P/\rho\}_z$  should drop even more than the  $\sim 20 - 25\%$  calculated above for  $\{P/\rho\}_z$  if a jet or centrally-concentrated dissipation profile is present.

To summarize our ideas in a more physical way and explicitly relate them to a model for a jet, it is useful to temporarily abandon our separate consideration of the dimensionless number  $\{\alpha P/\rho\}_z$  and instead return to the *full* turbulent stress term that appears in equations (3.62) through (3.65), which is simply  $\int \alpha P dz$  (in other words, the vertical integration of the turbulent stress). Clearly, the effect we have discussed above—in which the presence of a jet leads the value of  $\alpha$  to decrease at large heights while remaining roughly similar on the equatorial plane—will act to decrease  $\int \alpha P dz$ , although we caution that this effect may depend on the assumption of a polytrope model for the vertical structure. However, there also appears to be a second effect, which occurs even if the jet has *no* influence on the vertical distribution of  $\alpha$  and therefore does not depend on our polytrope assumption. A jet, at its most fundamental level, is a structure that transfers material that would otherwise be near the equatorial plane to a larger distance away from the main body of the disk. The material that is “elevated” in this way is at a location where the orbital shear that drives the turbulence is smaller than its value on the equatorial plane; therefore, a disk with a jet should have less turbulent dissipation than a disk without a jet. As long as the turbulence behaves in some way like a molecular viscosity, so that the  $\alpha$  we have defined here depends on the orbital shear (Chen & Taam 1993) and therefore decreases with height, this effect should occur—mathematically, we are simply saying that since

$\alpha$  decreases with height,  $\int \alpha P dz$  must get smaller when  $P$  is redistributed so that it is less centrally-concentrated. (Note the crucial difference between *redistributing*  $P$  to change the vertical structure, and changing the vertical structure by adding high-pressure material at large distances from the equatorial plane while leaving the material on the equatorial plane untouched. The second method will actually increase  $\int \alpha P dz$ , but we believe the first one is a more realistic description of what a jet does. For example, if we assume that a jet forms via the accumulation of a large-scale magnetic field in a highly-conducting disk, the jet will not be able to heat the gas, and therefore all it can do is vertically redistribute the internal energy  $E$ . Since  $P \sim E$ , the vertical integration of  $P$  should also be roughly conserved by the jet.)

Finally, we note that more accurate models for the disk's vertical structure (beyond the polytropic assumption) have been discussed and investigated numerically by many authors (e.g., Meyer & Meyer-Hofmeister 1982, 1983; Cannizzo & Wheeler 1984; Cannizzo & Cameron 1988; Cannizzo 1992). These models, however, are generally too detailed for our purposes; we believe it is sufficient to use simple models such as polytropes or compare to the results of time-dependent numerical simulations (e.g., Hirose et al. 2006) in order to get the level of accuracy that we desire in our determination of  $\{u\}_z$ .

### 3.3.2.3 The Vertical Scale Height

The vertical scale height  $H$  that appears in equations (3.63) through (3.65) can be calculated from equation (3.59), which expresses vertical momentum balance in the disk. An order of magnitude estimate for a steady disk with  $v_z = 0$  (e.g., equation 3.74) yields  $H \sim c_s/\Omega_c$ , where  $c_s = \sqrt{P_0/\rho_0}$  is a rough measure of the

“sound speed” on the equatorial plane (although it is not equal to the actual speed of sound waves, and not even close to the actual value if the disk is radiation pressure dominated).

Most treatments (e.g., Kato et al. 1998) implicitly assume  $v_z = 0$ , but here we expand on this standard picture by also allowing for the contribution of a steady-state jet to the vertical equilibrium. To first order, we assume that the ram pressure of the jet  $\rho v_z^2$  is more important than the other “jet-related” terms in equation (3.59) that involve the transport of the jet’s momentum in the  $r\phi$  plane (in other words, we assume a roughly constant jet structure). This leads to the following equation:

$$\frac{\partial}{\partial z} (P + \rho v_z^2) = -\rho \Omega_c^2 z, \quad (3.77)$$

which approximately evaluates to  $H \sim \sqrt{c_s^2 + v_z^2}/\Omega_c$ ; in other words, the ram pressure of the jet augments the gas and radiation pressure and makes the disk thick. Note that this is similar to the equation obtained by Shakura et al. (1978), where the velocity of convective turbulent transport was considered instead of the velocity of a jet.

We must be slightly careful when interpreting  $v_z$  in equation (3.77). We can loosely refer to it as the “jet velocity,” but it is unlikely to be equal to the relativistic velocities that have been measured for jets at large distances from the accretion disk. The reason is that  $v_z$  is a fluid velocity and thus represents an *average* over the velocities of all the particles *within* the disk. Setting  $v_z$  equal to the downstream jet velocity would be equivalent to saying that the “entire” disk is being ejected, which may be a realistic model for transient jets, but it is unlikely to be sustainable in the steady state because it would require an enormous rate of mass supply. Although we do not know enough about the physics of jet generation to understand exactly

how the jet is formed within the disk, it may be helpful to once again make an analogy with convection, where fast-moving turbulent bubbles can coexist with slower moving fluid and lead to a net upward pressure even though the entire fluid is not participating in the motion. In the case of a steady jet, we can imagine the formation of similar bubbles near the equatorial plane, which eventually transfer their energy to fewer and fewer particles at higher heights above the disk until a small number of particles are ejected with very high speeds. The details of this procedure do not concern us, as long as we remember that the value of  $v_z$  used here is an appropriately averaged fluid velocity that is likely to be much smaller than what we think of as the downstream “jet velocity.” The only way in which this might be a concern is that by assuming that the jet velocity contributes to the pressure balance (as we do in this section), we are effectively assuming that the origin of the jet is deep within the disk. If the jet is formed near the disk surface, on the other hand, then it may carry energy out of the disk, but will not contribute significantly to the pressure balance or to increasing the vertical scale height.

Keeping the above caveats in mind, if we specify the vertical dependence of  $v_z$ , it is possible to integrate equation (3.77) for either a polytrope or Gaussian vertical distribution of density and pressure and thereby obtain a more exact expression for  $H$  than the rough estimate we found previously. For a polytrope, if we assume that the Mach number of the jet<sup>10</sup> is constant with height (similar to the assumption made by Shakura et al. 1978 when integrating this equation for the case of convective transport rather than a jet), we find

$$H = \left[ \frac{2^{2N} [\Gamma(N+1)]^2 \sqrt{2(N+1)}}{\Gamma[2(N+1)]} \right] \frac{\sqrt{c_s^2 + v_{z0}^2}}{\Omega_c}, \quad (3.78)$$

---

<sup>10</sup>Technically, we are actually referring to  $v_z \sqrt{\rho/P}$ , which differs considerably from the true Mach number in a radiation pressure dominated region.

where  $v_{z0}$  refers to the value of  $v_z$  on the equatorial plane. The number in brackets is a slowly varying function of  $N$ , decreasing from  $\sim 1.4$  (for  $N = 0$ ) to  $\sim 1.3$  (for  $N = 4$ ). Thus, the scale height of the disk is not significantly affected by the turbulent heating profile, but the general trend is for a disk with more centrally-concentrated turbulent dissipation to be slightly thicker (probably due to the larger radiation pressure gradient that would be expected in this case). Clearly, however, this equation shows that if a jet is present, the ram pressure will have the largest effect on increasing  $H$ , regardless of the turbulent dissipation.

The assumption of a constant Mach number is in many ways restricting. Because the sound speed varies with height in a polytrope as  $c_s(z) \propto (1 - z^2/H_N^2)^{1/2}$ , we are effectively assuming that the jet decelerates near the top of the disk; our analysis might therefore be more appropriate for handling a confined structure such as a corona than an actual jet.

A Gaussian vertical distribution of density and pressure allows for a much more realistic calculation of the effect of a jet. In this case, if we assume that  $v_z$  itself is constant with height (rather than the Mach number), we can easily derive

$$H = C_H \frac{\sqrt{(c_s/h_P)^2 + v_z^2}}{\Omega_c}, \quad (3.79)$$

where  $C_H$  is a dimensionless number that is equal to  $(\pi/2)^{1/2}$  in this model and  $h_P$  is the ratio of the pressure scale height to the density scale height (see equation 3.68 and the surrounding discussion). This equation does not make the same restrictive assumptions as the polytrope model, and we can clearly see the interplay between the jet speed and the pressure scale height in determining the density scale height  $H$  of the disk. Specifically, increasing the jet speed tends to increase the ram pressure, but if the jet does not have much mass ( $h_P \gg 1$ ), then the effect will be

weakened. We can therefore use equation (3.79) to measure the effect of varying degrees of energy and mass content in jets of different speeds on the scale height of the combined disk/jet system.

We can derive a more useful expression for the vertical scale height by combining equation (3.79) with the definitions  $c_s = \sqrt{P_0/\rho_0}$  and  $\rho_0 = \Sigma/2H$  and the equation of state (3.13) to obtain a quadratic equation for  $H$ , which can be solved to give

$$H = \frac{1}{2}H_{rad} + \sqrt{\left(\frac{1}{2}H_{rad}\right)^2 + H_{gas}^2 + H_{jet}^2}, \quad (3.80)$$

where  $H_{rad}$ ,  $H_{gas}$  and  $H_{jet}$  are the scale heights that would occur if the disk were supported by radiation, gas or jet ram pressure alone. They are given by:

$$H_{rad} = \frac{2C_H^2}{h_P^2 \Sigma \Omega_c^2} \left( \frac{1}{3} a_{rad} T_0^4 \right) \quad (3.81)$$

$$H_{gas} = \frac{C_H}{h_P \Omega_c} \sqrt{\frac{kT_0}{\mu m_p}} \quad (3.82)$$

$$H_{jet} = \frac{C_H v_{z0}}{\Omega_c}. \quad (3.83)$$

Equation (3.80) is useful because it relates  $H$  to two physical variables ( $T_0$  and  $\Sigma$ ) that are fundamental for studying the local energy balance of the disk (§3.4.2). It can also easily be applied to different models for the vertical structure other than the Gaussian distribution, simply by choosing particular values of  $h_P$  and  $C_H$  (e.g., for a polytrope, we choose  $h_P = 1$  and set  $C_H$  equal to the dimensionless number that appears in front of equation 3.78).

Finally, we present here a useful set of equalities relating the pressure and vertical scale height that can be derived from equation (3.79) and other information in this section:

$$PH = \frac{h_P^2}{2C_H^2} \Sigma \Omega_c^2 (H^2 - H_{jet}^2) = \frac{h_P^2}{2C_H^2} \Sigma \Omega_c^2 (H_{gas}^2 + H_{rad}H). \quad (3.84)$$

### 3.3.2.4 The Surface Flux and Luminosity

The final important issue that we must deal with regarding the vertical structure is the surface term involving  $[F_{rad,z}]_{bottom}^{top}$  that appears in the energy equation (3.65). For notational convenience, we write this term as  $2F_{surf}$  for a symmetric disk, where  $F_{surf}$  is the flux of radiative energy at the photosphere, equal to one half the luminosity per unit surface area measured by an outside observer.

Our starting point for evaluating  $F_{surf}$  is the  $z$  component of equation (3.50), which we write here:

$$F_{rad,z} = -\frac{c}{\rho\kappa} \frac{\partial P_{rad}}{\partial z}. \quad (3.85)$$

An order of magnitude estimate therefore gives  $F_{surf} \sim cP_{rad}/(\kappa\Sigma) \sim \sigma T^4/(\kappa\Sigma)$  as a way to relate the surface flux to variables  $T$  and  $\kappa$  on the equatorial plane. Technically, we could attempt to solve this equation for  $F_{surf}$  more exactly by inserting the appropriate boundary values of the variables that appear on the right hand side of this equation, based on our model for the vertical structure. However, most of our simple models (e.g., polytropes) become highly inaccurate near the photosphere, with many quantities tending towards zero. Therefore, a better way to proceed is to integrate equation (3.85) from the equatorial plane to the disk surface and use an approximation borrowed from stellar astrophysics (the two-stream approximation or one of its variants) to substitute for the “surface” value of  $P_{rad}$  (e.g., Spruit 2001). This method frees us from having to calculate derivatives at the photosphere; we instead replace them with integrals over the disk.

The literature contains many different approximations that have been made to carry out these integrals, but here we attempt to provide a more general method based on the vertical parameters  $\{u\}_z$  we have defined previously. Specifically, we

parametrize the vertical distribution of the radiative energy flux by a dimensionless function  $f_{rad}(z)$ , defined so that

$$F_{rad,z} = f_{rad}(z) F_{surf}; \quad (3.86)$$

we require  $f_{rad} \rightarrow 1$  at the photosphere but otherwise do not make any assumptions about its value. If the disk is steady and thin, however, then  $f_{rad}$  is related to the vertical profile of turbulent heating by  $Q_{heat}(z) = (\partial f_{rad}/\partial z) F_{surf}$ . If we further set the surface value of  $P_{rad}$  to  $\sim \tau_{surf} F_{surf}/c$  via the two-stream approximation, where a typical value of the ‘‘surface’’ optical depth is  $\tau_{surf} = 2/3$  (e.g., Ostlie & Carroll 1996), then the vertical integration of equation (3.85) gives

$$F_{surf} = \frac{2cP_{rad}}{\{\kappa f_{rad}\}_z \kappa \Sigma} \left( 1 + \frac{2\tau_{surf}}{\{\kappa f_{rad}\}_z \kappa \Sigma} \right)^{-1}, \quad (3.87)$$

where  $\kappa$  and  $P_{rad}$  now refer to their values on the equatorial plane, according to our usual convention for equations involving  $\{u\}_z$ . This equation can be evaluated for specific models of the vertical structure and for a mix of scattering and Kramers absorption using the following identity:

$$\{\kappa f_{rad}\}_z \kappa = \{f_{rad}\}_z \kappa_{sc} + \{\rho T^{-7/2} f_{rad}\}_z \kappa_{abs}. \quad (3.88)$$

As an illustration of the possible range of values for  $\{\kappa f_{rad}\}_z$  and therefore for  $F_{surf}$ , we consider two extreme cases.

In the first case, we assume that the turbulent heating is entirely concentrated on the equatorial plane, so that no heating takes place in the body of the disk and therefore  $f_{rad} = 1$ . For a scattering-dominated disk, the relevant integrals can be calculated trivially; we obtain  $\{\kappa f_{rad}\}_z = 1$ , or  $F_{surf} \approx 2cP_{rad}/(\kappa \Sigma)$  for an optically thick disk (where  $\kappa \Sigma \gg \tau_{surf}$ ). This agrees with the results given by Spruit (2001) using a less general method than ours. If we include absorption,

$F_{surf}$  is likely to decrease somewhat below this value, due to the low temperature and high absorption opacity expected in the upper layers of the  $N \approx 0$  disks that are appropriate for centrally-concentrated turbulent heating.

In the second case, we assume that the turbulent heating rate per unit volume is constant throughout the height of the disk ( $f_{rad} \propto z$ ). As discussed in §3.3.2.2, this corresponds to an  $N = 3$  polytrope if the disk is scattering-dominated (and still roughly corresponds to an  $N = 3$  polytrope even when absorption becomes important). When we evaluate the relevant integrals, we obtain  $\{f_{rad}\}_z = 35/128$  and  $\{\rho T^{-7/2} f_{rad}\}_z = 35/112$ . This leads to a value of  $F_{surf}$  in the optically thick limit that is  $\sim 3.5$  times larger than in the case considered above, where all the turbulent heating took place on the equatorial plane. Kato et al. (1998) also calculated  $F_{surf}$  for an  $N = 3$  polytrope, and our results agree with theirs in the scattering-dominated case, but differ when absorption becomes important. However, they used a different method for evaluating  $F_{surf}$  in which they had to assume that  $\kappa$  is constant with height, an assumption that is certainly violated for the absorption opacity. Our method does not make any assumptions about  $\kappa$  and therefore should be more accurate than the method of Kato et al. (1998).

By comparing the above two cases, we see that  $F_{surf}$  can increase by a factor of  $\sim 3.5$  or more as the turbulent heating changes from being entirely concentrated on the equatorial plane to being distributed uniformly throughout the disk. This makes sense physically; the more heating that takes place in the upper atmosphere, the more radiation the disk emits.

The above formulas can be used directly in the case where the gas and radiation are in equilibrium and  $P_{rad} = \frac{1}{3}a_{rad}T^4$ . However, it is possible to derive corrections to this equation that apply at low optical depths, when the gas and radiation will

not be perfectly in equilibrium. This approach was taken by Hubeny (1990) and Artemova et al. (1996a) and requires only a minimal amount of work beyond that which we have already done (see also Malbet et al. 2001). We begin by writing the energy equation for radiation alone (Mihalas & Mihalas 1984; Stone et al. 1992), under the assumptions of steady-steady diffusion, a gray atmosphere (i.e., where  $\kappa$  does not depend on frequency), a thin disk (where the divergence can be replaced by  $\partial/\partial z$ ) and assuming that the emission is thermal and the Eddington approximation holds. In our notation, this equation reads

$$\left(\frac{\partial f_{rad}}{\partial z}\right) F_{surf} = \rho \kappa_{abs} c (a_{rad} T^4 - 3P_{rad}), \quad (3.89)$$

where  $T$  is the gas temperature. It is a simple matter to solve this equation for  $P_{rad}$  and substitute it into equation (3.87). We obtain

$$F_{surf} = \frac{4\sigma T^4}{3 \{\kappa f_{rad}\}_z \tau_0} \left(1 + \frac{\tau_{surf}}{\{\kappa f_{rad}\}_z \tau_0} + \frac{1}{3 \{\kappa f_{rad}\}_z \tau_{eff}^2} \frac{\partial f_{rad}}{\partial \tilde{z}}\right)^{-1}, \quad (3.90)$$

where  $\tilde{z} \equiv z/H$  is the dimensionless height variable,  $\tau_0 \equiv \kappa_0 \Sigma/2$  is a rough measure of the optical depth from the equatorial plane to the surface (not exactly equal to the optical depth because it uses the opacity measured at the equatorial plane rather than integrated over the disk height), and  $\tau_{eff} \equiv \sqrt{\tau_0 \tau_{abs,0}}$  is a rough measure of the effective optical depth, defined using equatorial plane values in a similar way as  $\tau_0$ .

Equation (3.90) is a general result that appears to agree with that of Hubeny (1990), who, like us, considered the possibility of an arbitrary vertical profile of turbulent heating and therefore produced similar formulas to those we have derived here. However, it is much more general than the formula derived by Artemova et al. (1996a) that has been used to study the transition between optically thick and optically thin regimes in a disk. Our result reduces to that of Artemova et al.

(1996a) in the very specific case they consider, in which  $\tau_{surf} = 2/3$  and, more important, the density, opacity and turbulent heating rate per unit volume were all assumed to be constant with height in the disk, so that  $f_{rad} \propto z$  and  $\{\kappa f_{rad}\}_z = 1/2$  (they claimed to assume that the heating rate per unit *mass* was constant with height, but since their model also assumed  $\rho$  was constant, it is not clear that this assumption has much meaning; what they really assumed was  $f_{rad} \propto z$ , which arises more generally from a constant heating rate per unit volume).

Equations (3.87) and (3.89) can also be used to calculate an expression for the radiation energy density on the equatorial plane that corrects for the slight departure from equilibrium that occurs at moderate optical depths. We obtain

$$\frac{E_{rad}}{a_{rad}T^4} = \left[ 1 + \frac{\tau_{surf}}{\{\kappa f_{rad}\}_z \tau_0} \right] \left[ 1 + \frac{\tau_{surf}}{\{\kappa f_{rad}\}_z \tau_0} + \frac{1}{3 \{\kappa f_{rad}\}_z \tau_{eff}^2} \frac{\partial f_{rad}}{\partial \tilde{z}} \right]^{-1}, \quad (3.91)$$

which again reduces to the formula in Artemova et al. (1996a) under the specific set of assumptions they made. For our purposes, equations (3.90) and (3.91) are most useful not as a direct input to the numerical simulations (since a large number of assumptions about the radiation energy equation were used to derive them), but rather as a method to estimate the errors introduced when we *assume* an optically thick disk in the simulations but the disk happens to reach a situation where the optical depth becomes relatively small ( $\gtrsim 1$ ).

### ***Dynamic Diffusion***

There is one final issue that can arise in determining  $F_{surf}$  in a vertically-integrated disk. In calculating  $F_{surf}$ , we effectively assumed that the radiation that emerges from a small patch of the disk can be related to the conditions in the equatorial plane below that patch. This is only true if the radiation is able to escape the disk before conditions within the disk change significantly. Otherwise,

many subtle effects can occur, including advection of the radiation with the gas, so that the radiation emitted by a particular point on the equatorial plane emerges from the disk at a different location.

These effects begin to set in whenever the accretion rate or optical depth are so high that the fluid speeds are faster than the radiation diffusion timescale  $\sim c/\tau$ ; this is effectively the condition for “dynamic diffusion” discussed by Mihalas & Mihalas (1984), but applied to an accretion disk based on our particular definition of  $\tau$ . Since there exist conditions in which the radiation diffusion is dynamic ( $v/c \gtrsim 1/\tau$ ) but the gas and radiation are still in equilibrium ( $v/c \ll \tau_{abs}$ ; see equation 3.2), this condition should be independently checked; there are times when it might effectively constitute the “upper  $\dot{M}$  limit” (where  $\dot{M}$  is the accretion rate) of the vertically-integrated disk models we are considering here. Ohsuga et al. (2002, 2005) consider what happens to an accretion disk in which the radiation *is* allowed to be advected by the gas, sometimes so much so that it falls into the black hole before it is able to diffuse out of the disk and reach an observer (Narayan & Yi 1994); they find that this phenomenon leads to disk luminosities that are lower than what would otherwise be predicted, although still on the order of the Eddington value.

### 3.4 Steady State Disk Models

In this section, we provide a steady state solution to our modified accretion disk equations that can be used as a starting point for numerical simulations. We also derive an equation expressing local energy balance in the disk (which can determine whether a small patch of the disk is heating up, cooling down or, in the steady state, staying at the same temperature), and we provide estimates of the typical

timescales on which the steady state disk evolves. Both of the latter will be useful in comparing to the results of numerical simulations.

Our starting point for this work is the vertically integrated equations for an axisymmetric disk, which can be found in equations (3.62) through (3.65). In the steady state, we can set  $\partial/\partial t = 0$  in these equations, and we also make several additional simplifications that are valid only when a local analysis of the disk is appropriate (with the understanding that truly global phenomena can only be studied by numerical simulations). Thus, we assume nearly circular orbits ( $v_r \ll v_\phi$ ) so that we can ignore the radial momentum equation (3.63) and set  $v_\phi = v_c$  everywhere on the equatorial plane, and we also ignore  $F_{rad,r}$  (the radiative energy flux in the radial direction). Note that the energy balance equation derived by Abramowicz et al. (1988) was integrated numerically and therefore did *not* need to assume circular orbits, but Mayer & Pringle (2006) have argued that even in the advection dominated regime, circular orbits are likely to only introduce an error of several percent.

In addition, we set the vertical flux of mass and momentum (although not energy) to zero at the surface of the disk. We are therefore unable to study the mass and momentum removed by a steady state jet. However, if the jet is relativistic, the mass and momentum that it removes are likely to be insignificant compared to the energy (Nayakshin et al. 2000). This conclusion is less likely to be valid for a transient jet (where the disk can be imagined to “store up” a large amount of mass which it later ejects), but since we are studying the steady state structure in this section, it appears to be a reasonable assumption.

We parametrize the energy removed by a jet in a similar, although slightly different and more precise, way as Svensson & Zdziarski (1994) parametrized the

energy dissipated in a corona. However, unlike papers that discuss the dissipation of energy into a corona (e.g., Ionson & Kuperus 1984; Svensson & Zdziarski 1994), we also take into account the possible contribution of the jet ram pressure to hydrostatic balance in the disk (§3.3.2.3). Effectively, this means that we are taking into account the “vertical transport” of  $\rho v_z$  (i.e., the ram pressure), even though we ignore the vertical transport of the other components of momentum density. The reason this is consistent is that we have integrated our equations over height, so we only care about the *surface* flux of the other momentum components, which is likely to be small, whereas the equation of vertical momentum balance is not integrated over height and therefore remains sensitive to the value of  $\rho v_z$  *within* the disk, as we have discussed in §3.3.2.3.

After making the above approximations, we solve the resulting equations in the usual way (e.g., Shakura & Sunyaev 1973; Kato et al. 1998; Frank et al. 2002), but here we explicitly keep the vertical integration terms  $\{u\}_z$  in the equations, unlike in the standard treatments. This will allow us to see how these terms affect the steady state disk solution, energy balance equation, and typical evolution timescales of the accretion disk.

We start by deriving some simple steady state conditions based on the mass and angular momentum equations (§3.4.1), then move on to derive the equation for energy balance (§3.4.2), the timescales on which the disk can evolve (§3.4.3) and the full, steady state disk solution (3.4.4).

### 3.4.1 Mass and Angular Momentum Conservation

The equations for mass conservation (3.62) and angular momentum conservation (3.64) in the steady state are important for deriving the rest of the steady state

equations, so we deal with them first. Under the simplifications discussed above, these equations become

$$\frac{\partial}{\partial r} \left[ \{v_r\}_z r \Sigma v_r \right] = 0 \quad (3.92)$$

$$\frac{\partial}{\partial r} \left[ r^2 \left( 2 \{ \alpha P / \rho \}_z \alpha P H + \{v_r v_\phi\}_z \Sigma v_r v_\phi \right) \right] = 0. \quad (3.93)$$

If we introduce the mass accretion rate  $\dot{M}$ , which is the local rate at which mass moves inward through radius  $r$  within the accretion disk and is given by

$$\dot{M} \equiv -2\pi r \{v_r\}_z \Sigma v_r, \quad (3.94)$$

then equation (3.92) can be easily solved to give

$$\dot{M} = \text{constant}. \quad (3.95)$$

This equation shows that if any region of the disk is in a steady state,  $\dot{M}$  throughout that region must be constant and, therefore, equal to the rate  $\dot{M}_{input}$  at which mass is being supplied to the region from the outside.

With the help of equation (3.94), equation (3.93) becomes

$$r^2 \left( 2 \{ \alpha P / \rho \}_z \alpha P H - \frac{\{v_r v_\phi\}_z}{\{v_r\}_z} \frac{\dot{M}}{2\pi r} v_\phi \right) = \text{constant}, \quad (3.96)$$

or, substituting  $v_\phi = v_c = (f_{grav} GM/r)^{1/2}$  (where  $f_{grav}$  is the ratio of the equatorial plane value of  $g_r$  to the value that would be given by Newtonian gravity, which can be derived from the equations in §3.2.4.2) and solving for the constant in equation (3.96) at the point in the disk where the turbulent stress goes to zero, we are able to obtain another expression for  $\dot{M}$ , which is

$$\dot{M} = \left[ \frac{4\pi V_{stress}}{f_{stress}} \right] \frac{\alpha P H}{\Omega_c}, \quad (3.97)$$

where

$$V_{stress} = \frac{\{v_r\}_z \{ \alpha P / \rho \}_z}{\{v_r v_\phi\}_z} \quad (3.98)$$

is a ratio of the dimensionless vertical integration parameters and

$$f_{stress} = 1 - \left[ \frac{(f_{grav,in})^{1/2} \{v_r v_\phi\}_{z,in} \{v_r\}_z}{(f_{grav})^{1/2} \{v_r v_\phi\}_z \{v_r\}_{z,in}} \right] \left( \frac{r_{in}}{r} \right)^{1/2} \quad (3.99)$$

is a dimensionless number that approaches 1 far away from the inner edge of the disk. In the above equation, we have used the subscript “in” to refer to quantities evaluated at the point in the disk where the turbulent stress goes to zero. However, we note that  $r_{in}$  is not necessarily the inner “physical” edge of the disk in any sense except that the stress is assumed to be zero there. In particular, we do not necessarily assume that it coincides with the radius of the innermost stable circular orbit (Krolik & Hawley 2002). This formalism allows us to obtain a steady state solution even for disks where the stress on the inner edge of the computational region is not equal to zero.

### 3.4.2 Local Energy Balance

To express the local energy balance, we do not work with the energy equation (3.65) directly, but rather use an energy equation in which the kinetic energy has been subtracted out. Specifically, we subtract equation (3.42) from equation (3.48) and ignore electromagnetic terms as usual, so that we obtain:

$$\frac{\partial E}{\partial t} = -\nabla \cdot (E\mathbf{v}) - (\mathbf{P} + \mathbf{t}) \cdot \nabla \mathbf{v} - \nabla \cdot \mathbf{F}_{rad}, \quad (3.100)$$

where  $E = E_{gas} + E_{rad}$  now represents the internal energy density only. Under the steady state assumptions we make in this section, this equation becomes (after some rearranging of terms)

$$-t_{r\phi} r \frac{\partial \Omega}{\partial r} = \left[ \frac{1}{r} \frac{\partial}{\partial r} (r E v_r) + P \frac{\partial v_r}{\partial r} \right] + \frac{\partial F_{rad}}{\partial z} + \frac{\partial}{\partial z} (E v_z), \quad (3.101)$$

where the term on the left represents the turbulent heating rate per unit volume<sup>11</sup> ( $Q_{heat}$ ), and the terms on the right represent the advective, radiative and jet cooling rates per unit volume ( $Q_{adv}$ ,  $Q_{rad}$  and  $Q_{jet}$ , respectively). We can therefore also write this equation in “shorthand” as

$$Q_{heat} = Q_{adv} + Q_{rad} + Q_{jet}. \quad (3.102)$$

The turbulent heating rate can be transformed by substituting the relation  $\Omega = \Omega_c = (f_{grav})^{1/2} \Omega_k$ , which is valid for circular orbits. We obtain

$$Q_{heat} = \frac{3}{2} f_{deriv} \alpha P \Omega_c, \quad (3.103)$$

where  $f_{deriv}$  is another dimensionless number, given by

$$f_{deriv} = 1 - \frac{1}{3} \frac{\partial \ln f_{grav}}{\partial \ln r}. \quad (3.104)$$

The advective cooling rate, meanwhile, can be transformed using simple manipulations and equation (3.94) to read

$$Q_{adv} = \frac{\xi}{\{v_r\}_z} \frac{\dot{M}}{2\pi r^2} \frac{P}{\Sigma}, \quad (3.105)$$

where

$$\xi = -\frac{E}{P} \left[ 1 + \frac{\partial \ln (E v_r)}{\partial \ln r} + \left( \frac{P}{E} \right) \frac{\partial \ln v_r}{\partial \ln r} \right]. \quad (3.106)$$

We therefore see that the advective energy term is fundamentally nonlocal, because it contains derivatives of time-dependent physical quantities with respect to  $r$ . It is possible to develop the equation for  $\xi$  further by using equation (3.16) to substitute for  $E/P$  in terms of the gas pressure fraction  $\beta$ . The lengthy calculations involved in doing this are outlined in Kato et al. (1998); however, we do not

---

<sup>11</sup>Note that in deriving this term, we use the identity  $\mathbf{t} \cdot \nabla \mathbf{v} = \nabla \cdot (\mathbf{t} \cdot \mathbf{v}) - \mathbf{v} \cdot (\nabla \cdot \mathbf{t})$  and the results of Appendix A, which leads to the appearance of  $r \partial \Omega / \partial r$  (as opposed to  $\partial v_\phi / \partial r$ , which one might naively expect).

find it worthwhile to pursue this level of detail because the local energy balance analysis that we are undertaking is, by definition, only an approximate method in the regime where the advection term is important. Since the advective energy transport only becomes significant when radiation pressure dominates, we can estimate  $\xi$  by assuming  $E = 3P$  (the equation of state for radiation) and using the steady-state Shakura & Sunyaev (1973) solution for a radiation pressure dominated disk to calculate the logarithmic derivatives in equation (3.106); doing this, we find  $\xi \approx 23/2$ . Again, this is only an approximation (since the Shakura & Sunyaev 1973 solution ignores advection), but it should at least give us a roughly accurate estimate for the local energy balance in the regime where advection is just starting to become important, which turns out to be the regime that is most interesting for studying the time-evolution of the disk.

If we substitute the expressions for  $Q_{heat}$  and  $Q_{adv}$  into equation (3.101) and integrate over height (replacing the integral over  $Q_{rad}$  with the optically thick limit of the expression for  $2F_{surf}$  in equation 3.90), we obtain

$$\left(3V_{heat}f_{deriv}\right)\Omega_c\alpha PH = \left(\frac{V_{adv}\xi}{\pi}\right)\frac{\dot{M}}{\Sigma r^2}PH + \left(\frac{16}{3V_{rad}}\right)\frac{\sigma T^4}{\kappa\Sigma} + F_{jet}, \quad (3.107)$$

where

$$V_{heat} = \{f_{deriv}\Omega_c\alpha P/\rho\}_z \quad (3.108)$$

$$V_{adv} = \frac{\{P/\rho\}_z}{\{v_r\}_z} \quad (3.109)$$

$$V_{rad} = \{\kappa f_{rad}\}_z \quad (3.110)$$

are combinations of the dimensionless vertical integration parameters, and the flux of internal energy carried away by the jet is given by  $F_{jet} = 2(Ev_z)_{surf}$ . Note that we did not bother to include  $\xi$  in the integration over height because, as previously discussed, the expression we have obtain for  $\xi$  is by necessity only a very rough

approximation; attempting to derive its dependence on height would therefore be of little use.

At times, we will find it useful to alter equation (3.107) by removing  $F_{jet}$  and instead multiplying the radiative cooling term (second to last term in that equation) by

$$1 + f_{jet} = 1 + \left( \frac{Ev_z}{F_{rad}} \right)_{surf}, \quad (3.111)$$

where  $f_{jet}$  is the ratio of the internal energy carried by the jet to the energy carried by radiation, measured at the surface of the disk. Our parametrization of the energy carried by the jet is similar in concept to the parametrization used by Janiuk et al. (2002) in their numerical simulations and to the parametrization used by other authors for the energy dissipated in a steady state corona. However, these authors generally parametrize the coronal or jet dissipation as a fraction of the turbulent heating rate, or (in less precise language) as a fraction of the dissipated gravitational energy (e.g., Haardt & Maraschi 1991; Svensson & Zdziarski 1994; Chen 1995; Janiuk et al. 2002). This is in some ways a more physical definition (since it allows us to write the energy carried away by the jet as a fraction of the energy produced locally and available for it to carry away), but it would cause  $f_{jet}$  to have a much more complicated expression that would depend on vertical integration and other model-dependent parameters. For our purposes, it is more natural to define the fraction of energy carried away by the jet in terms of the other surface term in the energy equation, which is the fraction carried away by radiation. This lets us directly specify the extent to which the energy release is dominated by the jet or dominated by the radiation. We note that if advection is unimportant, our definition of  $f_{jet}$  can be related to the more “standard” definition  $f'_{jet}$  (written in terms of the turbulent heating rate) by  $1 + f_{jet} = (1 - f'_{jet})^{-1}$ .

We are now in a position to write equation (3.107) in terms of only two primary physical variables. We first combine equations (3.84) and (3.97) to obtain

$$\dot{M} = \left( \frac{2\pi \{v_r\}_z \{\alpha P/\rho\}_z h_P^2}{\{v_r v_\phi\}_z f_{stress} C_H^2} \right) \alpha \Omega_c \Sigma (H^2 - H_{jet}^2), \quad (3.112)$$

which allows us to write the local energy balance equation (3.107) in its final form:

$$D_{heat} \alpha \Omega_c^3 \Sigma (H^2 - H_{jet}^2) = D_{adv} \frac{\alpha \Omega_c^3 \Sigma}{r^2} (H^2 - H_{jet}^2)^2 + D_{rad} \frac{\sigma T^4}{\kappa \Sigma} + F_{jet}, \quad (3.113)$$

where

$$D_{heat} = \frac{3 \{f_{deriv} \Omega_c \alpha P/\rho\}_z f_{deriv} h_P^2}{2 C_H^2} \quad (3.114)$$

$$D_{adv} = \frac{\{P/\rho\}_z \{\alpha P/\rho\}_z \xi h_P^4}{\{v_r v_\phi\}_z f_{stress} C_H^4} \quad (3.115)$$

$$D_{rad} = \frac{16}{3 \{\kappa f_{rad}\}_z}. \quad (3.116)$$

These equations allow us to make a plot of the local energy balance on the  $T_0$ - $\Sigma$  or  $\dot{M}$ - $\Sigma$  plane at a given radius  $r$  (with a corresponding value of  $\Omega_c$  calculated from  $r$ ,  $M$  and, in some cases, the black hole spin parameter  $a$ ), as long as we are given the following additional information:

- The value of  $\alpha$  on the equatorial plane.
- A particular set of assumptions about the disk's vertical structure, including the vertical profile of turbulent heating.
- The contribution of the jet to the vertical pressure balance (parameterized by  $H_{jet}$ ).
- The internal energy flux at the surface of the disk due to the jet (parametrized by  $F_{jet}$ ).

The procedure for making this plot is quite simple; for a particular combination of  $T_0$  and  $\Sigma$ , calculate  $H$  via equation (3.80), and then use these quantities to calculate the opacity  $\kappa$  via equation (3.51). Since the heating rate is generally a monotonic function of  $\Sigma$ , we can solve equation (3.113) for a given value of  $T_0$  using a simple root finding method (in practice, we use the simplest method: bisection). Finally, we can determine the solution on the  $\dot{M}$ - $\Sigma$  plane using equation (3.112).

The effect of choosing different values of  $r$  and  $\alpha$  on the solution of the energy balance equation is well-known (e.g., Abramowicz et al. 1988; Janiuk et al. 2002). Abramowicz et al. (1988) were among the first to notice the importance of the advection term in equation (3.113) that gives rise to an “S-shaped” energy balance curve and allows for the possibility of limit-cycle oscillations. Janiuk et al. (2002) also investigated the effect of  $F_{jet}$ , but only under a very specific model in which the jet was assumed to be more active at higher values of  $\dot{M}$ ; this model is contradicted by the most recent observational data (e.g. Fender et al. 2004). The effect of the other two variables in the equation we have derived (those involving the jet ram pressure and the vertical structure of the disk) do not appear to have been previously considered at all.

We show the effects of these terms in Figures 3.7 through 3.9. Figure 3.7 shows the effect of including the ram pressure due to a jet in an otherwise normal disk with  $\alpha = 0.05$ ; the energy balance curve is shown for  $r = 10r_g$  and a  $14M_\odot$  black hole, assuming Newtonian gravity. The jet fluid velocity within the disk was chosen to be  $\sim 1/20$  of the orbital speed at that radius. As can be seen from Figure 3.7, the inclusion of the jet ram pressure significantly changes the structure of the energy curve around the lowermost “bend,” which is the region of the curve that a disk will traverse right before going into an outburst. A significant region

of parameter space that was previously to the left of the curve (and therefore undergoing cooling) is now to the right (and therefore undergoing heating, leading to an unstable outburst). This is due to the fact that the jet ram pressure thickens the disk, thereby increasing the height-integrated gas and radiation pressure (and therefore the turbulent heating) as compared to a disk without a jet. In the upper part of the curve, the value of  $H_{jet}$  we have chosen has a very small effect because the radiation pressure scale height is dominant, whereas in the lower part of the curve, the effect is also small, because the  $H^2 - H_{jet}^2$  term in equation (3.113) can also be written as  $H_{gas}^2 + H_{rad}H$ , which means that the jet ram pressure requires a significant  $H_{rad}$  in order to have an effect.

Figure 3.7 suggests that the shape of an outburst light curve depends on the jet ram pressure. A disk with a strong jet would appear to have a sharper change in the accretion rate and temperature before going into outburst than a disk with a weaker jet. It is interesting to note that in GRS 1915+105, there is a wide range of behavior that can occur during the long “dips” in the  $\sim 30$  minute light curves that precede an outburst, encompassing both spectral and temporal differences (Mikles, Eikenberry, & Rothstein 2006), and we speculate that this may be partially due to differences in the jet structure on the lower part of the energy balance curve, as shown in Figure 3.7.

Another interesting effect seen in Figure 3.7 is that the disk becomes more radiation pressure dominated at a given local accretion rate (and temperature) when a jet is included. This is due to the fact that the jet increases the scale height of the disk, thereby spreading out the disk material and decreasing the density and gas pressure. This effect could be relevant for the wide variety of models in which  $\alpha$  is assumed to respond less strongly to the radiation pressure

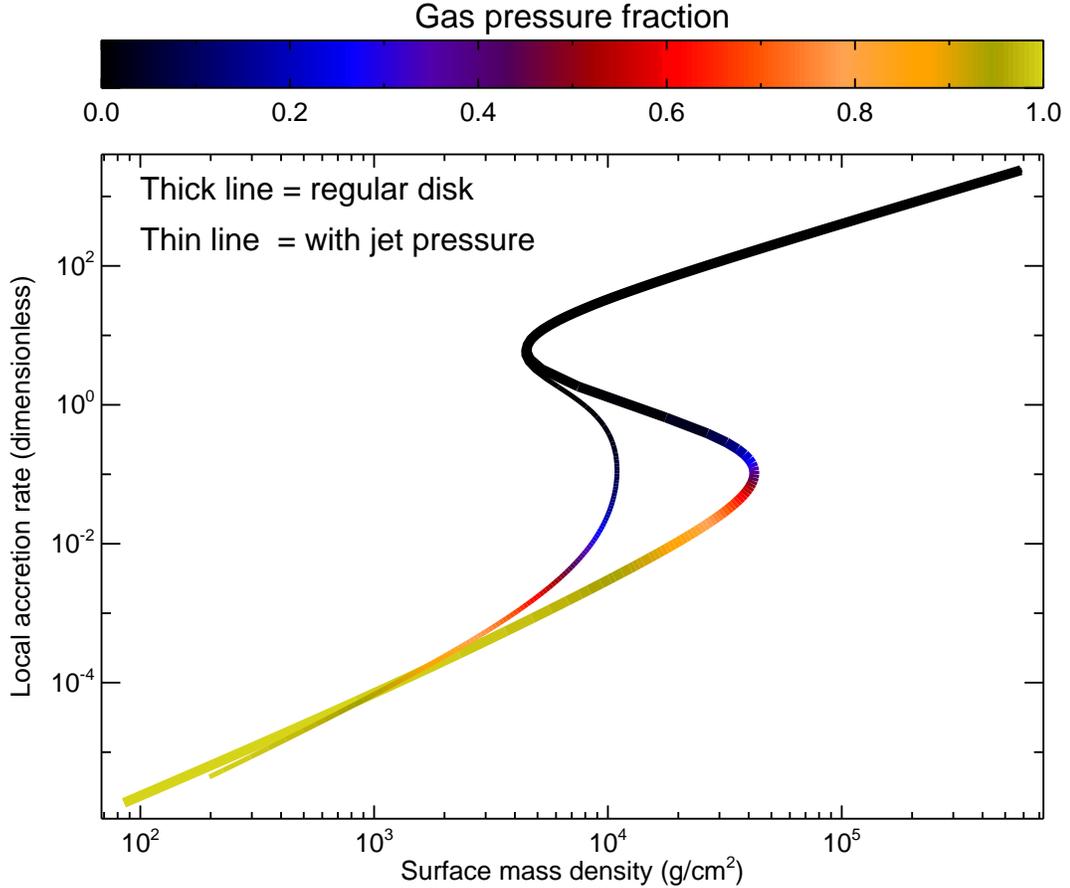


Figure 3.7: The effect of the jet ram pressure on the local energy balance curve of an accretion disk (the vertical axis is the local mass accretion rate  $\dot{M}$ , plotted in units of the Eddington luminosity divided by  $c^2$ ). The jet fluid velocity within the disk was chosen to be  $\sim 1/20$  of the orbital speed at the radius for which the energy balance curve is plotted ( $10 r_g$ ). We also chose  $\alpha = 0.05$ , a black hole mass of  $14 M_\odot$ , and we assumed Newtonian gravity. It can be seen that the jet ram pressure significantly changes the shape of the energy balance curve and causes regions of parameter space that would otherwise be stable to become unstable due to heating. It also has a significant effect on the gas pressure fraction  $\beta$ , causing the gas pressure to weaken at the expense of radiation pressure.

than the gas pressure (e.g., Szuszkiewicz 1990; Watarai & Mineshige 2003).

We caution that Figure 3.7 is only a rough illustration of the effect of a jet; for example, it assumes that the jet speed is the same regardless of the disk’s position on the energy balance curve, which is not likely to be the case, especially if jets are more prevalent in the low state. Also, it is difficult to use this figure to determine how a disk would behave if a jet were *quickly* added or removed (as in a major ejection event), since the presence of a jet likely would change the local mass accretion rate and equatorial plane temperature, as well as cause other changes to the vertical structure (such as those discussed in Figures 3.8 and 3.9).

Figure 3.8 shows the effect of including the energy removed by a jet on the local energy balance curve. The basic parameters are the same as for the regular disk in Figure 3.7, but now we investigate the effect of the energy removed by the jet at the disk surface. We set  $f_{jet} = 1$  (i.e., the jet energy flux is always equal to the radiative energy flux). This assumption is likely to be true somewhere in the vicinity of the rightmost “bend” in the curve, since jets are likely to dominate the energy equation at low accretion rates but be insignificant during the outburst (Fender et al. 2003; K rding et al. 2006). Our rough model in Figure 3.8 shows that the energy removed by a jet has the opposite effect of the jet ram pressure shown in Figure 3.7, although the jet ram pressure effect appears to be bigger (at least for the very approximate fiducial values we have chosen). The degree to which these two effects cancel out or not would depend on the details of the jet formation process, in particular on how “deep” within the disk the jet is formed. We also note that our plot in Figure 3.8 looks significantly different from the plot in Janiuk et al. (2002), because their model for a jet only became significant at very high accretion rates, where either the radiation pressure or advective energy

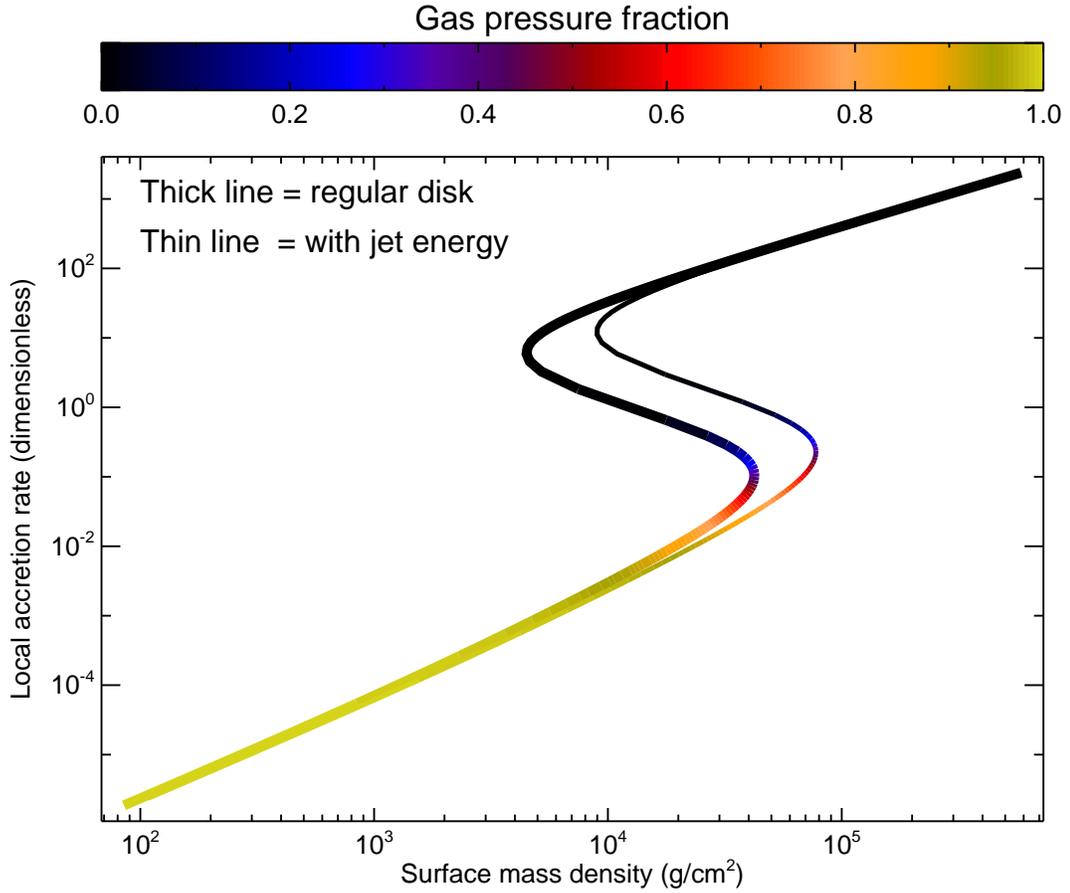


Figure 3.8: The effect of the jet energy transport on the local energy balance curve of an accretion disk. The parameters for the regular disk are the same as those in Figure 3.7, but we now investigate the effect of the energy transported out of the disk by the jet, assuming that this energy is always equal to the energy carried by radiation. The effect here is seen to be the opposite of that shown in Figure 3.7.

transport were already dominant. However, the general fact that the jet energy transport is seen to have a stabilizing effect in the vicinity of the “bend” in the energy balance curve is consistent with the specific results of Merloni (2003), who found that the energy dissipated in a corona (rather than a jet) can completely stabilize a disk that is moderately radiation pressure dominated, if the fraction of energy dissipated in the corona is very high and if a particular prescription for  $\alpha$  is chosen.

Finally, we show in Figure 3.9 the effect of changing the vertical parameters  $\{u\}_z$  in a manner that might be consistent with that of a jet. Specifically, based on our previous discussion, we model the jet by lowering all terms of the form  $\{\sim P/\rho\}_z$  by 80%, lowering all terms of the form  $\{\sim \alpha P/\rho\}_z$  by 50%, and raising  $\{\kappa f_{rad}\}_z$  by a factor of 2.5 (corresponding to a more centrally concentrated heating profile). As can be seen in Figure 3.9, the effects here are more subtle and appear to be smaller than those discussed previously; more work is needed to fully investigate the parameter space in this case. We note, however, that even if the effect of the  $\{u\}_z$  on the *shape* of the energy balance curve is small, these terms do have an important effect on the evolution timescales on which a disk *traverses* the energy balance curve, as we discuss further in the next section.

### 3.4.3 Evolution Timescales

In this section, we derive typical evolution timescales for the steady state accretion disk. Since our goal is to estimate how the vertical integration parameters  $\{u\}_z$  affect the timescales, we keep these parameters in our equations, but we are still free to ignore other dimensionless numbers of order  $\sim 1$  that are not affected by the disk’s vertical structure.

We first derive the accretion timescale (often called the “viscous timescale” in the literature). This is roughly equal to the time it takes for material at a particular location in the disk to accrete onto the central object. There are several ways this is often calculated that are equivalent when the vertical integration parameters are ignored, but for our purposes we must be careful to employ a fully three-dimensional method. We therefore define the accretion timescale at radius  $r$  as the total amount of mass in the vicinity of  $r$  divided by the accretion rate, or

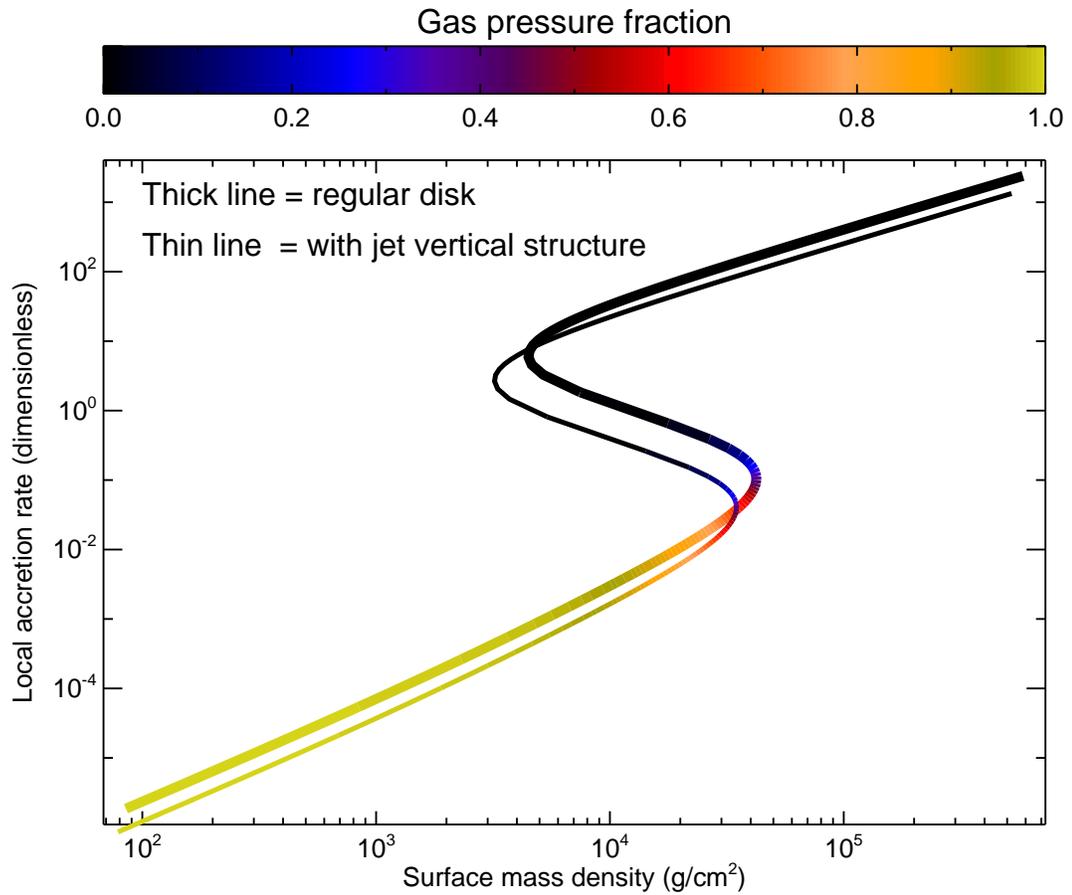


Figure 3.9: The effect of the jet vertical structure on the local energy balance curve of an accretion disk. The parameters for the regular disk are the same as those in Figure 3.7. We modify the vertical parameters  $\{u\}_z$  in a way that might be consistent with the presence of a jet, as discussed further in the text. The overall effect on the energy balance curve appears to be small.

$t_{accr} \sim \Sigma r^2 / \dot{M}$ . We write this in two equivalent ways, first using equation (3.97) and then using equation (3.112):

$$\begin{aligned}
 t_{accr} &\sim \frac{\{v_r v_\phi\}_z}{\{v_r\}_z} \frac{\Sigma r v_\phi}{\int \alpha P dz} \\
 &\sim \left[ \frac{\{v_r v_\phi\}_z}{\{v_r\}_z \{\alpha P / \rho\}_z h_P^2} \right] \frac{1}{\alpha \Omega_c} \left( \frac{r^2}{H^2 - H_{jet}^2} \right). \quad (3.117)
 \end{aligned}$$

The first expression is useful for studying the way in which the accretion timescale is affected by the vertical redistribution of material due to a steady jet or to changes in the vertical profile of turbulent heating. To first order, we can assume that a steady jet does not change the  $r$  and  $\phi$  momentum in the disk, only redistributes it vertically (in other words, the force on the disk produced by the jet is primarily in the vertical direction). This suggests that  $\{v_r v_\phi\}_z / \{v_r\}_z$  is not significantly affected by the presence or absence of a jet. On the other hand, as we have discussed in §3.3.2.2, a jet is likely to decrease the value of  $\int \alpha P dz$ , either because it redistributes pressure vertically to regions where the turbulent heating is less efficient, or because it causes the turbulent heating *itself* to become less efficient or more centrally concentrated. We therefore find that a steady jet is likely to increase the accretion timescale  $t_{accr}$ . Similarly, a disk with a more centrally concentrated turbulent heating rate is also likely to have larger values of  $t_{accr}$ .

The second expression in equation (3.117) is more difficult to interpret physically, but useful for comparing to the standard expression  $t_{accr} \sim (r/H)^2 / (\alpha \Omega_c)$  found in previous works (e.g., Kato et al. 1998; Frank et al. 2002).

We can perform a similar analysis for the cooling timescale, which we define to be roughly equal to the time it takes a local region of the disk to lose its internal energy due to radiative cooling or the energy carried away by advection or a jet. We can define this as  $t_{cool} \sim (\int E r^2 dz) / (\int Q_{heat} r^2 dz)$ ; in other words, we divide

the total internal energy in the vicinity of radius  $r$  by the cooling rate (where we can use  $Q_{heat}$  for the cooling rate per unit volume because the disk is assumed to be in a steady state; e.g., see equation 3.102). It is easy to evaluate this expression based on our previous work; we obtain

$$t_{cool} \sim \left[ \frac{\{E/\rho\}_z}{\{f_{deriv}\Omega_c\alpha P/\rho\}_z} \right] \frac{1}{\alpha\Omega_c}, \quad (3.118)$$

where the standard textbook expression is  $t_{cool} \sim 1/(\alpha\Omega_c)$ . It is more difficult to interpret the vertical integration parameters in this case, but we note that the appearance of a term  $\{\sim \alpha P/\rho\}_z$  in the denominator suggests that the cooling timescale, like the accretion timescale, will increase when a jet or centrally-concentrated turbulent heating profile is present. We note, however, that the features seen in the GRS 1915+105 light curves in Chapter 2 are likely controlled primarily by the viscous timescale (e.g., Belloni et al. 1997a,b), since the cooling timescale is generally too short.

### 3.4.4 Steady State Solution

In this section, we solve the equations derived previously in order to obtain a steady state disk structure similar to that of Shakura & Sunyaev (1973). We ignore advection (as Shakura & Sunyaev 1973 did) but allow for the presence of a jet, recalling that observations indicate that jets are most prevalent when radiative cooling is inefficient (e.g. Fender et al. 2004) and when, therefore, advection can usually be ignored.

We solve these equations using the usual method. A concise way to illustrate this is to combine equations (3.84) and (3.97) to get a relation between  $P$  and  $\rho$ ,

given by

$$P = \left[ \left( \frac{h_P f_{stress}}{4\pi V_{stress} C_H} \right)^{2/3} f_{jetpr}^{1/3} \right] \left( \frac{\dot{M} \Omega_c^2}{\alpha} \right)^{2/3} \rho^{1/3}, \quad (3.119)$$

where  $f_{jetpr} = 1 - H_{jet}^2/H^2$ . We then combine equations (3.97) and (3.107)—ignoring advection in the latter—to get another relation between  $P$  and  $\rho$  that also involves  $T$  and  $\kappa$ , which is

$$P = \frac{9}{128\pi^2} \left[ \frac{V_{rad} V_{heat} f_{deriv} f_{stress}^2}{V_{stress}^2 (1 + f_{jet})} \right] \left( \frac{\dot{M}^2 \Omega_c^3}{\alpha} \right) \left( \frac{\kappa}{\sigma T^4} \right) \rho. \quad (3.120)$$

These equations can be combined with the equation of state (3.13) and opacity prescription (3.51) to obtain four equations in four unknowns and thereby derive a solution.

As an example, we give our solution for the “middle region” of the Shakura & Sunyaev (1973) disk, where gas pressure and scattering opacity dominate. This is the most important steady state solution for our purposes, since a radiation pressure dominated disk is unstable (Lightman & Eardley 1974; Shakura & Sunyaev 1976) and therefore the system should never reach a steady state when it is dominated by radiation (unless  $\alpha$  responds more strongly to the gas pressure than the radiation pressure; Szuszkiewicz 1990; Merloni & Nayakshin 2006). A disk in which absorption opacity dominates, meanwhile, is usually too cool to be of interest for comparing to our observations.

Our solution for the temperature is

$$kT = \underbrace{\left[ \frac{3}{2} \left( \frac{5}{3\pi^3} \right)^{1/5} \alpha_e^{-2/5} \left( \frac{m_p}{m_e} \right)^{1/5} \left( \frac{m_p}{\alpha_g M_\odot} \right)^{1/5} m_e c^2 \right]}_{\text{Fiducial Value}} \underbrace{\left[ \mu (1 + X) \right]^{1/5}}_{\text{Composition}} \times \\ \times \underbrace{\left[ \frac{\{\kappa f_{rad}\}_z \{f_{deriv} \Omega_c \alpha P / \rho\}_z \{v_r v_\phi\}_z^2}{\{\alpha P / \rho\}_z^2 \{v_r\}_z^2} \right]^{1/5}}_{\text{Vertical Structure}} \underbrace{\left[ 1 + f_{jet} \right]^{-1/5}}_{\text{Jet Cooling}} \underbrace{\left[ f_{deriv}^{1/5} f_{grav}^{3/10} \right]}_{\text{Gravity}} \times$$

$$\times \underbrace{\left[ f_{stress}^{2/5} \left( \frac{M}{M_\odot} \right)^{-1/5} \left( \frac{\dot{M}}{\dot{M}_{crit}} \right)^{2/5} \left( \frac{r}{r_g} \right)^{-9/10} \alpha^{-1/5} \right]}_{\text{Original Shakura \& Sunyaev (1973) Dependence}}, \quad (3.121)$$

where  $\alpha_e$  is the fine-structure constant,  $\alpha_g$  is the gravitational coupling constant,  $m_e$  and  $m_p$  are the electron and proton masses, respectively, and  $\dot{M}_{crit} \equiv L_{edd}/c^2$  (where  $L_{edd}$  is the Eddington luminosity). The fiducial value of the temperature is  $kT \approx 0.092m_e c^2 \approx 47$  keV (note that this is a bit higher than typical values found in actual disks, because the disk does not extend all the way down to  $r_g$ ).

Our solution for the surface mass density is

$$\begin{aligned} \Sigma &= \underbrace{\left[ \frac{4}{3} \left( \frac{5}{3\pi^3} \right)^{-1/5} \alpha_e^{2/5} \left( \frac{m_p}{m_e} \right)^{4/5} \left( \frac{m_p}{\alpha_g M_\odot} \right)^{-1/5} \frac{m_p}{\sigma_{th}} \right]}_{\text{Fiducial Value}} \underbrace{\left[ \mu^{4/5} (1+X)^{-1/5} \right]}_{\text{Composition}} \times \\ &\times \underbrace{\left[ \left( \frac{1}{\{\kappa f_{rad}\}_z \{f_{deriv} \Omega_c \alpha P / \rho\}_z \{\alpha P / \rho\}_z^3} \right) \frac{\{v_r v_\phi\}_z^3}{\{v_r\}_z^3} \right]}_{\text{Vertical Structure}}^{1/5} \underbrace{\left[ 1 + f_{jet} \right]}_{\text{Jet Cooling}}^{1/5} \times \\ &\times \underbrace{\left[ \left( \frac{f_{grav}}{f_{deriv}} \right)^{1/5} \right]}_{\text{Gravity}} \underbrace{\left[ f_{stress}^{3/5} \left( \frac{M}{M_\odot} \right)^{1/5} \left( \frac{\dot{M}}{\dot{M}_{crit}} \right)^{3/5} \left( \frac{r}{r_g} \right)^{-3/5} \alpha^{-4/5} \right]}_{\text{Original Shakura \& Sunyaev (1973) Dependence}}, \quad (3.122) \end{aligned}$$

where  $\sigma_{th}$  is the Thomson cross section for electron scattering. Here, the fiducial value is  $\Sigma \approx 4.0 \times 10^4 (m_p / \sigma_{th}) \approx 1.0 \times 10^5$  g cm<sup>-2</sup>, thereby showing that the disk is optically thick for typical parameters (since  $\tau \sim \Sigma \sigma_{th} / m_p$ ).

### 3.5 Conclusions

In this chapter, we have revisited the foundations of stellar-mass black hole accretion disk theory, especially in its vertically-integrated form. In the first half of the chapter, we carefully laid out the assumptions that go into deriving the standard disk equations, thereby providing a way to strictly test the validity of simplified

disk models in various regimes and, in particular, in numerical simulations.

In the second half of the chapter, motivated by the observational evidence for jets that undoubtedly have a significant effect on the vertical structure of a disk, we have attempted to encompass our uncertainty about the jet physics and disk vertical structure into a series of (mostly) dimensionless numbers, similar to the  $\alpha$  parameter for turbulent stress introduced by Shakura & Sunyaev (1973). We then began a preliminary investigation of the way in which a jet might alter some of these parameters, through the effect of the jet ram pressure, energy transport, and the vertical redistribution of material within the disk. We found, in particular, that a jet can lead to significant changes in the energy balance equation and evolution timescales of a disk.

The true effect of the variety of parameters defined in this chapter on the vertical structure of the disk can only be investigated via numerical simulations. However, by carefully deriving the dimensionless numbers and their plausible ranges in the way we have done here, we hope to provide guidance for future numerical work that can begin to map out the wide range of parameter space that is possible and compare the numerical results, either in a statistical or direct way, to the wide variety of X-ray observations that have been undertaken of GRS 1915+105 and other objects.

## Chapter 4

# FRIENDLY: A New Code for Integrating Time-Dependent Equations to Arbitrary Orders of Accuracy

### 4.1 Summary

In this chapter, we describe FRIENDLY, a new FORTRAN 95 program for numerical integration that was designed, written and tested during the course of this thesis. The primary useful feature of the code is that it is designed to entirely separate the computational algorithms from the physical equations that are being simulated; currently, it is possible to change the underlying physical model simply by altering one file and recompiling the code. FRIENDLY also allows a wide variety of other features to be changed by the user at runtime (including the number of dimensions being simulated, the boundary conditions, the grid spacing, which can be regular or irregular, and the order of accuracy to which the equations are evolved).

The code uses a “method of lines” explicit integration approach that is very flexible and allows an arbitrary set of equations in an arbitrary number of dimensions to be integrated to arbitrary orders of spatial and temporal accuracy (at present, second and fourth order methods are implemented in the code), and it

also uses powerful FORTRAN 95 array features to allow the code to operate in any number of dimensions from one to three, while ensuring that as many operations as possible are done outside of the main integration loop and that the code can, in the future, be used in a parallel computing environment.

In the following sections, we discuss the motivation for writing a new code, describe some of its features, show preliminary results, and also derive specific, new algorithms that can be used within the method of lines environment to calculate spatial derivatives to any arbitrary order of accuracy on any arbitrarily-spaced grid, and also to preserve conservation properties of the numerical scheme, if so desired.

## 4.2 Motivation

There are many numerical codes currently in use that implement a wide variety of procedures for integrating time-dependent equations. In astrophysics, the most popular such code is probably ZEUS (Stone & Norman 1992a,b; Stone et al. 1992). We are interested in creating a new, publicly available code for three main reasons.

First, open source FORTRAN 95 compilers have become available in the last few years, suggesting that FORTRAN 95 is poised to truly become the “new” standard language in scientific computing. FORTRAN 95 provides many modern features that older codes were unable to take advantage of, including powerful array operations that simplify the handling of multiple dimensions, and built-in compatibility with parallel computing environments. FRIENDLY is written in FORTRAN 95 and takes advantage of these features extensively.

Second, there are situations in which it is desirable to be able to integrate equations to higher orders of accuracy than the second order methods that are

commonly in use. The “method of lines” code that we have implemented in FRIENDLY (which consists of combining a high order spatial derivative scheme with a Runge-Kutta time integration scheme) can easily be extended to any desired order of accuracy. A particular application of this could be in simulating magnetic fields using the vector potential  $\mathbf{A}$  rather than the magnetic field  $\mathbf{B}$  directly (where  $\mathbf{B} = \nabla \times \mathbf{A}$ ); such schemes require high orders of accuracy but have the advantage that they automatically give  $\nabla \cdot \mathbf{B} = 0$  (Brandenburg 2003), while other codes need to specifically enforce this physical condition. Thus, the method of lines implemented in FRIENDLY furthers the goal of separating the physical and computational aspects of the problem.

Finally, there is a larger goal. As computers become faster, numerical results are becoming more important in many fields of science, including those that have important public policy implications (e.g., global warming). It is therefore vital to have an extremely flexible, publicly available, and, first and foremost, user-friendly code for performing numerical integration. This would of course be beneficial for scientists in certain applications, but it would be even more beneficial in order to make the work that scientists do as transparent as possible to the general public. Although FRIENDLY is not yet at the stage where it could be used by the general public, it would not be hard to extend it to meet this goal; because the numerical details are completely separate from the physical equations being simulated, it would not be difficult to write a user-friendly wrapper program that actually writes the necessary FORTRAN code for a new physical model “on the fly.” Thus, one could imagine allowing students or other members of the general public to design their own “physical system” (perhaps in a graphical way) that could then be automatically evolved in time using FRIENDLY.

### 4.3 Separating Space and Time Derivatives with the Method of Lines

Any numerical scheme that is designed for fluid dynamics essentially consists of an underlying physical choice as to how the fluid should be represented in a discrete way. We focus here on explicit, Eulerian finite-difference schemes, which essentially consist of treating the fluid as a series of discrete particles fixed in space and connected by a massless medium; this means that entropy generation in, for example, a shock, can directly manifest itself as oscillations on the discrete grid (Richtmyer & Morton 1967). Other choices include Godunov methods (which treat the fluid like a series of small fluid cells and therefore can more naturally handle shocks), Lagrangian methods, and implicit methods. The explicit methods, however, are the most flexible, so we have chosen to focus on them.

A wide variety of explicit, finite-difference integration schemes have been derived (see Hirsch 1988 and Press et al. 1992 for some examples), and the subtle differences between them can often be bewildering. In choosing between these possibilities, our goal is to find schemes that are first and foremost *flexible*, in the sense that they can be applied to a wide variety of equations (not just equations in conservative form), and also in the sense that they can easily be generalized from a single spatial dimension all the way up to a full three-dimensional problem. Based on these criteria, we have decided to focus on the broad class of integration schemes that fall within the domain of the “method of lines.”

The method of lines is less an integration scheme itself than it is a philosophical approach to solving partial differential equations, in which the goal is to reduce the equations to *ordinary* differential equations in one variable (for a review, see

Schiesser 1991 and Jones et al. 1972, who also refer to earlier works discussing the method's origins in the Soviet Union<sup>1</sup>).

Although the name “method of lines” is quite obscure, the basic approach is very simple: Choose one variable that we want to have in our final ordinary differential equation (in our case, this will always be time). Then, search through the equations, find all partial derivatives with respect to *other* variables (in our case, the space variables; e.g.,  $x$ ,  $y$  and  $z$ ), discretize these variables onto a grid and replace their derivatives with finite-difference approximations.<sup>2</sup> We are then left with equations that only have derivatives with respect to time. In some cases, these equations can be solved analytically, but more generally we can solve them with one of the many numerical methods that have been developed for ordinary differential equations (e.g., Runge-Kutta methods; see Press et al. 1992).

Low order versions of the method of lines are generally similar (or even identical) to integration methods that have been devised in other contexts; in fact, as we will see below, popular second-order methods that were primarily developed for conservative equations such as the two-step Lax-Wendroff and “staggered leapfrog” schemes can really be viewed as “modifications” of the basic method of lines. The difference between the method of lines and other methods, however, becomes apparent at high orders of accuracy. It is very simple conceptually to construct a method of lines scheme that is accurate to *any* desired order in space and time; simply obtain a high order finite-difference approximation for the spatial deriva-

---

<sup>1</sup>Also see the discussion by Hirsch (1988) of “central schemes with independent time integration” that are very similar in spirit to the method of lines.

<sup>2</sup>The name “method of lines” comes from the fact that if we apply it to a problem that *only* contains spatial derivatives, then after we have discretized all but one of these dimensions, the resulting grid will look like a bunch of long, thin sticks (or “lines”) pointing along the one dimension that remains continuous (Jones et al. 1972; Schiesser 1991).

tives (see §4.6), and then apply a high-order Runge-Kutta method to integrate the resulting ordinary differential equation in time.

Though our program currently only implements the second and fourth order versions of the method of lines (as described in the next section), we would only need to change a few small bits of code to add the ability to do arbitrarily higher orders of accuracy. Organizing our program around the method of lines concept allows the desired order of accuracy to essentially be a *parameter* input by the user, rather than a defining feature of the computer program that is inseparable from the choice of integration scheme.

## 4.4 Choosing an Order of Accuracy for the Space and Time Derivatives

In this section, we describe the levels of accuracy chosen to implement in our method of lines code. For simplicity, consider a one-dimensional conservative equation,  $\partial\rho/\partial t = \partial f/\partial x$ . Our goal in numerical integration is to find the actual change  $\Delta\rho$  that occurs in a given time step  $\Delta t$ . This can be expressed as:

$$\Delta\rho = \Delta t \left( \frac{\partial f}{\partial x} \right)_{\tau}, \quad (4.1)$$

where the Mean Value Theorem of calculus guarantees this equality to be true at some time  $\tau$  during the time step.

There are two sources of error in our numerical estimate of this equation. The first is that we do not know the value of  $\tau$ ; instead, the Runge-Kutta methods typically seek to estimate  $(\partial f/\partial x)_{\tau}$  by taking linear combinations of  $\partial f/\partial x$  evaluated at various times during the step  $\Delta t$ . It is traditional to define an *n*th order Runge-Kutta scheme as one where the particular linear combination chosen produces an

error added to  $\Delta\rho$  that is of order  $O(\Delta t^{n+1})$ .

The second source of error is the finite difference method used in each calculation of  $\partial f/\partial x$ ; in this case, an  $m$ th order finite difference method is defined as one where the resulting error added to  $\partial f/\partial x$  is  $O(\Delta x^m)$  (note the somewhat confusing difference between this definition and the one above).

Putting all the above information together, we find:

$$(\Delta\rho)_{estimated} \approx \Delta\rho + O(\Delta x^m \Delta t) + O(\Delta t^{n+1}), \quad (4.2)$$

where the first error term is associated with the finite difference method and the second is associated with the Runge-Kutta method (note that the error terms add linearly because they are not statistical errors but rather systematic deviations from the true  $\Delta\rho$ ). The fact that these error terms add together means that we would be wasting our time if we tried to make one of our methods much more accurate than the other; the error from the less accurate method would dominate (Schiesser 1991). Thus, whatever order of accuracy we choose, we should strive to make the two error terms roughly equal. Since the Courant stability condition typically requires  $\Delta t \lesssim \Delta x$  in dimensionless units, our condition boils down to  $m = n$ , which means that an  $n$ th order finite difference method should be accompanied by an  $n$ th order Runge-Kutta method.

Second and fourth order Runge-Kutta methods are the most straightforward to implement and generally sufficient for most applications, so we have initially implemented those methods in our code. We have specifically checked that the second and fourth order methods implemented in FRIENDLY work to the expected level of accuracy, by testing the code on a model in which the space and time derivatives of our input function are equal to polynomials in space and time.

As an example of what FRIENDLY can do, we show in Figure 4.1 a comparison

of our second and fourth order methods for the evolution of a sound wave in an adiabatic gas. The initial condition is similar to the classic “shock tube” problem in which a barrier separating a high pressure fluid from a low pressure fluid is removed, and the fluid is allowed to evolve (although strictly speaking, the example shown Figure 4.1 is not an actual shock because the Mach number is too low). As can be seen, the fourth order (black lines) and second order (red lines) methods produce very similar results. In this case, the fourth order method does not actually improve the accuracy noticeably, because the physical situation is so simple (in fact, the fourth order method appears to be worse in the second and third panels of Figure 4.1, but that is primarily because the characteristic grid speed for a fourth order method is twice as fast as that for a second order method, and thus the oscillations which are a natural result of the initial sharp discontinuity in the pressure can travel faster in the fourth order case). In addition to the two sound waves that move outward from the initial fluid boundary, also visible in Figure 4.1 are the contact discontinuities and rarefaction waves expected in this situation; the two rarefaction waves actually collide with each other in the final panel.

In general, we have tested our code primarily using an adiabatic gas with discontinuities in the initial conditions. The initial discontinuity will always lead to the presence of oscillations, and thus these tests allow us to see how our code behaves under extreme conditions beyond that which a finite difference method is typically designed to handle. Since our code includes an artificial viscosity (e.g., Richtmyer & Morton 1967), we do not expect oscillations to be a major problem in a real computation, since any shocks that happen to arise will be largely smoothed out by the artificial viscosity.

We show an example of the two-dimensional capabilities of the code in Figure

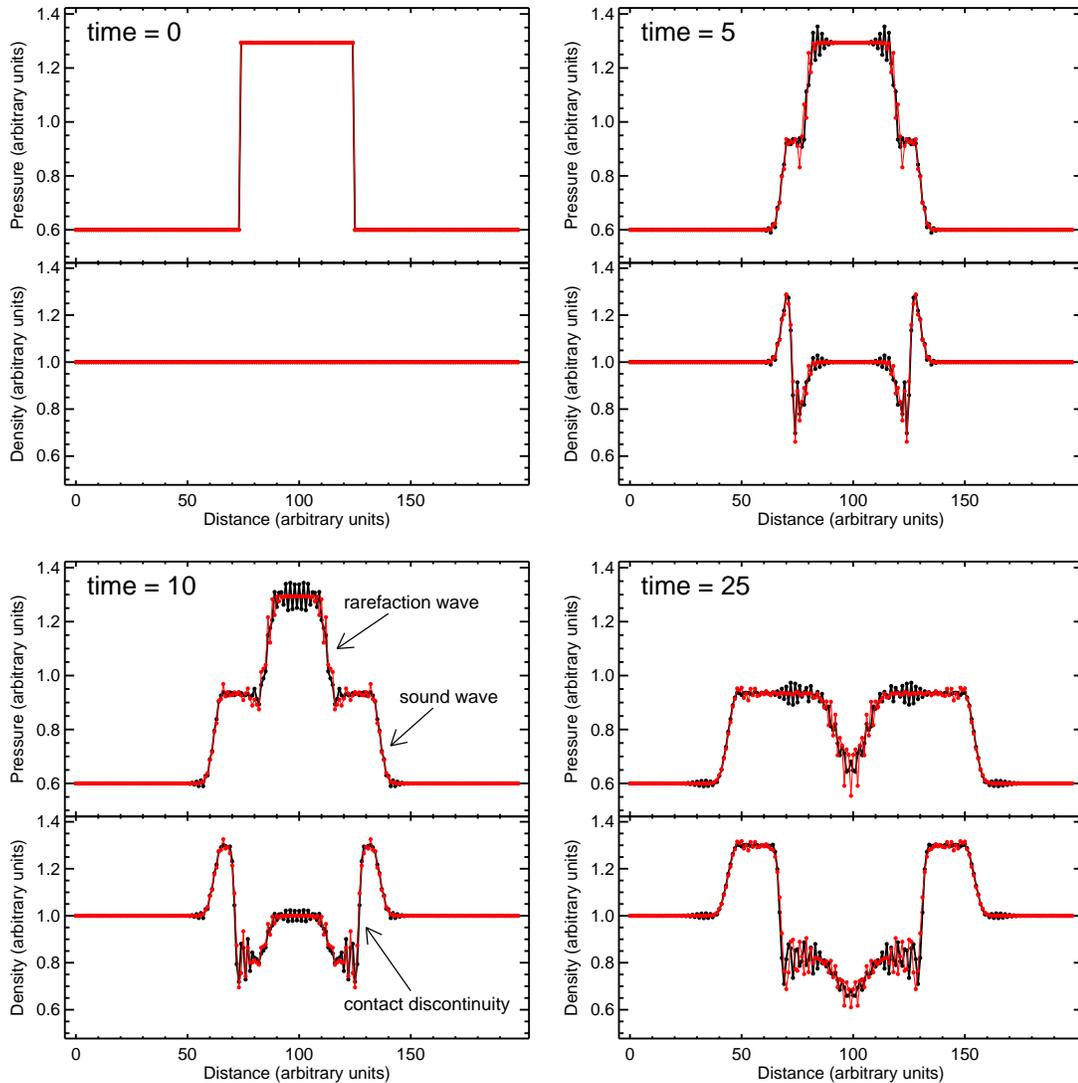


Figure 4.1: Four snapshots of the numerical evolution of a sound wave in a one-dimensional adiabatic gas made using FRIENDLY, from an initial condition in which a high pressure gas is allowed to interact with initially low pressure material of the same density. The red lines represent our code’s second order computation of the problem, and the black lines represent the fourth order computation. Characteristic features visible in these figures include sound waves, contact discontinuities and rarefaction waves.

4.2. There, a similar initial condition as in Figure 4.1 is set up as a rectangular region in two dimensions, thereby sending off a two-dimensional sound wave. In the simulation in Figure 4.2, we have chosen to use periodic boundary conditions, which allows the sound wave to interact with itself multiple times throughout the simulation run. We therefore see that the multidimensional capabilities of FRIENDLY appear to work as expected, although we have not yet tested these as thoroughly as we have tested the one-dimensional cases.

## 4.5 Second Order Method of Lines in Comparison to More Specialized Alternatives

Our second order, method of lines Runge-Kutta scheme proceeds in two steps: the first integrates over a half time step  $\Delta t/2$  (giving a first order result), and the second uses the spatial derivatives calculated *at the half step* to go back and integrate over the full step  $\Delta t$  (giving a second order result).

This method, in fact, is extremely similar to other, more specialized second order methods that have been developed, such as the two-step Lax-Wendroff and “staggered leapfrog” schemes. The close relationship between these methods and the method of lines do not generally appear to have been noted in the literature,<sup>3</sup> so we discuss them here in order to help us decide whether it is worth replacing or augmenting our second order scheme with one of the more specialized versions.

---

<sup>3</sup>However, Hirsch (1988) does note in passing the relationship between the staggered leapfrog method and the “central schemes with independent time integration” that are closely related to the method of lines.

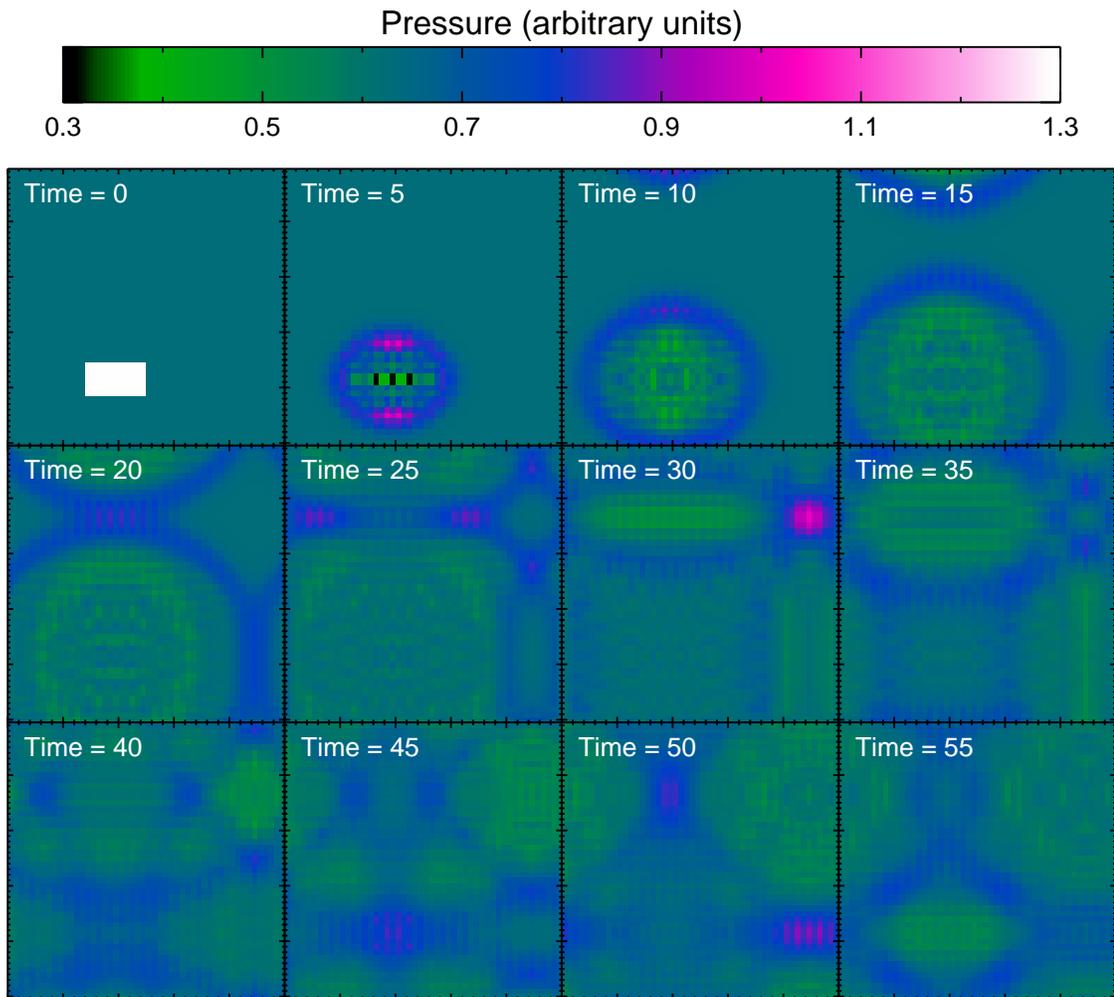


Figure 4.2: Twelve snapshots of the numerical evolution of a sound wave in a two-dimensional adiabatic gas, made using FRIENDLY. The initial region of elevated pressure spreads out in a sound wave which interacts with itself multiple times on the grid, due to the periodic boundary conditions being used.

### 4.5.1 The Two-Step Lax-Wendroff Scheme

The two-step Lax-Wendroff scheme (Richtmyer & Morton 1967) proceeds exactly like the method of lines, only before each numerical integration, the conserved quantities are interpolated halfway between the existing grid points; the first Runge-Kutta step ( $\Delta t/2$ ) is taken from the interpolated points, while the second step ( $\Delta t$ ) is taken from the original points.<sup>4</sup> The end result is that the spatial derivatives are effectively calculated on a “tighter” grid (over a distance  $\sim \Delta x$  rather than  $\sim 2\Delta x$ ), which should, when compared to the method of lines, allow us to compute more accurate spatial derivatives without having to increase our grid resolution, and at the expense of only a small increase in the number of computational operations required (due to the extra interpolation step).

This improvement appears to primarily be useful in cases where we are looking to push the time step  $\Delta t$  smaller than the Courant value for reasons of physical interest (in other words, if we want to output data about our simulation at a faster rate). In these cases, the discussion accompanying equation (4.2) shows that the error term associated with the spatial derivative becomes dominant, and we therefore might be motivated to decrease the size of this term as much as we can. Alternatively, the two-step Lax-Wendroff method could allow us to save computational operations by cutting down the number of points in our grid (i.e., decreasing the resolution); this will allow us to run the code faster without sacrificing much accuracy, although it certainly will sacrifice *information* because it

---

<sup>4</sup>There are also a family of schemes known as Lerat-Peyret that involve generalizing this concept by relaxing the requirement that the spatial interpolation and initial time step be “halfway”; instead, the spatial interpolation can take place at an arbitrary location between the grid points, and the initial time step can be an arbitrary amount less than  $\Delta t$  (Hirsch 1988). These schemes are also referred to as “predictor-corrector” schemes (in other words, the first step is the predictor, and the second step is the corrector, much like the concept behind the Runge-Kutta method).

limits the spatial scales that our simulation can resolve down to.

In summary, the two-step Lax-Wendroff scheme appears most useful in cases where we care much more about achieving the highest possible time resolution than we do about achieving the highest possible space resolution. Given that our goal in the current work is to study the evolution of GRS 1915+105 on extremely long timescales (compared to the typical timescales in the system; e.g., the orbital time), it does not appear that the two-step Lax-Wendroff method is worth implementing at this time.<sup>5</sup>

### 4.5.2 The Staggered Leapfrog Scheme

The staggered leapfrog scheme (e.g., Press et al. 1992) is also very similar to the second order method of lines, but in this case, instead of calculating the initial  $\Delta t/2$  step at first order accuracy and throwing it away, the special symmetry of the second order Runge-Kutta method (i.e., where the intermediate step  $\Delta t/2$  is exactly halfway along the full step) is taken advantage of; the “intermediate step” is not thrown away, but rather treated exactly the same as the full step. In other words, *every*  $\Delta t/2$  timestep that the program takes is calculated using the spatial derivatives from the previous time step to advance the solution from *two* previous

---

<sup>5</sup>A more general concern is that the two-step Lax-Wendroff scheme appears to be much less flexible than the method of lines. We have found that the interpolation required by the two-step Lax-Wendroff scheme would create significant complexity in our computer code, particularly in the case of applying boundary conditions in multiple dimensions (see Hirsch 1988 for a discussion of the overall application of the two-step Lax-Wendroff scheme in multiple dimensions). Also, the two-step Lax-Wendroff method is usually discussed in the context of conservative equations on a regular grid, and it does not appear easy to extend this method to more complicated situations. Because the grid at each half time step is “shifted” with respect to the original grid, we lose the ability to do a *three* point calculation of spatial derivatives (i.e., those that use the two neighboring points *plus the point itself*), which is necessary to achieve second order accuracy on an irregular spatial grid and/or for equations involving second derivatives.

time steps. The distinction between a “half step” and “full step” is therefore completely lost, and the result is that staggered leapfrog with  $\Delta t/2$  requires the same number of computational operations as the method of lines with  $\Delta t$ .

For our particular problem, we are trying to keep the time step as high as possible (as discussed above), so getting “free” information at  $\Delta t/2$  is not particularly appealing; on the other hand, we could easily use the more efficient computation provided by the staggered leapfrog scheme to decrease the number of computational operations required for the same  $\Delta t$  and thereby speed up our program execution. It therefore appears to be a competitive method that we should consider adding as an option to FRIENDLY in the future.

On the other hand, the subtle difference between staggered leapfrog and method of lines does lead to one potential problem: a “chess board” or “mesh drifting instability” (Press et al. 1992). Specifically, a point in the simulation at space and time location  $(x_j, t_{n+1})$  is only calculated using information at  $(x_{j\pm 1}, t_n)$  and  $(x_j, t_{n-1})$ ; a space vs. time plot of the numerical scheme can therefore be colored like a chess board, where each black or white square is only influenced by (and only has influence on) other squares of the same color (Press et al. 1992). This problem does not arise in the second order method of lines, because the first order  $\Delta t/2$  step ensures that each point in the grid is at least partially influenced by *all* of its neighbors (including itself) at the previous time step. On the other hand, Press et al. (1992) suggest that the mesh drifting instability can be fixed through the use of artificial viscosity, which allows the “chess board squares” to communicate with their neighbors.

## 4.6 Calculating Spatial Derivatives on Irregular Grids to Arbitrary Orders of Accuracy

Though the Runge-Kutta methods are well known, methods for calculating spatial derivatives to high order are less commonly discussed. Here, we provide a simple method that can be used to calculate spatial derivatives to any arbitrary order of accuracy on an irregular grid; this method is implemented in FRIENDLY. Surprisingly, this method does not appear to have been previously presented in the literature.

In order to convert a regular grid with spacing  $\Delta x$  into an arbitrary irregular one, we can imagine “stretching” or “squeezing” each grid interval—that is, multiplying each  $\Delta x$  by a dimensionless number whose value depends on location in the grid, with the end result being that each grid point  $x_j$  is associated with a spacing  $\Delta x_j$ . From a notational standpoint, a convenient way to refer to our new grid is to create a label for the spacing between *any* two grid points  $x_j$  and  $x_n$ , not just adjacent ones. We can write this spacing as  $\Delta x_{jn}$  and define it to be equal to  $x_n - x_j$  (in the case of a regular grid, we have  $\Delta x_{jn} = (n - j) \Delta x$ ).

In order to calculate the spatial derivatives of a known function  $f(x)$  at  $x_j$ , we first Taylor expand the function about this point. If we extend the expansion to the point  $x_n$ , we can write the Taylor expansion as:

$$f_n - f_j = \sum_{k=1}^{\infty} D_{jk} (\Delta x_{jn})^k, \quad (4.3)$$

where  $D_{jk} \equiv (1/k!) (\partial^k f / \partial x^k)_j$ . The only unknowns in this equation are the  $D_{jk}$ 's, which are directly related to the spatial derivatives at  $x_j$  that we are trying to calculate.

If we write the above equation for several different  $x_n$ , we can think of it as a

matrix equation that takes a  $k$ -element vector  $D_{jk}$  as input and spits back an  $n$ -element vector  $f_n - f_j$  as output. Inverting this matrix therefore allows us to solve for the spatial derivatives at our grid point of interest (the  $D_{jk}$ 's) as a simple linear combination of the function values at other grid points (the  $f_n$ 's). The problem we face in doing this, however, is that the sum over  $k$  extends to infinity—in other words, the matrix that we need to invert has an infinite number of columns! In order to “solve” the equation, therefore, we must truncate the sum in the above equation after a particular number of terms (equal to the number of  $f_n$ 's we want to use) and live with the resulting inaccuracy in our solution. For example, using two terms at points  $x_n$  and  $x_m$  gives:

$$\begin{bmatrix} f_n - f_j \\ f_m - f_j \end{bmatrix} = \begin{bmatrix} \Delta x_{jn} & (\Delta x_{jn})^2 \\ \Delta x_{jm} & (\Delta x_{jm})^2 \end{bmatrix} \begin{bmatrix} D_{j1} \\ D_{j2} \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}, \quad (4.4)$$

which can be inverted to solve for  $D_{j1}$  and  $D_{j2}$  and therefore the first and second derivatives at  $x_j$ . The error terms  $\epsilon_i$  in the above equation are of order  $O(\Delta x^3)$ .

The solution of the above matrix gives a relatively trivial result, but the power of the method is that it can easily be extended to any size matrix and therefore any arbitrary order of accuracy. Solving these matrices using a symbolic mathematics package leads to a simple algebraic formula that converts a particular set of grid spacings  $\Delta x_{jn}$  into coefficients that need to multiply each  $f_n$  in order to calculate any necessary derivative at  $x_j$  to any necessary order of accuracy.

As mentioned, the formulation of this method as a matrix problem does not appear to have been noticed in the literature; for example, Schiesser (1991) goes through a long series of mathematical manipulations needed to solve for the above coefficients in particular cases, apparently without realizing that the mathematical manipulations in question are simply those needed to solve a matrix equation! The

best method we have found presented in the literature to calculate coefficients for spatial derivatives is that due to Fornberg (1988), who presented a general algorithm that can be used to calculate the coefficients based on fitting a polynomial to the adjacent grid points. This method was specifically designed for situations where speed is important (such as in adaptive grid methods, where the grid is constantly being resampled throughout an integration), and we have not investigated whether our method is faster. However, the fact that we can derive a simple analytic formula for any order of accuracy certainly suggests that our method is a more straightforward one.

## 4.7 Conservative Integration Schemes for Irregular Grids

Second order conservative integration schemes for regular grids have been discussed by many authors (e.g., Hirsch 1988). In this section, we expand the discussion to irregular grids and arbitrary orders of accuracy and derive the method that FRIENDLY uses to provide a fully conservative numerical scheme if requested by the user at runtime. Along the way, we derive formulas that will be useful in implementing specific boundary conditions in the fourth order (or higher) case.

A numerical integration scheme is “conservative” whenever it properly conserves quantities that are conserved in the underlying, continuous physical model. This means that the total amount of the quantity contained within the computational grid must only change due to the flux of material through the boundary of the grid, or, for periodic boundary conditions, must not change at all.

If a quantity  $u$  is locally conserved on a discrete spatial grid, we can write the

following equation at any grid cell  $j$ :

$$\left(\frac{\partial u}{\partial t}\right)_j = \dot{u}_{j,in} - \dot{u}_{j,out}, \quad (4.5)$$

where  $\dot{u}_{j,in}$  is the rate at which  $u$  enters the lower boundary of the cell and  $\dot{u}_{j,out}$  is the rate at which it exits the upper boundary (this equation could easily be generalized to multiple dimensions, of course). Global conservation requires that the rate at which  $u$  leaves the upper boundary of cell  $j$  equal the rate at which it enters the lower boundary of cell  $j + 1$ , or, mathematically,  $\dot{u}_{j,out} = \dot{u}_{j+1,in}$ . This requirement suggests a change in notation; we rewrite  $\dot{u}_{j,in}$  as  $\dot{u}_{j-1/2}$  and  $\dot{u}_{j,out}$  as  $\dot{u}_{j+1/2}$ , to indicate that the midcell values of  $\dot{u}$  depend on the location on the grid at which they are being evaluated, not on which neighboring cell (lower or upper) they are being evaluated for.

We can convert this equation to one that can be compared with our numerical scheme by defining  $\rho$  as the density of  $u$  (i.e.,  $u$  divided by the volume of the cell) and  $F$  as the flux of  $u$ , so that, for example,  $F_{j+1/2}$  is the flux of  $u$  through the upper cell boundary (often referred to as the midcell flux or numerical flux). Using these definitions, we obtain:

$$\left(\frac{\partial \rho}{\partial t}\right)_j = \frac{F_{j-1/2} - F_{j+1/2}}{\Delta x_j}, \quad (4.6)$$

where  $\Delta x_j$  is the width of the cell in the  $x$  direction. If we define each cell boundary to be located exactly halfway between adjacent locations  $x_j$  at which the physical quantities are evaluated, then  $\Delta x_j = (x_{j+1} - x_{j-1})/2$  for a grid with arbitrary spacing.

The above equation shows the relationship that  $\partial \rho / \partial t$  must satisfy for a conservative integration scheme. For  $u$  to be conserved *physically*, however, we require

$\partial\rho/\partial t = -\partial F/\partial x$ . Equating these values, we obtain:

$$\left(\frac{\partial F}{\partial x}\right)_j = \frac{F_{j+1/2} - F_{j-1/2}}{\Delta x_j}. \quad (4.7)$$

A conservative numerical approximation for the derivative is one that, when written in the above form, leads to a formula for  $F_{j\pm 1/2}$  that is globally conservative (in the manner discussed above for  $u_{j\pm 1/2}$ ). Another important requirement, of course, is that the formulas for  $F_{j\pm 1/2}$  (the midcell fluxes on the boundaries) should have some reasonable relationship to  $F_j$  (i.e., the flux calculated within the adjacent cell). This is effectively saying that the integration scheme is a “good” one that accurately represents the underlying physics. One way to define this condition more rigorously is to say that when all the  $F_k$ ’s that go into calculating  $F_{j\pm 1/2}$  have the same value,  $F_{j\pm 1/2}$  should also have that value (Hirsch 1988). If a numerical scheme calculates  $F_{j\pm 1/2}$  as a linear combination of nearby  $F_k$ ’s with coefficients  $a_k$  (as is common), then the condition reduces to  $\sum a_k = 1$ .

If a numerical approximation for the derivative can be written in the form of equation (4.7) and meet both the above conditions (“global conservation” and “physical accuracy”), then the integration scheme in question is conservative. Furthermore, such a scheme will also handle conservative equations with source terms correctly; the rate at which the total amount of  $u$  on the grid changes will be the sum of the source term contributions within each cell, plus any fluxes on the outside boundary of the grid.

#### 4.7.1 Nth Order Conservative Integration Schemes

To test whether our integration scheme discussed in §4.6 is conservative, we substitute the expression derived there for  $(\partial F/\partial x)_j$  into equation (4.7) and see if we

can obtain consistent solutions for  $F_{j\pm 1/2}$ . The expression for  $(\partial F/\partial x)_j$  can be found, for example, from equation (4.4), or a similar equation for higher orders of accuracy. We write:

$$(\partial F/\partial x)_j = \sum_n d_{jn} (F_{j+n} - F_j), \quad (4.8)$$

where the  $d_{jn}$  are the coefficients (with units  $\sim 1/\Delta x$ ) in the linear combination of  $F$ 's that are used to approximate the derivative; for each  $j$ , the  $d_{jn}$  are obtained by taking the first row of the inverse of a matrix like that found on the right hand side of equation (4.4). Note that in this section, we are also redefining  $n$  to be the *relative* location of each neighbor point of  $j$  that is used in calculating the derivative, rather than the absolute location. For the scheme to be conservative, we now must have:

$$\frac{F_{j+1/2} - F_{j-1/2}}{\Delta x_j} = \sum_n d_{jn} (F_{j+n} - F_j). \quad (4.9)$$

We can now see that obtaining a globally conservative solution for  $F_{j\pm 1/2}$  is a difficult problem in the general case, because the solution for each numerical flux  $F_{j+1/2}$  is coupled to the solution for its neighbor  $F_{j-1/2}$ , and therefore, ultimately, all the numerical fluxes on the spatial grid are coupled together through a  $\sim J \times J$  matrix equation (where  $J$  is the number of points on the spatial grid that the  $j$  index runs over). This matrix is not, in the general case, invertible, which means that it is possible to produce grid spacings for which the integration scheme we have defined above is not conservative.

To find the conditions under which the integration scheme *is* conservative, it is instructive to not look at the matrix equation directly, but rather to realize that if *one* of the midcell fluxes on the grid is known (say  $F_{j-1/2}$ ), then it is possible to

iteratively solve for all others by rewriting equation (4.9) as:

$$F_{j+1/2} = F_{j-1/2} + \Delta x_j \sum_n d_{jn} (F_{j+n} - F_j). \quad (4.10)$$

This gives  $F_{j+1/2}$ , which can be plugged into a similar equation at  $j + 1$  to solve for  $F_{j+3/2}$ , and so on down the line.

As an example, we consider the case of periodic boundary conditions, which impose a consistency requirement on the above set of equations whereby the initial value of  $F_{j-1/2}$  that starts off the solution chain must be equal to the value derived once we have wrapped the chain around the grid and approached  $x_{j-1/2}$  from the opposite direction (this is effectively the global conservation requirement). In this case, repeated application of equation (4.10) leads to the following requirement:

$$\sum_j \Delta x_j \sum_n d_{jn} (F_{j+n} - F_j) = 0, \quad (4.11)$$

where the sum over  $j$  runs over the entire spatial grid.

We have found that a sufficient condition for this equation to be satisfied is that the “dimensionless derivative coefficients”  $\Delta x_j d_{jn}$  do not depend on the location on the grid (i.e., do not depend on  $j$ ). In that case, these coefficients can be pulled out of the sum over  $j$ , leaving

$$\sum_n \Delta x_j d_{jn} \left[ \sum_j (F_{j+n} - F_j) \right] = 0, \quad (4.12)$$

which is always true for periodic boundary conditions because both terms within the square brackets are equal to the sum of  $F_j$  over all points on the grid (the shift  $n$  is arbitrary and just indicates the location on the grid at which the sum is “started”).

On a periodic grid, the only realistic way to force  $\Delta x_j d_{jn}$  to not depend on  $j$  is to have the grid spacing be uniform, which is usually the case anyway (since

periodic grids do not tend to have any “special” locations where higher resolution is required). On non-periodic grids, however, there is another type of useful grid spacing that has this property: any grid in which each cell width is a constant multiple of the previous cell width,  $\Delta x_{j+1} = g\Delta x_j$ , where  $g$  does not depend on  $j$ . (A special example of this kind of spacing is logarithmic spacing,  $\Delta x_j \propto x_j$ .)

In fact, we can easily extend the above analysis to non-periodic grids. Suppose we wish to specify particular midcell fluxes  $F_{left}$  and  $F_{right}$  just outside the boundaries of a non-periodic grid (i.e., at  $x_{1/2}$  and  $x_{J+1/2}$  if the grid runs from  $j = 1$  to  $J$ ). Equation (4.11) therefore has its right hand side replaced with  $F_{right} - F_{left}$ , which again cannot be satisfied for an arbitrary integration scheme. However, if we once again assume that  $\Delta x_j d_{jn}$  does not depend on  $j$ , the equivalent of equation (4.12) no longer has *all* of its terms cancel out, but rather has some neighboring points remain near both edges of the grid. For a second order integration method, for example (where  $n$  ranges from -1 to 1), we obtain:

$$F_{right} - F_{left} = a_{-1}(F_0 - F_J) + a_1(F_{J+1} - F_1), \quad (4.13)$$

while a fourth order method has additional terms for  $a_{-2}$  and  $a_2$ . (Here, we are using the simplified notation  $a_n = \Delta x_j d_{jn}$ , because we have assumed that this term does not depend on  $j$ .) Note that  $F_0$  and  $F_J$  are “ghost cell” fluxes that occur outside of the computational grid.

This equation can be used to solve for  $F_{left} = a_1 F_1 - a_{-1} F_0$  and  $F_{right} = a_1 F_{J+1} - a_{-1} F_J$  separately (under the physically reasonable assumption that  $F_{left}$  and  $F_{right}$  should only depend on the grid points near them), and therefore, since we are assuming that the  $a_n$  are the same everywhere on the grid, it can also be

used to obtain the following general formula:

$$F_{j+1/2} = a_1 F_{j+1} - a_{-1} F_j. \quad (4.14)$$

It is easy to verify, however, that this formula for the midcell flux  $F_{j+1/2}$  is not automatically physically accurate (the sum of the coefficients multiplying each  $F$  does not, in general, equal 1). Only uniform grid spacing (where  $a_{-1} = -0.5$  and  $a_1 = 0.5$ ) is physically accurate, and a similar result holds for the fourth-order case. Assuming uniform spacing, we therefore obtain

$$F_{j+1/2} = \frac{F_j + F_{j+1}}{2} \quad (4.15)$$

for second order (a trivial result), and

$$F_{j+1/2} = \frac{-F_{j-1} + 7F_j + 7F_{j+1} - F_{j+2}}{12} \quad (4.16)$$

for fourth order (a less trivial result), where  $F_{left}$  can be obtained from these equations by setting  $j = 0$ , and  $F_{right}$  can be obtained by setting  $j = J$ .

## 4.7.2 Applications to User-Specified Boundary Conditions

What was the point of doing the above exercise if all it did was strongly suggest that uniform spacing is the only way to get a conservative integration scheme? One of the main benefits is that it provides a method to numerically set desired fluxes on the grid boundaries. Specifically, when equations (4.15) or (4.16) are evaluated for  $F_{left}$  or  $F_{right}$ , the formula will contain terms involving the flux at “ghost cells” off the edge of the grid (for example,  $F_0$  and  $F_J$ , as noted above). If the formula is arranged to solve for these ghost cell values, we immediately obtain a method for numerically specifying *any* desired value of  $F_{left}$  or  $F_{right}$  entering the boundary of a globally conservative uniform grid, to second or fourth order accuracy: Simply

set the ghost cell fluxes to the indicated values whenever spatial derivatives are being taken.

Specifically, for second order we obtain  $F_0 = 2F_{left} - F_1$  on the left boundary, and a similar equation for the right boundary (once again, a trivial result). For fourth order, however, equation (4.16) has two ghost cells contained within it, and therefore we cannot uniquely determine the value of each. The apportioning of flux between each ghost cell is *not* completely arbitrary; the ghost cell located two spaces away from the boundary is only used to compute the derivative at the closest grid point to the boundary, but the ghost cell located only *one* space away from the boundary is also used to compute the derivative at the second-closest grid point to the boundary. We have chosen to apportion flux between these ghost cells by requiring that  $F_{3/2} = (F_1 + F_2)/2$  (and a similar equation for the right boundary); in other words, we require that the midcell flux between the first two grid points contains no dependence on the ghost cells, and instead is given by the appropriate second order formula using only points interior to the grid. This is physically reasonable, since the ghost cell is not supposed to be approximating any “real” point off the edge of the grid; its sole function is to specify the flux entering the computational grid at its extreme edge, and therefore its value should not affect anything except the edge of the grid. Adding this constraint, we obtain the following solution for the left boundary:

$$F_0 = F_1 + F_2 - F_3 \tag{4.17}$$

$$F_{-1} = -12F_{left} + 14F_1 + 6F_2 - 7F_3, \tag{4.18}$$

with an analogous solution on the right boundary. Although the specific apportioning does not appear to have a huge effect on the integration, we have found that

in some cases, the apportioning we have chosen produces a slightly more stable integration scheme than other possible choices (e.g., requiring  $F_{-1} = F_0$ ).

### 4.7.3 Conservation on Non-Uniform Grids

Finally, what should we do about the case of a non-uniform grid? Morinishi et al. (1998) and Vasilyev (2000) have considered the question of fourth order accurate schemes in the context of staggered mesh grids (where velocities and other vector quantities are defined on the cell boundaries, and scalar quantities are defined at the cell centers) and have suggested, as we do, that it is impossible to create a conservative integration scheme while preserving fourth order accuracy. Nonetheless, these authors found that it is possible to create fourth order accurate schemes that do a relatively good job conserving physical quantities, although not doing so to the level of machine accuracy. Indeed, we have found in our case (where we are using a regular, *non-staggered* mesh) that our second and fourth order methods conserve quantities relatively well, as long as the grid is relatively close to uniform (e.g., in the adiabatic gas case, quantities integrated over the grid appear to be within  $\sim 0.5\%$  of the expected value).

In our code, we also provide an option to *force* the integration scheme to be conservative, at the expense of a small loss of accuracy from the formal second or fourth order specification. Essentially, this method *assumes* that the midcell fluxes are given by the values in equations (4.15) and (4.16), with the spatial derivative of  $F$  given by a simple difference of the midcell fluxes (i.e., “flux in” minus “flux out”) divided by the cell width  $\Delta x_j$ , without regard to where the cell boundaries are located. Another, equivalent way to look at this method is to say that the grid spacing is assumed to be “locally uniform”—that is, constant and equal to  $\Delta x_j$ —

within each group of  $n$  points used to calculate the derivative at  $x_j$ , but allowed to be different at different grid points  $j$ . This method should work quite well if the grid spacing is close to uniform, so that adjacent points  $n$  have relatively similar spacings; note also that it is very similar to the methods described by Verstappen & Veldman (1997) and Morinishi et al. (1998) for nonuniform, staggered mesh grids.

## 4.8 Boundary Conditions

FRIENDLY currently provides a wide variety of boundary conditions that can be chosen by the user at runtime. In addition to the ability to specify a particular boundary flux (discussed in §4.7.2), the user can choose periodic boundary conditions, reflective boundary conditions, “lopsided” boundary conditions (in which the derivative on the boundary is computed using points within the grid only), fixed boundary conditions, and two methods of extrapolation using any number of interior points on the grid (extrapolation via fitting a polynomial to the interior points, or extrapolation using the average of the interior points). Furthermore, through the use of FORTRAN 95 array notation, these boundary conditions can easily be applied separately on any boundary in any dimension, and to any *number* of points on each boundary (for example, the fourth order method requires two “ghost points” past the edge of the grid; the user can choose to specify the first point using a particular method, and then use a “lopsided” derivative formula in place of specifying the other). Thus, FRIENDLY is very flexible in being able to work with a wide variety of different physical conditions.<sup>6</sup>

---

<sup>6</sup>One current limitation of the code, however, is that each equation in the physical model must use the same boundary conditions (on a particular face of the computational grid) as all the other equations.

As an example of one of the most important boundary conditions, in which the user specifies a particular flux of the relevant physical quantities on the boundary, we show in Figure 4.3 a comparison between two identical sound waves, one of which begins on the grid, and the other of which arises entirely from a user-specified boundary flux of mass, momentum and energy. As can be seen, these two sound waves have the same properties and also travel at the same physical speed regardless of their location on the irregular grid used here, as expected.

Additionally, we provide in FRIENDLY a method to enforce certain conditions on the boundary fluxes but still allow these boundary fluxes to be calculated using one of the normal boundary conditions. An example of a condition that might be imposed is to mimic a black hole by requiring that the flux of mass, momentum and energy at the inner boundary always be negative. This can be simply enforced by adding or subtracting a particular amount from the otherwise-specified ghost cell fluxes. Equation (4.18) and the surrounding discussion shows that in order to add a flux  $F_{add}$  to the boundary flux in the fourth order case (i.e., to replace the boundary flux  $F_{left}$  with  $F_{left} + F_{add}$ ), we must subtract  $12F_{add}$  from the ghost cell located two spaces away from the boundary; in the second order case, meanwhile, we must add  $2F_{add}$  to the ghost cell (once again, the second order case is a trivial result, whereas the fourth order case is not).

## 4.9 Conclusions

In this chapter, we have described the development and some of the unique features of FRIENDLY, a new code for integrating time-dependent equations that is designed to be user-friendly, flexible and (eventually) publicly available; it achieves the first two goals primarily by using a “method of lines” approach to the numeri-

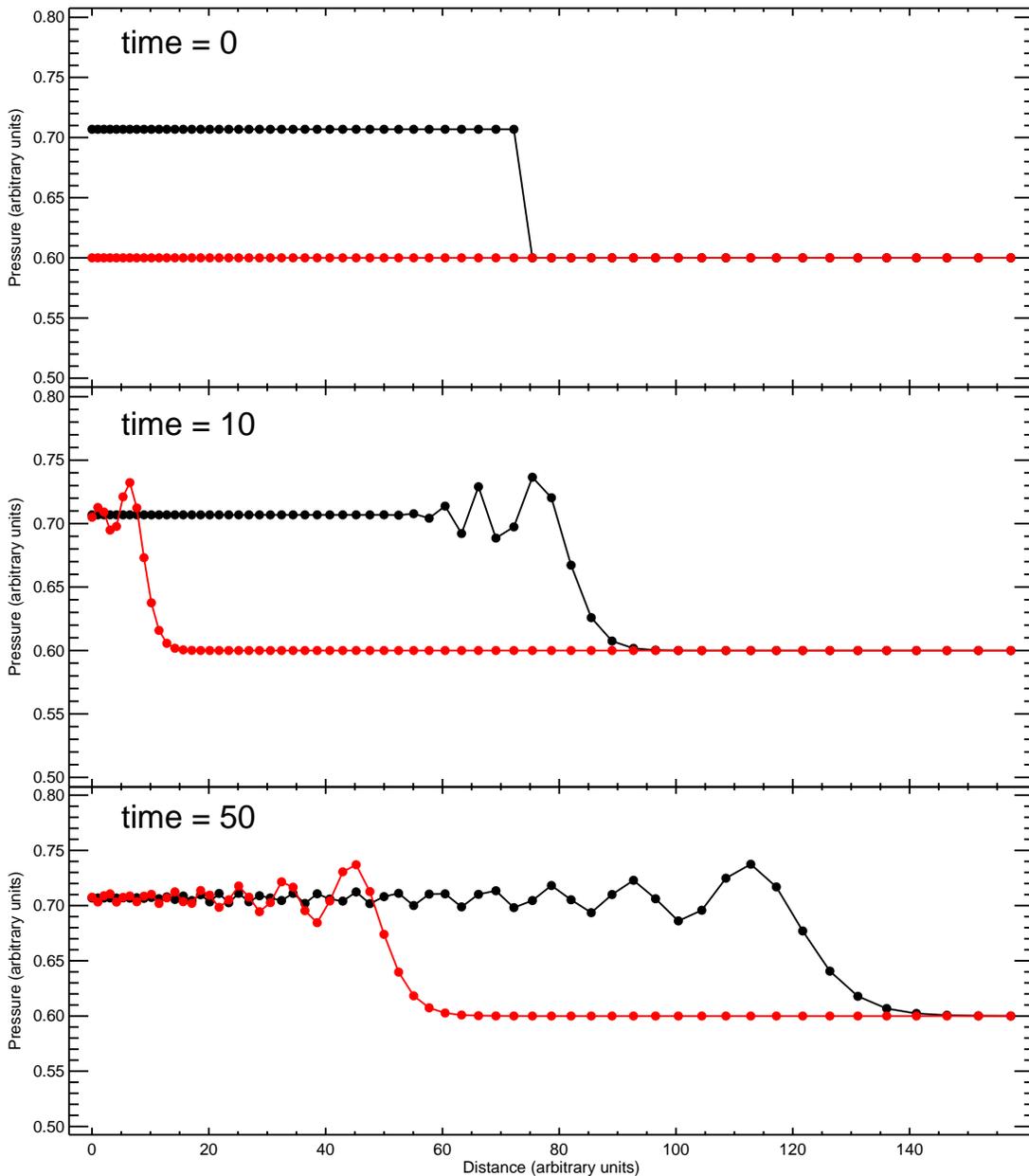


Figure 4.3: A comparison of the evolution of two identical sound waves, one of which begins on the spatial grid and the other of which is initially imposed on the grid boundary, made using FRIENDLY. As expected, the two sound waves have virtually the same properties once they are both on the grid (although note that the oscillations associated with the initial discontinuity are larger for the sound wave in the region of the irregular grid with larger spacing; this is because the oscillations are a grid phenomenon, not a physical one). Also, the evolution is performed on an irregular grid, but, as should be the case, the sound waves maintain a constant physical speed (and therefore a constant physical separation) throughout the run.

cal integration that allows for an almost complete separation between the physics and the underlying computational algorithms, as well as by taking advantage of powerful FORTRAN 95 array manipulation features. In this chapter, we have described some results of the initial implementation of FRIENDLY, as well as some of the theoretical work that needed to be done to allow FRIENDLY to achieve fourth-order accurate spatial derivatives (and/or conservation to the level of machine accuracy) for an arbitrarily-spaced irregular grid; the procedures described here could, in theory, be easily extended to higher orders of accuracy as well. It is hoped that the code will soon be ready to make available to the larger community.

# Chapter 5

## Simulations of Accretion Disk Evolution in GRS 1915+105

### 5.1 Summary

In this chapter, we use the numerical code described in Chapter 4 to begin a preliminary investigation of accretion disk and jet evolution in GRS 1915+105. We focus on allowing  $\alpha$ , the Shakura & Sunyaev (1973) parameter describing the strength of turbulence, to change rapidly in time, and we describe how this change can lead to an outburst in the disk. Many models for jets involve large scale magnetic fields that should inhibit the magnetorotational instability believed to drive turbulence in the disk, and therefore these models should naturally lead to a rapid increase in  $\alpha$  when the magnetic field is ejected from the system. We find that in our simulations, the timescale and luminosity for the outburst can be controlled by the manner in which  $\alpha$  is allowed to change, which is a promising start to our goal of explaining the complex, variable behavior of GRS 1915+105 and other objects.

### 5.2 Introduction

In Chapter 3, we derived a wide variety of dimensionless parameters that might be expected to depend on the presence of a jet or other aspects of the vertical structure

of an accretion disk. Here, as an illustration of the way in which these kinds of parameters can affect the observed light curves of GRS 1915+105 and other black hole candidates, we present numerical simulations that show how rapid changes in the  $\alpha$  parameter of Shakura & Sunyaev (1973) can lead to interesting behavior in the accretion disk.

There are two main reasons why we choose to focus on  $\alpha$  initially. First, the value of  $\alpha$  has a very significant effect on both the evolution timescales and the shape of the local energy balance curve of an accretion disk. In particular, the thermal and viscous timescales are both inversely proportional to  $\alpha$ , while the turbulent heating rate is directly proportional to  $\alpha$ , in contrast to the radiative cooling rate which has no dependence. Thus, a change in  $\alpha$  could lead to a change in the outburst timescale of the disk, as well as a change in the detailed track the disk takes as it traverses the energy balance diagram.

Second, many theoretical models suggest, usually indirectly, that the value of  $\alpha$  should be different when a jet is and is not present. The basic picture that many models invoke for a steady jet is one in which a large-scale, ordered magnetic field is present that can channel matter and energy away from the accretion disk (e.g., Blandford & Payne 1982; Eikenberry & van Putten 2003; Livio et al. 2003; Tagger et al. 2004). Although the details of the magnetic field structure in each individual model are different, the common theme is clearly that a large-scale magnetic field is responsible for the jet. There is some debate about how strong these magnetic fields can grow via advection from the outer disk, but there is at least some theoretical work that suggests they can grow to be stronger than equipartition, at which point there is a large range of magnetic field values for which the magnetic structure would be stable (Stehle & Spruit 2001) and therefore amenable to the presence

of a steady jet. In particular, the magnetorotational instability that is thought to drive accretion during the thermal dominant state of an accretion disk is not generally expected to operate when the magnetic field reaches equipartition value or higher (Balbus & Hawley 1991).

In this spirit, Tagger et al. (2004) proposed a model for GRS 1915+105 in which the transition from the low state to the period of oscillations seen in this object's light curve (e.g., Chapter 2) involves the destruction of a large-scale magnetic field via a reconnection event, during which the magnetic field eventually becomes low enough that the magnetorotational instability can turn on. We suggest that the simplest way to approximate this removal of magnetic field is to allow  $\alpha$  to increase sharply over some region of the disk, thereby producing more efficient turbulent transport and energy dissipation which could be due to the magnetorotational instability suddenly becoming operable. However, we do recall from Chapter 3 that other dimensionless parameters in the disk might be expected to change as the large scale jet evolves or material is ejected into it; thus, the work presented in this chapter is only a preliminary investigation of the possible parameter space.

We note that an overwhelming number of possible scenarios for a variable value of  $\alpha$  have been suggested in the literature, usually where  $\alpha \equiv t_{r\phi}/P$  (as we have defined it in this thesis) has some dependence on other disk parameters. For example, dwarf novae outbursts have long been modeled using different values of  $\alpha$  on the upper and lower branches of the thermal stability curve in an attempt to better match observations (Smak 1984; Cannizzo 1993), with the time-variable values of  $\alpha$  sometimes even chosen based on a direct comparison to simulations of the magnetorotational instability in this regime (Menou 2000). The same basic idea has also been explored for the cool outer disks of X-ray binaries (e.g., Menou

et al. 2000; Lasota 2001) and specifically for the long term ( $\sim$  months) outbursts of GRS 1915+105 by Truss & Wynn (2004). Note that all these simulations involve a different limit cycle instability than the one discussed in this thesis, occurring on a much lower temperature portion of the energy balance curve, where the curve assumes an “S shape” due to changes in the opacity as hydrogen becomes partially ionized. Furthermore, a wide variety of disk models propose values of  $\alpha$  that depend on other parameters in the disk (e.g., Takeuchi & Mineshige 1997; Merloni 2003; Meier 2005), often depending on specific assumptions about the magnetic field.

The main difference between the idea in this work and previous ideas about  $\alpha$  is that we view a change in  $\alpha$  to be a possible *cause* of an outburst related to the ejection of a jet, rather than a consequence of the evolution of other disk parameters. Furthermore, we change  $\alpha$  rapidly in our models. In fact, we change it “instantaneously” over a wide region of the disk, which would appear to violate causality; however, as long as the true rate at which  $\alpha$  changes is much faster than other typical timescales in the system (which appears to be quite reasonable based on the results of this chapter, and also if the change in  $\alpha$  propagates at the Alfvén speed in a magnetically dominated disk), this is not problematic. We note that in order to change  $\alpha$  faster than the thermal timescale at a particular location in the disk (i.e., in order to have the local energy balance curve change faster than the disk can keep up with it), the “jet speed”  $v_{jet}$  that removes magnetic field from the disk must roughly obey  $v_{jet} \gg \alpha c_s$ , where  $c_s$  is the sound speed; this constraint is not very hard to meet. We also note that global observations of accretion disks show a hint of evidence for the jet timescales being faster than the disk ones, based on observations by Kalemci et al. (2006) of the power law and disk components

of the X-ray spectrum of the black hole candidate H1743-322 as it came out of an outburst.

### 5.3 Simulations

We used the FRIENDLY code described in Chapter 4 to simulate a one-dimensional accretion disk around a black hole of mass  $14M_{\odot}$ , appropriate for GRS 1915+105 (Greiner et al. 2001a; Harlaftis & Greiner 2004). We used a simpler set of time-dependent equations than those discussed in Chapter 3. Specifically, we assumed that orbits were always nearly circular ( $v_r \ll v_{\phi}$ ), so that the radial momentum equation can be ignored and the angular momentum and mass conservation equations can be combined (Pringle 1981); the appropriate energy equation to use in this case is the internal energy equation, rather than the full conservation of energy equation which contains kinetic energy terms. In general, the simplified equations we used should still be relatively accurate, even when advection dominates the energy transport in the disk (Mayer & Pringle 2006).

We originally tried to simulate the full set of equations, and that is our eventual goal. However, our attempts to do so using the Shakura & Sunyaev (1973) solution as an initial condition were unsuccessful, quickly becoming unstable in the inner part of the disk. In retrospect, it is possible to see why this is the case. Although the Shakura & Sunyaev (1973) solution assumes that the radial balance equation is insignificant, it does not assume that each individual *term* within that equation is negligible. For example, it is possible to derive the typical timescale for the radial momentum to change when acted upon by the pressure force in a Shakura & Sunyaev (1973) disk; we find that this timescale is  $\sim \alpha/\Omega_c$  (in other words, much faster than the dynamical timescale!). This somewhat surprising re-

sult indicates that our attempts to simulate disks in this manner were doomed to fail; in the initial condition for the Shakura & Sunyaev (1973) disk, the centrifugal and gravitational forces balance out, so that the pressure force drives material out of the inner disk much faster than the code is able to handle for any reasonable assumed Courant time. To proceed further, we would need to set the initial condition in our simulations using a numerical solution of the full set of equations in the steady state, as has been done by some studies in the literature (e.g., Watarai & Mineshige 2003); it is our goal to pursue this work in the future. However, other studies use the simplified assumptions that we have made (Janiuk et al. 2002; Mayer & Pringle 2006), and overall we should be able to compare our results to most one-dimensional disk models in the literature, as long as we are not interested in obtaining very high accuracy in the extreme inner disk.

All of our simulations begin with a gas pressure dominated Shakura & Sunyaev (1973) disk with a dimensionless accretion rate  $\dot{M} = 0.1$  (measured in units of the Eddington luminosity divided by the speed of light squared). For these parameters, the entire disk in the region of our simulation (typically  $7r_g$  out to a few hundred  $r_g$ ) is initially stable, although there is occasionally some initial settling down of the solution as the disk finds its true steady state (which may differ from the Shakura & Sunyaev 1973 solution if, for example, there is a mix of gas and radiation pressure).

For some combinations of parameters, the code fails during times when the disk is in an outburst state. In particular, there is a tendency for the surface mass density and internal energy to become negative as the disk exits the outburst phase; thus, we cannot generally follow the evolution of more than one successive disk outburst. We have not yet been able to fix this problem, which does not appear to be a problem with the FRIENDLY code but rather a more fundamental

issue which arises from the difficult problem of simulating a highly time-dependent inner boundary with significant outflow from the computational region into the black hole. In the present study, we find that the best results can usually be found by using a Paczyński & Wiita (1980) gravitational potential and the option in FRIENDLY that enforces the condition that no mass, momentum or energy are allowed to emerge from inside the inner disk boundary (see §4.8); these options generally, although not always, prevent instabilities from forming in the extreme inner computational regime.

Most of our simulations begin with a value of  $\alpha = 0.01$ , which is subsequently increased to  $\alpha = 0.1$  to mimic the effect of an ejection event that removes the large-scale magnetic field. In general, simulations of the magnetorotational instability suggest that  $\alpha \sim 0.1$  is an appropriate value when the instability is operating, although the fact that Maxwell stress dominates in these simulations means that  $\alpha$  roughly tracks the magnetic energy density as well (Hawley 2000; Blackman et al. 2006). This effect may also lead  $\alpha$  to have some dependence on the local value of the gas pressure (Sano et al. 2004). In the current work, we do not take into account these possible variations of  $\alpha$  while the magnetorotational instability is operating. When radiation pressure is important, meanwhile, the magnetorotational instability appears to respond similarly to the radiation pressure as it does the gas pressure, provided that the opacity is high enough for the gas and radiation to be well-coupled (Turner et al. 2003).

## 5.4 Results

Figure 5.1 shows the surface mass density evolution for a run in which the disk was initially held in a steady state with  $\alpha = 0.01$  for the first 100 seconds, and then

$\alpha$  was increased to a value of 0.1 in the inner part of the disk. We find that the change in  $\alpha$  drives the disk into outburst; the inner disk becomes nearly evacuated of matter, and a density wave propagates outward before eventually stalling when it reaches the region where  $\alpha$  is still  $\sim 0.01$ . The propagation of this density wave is shown more clearly in Figure 5.2.

Figure 5.3 shows the local energy balance curve, indicating that the inner regions of the disk do go into outburst and reach the advection dominated branch; the extra heating when  $\alpha$  is increased means that these regions of the disk are now in an unstable portion of parameter space. Figure 5.3 also shows that the outburst has a different character further out in the disk; there, the disk heats up when  $\alpha$  is increased but does not immediately become unstable; instead, the instability is triggered when the density wave arrives and the extra mass pushes this region of the disk “over the edge.”

The overall outburst light curve is shown in Figure 5.4 (black points). We see that the disk emits radiation at a super-Eddington rate as long as the transition wave is propagating through the disk and supplying mass to trigger new outbursts at large radii and simultaneously sustain the original outburst at the inner radii. When the transition wave stalls, however, the inner radii stop receiving mass, and eventually the outburst dies down as the disk is able to cool.

The transition wave feature seen here is not new; it was observed in some of the earliest one-dimensional simulations of accretion disk outbursts (e.g., Honma et al. 1991) and in many others works since then. In these simulations, outbursts in an  $\alpha = 0.1$  disk were driven not by a change in  $\alpha$ , but rather by an external accretion rate  $\dot{M}$  that was high enough so that the equilibrium state of the disk would put it in the unstable portion of the local “S-shaped” energy balance curve.

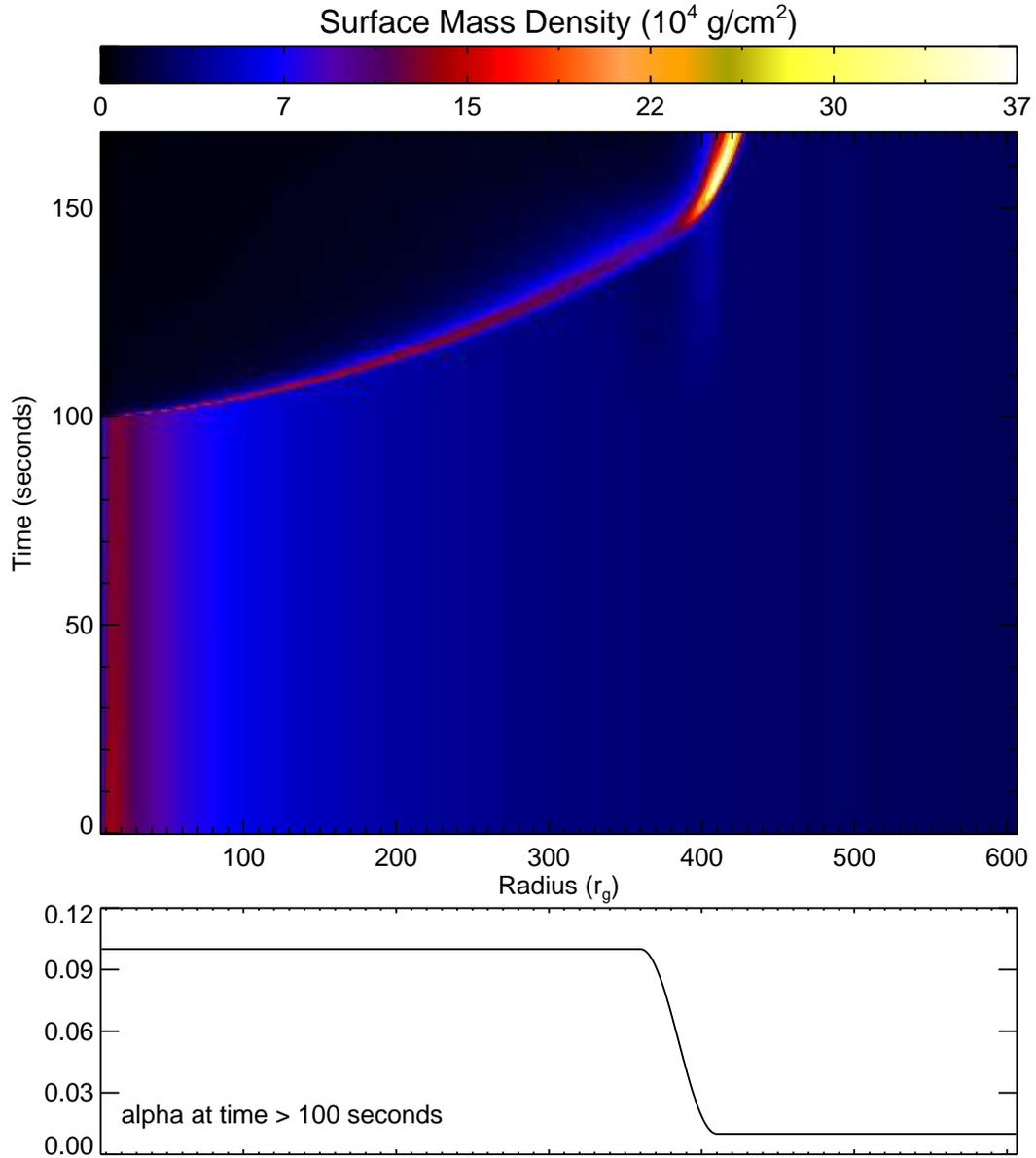


Figure 5.1: The surface mass density evolution for a run in which the value of  $\alpha$  is raised from 0.01 to 0.1 in the inner region of the disk ( $r \lesssim 400r_g$ ; bottom panel). A transition to the high state is initiated in the inner region, largely evacuating the disk and sending a wave of mass outwards, which eventually stalls when it reaches the region where  $\alpha$  is still equal to 0.01.

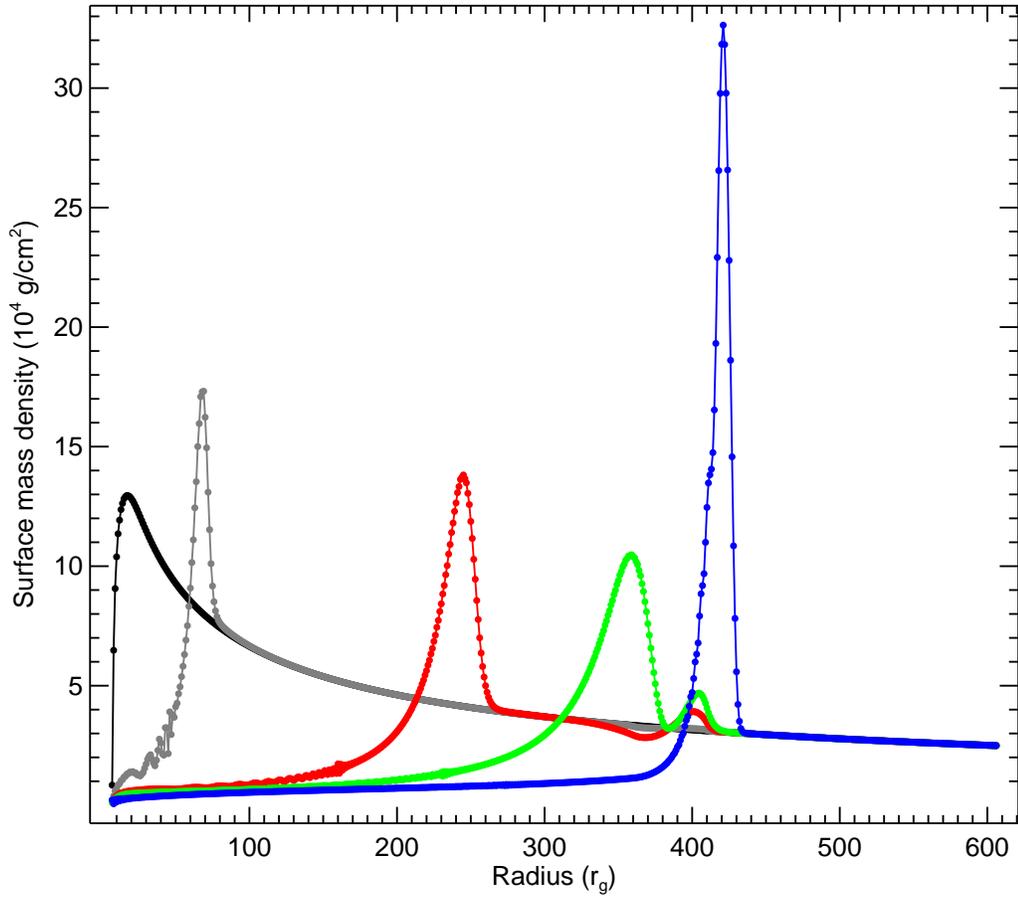


Figure 5.2: The profile of the surface mass density at five times from Figure 5.1: 100 seconds (black; right before the change in  $\alpha$  is initiated), 102.5 seconds (gray), 120 seconds (red), 140 seconds (green), and 168 seconds (blue).

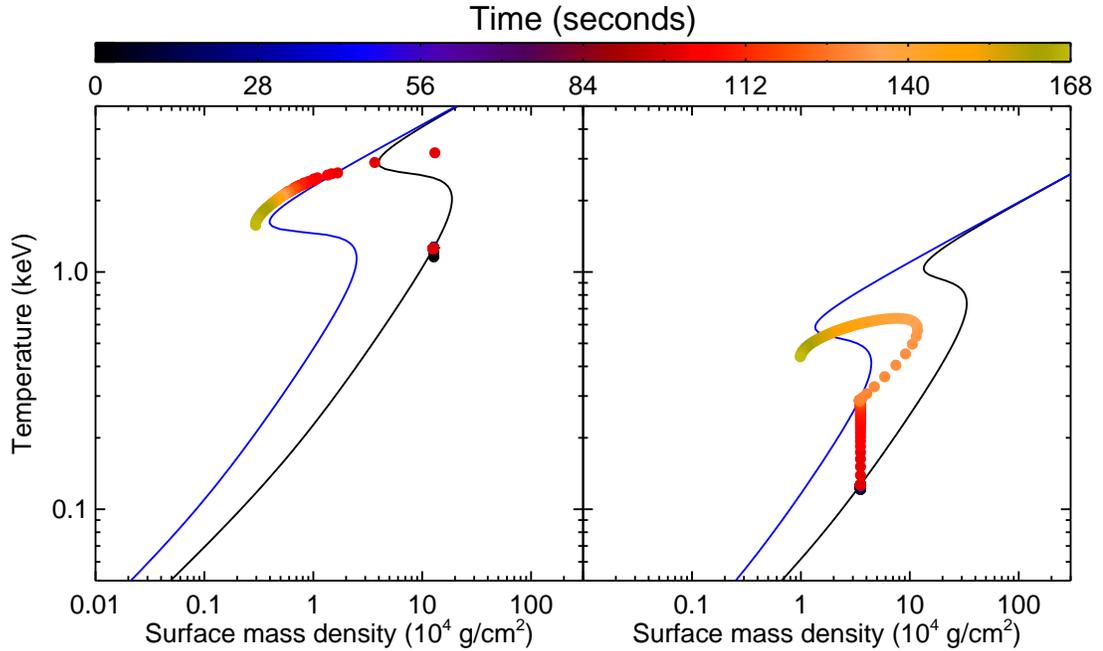


Figure 5.3: The evolution on the local energy balance curve at two radii, when  $\alpha$  is raised from 0.01 to 0.1 in the inner region of the disk. The left panel shows the evolution at  $r = 22r_g$ ; the energy balance curve for  $\alpha = 0.01$  is shown in black, and that for  $\alpha = 0.1$  is shown in blue. When  $\alpha$  changes, the material at this location in the disk is suddenly in an unstable region of parameter space, and an outburst is triggered. At a radius of  $r = 327r_g$ , meanwhile (right panel), the disk heats up when  $\alpha$  changes, but it still remains on a stable part of the energy balance curve. Only when the transition wave from the inner part of the disk reaches this radius (at  $\sim 140$  seconds) is enough mass added to drive it into an outburst.

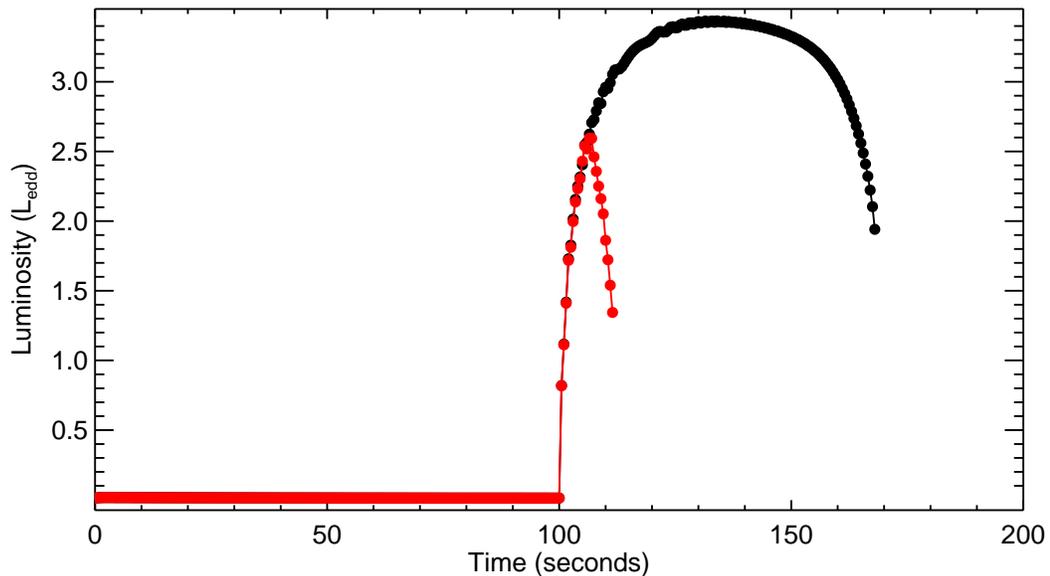


Figure 5.4: The outburst light curve for the run shown in Figures 5.1 through 5.3 (black), where the value of  $\alpha$  is raised for  $r \lesssim 400r_g$ , compared to the outburst light curve for a run in which the value of  $\alpha$  is only raised for  $r \lesssim 100r_g$  (red).

In these simulations, the transition wave stalls at a radius smaller than  $\sim 150 - 200r_g$ , leading to an outburst of duration  $\sim 20 - 30$  seconds (Honma et al. 1991; Szuszkiewicz & Miller 1998, 2001).

In our simulations, there are several important differences. First, we have found that the timescale and radial extent of the outburst are not fixed, but rather depend on the size of the region in which  $\alpha$  is changed. We can obtain an outburst that is much longer than normal, as in the case discussed above, or we can obtain a shorter outburst by only allowing  $\alpha$  to change over a small region of the disk (as shown by the red points in Figure 5.4). Within a large range of radii over which  $\alpha$  might be allowed to change, the transition wave will always stall near the outer boundary of the region where  $\alpha$  is altered (i.e., the outer boundary of the region where the “jet ejection” takes place). The reason that our model can give longer outbursts than the original accretion disk simulations can be appreciated

by looking at the right panel of Figure 5.3; by increasing  $\alpha$  rapidly, we effectively obtain an  $\alpha = 0.1$  disk that has the surface mass density of an  $\alpha = 0.01$  disk. In other words, after heating up on a thermal timescale and finding its new location on the energy balance curve, the disk is much more dense than would normally be the case. At any given radius, therefore, the disk is very close to being near the critical bend in the energy balance curve at which it becomes unstable, and the transition wave only needs to add a small amount of mass to trigger a new outburst at that radius. Thus, we should be able to get outbursts quite far out in the disk, as long as  $\alpha$  is increased in those regions, whereas in the normal case, the transition wave will peter out much earlier because it quickly reaches a point in the disk where the surface mass density is too low to undergo an outburst, even with the added mass from the transition wave.

These results have important observational implications. First, there are times when GRS 1915+105 remains in the high state for a significant period of time. Nayakshin et al. (2000) have discussed the difficulty of reproducing these observations. Here, we show that it is possible to do so by removing a jet from a large enough portion of the disk (however, our code limitations make it impossible for us to currently investigate one of the main issues raised by Nayakshin et al. 2000, which is the difficulty in reproducing the large *duty cycles* of the GRS 1915+105 outbursts).

In addition, Belloni et al. (1997b) found a correlation during certain observations of GRS 1915+105 between the observed duration of the outburst and the amount of the inner disk that appeared to be “missing” in spectral fits of the quiescent time between outbursts. They attributed this to a model in which a certain portion of the disk was removed or became unobservable, and the larger

this region was, the more time it took the disk to refill on the viscous timescale. At least qualitatively, our simulations would appear to reproduce this correlation by relating both the outburst timescale and size of the “missing” part of the disk to the size of the region of the disk from which the jet was ejected. However, we would need to be able to simulate multiple repeating events in order to truly test this correlation.

Finally, and perhaps most important, the original simulations of Honma et al. (1991) and others predicted continuously repeating outburst cycles, because the disks they studied were inherently unstable. GRS 1915+105, however (as well as other black hole candidates), frequently undergoes an outburst after a long period of quiescence. Our simulations can reproduce this behavior easily, because it is the change in  $\alpha$  (due, we posit, to the removal of magnetic field by a jet) that drives an *otherwise stable disk* into an outburst. Furthermore, if the value of  $\dot{M}$  is such that the disk is stable for both the initial and final values of  $\alpha$ , then we would expect to see a single outburst, with the disk returning to a stable state on the “new” energy balance curve after the outburst is complete. For a certain range of  $\dot{M}$ , however, it is possible to have the disk be stable for the low value of  $\alpha$  but unstable for the high value (this is essentially due to the fact that a low  $\alpha$  disk has a higher surface mass density and therefore a higher gas pressure fraction if all else is equal, which tends to keep the disk stable). In this case, we would expect the disk to undergo repeated oscillations following the initial outburst. Furthermore, we speculate that these subsequent oscillations might be of the more “normal” variety found by Honma et al. (1991) and others, because the disk at this point will be in a normal  $\alpha = 0.1$  state. Thus, one could imagine light curves with some resemblance to the “class  $\beta$ ” light curves in Chapter 2, where the initial outburst state appears to be

on a longer timescale than the subsequent oscillations. However, a full simulation of this phenomenon would be necessary to see if this behavior really does occur.

We show in Figures 5.5 and 5.6 two runs in which the value of  $\alpha$  is only changed in the middle part of the disk, not the inner part. In both figures, the inner boundary of the region where  $\alpha$  is changed is the same, but in Figure 5.5, the outer boundary is farther away. We see similar behavior in both figures, with the change in  $\alpha$  sending mass primarily towards the inner region, where it builds up without penetrating to the extreme inner regions of the disk. Significantly more mass builds up in Figure 5.5, which would be expected because more mass is available; however, in neither case is there a prominent transition wave at the moment  $\alpha$  is changed. Rather the disk must wait until mass builds up before a noticeable effect is achieved.

We show the outburst light curves from these two runs in Figure 5.7. The first case has a much brighter light curve than the second, which may seem surprising since the only difference between these runs was in the value of  $\alpha$  at very large radii. Figure 5.8, however, shows that there actually is a significant difference at the radii near where the mass builds up; the extra mass available in the case of Figure 5.5 is just enough to push the disk over the edge of stability and allow it to go into an outburst at this radius. In the case of Figure 5.6, however, not enough mass is received, and the disk at this radius always stays on or near the lower, stable portion of the energy balance curve.

In Figure 5.9, we show that the outburst light curve for the case of Figure 5.5 is indeed due to emission at the middle radii rather than the inner portion of the disk, by plotting the luminosity per unit radius in the disk as a function of time. Thus, it is possible to get an observable outburst in the disk without

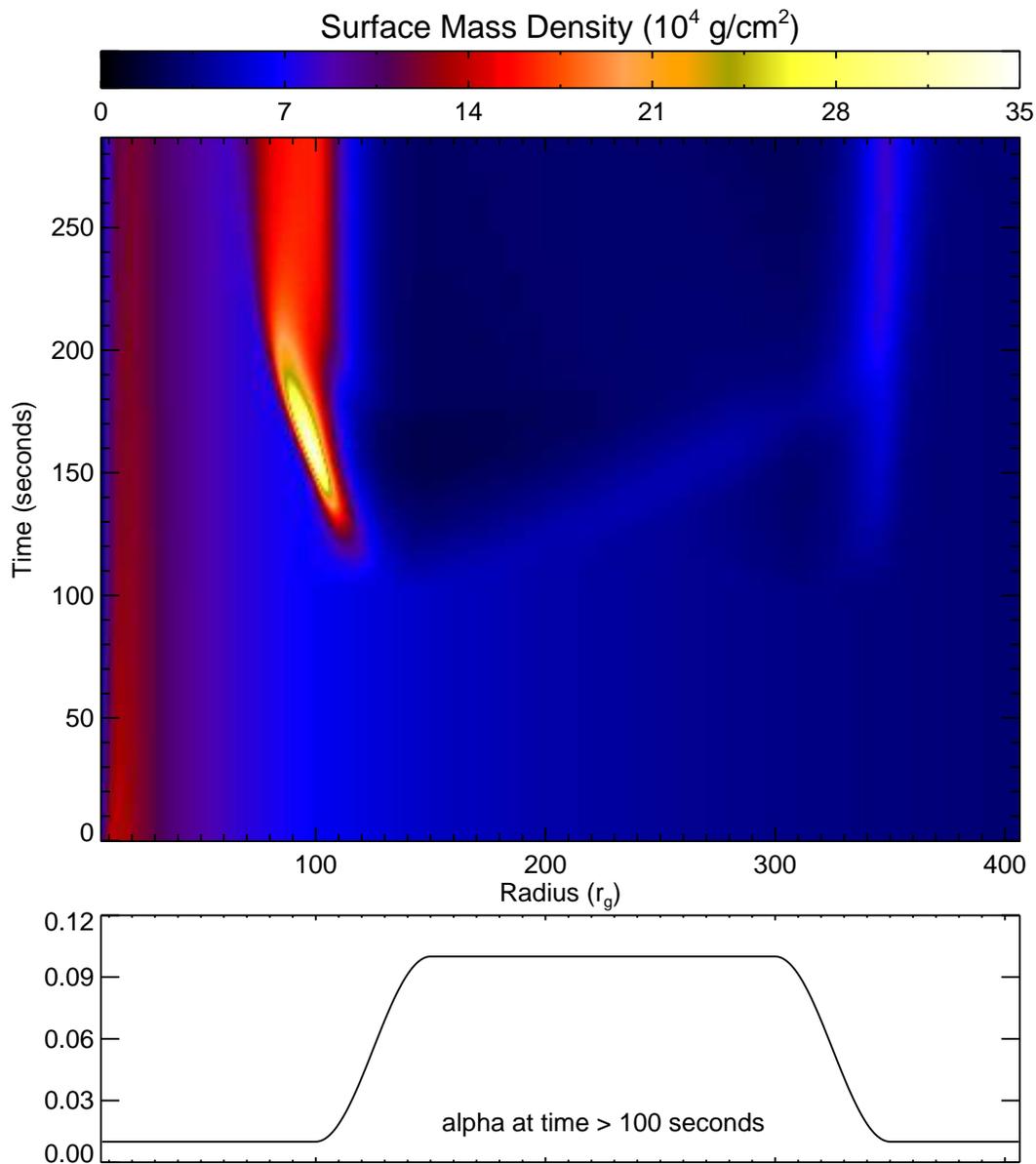


Figure 5.5: The surface mass density evolution for a run in which the value of  $\alpha$  is raised from 0.01 to 0.1 in the middle part of the disk (cf. Figure 5.1). In this case, mass builds up around the boundary of the region where  $\alpha$  is changed and penetrates it slightly, but never reaches the inner part of the disk. Also, unlike Figure 5.1, there is no prominent transition wave initiated when  $\alpha$  changes.

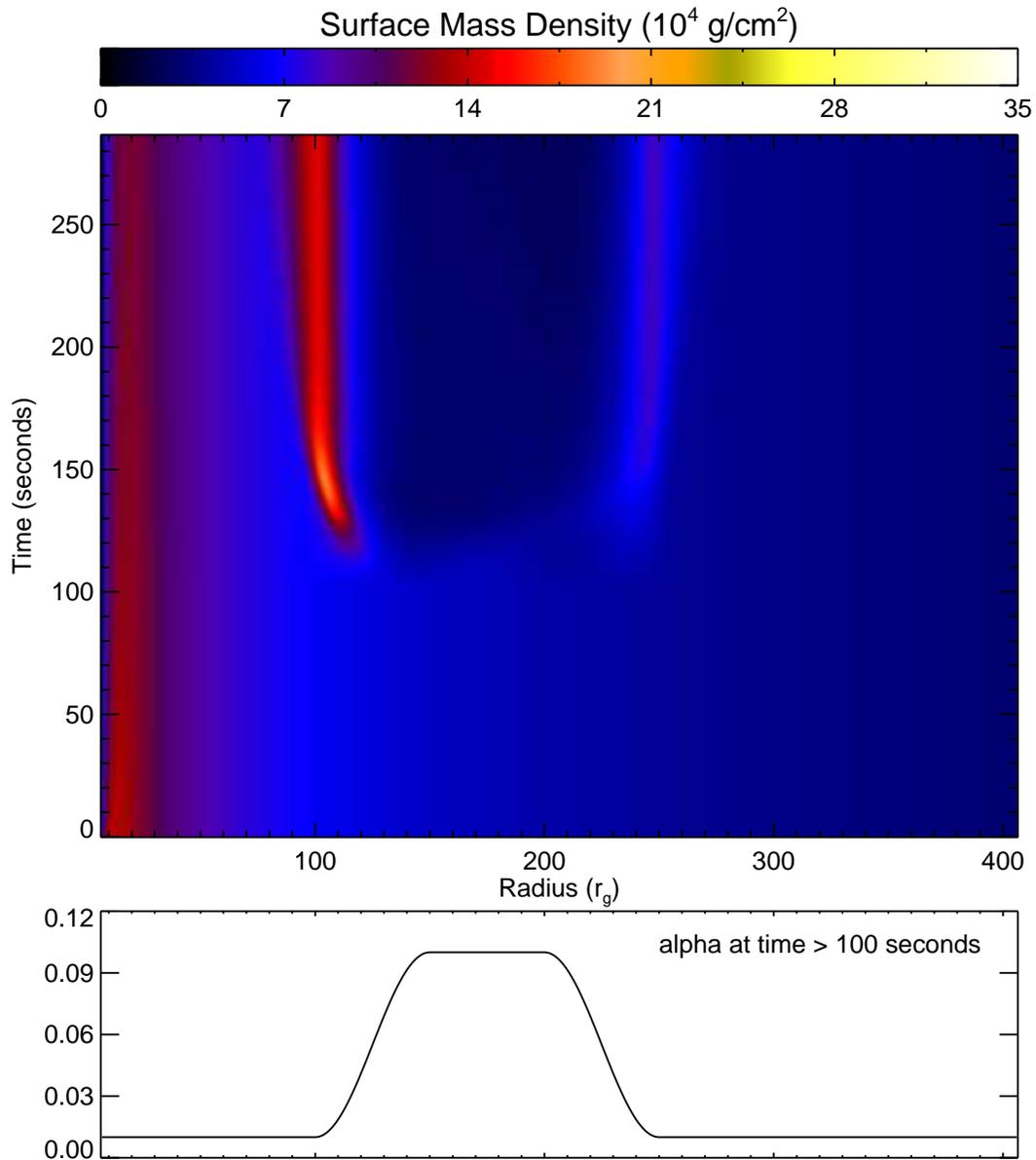


Figure 5.6: Same as Figure 5.5, but the outer boundary of the region where  $\alpha$  changes is now  $100r_g$  closer to the black hole. In this case, much less mass builds up in the inner region. (Note that the color scale in this figure is the same as the color scale in Figure 5.5.)

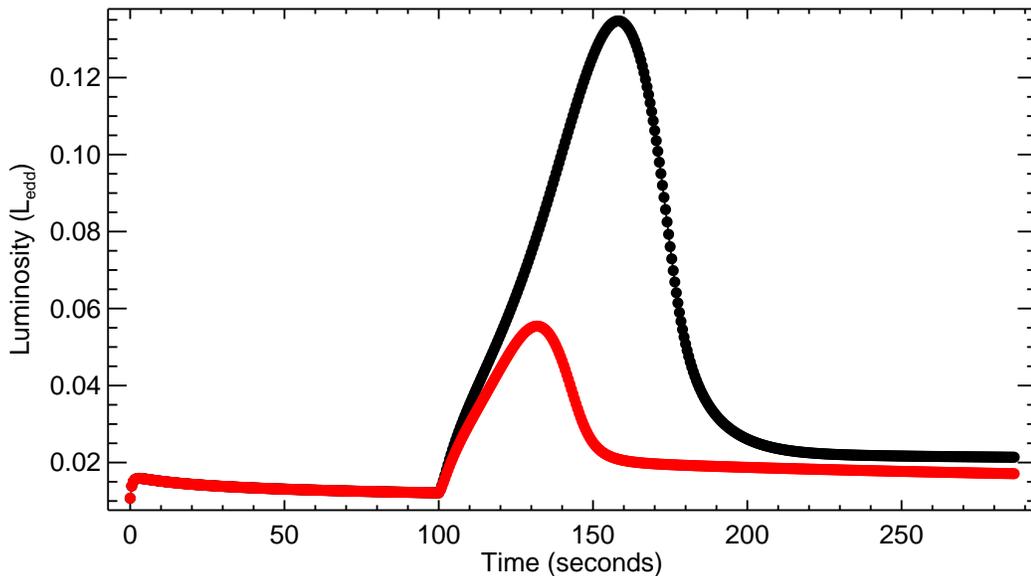


Figure 5.7: The outburst light curves for the runs shown in Figure 5.5 (black) and 5.6 (red). The extra mass that accumulates in the inner region of the disk in the case of Figures 5.5 appears to trigger an outburst that is brighter and longer, and whose peak is significantly delayed.

the inner radii participating. This situation is similar to the one investigated by Lovelace et al. (1994) in simulations of a magnetized disk, where a disturbance in the magnetic field at large radii led to an ejection as well as significant accretion into the inner disk. Here, we show that the change in  $\alpha$  within the disk that might be expected if material were ejected at a large enough radius is enough to initiate similar accretion behavior, regardless of the power that goes into the jet. We also find that the outburst does not necessarily have to reach the extreme inner disk but can be confined to the region in which the jet is ejected. Finally, we note that the delayed outburst light curves shown in 5.7 could be relevant for observations of GRS 1915+105 discussed in Chapter 2, where infrared activity in the jet sometimes appears to precede X-ray activity from the disk.

Finally, we show in Figures 5.10 and 5.11 measurements of the degree to which

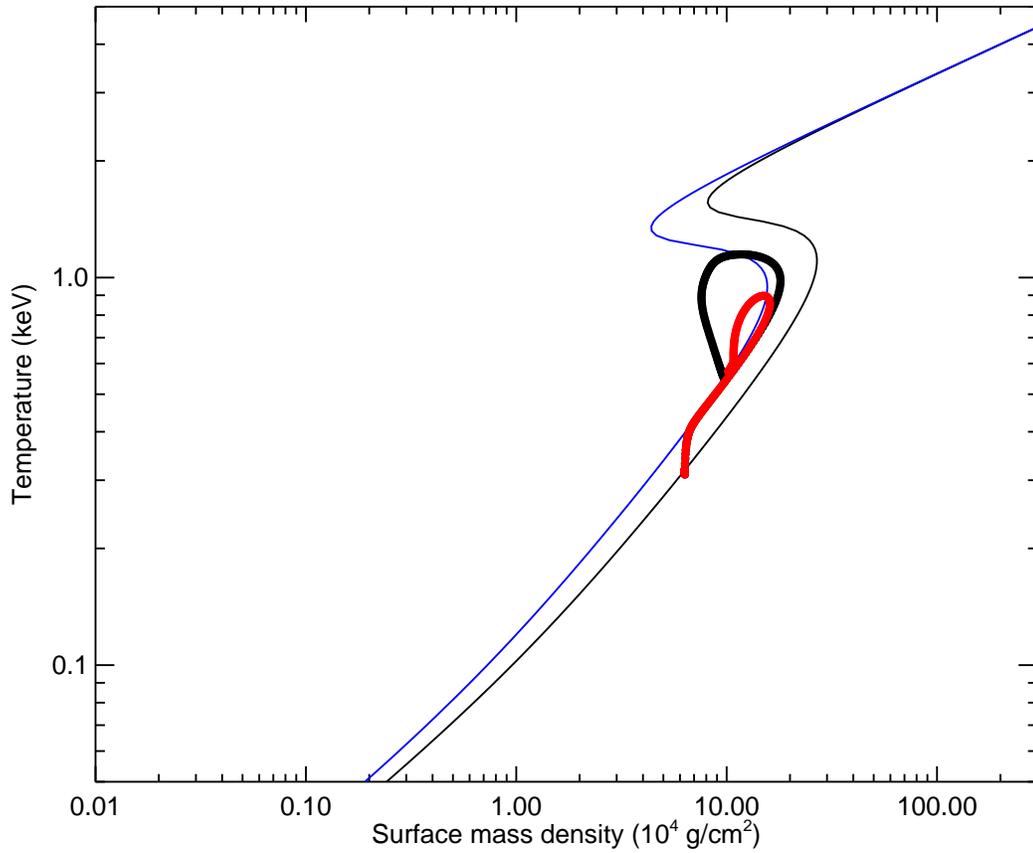


Figure 5.8: The evolution on the local energy balance curve at a radius  $r = 110r_g$ , for the run shown in Figure 5.5 (thick black line) and the run shown in Figure 5.6 (thick red line). The energy balance curves at this radius before and after  $\alpha$  is changed are represented by thin black and blue lines, respectively. In the first case (thick black line), the material at this radius receives just enough mass from farther out in the disk to undergo a small outburst, while in the second case (thick red line), it does not.

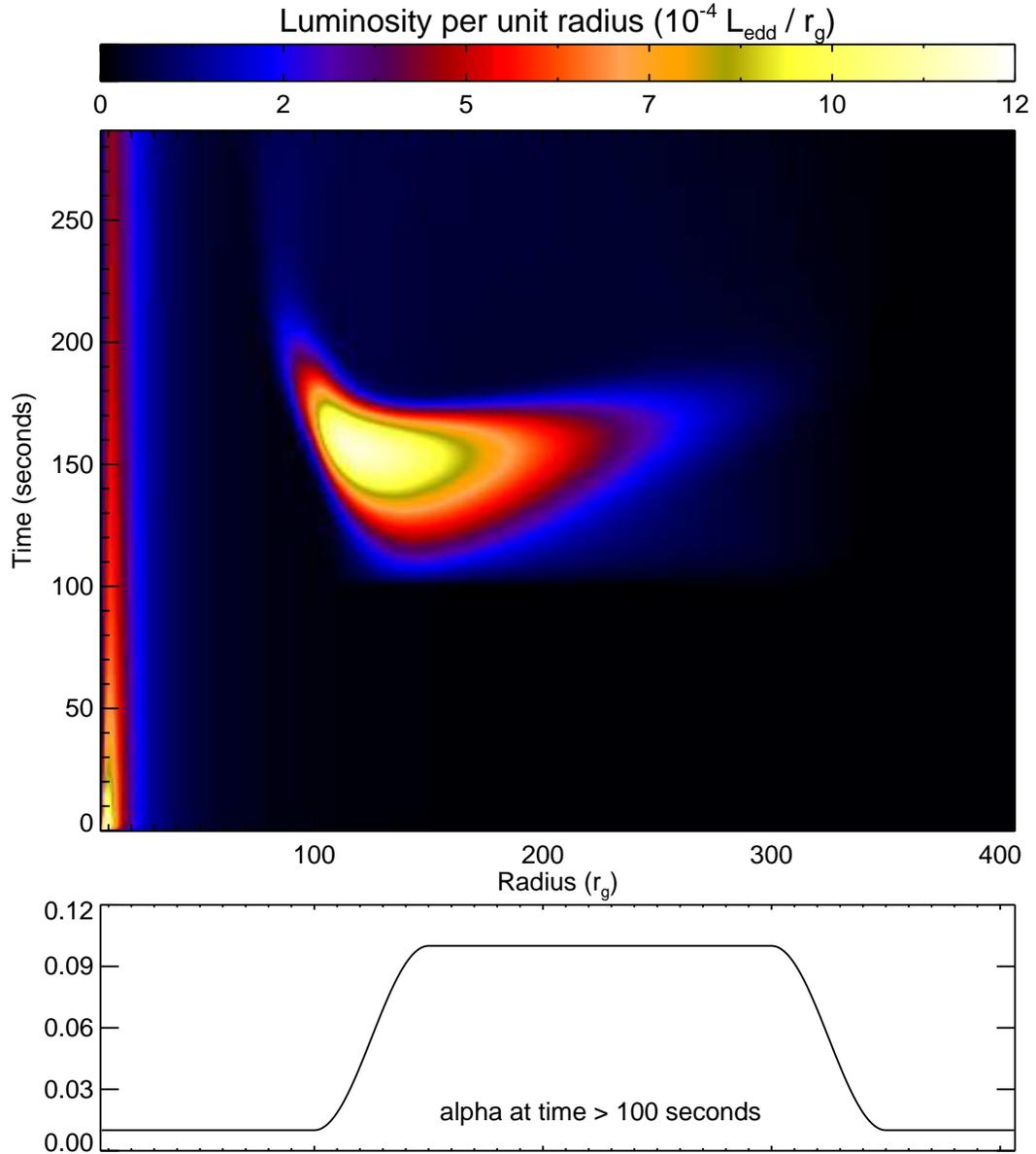


Figure 5.9: The evolution of the luminosity per unit radius for the run shown in Figure 5.5. It is clear that the outburst light curve is produced almost entirely via emission from the middle part of the disk, and the inner disk does not participate in any significant way.

various assumptions made in our one-dimensional disk model break down in the “inner disk outburst” from Figure 5.1 and the “middle disk outburst” from Figure 5.5. Checking these conditions is an important step in any one-dimensional disk calculation, but it is not done in papers published in the literature. As described in the captions of Figures 5.10 and 5.11, we find that our simulations of outbursts originating in the middle region of the disk are generally reliable, while those for an outburst originating in the inner region of the disk do violate the assumptions in some ways; in particular, the assumption that the gas and radiation are always in equilibrium appears to be invalid during the outburst. Thus, we should be suspicious of some of the detailed behavior seen in the disk simulations during this time (although we believe that our overall conclusions about this outburst are still valid). The brightest portions of the light curve in this case would be a good candidate for two-dimensional simulations with the gas and radiation energy equations evolved separately.

## 5.5 Conclusions and Future Work

In this chapter, we have presented preliminary results on a study of how a rapid change in  $\alpha$  might be able to cause outbursts in a disk that roughly match some of the observations of GRS 1915+105. Future work should involve investigating a wider range of parameters, tracking a complete cycle of several outbursts, investigating key parts of the light curve using two-dimensional simulations, and, finally, comparing the simulations to the data in a more comprehensive way, either statistically or by attempting to actually “fit” an observational light curve of GRS 1915+105 by allowing  $\alpha$ , as well as other parameters which were discussed and introduced in Chapter 3, to change in space and time.

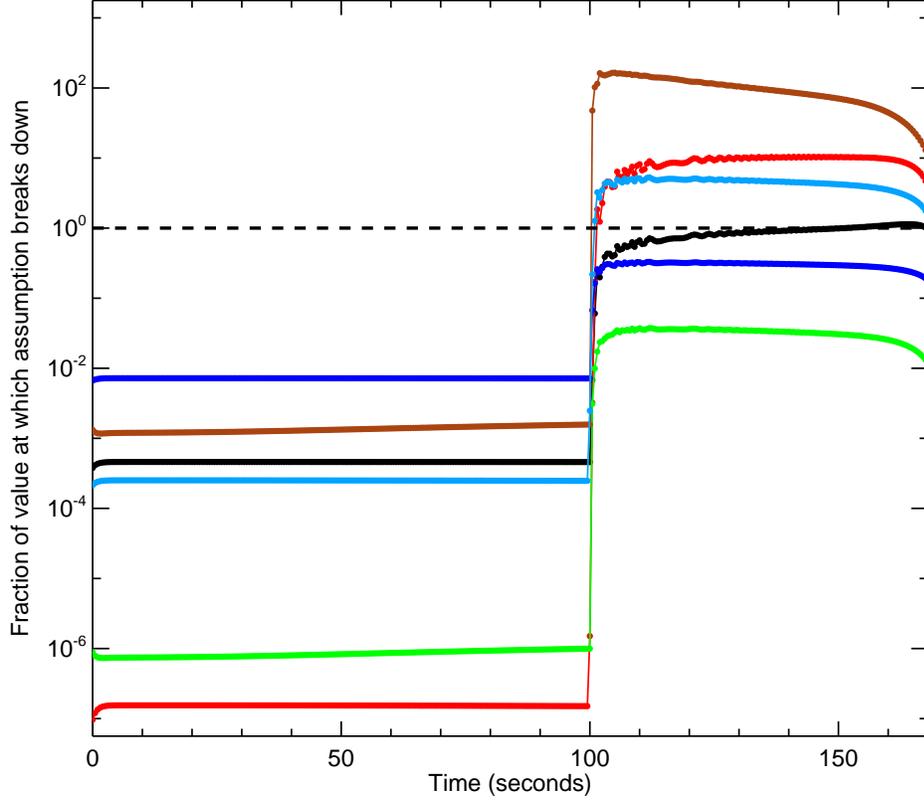


Figure 5.10: A rough measure of the degree to which various assumptions made for our one-dimensional disk model break down during the inner disk outburst shown in Figure 5.1, at a radius of  $r = 22r_g$ . Each parameter plotted should be  $\ll 1$  for the assumption to be valid (i.e., well below the dashed line). The parameters plotted include  $1/\tau_{eff}$ , which measures the optical thinness of the disk (black points);  $(\alpha/\tau_{abs})(H/r)(v_\phi/c)$ , which is the ratio of the thermalization to thermal timescales and therefore measures the extent to which the gas and radiation are locally decoupled (red points);  $H/r$ , which measures the thickness of the disk (dark blue points);  $v_r/v_\phi$ , which measures the degree to which the orbits are noncircular (green points);  $\tau(v_r/c)(H/\Delta r)$ , where  $\Delta r$  is the grid spacing, which measures the extent to which material on the equatorial plane can accrete to an adjacent grid cell faster than radiation can diffuse vertically (brown points); and  $\alpha(H/\Delta r)^2$ , which measures the degree to which different grid cells can communicate horizontally faster than the disk can come to vertical hydrostatic equilibrium (light blue points). We find that after  $\alpha$  is increased, the worst assumptions are that the gas and radiation are in equilibrium in the inner disk (red and black points). Also, our grid spacing is “too high resolution” for structure on the order of  $\Delta r$  to be believable, because radiation will diffuse out of the disk at a different location than it was produced (brown points), and because horizontal communication across a distance  $\Delta r$  can happen before the disk has come to vertical equilibrium (light blue points).

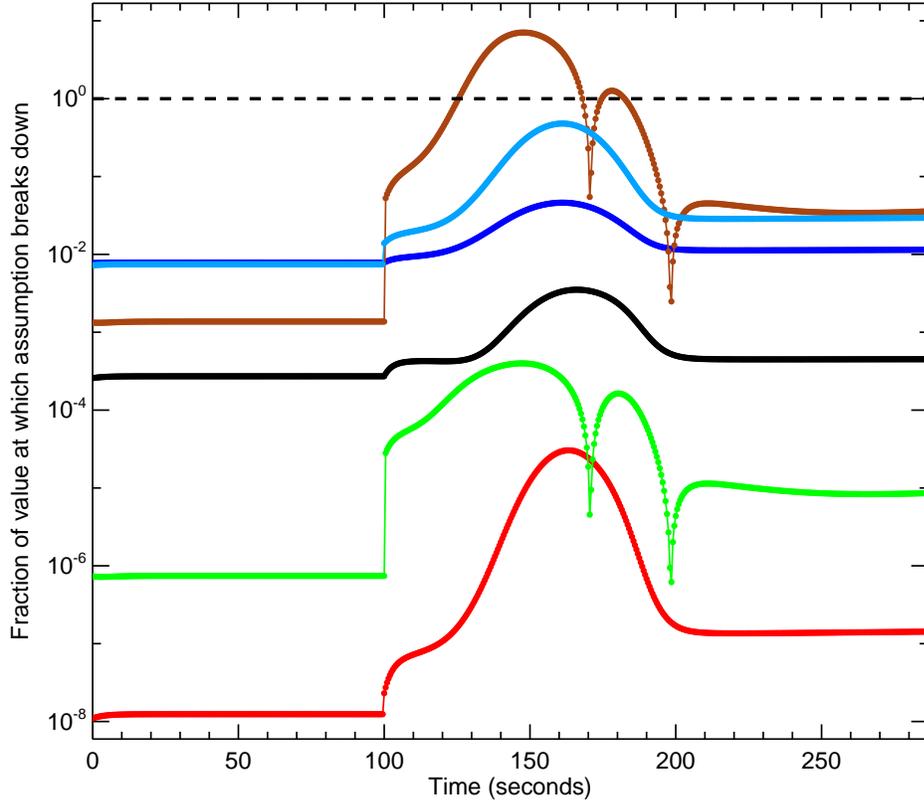


Figure 5.11: Same as Figure 5.10, but for the middle disk outburst shown in Figure 5.5, at a radius of  $r = 110r_g$  (where a significant amount of the outburst radiation is emitted; cf. Figure 5.9). In this case, the one-dimensional disk assumptions generally do *not* break down, although our grid spacing  $\Delta r$  is still “too high resolution” to believe that all the radiation produced at this radius is actually emitted from the disk there (brown points).

# Appendix A

## The Divergence of a Tensor in Cylindrical Coordinates

In this appendix, we compute the divergence of a tensor in cylindrical coordinates, for use in the accretion disk equations in Chapter 3. We first derive equations that can be used to compute the divergence of a tensor in any orthonormal coordinate system (§A.1), and we then specialize to the case of cylindrical coordinates (§A.2).

We begin by reviewing a few basic definitions. For an arbitrary tensor  $T$ , the transformation from a coordinate system labeled by the indices  $g$  and  $h$  to a coordinate system labeled by the indices  $u$  and  $v$  is given by

$$T_{uv} = \sum_{g,h} T_{gh} \left( \frac{\partial \hat{x}_g}{\partial \hat{x}_u} \right) \left( \frac{\partial \hat{x}_h}{\partial \hat{x}_v} \right), \quad (\text{A.1})$$

where each  $\hat{x}$  represents a basis vector in the indicated coordinate system. All tensors, by definition, must transform according to this equation (we ignore here the difference between covariant and contravariant tensors, which is not relevant for our purposes). An ordinary vector  $\mathbf{v}$ , meanwhile, transforms from a coordinate system labeled by  $g$  to one labeled by  $u$  according to

$$v_u = \sum_g v_g \left( \frac{\partial \hat{x}_g}{\partial \hat{x}_u} \right). \quad (\text{A.2})$$

In the specific case of a transformation between Cartesian coordinates  $(x, y, z)$  and cylindrical coordinates  $(r, \phi, z)$ , the basis vectors are related by the following

equations:

$$\hat{r} = (\cos \phi) \hat{x} + (\sin \phi) \hat{y} \quad (\text{A.3})$$

$$\hat{\phi} = (-\sin \phi) \hat{x} + (\cos \phi) \hat{y} \quad (\text{A.4})$$

$$\hat{x} = (\cos \phi) \hat{r} + (-\sin \phi) \hat{\phi} \quad (\text{A.5})$$

$$\hat{y} = (\sin \phi) \hat{r} + (\cos \phi) \hat{\phi}, \quad (\text{A.6})$$

which can be used to compute the partial derivatives in equations (A.1) and (A.2).

## A.1 Evaluation of the Divergence

In Cartesian coordinates, the divergence of a tensor is a vector whose components equal the divergence of each row of the tensor. We therefore have

$$(\nabla \cdot \mathbb{T})_j = \sum_i \frac{\partial}{\partial x_i} (T_{ij}), \quad (\text{A.7})$$

where the indices  $i$  and  $j$  run over Cartesian coordinates. With the help of equation (A.1), we can convert  $T_{ij}$  to a new coordinate system as follows:

$$(\nabla \cdot \mathbb{T})_j = \sum_{i,c,d} \frac{\partial}{\partial x_i} \left( T_{cd} \frac{\partial \hat{x}_c}{\partial \hat{x}_i} \frac{\partial \hat{x}_d}{\partial \hat{x}_j} \right), \quad (\text{A.8})$$

where we use the indices  $c$  and  $d$  to refer to any orthonormal coordinate system.

We can now use equation (A.2) to transform this vector to our new coordinate system, which gives

$$(\nabla \cdot \mathbb{T})_b = \sum_{i,j,c,d} \frac{\partial \hat{x}_j}{\partial \hat{x}_b} \frac{\partial}{\partial x_i} \left( T_{cd} \frac{\partial \hat{x}_c}{\partial \hat{x}_i} \frac{\partial \hat{x}_d}{\partial \hat{x}_j} \right), \quad (\text{A.9})$$

where we have introduced an index  $b$  that runs over the new coordinate system.

Next, we can use the chain rule to transform the derivative with respect to  $x_i$  to a derivative with respect to our new coordinate system. We obtain

$$(\nabla \cdot \mathbb{T})_b = \sum_{i,j,a,c,d} \frac{\partial \hat{x}_j}{\partial \hat{x}_b} \frac{\partial \hat{x}_a}{\partial \hat{x}_i} \frac{\partial}{\partial x_a} \left( T_{cd} \frac{\partial \hat{x}_c}{\partial \hat{x}_i} \frac{\partial \hat{x}_d}{\partial \hat{x}_j} \right), \quad (\text{A.10})$$

where we have introduced another index  $a$  that runs over the new coordinate system. Applying the chain rule to the term in parentheses gives

$$(\nabla \cdot \mathbb{T})_b = \sum_{i,j,a,c,d} \left[ \frac{\partial \hat{x}_j}{\partial \hat{x}_b} \frac{\partial \hat{x}_a}{\partial \hat{x}_i} \frac{\partial \hat{x}_c}{\partial \hat{x}_i} \frac{\partial \hat{x}_d}{\partial \hat{x}_j} \frac{\partial}{\partial x_a} T_{cd} + T_{cd} \frac{\partial \hat{x}_j}{\partial \hat{x}_b} \frac{\partial \hat{x}_a}{\partial \hat{x}_i} \frac{\partial}{\partial x_a} \left( \frac{\partial \hat{x}_c}{\partial \hat{x}_i} \frac{\partial \hat{x}_d}{\partial \hat{x}_j} \right) \right]. \quad (\text{A.11})$$

The long chain of partial derivatives in the first term of this equation can be simplified considerably. We can arrange the order of the sum so that we first evaluate

$$\sum_i \frac{\partial \hat{x}_a}{\partial \hat{x}_i} \frac{\partial \hat{x}_c}{\partial \hat{x}_i} = \sum_i \frac{\partial \hat{x}_i}{\partial \hat{x}_a} \frac{\partial \hat{x}_c}{\partial \hat{x}_i} = \sum_i \frac{\partial \hat{x}_c}{\partial \hat{x}_a} = \delta_{ac}, \quad (\text{A.12})$$

where the first step uses  $\partial \hat{x}_a / \partial \hat{x}_i = \partial \hat{x}_i / \partial \hat{x}_a$  (which is true for orthogonal coordinate systems because both can be expressed as  $\hat{x}_i \cdot \hat{x}_a$ ), the second step uses the chain rule, and the final step uses the fact that the coordinate system specified by  $a$  and  $c$  is orthonormal. Similar manipulations can be performed for the sum over  $j$ , which replaces the remaining basis vector partial derivatives with  $\delta_{bd}$ . The first term of equation (A.11) therefore becomes

$$\sum_{a,c,d} \delta_{ac} \delta_{bd} \frac{\partial}{\partial x_a} T_{cd} = \sum_a \frac{\partial}{\partial x_a} T_{ab}. \quad (\text{A.13})$$

We can now rewrite equation (A.11) as

$$(\nabla \cdot \mathbb{T})_b = \sum_a \frac{\partial}{\partial x_a} T_{ab} + \sum_{i,j,a,c,d} T_{cd} \frac{\partial \hat{x}_j}{\partial \hat{x}_b} \frac{\partial \hat{x}_a}{\partial \hat{x}_i} \frac{\partial}{\partial x_a} \left( \frac{\partial \hat{x}_c}{\partial \hat{x}_i} \frac{\partial \hat{x}_d}{\partial \hat{x}_j} \right). \quad (\text{A.14})$$

The first sum in this equation is structured exactly like the divergence in Cartesian coordinates. We will therefore only focus on the second sum, which represents the “extra” terms that arise in our coordinate system. Applying the chain rule to the derivative inside this sum and performing similar simplifications as above, we obtain

$$(\nabla \cdot \mathbb{T})_{b,extra} = \sum_{i,j,a,c,d} \left[ T_{cb} \frac{\partial \hat{x}_a}{\partial \hat{x}_i} \frac{\partial}{\partial x_a} \left( \frac{\partial \hat{x}_c}{\partial \hat{x}_i} \right) + T_{ad} \frac{\partial \hat{x}_j}{\partial \hat{x}_b} \frac{\partial}{\partial x_a} \left( \frac{\partial \hat{x}_d}{\partial \hat{x}_j} \right) \right]. \quad (\text{A.15})$$

Inspection of the above equation shows that  $j$  and  $d$  are both dummy indices, since they only appear in the second term of the sum; we therefore can replace them with the equivalent indices  $i$  and  $c$ , which gives

$$(\nabla \cdot \mathbb{T})_{b,extra} = \sum_{i,a,c} \left[ T_{cb} \frac{\partial \hat{x}_a}{\partial \hat{x}_i} \frac{\partial}{\partial x_a} \left( \frac{\partial \hat{x}_c}{\partial \hat{x}_i} \right) + T_{ac} \frac{\partial \hat{x}_i}{\partial \hat{x}_b} \frac{\partial}{\partial x_a} \left( \frac{\partial \hat{x}_c}{\partial \hat{x}_i} \right) \right], \quad (\text{A.16})$$

or, rewriting this equation to collect all derivatives in one place, applying the orthogonality condition and reintroducing the index  $d$  to run over our new coordinate system, we finally obtain

$$(\nabla \cdot \mathbb{T})_{b,extra} = \sum_{a,c,d} (T_{cb} \delta_{ad} + T_{ac} \delta_{bd}) \sum_i \frac{\partial \hat{x}_d}{\partial \hat{x}_i} \frac{\partial}{\partial x_a} \left( \frac{\partial \hat{x}_c}{\partial \hat{x}_i} \right). \quad (\text{A.17})$$

## A.2 Specialization to Cylindrical Coordinates

If we specialize to cylindrical coordinates, inspection of equations (A.3) through (A.6) shows that the basic vectors only depend on  $\phi$ ; therefore, the only nonzero  $\partial/\partial x_a$  derivatives in equation (A.17) will be those with respect to  $\phi$ . We therefore can rewrite equation (A.17) as

$$(\nabla \cdot \mathbb{T})_{b,extra} = \frac{1}{r} \sum_{c,d} (T_{cb} \delta_{\phi d} + T_{\phi c} \delta_{bd}) \sum_i \frac{\partial \hat{x}_d}{\partial \hat{x}_i} \frac{\partial}{\partial \phi} \left( \frac{\partial \hat{x}_c}{\partial \hat{x}_i} \right). \quad (\text{A.18})$$

There are only four possible nonzero values of the sum over  $i$  in this equation (any term involving  $\hat{z}$  can be shown to be trivially zero, so there are two possibilities each for  $c$  and  $d$ ). If we write these sums using the notation  $S_{cd}$  to indicate the particular values of  $c$  and  $d$  that the sum is calculated for, the above equation simplifies to

$$(\nabla \cdot \mathbb{T})_{b,extra} = \frac{1}{r} (T_{rb} S_{r\phi} + T_{\phi b} S_{\phi\phi} + T_{\phi r} S_{r\phi} + T_{\phi\phi} S_{\phi b}). \quad (\text{A.19})$$

The values of  $S_{cd}$  can be calculated using equations (A.3) through (A.6). We obtain the following results:

$$S_{rr} = -\sin \phi \cos \phi + \sin \phi \cos \phi = 0 \quad (\text{A.20})$$

$$S_{r\phi} = \sin^2 \phi + \cos^2 \phi = 1 \quad (\text{A.21})$$

$$S_{\phi r} = -\sin^2 \phi - \cos^2 \phi = -1 \quad (\text{A.22})$$

$$S_{\phi\phi} = \sin \phi \cos \phi - \sin \phi \cos \phi = 0. \quad (\text{A.23})$$

We can therefore simplify equation (A.19) to read

$$(\nabla \cdot \mathbb{T})_{b,extra} = \frac{1}{r} (T_{rb} + T_{\phi r} S_{rb} + T_{\phi\phi} S_{\phi b}). \quad (\text{A.24})$$

Finally, we can evaluate this equation for each  $b$ . The results we obtain are as follows:

$$(\nabla \cdot \mathbb{T})_{r,extra} = \frac{T_{rr}}{r} - \frac{T_{\phi\phi}}{r} \quad (\text{A.25})$$

$$(\nabla \cdot \mathbb{T})_{\phi,extra} = \frac{T_{r\phi}}{r} + \frac{T_{\phi r}}{r} \quad (\text{A.26})$$

$$(\nabla \cdot \mathbb{T})_{z,extra} = \frac{T_{rz}}{r}. \quad (\text{A.27})$$

Inserting these terms into the full expression for the divergence, we obtain:

$$(\nabla \cdot \mathbb{T})_r = \frac{1}{r} \frac{\partial}{\partial r} (rT_{rr}) + \frac{1}{r} \frac{\partial}{\partial \phi} T_{\phi r} + \frac{\partial}{\partial z} T_{zr} - \frac{T_{\phi\phi}}{r} \quad (\text{A.28})$$

$$(\nabla \cdot \mathbb{T})_\phi = \frac{1}{r} \frac{\partial}{\partial r} (rT_{r\phi}) + \frac{1}{r} \frac{\partial}{\partial \phi} T_{\phi\phi} + \frac{\partial}{\partial z} T_{z\phi} + \frac{T_{\phi r}}{r} \quad (\text{A.29})$$

$$(\nabla \cdot \mathbb{T})_z = \frac{1}{r} \frac{\partial}{\partial r} (rT_{rz}) + \frac{1}{r} \frac{\partial}{\partial \phi} T_{\phi z} + \frac{\partial}{\partial z} T_{zz}. \quad (\text{A.30})$$

## REFERENCES

- Abramowicz, M. A., Czerny, B., Lasota, J. P., & Szuszkiewicz, E. 1988, *The Astrophysical Journal*, 332, 646
- Armitage, P. J. 2004, *Theory of Disk Accretion onto Supermassive Black Holes* (ASSL Vol. 308: Supermassive Black Holes in the Distant Universe), 89
- Artemova, I. V., Bisnovatyi-Kogan, G. S., Bjoernsson, G., & Novikov, I. D. 1996a, *The Astrophysical Journal*, 456, 119
- Artemova, I. V., Bisnovatyi-Kogan, G. S., Igumenshchev, I. V., & Novikov, I. D. 2001, *The Astrophysical Journal*, 549, 1050
- Artemova, I. V., Björnsson, G., & Novikov, I. D. 1996b, *The Astrophysical Journal*, 461, 565
- Balbus, S. A. & Hawley, J. F. 1991, *The Astrophysical Journal*, 376, 214
- Balbus, S. A. & Papaloizou, J. C. B. 1999, *The Astrophysical Journal*, 521, 650
- Bardeen, J. M., Press, W. H., & Teukolsky, S. A. 1972, *The Astrophysical Journal*, 178, 347
- Becker, P. A. & Le, T. 2003, *The Astrophysical Journal*, 588, 408
- Belloni, T., Klein-Wolt, M., Méndez, M., van der Klis, M., & van Paradijs, J. 2000a, *Astronomy and Astrophysics*, 355, 271
- Belloni, T., Mendez, M., King, A. R., van der Klis, M., & van Paradijs, J. 1997a, *The Astrophysical Journal*, 479, L145
- . 1997b, *The Astrophysical Journal*, 488, L109
- Belloni, T., Migliari, S., & Fender, R. P. 2000b, *Astronomy and Astrophysics*, 358, L29
- Bisnovatyi-Kogan, G. S. & Blinnikov, S. I. 1977, *Astronomy and Astrophysics*, 59, 111
- Bisnovatyi-Kogan, G. S. & Lovelace, R. V. E. 1997, *The Astrophysical Journal*, 486, L43
- . 2000, *The Astrophysical Journal*, 529, 978
- . 2001, *New Astronomy Review*, 45, 663
- Blackman, E. G., Penna, R. F., & Varniere, P. 2006, preprint (astro-ph/0607119)
- Blaes, O. & Socrates, A. 2001, *The Astrophysical Journal*, 553, 987

- . 2003, *The Astrophysical Journal*, 596, 509
- Blandford, R. D. & Payne, D. G. 1982, *Monthly Notices of the Royal Astronomical Society*, 199, 883
- Blandford, R. D. & Thorne, K. S. 2004, *Applications of Classical Physics* (in preparation; available at <http://www.pma.caltech.edu/Courses/ph136/yr2004>)
- Brandenburg, A. 2003, *Computational aspects of astrophysical MHD and turbulence* (*Advances in Nonlinear Dynamics*), 269
- Cannizzo, J. K. 1992, *The Astrophysical Journal*, 385, 94
- . 1993, *The Astrophysical Journal*, 419, 318
- Cannizzo, J. K. & Cameron, A. G. W. 1988, *The Astrophysical Journal*, 330, 327
- Cannizzo, J. K. & Wheeler, J. C. 1984, *The Astrophysical Journal Supplement Series*, 55, 367
- Castro-Tirado, A. J., Brandt, S., Lund, N., Lapshov, I., Sunyaev, R. A., Shlyapnikov, A. A., Guziy, S., & Pavlenko, E. P. 1994, *The Astrophysical Journal Supplement Series*, 92, 469
- Chapuis, C. & Corbel, S. 2004, *Astronomy and Astrophysics*, 414, 659
- Chen, X. 1995, *The Astrophysical Journal*, 448, 803
- Chen, X. & Taam, R. E. 1993, *The Astrophysical Journal*, 412, 254
- Collins, R. S., Kaiser, C. R., & Cox, S. J. 2003, *Monthly Notices of the Royal Astronomical Society*, 338, 331
- Dhawan, V., Mirabel, I. F., & Rodríguez, L. F. 2000, *The Astrophysical Journal*, 543, 373
- Done, C., Wardziński, G., & Gierliński, M. 2004, *Monthly Notices of the Royal Astronomical Society*, 349, 393
- Eikenberry, S. S., Matthews, K., Morgan, E. H., Remillard, R. A., & Nelson, R. W. 1998a, *The Astrophysical Journal*, 494, L61
- Eikenberry, S. S., Matthews, K., Munro, M., Blanco, P. R., Morgan, E. H., & Remillard, R. A. 2000, *The Astrophysical Journal*, 532, L33
- Eikenberry, S. S., Matthews, K., Murphy, Jr., T. W., Nelson, R. W., Morgan, E. H., Remillard, R. A., & Munro, M. 1998b, *The Astrophysical Journal*, 506, L31
- Eikenberry, S. S. & van Putten, M. H. H. M. 2003, preprint (astro-ph/0304386)

- Fender, R. P. 2006, in *Compact Stellar X-ray Sources*, ed. W. Lewin & M. van der Klis (Cambridge: Cambridge University Press), 381
- Fender, R. P. & Belloni, T. 2004, *Annual Review of Astronomy and Astrophysics*, 42, 317
- Fender, R. P., Belloni, T. M., & Gallo, E. 2004, *Monthly Notices of the Royal Astronomical Society*, 355, 1105
- Fender, R. P., Corbel, S., Tzioumis, T., McIntyre, V., Campbell-Wilson, D., Nowak, M., Sood, R., Hunstead, R., Harmon, A., Durouchoux, P., & Heindl, W. 1999a, *The Astrophysical Journal*, 519, L165
- Fender, R. P., Gallo, E., & Jonker, P. G. 2003, *Monthly Notices of the Royal Astronomical Society*, 343, L99
- Fender, R. P., Garrington, S. T., McKay, D. J., Muxlow, T. W. B., Pooley, G. G., Spencer, R. E., Stirling, A. M., & Waltman, E. B. 1999b, *Monthly Notices of the Royal Astronomical Society*, 304, 865
- Fender, R. P. & Pooley, G. G. 1998, *Monthly Notices of the Royal Astronomical Society*, 300, 573
- . 2000, *Monthly Notices of the Royal Astronomical Society*, 318, L1
- Fender, R. P., Pooley, G. G., Brocksopp, C., & Newell, S. J. 1997, *Monthly Notices of the Royal Astronomical Society*, 290, L65
- Feroci, M., Matt, G., Pooley, G., Costa, E., Tavani, M., & Belloni, T. 1999, *Astronomy and Astrophysics*, 351, 985
- Fornberg, B. 1988, *Mathematics of Computation*, 51, 699
- Frank, J., King, A., & Raine, D. 2002, *Accretion Power in Astrophysics*, 3rd edn. (Cambridge: Cambridge University Press)
- Fuchs, Y., Mirabel, I. F., & Claret, A. 2003, *Astronomy and Astrophysics*, 404, 1011
- Greiner, J., Cuby, J. G., & McCaughrean, M. J. 2001a, *Nature*, 414, 522
- Greiner, J., Cuby, J. G., McCaughrean, M. J., Castro-Tirado, A. J., & Mennickent, R. E. 2001b, *Astronomy and Astrophysics*, 373, L37
- Greiner, J., Morgan, E. H., & Remillard, R. A. 1996, *The Astrophysical Journal*, 473, L107
- Hōshi, R. 1977, *Progress of Theoretical Physics*, 58, 1191

- Haardt, F. & Maraschi, L. 1991, *The Astrophysical Journal*, 380, L51
- Hansen, C. J. & Kawaler, S. D. 1999, *Stellar Interiors* (New York: Springer)
- Harlaftis, E. T. & Greiner, J. 2004, *Astronomy and Astrophysics*, 414, L13
- Hawley, J. F. 2000, *The Astrophysical Journal*, 528, 462
- Hawley, J. F., Gammie, C. F., & Balbus, S. A. 1995, *The Astrophysical Journal*, 440, 742
- . 1996, *The Astrophysical Journal*, 464, 690
- Hirose, S., Krolik, J. H., & Stone, J. M. 2006, *The Astrophysical Journal*, 640, 901
- Hirsch, C. 1988, *Numerical Computation of Internal and External Flows*, Volume 2 (Chichester: John Wiley & Sons, Inc.)
- Honma, F., Kato, S., & Matsumoto, R. 1991, *Publications of the Astronomical Society of Japan*, 43, 147
- Hubeny, I. 1990, *The Astrophysical Journal*, 351, 632
- Ionson, J. A. & Kuperus, M. 1984, *The Astrophysical Journal*, 284, 389
- Jackson, J. D. 1999, *Classical Electrodynamics*, 3rd edn. (New York: John Wiley & Sons, Inc.)
- Janiuk, A., Czerny, B., & Siemiginowska, A. 2002, *The Astrophysical Journal*, 576, 908
- Jones, D. J., South, Jr., J. C., & Klunker, E. B. 1972, *Journal of Computational Physics*, 9, 496
- Kaiser, C. R., Sunyaev, R., & Spruit, H. C. 2000, *Astronomy and Astrophysics*, 356, 975
- Kalemci, E., Tomsick, J. A., Rothschild, R. E., Pottschmidt, K., Corbel, S., & Kaaret, P. 2006, *The Astrophysical Journal*, 639, 340
- Karzas, W. J. & Latter, R. 1961, *The Astrophysical Journal Supplement Series*, 6, 167
- Kato, S., Fukue, J., & Mineshige, S. 1998, *Black-Hole Accretion Disks* (Kyoto: Kyoto University Press)
- Kaufman, A. N. 1966, in *Plasma Physics in Theory and Application*, ed. W. B. Kunkel (New York: McGraw-Hill), 91

- Klein-Wolt, M., Fender, R. P., Pooley, G. G., Belloni, T., Migliari, S., Morgan, E. H., & van der Klis, M. 2002, *Monthly Notices of the Royal Astronomical Society*, 331, 745
- Kley, W. 1989, *Astronomy and Astrophysics*, 208, 98
- Körding, E. G., Fender, R. P., & Migliari, S. 2006, *Monthly Notices of the Royal Astronomical Society*, 369, 1451
- Krolik, J. H. & Hawley, J. F. 2002, *The Astrophysical Journal*, 573, 754
- Krolik, J. H., Hawley, J. F., & Hirose, S. 2005, *The Astrophysical Journal*, 622, 1008
- Landau, L. D. & Lifshitz, E. M. 1987, *Fluid Mechanics*, 2nd edn. (Oxford: Butterworth-Heinemann)
- Lasota, J.-P. 2001, *New Astronomy Review*, 45, 449
- Levermore, C. D. & Pomraning, G. C. 1981, *The Astrophysical Journal*, 248, 321
- Lightman, A. P. & Eardley, D. M. 1974, *The Astrophysical Journal*, 187, L1
- Livio, M., Pringle, J. E., & King, A. R. 2003, *The Astrophysical Journal*, 593, 184
- Lovelace, R. V. E., Romanova, M. M., & Newman, W. I. 1994, *The Astrophysical Journal*, 437, 136
- MacDonald, D. & Thorne, K. S. 1982, *Monthly Notices of the Royal Astronomical Society*, 198, 345
- Malbet, F., Lachaume, R., & Monin, J.-L. 2001, *Astronomy and Astrophysics*, 379, 515
- Markoff, S., Falcke, H., & Fender, R. 2001, *Astronomy and Astrophysics*, 372, L25
- Markoff, S., Nowak, M., Corbel, S., Fender, R., & Falcke, H. 2003, *Astronomy and Astrophysics*, 397, 645
- Masaki, I. 1971, *Publications of the Astronomical Society of Japan*, 23, 425
- Matsumoto, R., Kato, S., Fukue, J., & Okazaki, A. T. 1984, *Publications of the Astronomical Society of Japan*, 36, 71
- Mayer, M. & Pringle, J. E. 2006, *Monthly Notices of the Royal Astronomical Society*, 368, 379
- McClintock, J. E. & Remillard, R. A. 2006, in *Compact Stellar X-ray Sources*, ed. W. Lewin & M. van der Klis (Cambridge: Cambridge University Press), 157

- Meier, D. L. 2005, *Astrophysics and Space Science*, 300, 55
- Menou, K. 2000, *Science*, 288, 2022
- Menou, K., Hameury, J.-M., Lasota, J.-P., & Narayan, R. 2000, *Monthly Notices of the Royal Astronomical Society*, 314, 498
- Merloni, A. 2003, *Monthly Notices of the Royal Astronomical Society*, 341, 1051
- Merloni, A., Fabian, A. C., & Ross, R. R. 2000, *Monthly Notices of the Royal Astronomical Society*, 313, 193
- Merloni, A. & Nayakshin, S. 2006, preprint (astro-ph/0603159)
- Meyer, F. & Meyer-Hofmeister, E. 1982, *Astronomy and Astrophysics*, 106, 34
- . 1983, *Astronomy and Astrophysics*, 128, 420
- Migliari, S. & Belloni, T. 2003, *Astronomy and Astrophysics*, 404, 283
- Mihalas, D. & Mihalas, B. W. 1984, *Foundations of Radiation Hydrodynamics* (Oxford: Oxford University Press)
- Mikles, V. J., Eikenberry, S. S., & Rothstein, D. M. 2006, *The Astrophysical Journal*, 637, 978
- Mirabel, I. F., Dhawan, V., Chaty, S., Rodríguez, L. F., Martí, J., Robinson, C. R., Swank, J., & Geballe, T. 1998, *Astronomy and Astrophysics*, 330, L9
- Mirabel, I. F., Duc, P. A., Rodriguez, P. A., Teyssier, R., Paul, J., Claret, A., Auriere, M., Golombek, D., & Martí, J. 1994, *Astronomy and Astrophysics*, 282, L17
- Mirabel, I. F. & Rodríguez, L. F. 1994, *Nature*, 371, 46
- . 1999, *Annual Review of Astronomy and Astrophysics*, 37, 409
- Mitsuda, K., Inoue, H., Koyama, K., Makishima, K., Matsuoka, M., Ogawara, Y., Suzuki, K., Tanaka, Y., Shibasaki, N., & Hirano, T. 1984, *Publications of the Astronomical Society of Japan*, 36, 741
- Morinishi, Y., Lund, T. S., & Vasilyev, O. V. 1998, *Journal of Computational Physics*, 143, 90
- Mukhopadhyay, B. & Misra, R. 2003, *The Astrophysical Journal*, 582, 347
- Muno, M. P., Morgan, E. H., & Remillard, R. A. 1999, *The Astrophysical Journal*, 527, 321

- Muno, M. P., Remillard, R. A., Morgan, E. H., Waltman, E. B., Dhawan, V., Hjellming, R. M., & Pooley, G. 2001, *The Astrophysical Journal*, 556, 515
- Narayan, R. & Yi, I. 1994, *The Astrophysical Journal*, 428, L13
- . 1995, *The Astrophysical Journal*, 452, 710
- Nayakshin, S., Rappaport, S., & Melia, F. 2000, *The Astrophysical Journal*, 535, 798
- Nicholson, D. R. 1983, *Introduction to Plasma Theory* (New York: John Wiley & Sons, Inc.)
- Nowak, M. A. & Wagoner, R. V. 1991, *The Astrophysical Journal*, 378, 656
- Ogilvie, G. I. 2001, *Monthly Notices of the Royal Astronomical Society*, 325, 231
- Ohsuga, K., Mineshige, S., Mori, M., & Umemura, M. 2002, *The Astrophysical Journal*, 574, 315
- Ohsuga, K., Mori, M., Nakamoto, T., & Mineshige, S. 2005, *The Astrophysical Journal*, 628, 368
- Okuda, T., Teresi, V., Toscano, E., & Molteni, D. 2004, *Publications of the Astronomical Society of Japan*, 56, 547
- Ostlie, D. A. & Carroll, B. W. 1996, *An Introduction to Modern Stellar Astrophysics* (Reading, MA: Addison-Wesley Publishing Company, Inc.)
- Oswatitsch, K. 1956, *Gas Dynamics* (New York: Academic Press, Inc.)
- Paczynski, B. & Bisnovatyi-Kogan, G. 1981, *Acta Astronomica*, 31, 283
- Paczynski, B. & Wiita, P. J. 1980, *Astronomy and Astrophysics*, 88, 23
- Persson, S. E., Murphy, D. C., Krzeminski, W., Roth, M., & Rieke, M. J. 1998, *The Astronomical Journal*, 116, 2475
- Pooley, G. G. & Fender, R. P. 1997, *Monthly Notices of the Royal Astronomical Society*, 292, 925
- Poutanen, J. 1998, in *Theory of Black Hole Accretion Disks*, ed. M. A. Abramowicz, G. Björnsson, & J. E. Pringle (Cambridge: Cambridge University Press), 100
- Press, W. H., Flannery, Brian P. and Teukolsky, S. A., & Vetterling, W. T. 1992, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd edn. (Cambridge: Cambridge University Press)
- Pringle, J. E. 1981, *Annual Review of Astronomy and Astrophysics*, 19, 137

- Remillard, R., Muno, M., McClintock, J. E., & Orosz, J. 2002, in Proceedings of the Fourth Microquasar Workshop: New Views on Microquasars, ed. P. Durouchoux, Y. Fuchs, & J. Rodriguez (Kolkata, India: Center for Space Physics), 57
- Remillard, R. A. & McClintock, J. E. 2006, *Annual Review of Astronomy and Astrophysics*, 44, 49
- Richtmyer, R. D. & Morton, K. W. 1967, *Difference Methods for Initial-Value Problems*, 2nd edn. (New York: John Wiley & Sons, Inc.)
- Rodriguez, J., Varnière, P., Tagger, M., & Durouchoux, P. 2002, *Astronomy and Astrophysics*, 387, 487
- Rodríguez, L. F. & Mirabel, I. F. 1999, *The Astrophysical Journal*, 511, 398
- Rothstein, D. M. & Eikenberry, S. S. 2002, in Proceedings of the Fourth Microquasar Workshop: New Views on Microquasars, ed. P. Durouchoux, Y. Fuchs, & J. Rodriguez (Kolkata, India: Center for Space Physics), 341
- Rothstein, D. M., Eikenberry, S. S., & Matthews, K. 2005, *The Astrophysical Journal*, 626, 991
- Rybicki, G. B. & Lightman, A. P. 1979, *Radiative Processes in Astrophysics* (New York: John Wiley & Sons, Inc.)
- Sano, T., Inutsuka, S.-i., Turner, N. J., & Stone, J. M. 2004, *The Astrophysical Journal*, 605, 321
- Schiesser, W. E. 1991, *The Numerical Method of Lines: Integration of Partial Differential Equations* (San Diego: Academic Press, Inc.)
- Shakura, N. I. & Sunyaev, R. A. 1973, *Astronomy and Astrophysics*, 24, 337
- . 1976, *Monthly Notices of the Royal Astronomical Society*, 175, 613
- Shakura, N. I., Sunyaev, R. A., & Zilitinkevich, S. S. 1978, *Astronomy and Astrophysics*, 62, 179
- Shapiro, S. L., Lightman, A. P., & Eardley, D. M. 1976, *The Astrophysical Journal*, 204, 187
- Smak, J. 1984, *Acta Astronomica*, 34, 161
- Sonin, A. A. 1994, in Fay, J. A., *Introduction to Fluid Mechanics* (Cambridge: MIT Press)
- Spruit, H. C. 2001, in *The Neutron Star - Black Hole Connection*, ed. C. Kouveliotou, J. Ventura, & E. van den Heuvel, 111 (astro-ph/0003144)

- Stehle, R. & Spruit, H. C. 2001, *Monthly Notices of the Royal Astronomical Society*, 323, 587
- Stone, J. M., Mihalas, D., & Norman, M. L. 1992, *The Astrophysical Journal Supplement Series*, 80, 819
- Stone, J. M. & Norman, M. L. 1992a, *The Astrophysical Journal Supplement Series*, 80, 753
- . 1992b, *The Astrophysical Journal Supplement Series*, 80, 791
- Svensson, R. & Zdziarski, A. A. 1994, *The Astrophysical Journal*, 436, 599
- Szuskiewicz, E. 1990, *Monthly Notices of the Royal Astronomical Society*, 244, 377
- Szuskiewicz, E. & Miller, J. C. 1997, *Monthly Notices of the Royal Astronomical Society*, 287, 165
- . 1998, *Monthly Notices of the Royal Astronomical Society*, 298, 888
- . 2001, *Monthly Notices of the Royal Astronomical Society*, 328, 36
- Taam, R. E., Chen, X., & Swank, J. H. 1997, *The Astrophysical Journal*, 485, L83
- Tagger, M. & Pellat, R. 1999, *Astronomy and Astrophysics*, 349, 1003
- Tagger, M., Varnière, P., Rodriguez, J., & Pellat, R. 2004, *The Astrophysical Journal*, 607, 410
- Takeuchi, M. & Mineshige, S. 1997, *The Astrophysical Journal*, 486, 160
- Tassoul, J.-L. 1978, *Theory of Rotating Stars* (Princeton: Princeton University Press)
- Toomre, A. 1964, *The Astrophysical Journal*, 139, 1217
- Truss, M. R., Murray, J. R., Wynn, G. A., & Edgar, R. G. 2000, *Monthly Notices of the Royal Astronomical Society*, 319, 467
- Truss, M. R. & Wynn, G. A. 2004, *Monthly Notices of the Royal Astronomical Society*, 353, 1048
- Türler, M., Courvoisier, T. J.-L., Chaty, S., & Fuchs, Y. 2004, *Astronomy and Astrophysics*, 415, L35
- Turner, N. J., Blaes, O. M., Socrates, A., Begelman, M. C., & Davis, S. W. 2005, *The Astrophysical Journal*, 624, 267

- Turner, N. J. & Stone, J. M. 2001, *The Astrophysical Journal Supplement Series*, 135, 95
- Turner, N. J., Stone, J. M., Krolik, J. H., & Sano, T. 2003, *The Astrophysical Journal*, 593, 992
- Ueda, Y., Yamaoka, K., Sánchez-Fernández, C., Dhawan, V., Chaty, S., Grove, J. E., McCollough, M., Castro-Tirado, A. J., Mirabel, F., Kohno, K., Feroci, M., Casella, P., Trushkin, S. A., Castaneda, H., Rodríguez, J., Durouchoux, P., Ebisawa, K., Kotani, T., Swank, J., & Inoue, H. 2002, *The Astrophysical Journal*, 571, 918
- Vadawale, S. V., Rao, A. R., Naik, S., Yadav, J. S., Ishwara-Chandra, C. H., Pramesh Rao, A., & Pooley, G. G. 2003, *The Astrophysical Journal*, 597, 1023
- van der Laan, H. 1966, *Nature*, 211, 1131
- Vasilyev, O. V. 2000, *Journal of Computational Physics*, 157, 746
- Verstappen, R. W. C. P. & Veldman, A. E. P. 1997, *Journal of Engineering and Mathematics*, 32, 143
- Watarai, K., Fukue, J., Takeuchi, M., & Mineshige, S. 2000, *Publications of the Astronomical Society of Japan*, 52, 133
- Watarai, K. & Mineshige, S. 2003, *The Astrophysical Journal*, 596, 421
- Yadav, J. S. 2001, *The Astrophysical Journal*, 548, 876