Black Hole Mergers and Their Electromagnetic Counterparts

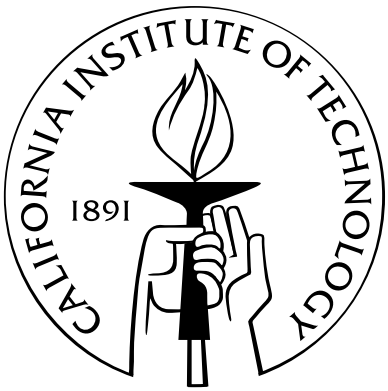
Thesis by

J. Nate Bode

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy



California Institute of Technology

Pasadena, California

2011

(Defended May 18, 2011)

Acknowledgements

During this most recent epoch of my life, my physics epoch, my gratitude for those who surround me has grown tremendously, both due to personal growth and due to the challenges that I have had to overcome during these times. It has been while overcoming these challenges that not only has the importance of my closest friends and family has become blatantly clear, but also the importance of those past persons who have led me to be the person who I am. I would like to take a bit of time to thank some of these people.

First, it would seem appropriate to begin the acknowledgements thanking my high school physics teacher, Mr. Stimson, who first instilled in me a love for physics, and who had a natural ability for making it interesting.

There is great gratitude for Prof. Schneider at Knox College, who made mathematics come alive for me, challenged me, was a pivotal role model for me, and ultimately helped give me the confidence to continue to pursue academia (eventually). Though never my professor, Prof. Bryan was a great friend and also a special role model. He was someone who exemplified kindness and decency. My thanks are belated, but not lessened.

When I first met Stephan Meyer, I had already met several prospective professors to work with. He was different. He seemed excited by the science, and appeared to genuinely enjoy it. There was no doubt who I wanted to work with, and obviously there are no regrets. In a subculture of stoics, finding someone who shows excitement is really wonderful and is largely what accelerated me along my path. I appreciate(d) both his time and advise, and attribute a large part of the completion of my Ph.D. to the effort he made in helping me succeed.

Scientifically, it has only been in the last several years under the guidance of Prof. Sterl Phinney that my own excitement for my research has been found. Sterl has provided many hours of help and advice, and countless points and ideas in the pursuit of finding my own interesting research questions. My time working with him has led me to an amazing quantity of information (when normalized to my memory capacity), a much greater ability to do astrophysics, and enjoy it in the process.

During the past several years Prof. Chris Hirata has read significant fractions of this thesis, provided invaluable professional advice, and has generally been more than generous with his small

quantity of free time. Along with many enjoyable conversations, his time and thoughts have always been greatly appreciated.

In the past year, Chris Wegg and I have collaborated on a couple amazing projects (one of which is given here in Chapter 4). Together we have embarked on a scientific journey which has many possible directions. It has been more fun scientifically than all of the other years together. Moreover, in reality, the completion of this thesis by the requisite date is as much due to Chris's work as mine. For us, the final two months have been a tour de force, and I am tremendously grateful to have had a collaborator who put in this quantity of effort over such an extended period of time.

There have been quite a few people who have taught me important lessons, which were vital both to the initial attempt to study physics well after leaving college, and to the ability to overcome the many challenges since, which have ultimately culminated in this thesis's successful completion. Thank you Sam for the somewhat ridiculous discussion about doing whatever it takes to do what you want to do, no matter how small. I'm still not sure I'd take it to the same extreme, but the power of the idea eventually became clear, and has been astoundingly important in my life choices. Along similar lines, I would like to extend generous thanks to Angus, who first made it clear that when failure is not an option, it is significantly less probable. Your words were salty, but true, and have ultimately changed how I view life and, in particular, challenges.

Since I began this physics odyssey I have met and enjoyed the company of many friends, who have helped make this epoch of my life incredibly fulfilling. With some of them I am still in contact, with others I am not. But I am grateful to all of them: never in my life have I been surrounded by so many people with whom I feel so comfortable and at home. At the Univ. of Chicago I shared wonderful moments with Seth Goldman, Dave Michaeli, Yonit Fuhrmann, Martin Becker, and Tristan Smith. Seth and I shared our first physics courses as lab mates and lunch mates. He single handedly ensured that I remained nourished with his awesome lunch company and cooking setup. Enjoyable conversations, meals, and occasional bike rides down Lakeshore are not forgotten. Great conversations and meals with Dave Michaeli also helped make these two years spectacular. My first year I lived at the International House where I met a huge number of wonderful people, but two of which stand out: Yonit and Martin. Random evening conversations about random things; good humor and lots of laughter; eating meals in the common area; being taught everything I don't know about all of the things I should know about things that are read... Yonit helped make the I-House much more than a place just to sleep and work. It was the thoughtful conversations about life and the mind that really made Martin an amazing friend. With a tremendous sense of humor and a comparable amount of modesty, Martin has always been a pleasure to be with. One person who single-handedly ensured that I didn't learn too much physics and didn't get too much done was Krzysztof Wolek. There were always one-too-many games of foosball and pool, and conversations at the dinner table. Of course, not only did he ensure that I didn't succeed with too much glory,

but he also ensured that I did it in the most fun way possible. Truly great times. The list could go on, because ironically the Univ. of Chicago was an amazing place for me socially.

A special bit of gratitude should be expressed for Tristan Smith (a.k.a., O Fro; Natuguese for "The Fro" and pronounced as "oo frō"), who has been a part of my epic physics odyssey since the first summer at the Univ. of Chicago. Not only were we fellow students in our courses, but we worked together throughout the two years and ultimately have become great friends. He convinced me to apply to Caltech, the university which we both eventually attended for our Ph.D's, and he has been there through every key point of this journey, as both a special friend and confidant. He is youthful and idealistic, positive, good-natured, fun-loving, thoughtful, intelligent, and adventuresome (i.e. the perfect companion for any epic journey like life). I look forward to many more years of great friendship.

Since being at Caltech, I have found myself surrounded by a plethora of great friends. Far more than I have ever deserved to have. Since the first year, Tristan, Hernan Garcia, Eric Peterson and I shared many amazing times. The three amigos are some of the best amigos I've ever had. Hernan, the non-Argentinian Argentinian. The only person I have known to learn the art of driving in the chaos of Buenos Aires ("chaos" used as a euphemism for more appropriate descriptors), and still only have an accident for defensive driving. One of the most driven, hard working people I know he is an amazing combination of neurons, decency, and good old-fashioned elbow grease, with a good sense of humor to boot. Eric, most likely the best by-product of nuclear testing at Los Alamos, has been one of the most genuine and sincere people I have met. Always up for a good laugh, and some good, clean fun he has been a blast through the years.

Also, of special note have been the many amazing officemates I've had over my years at Caltech. One should not underestimate the importance of officemates and how they affect our lives. I am grateful beyond my words to have had so many great ones: Alison Farmer, Anthony Pullen, Yacine Ali-Haïmoud, and Dan Grin. Anthony, Yacine, Dan and I spent many a year together "poking away" at our various projects. They have been hugely important for many reasons, the greatest of which have been the scientific discussions, the camaraderie, and the general long-term great energy in the office. It really is a wonderful way to work. Over the years we have all become good friends. In particular, Dan and I have had the pleasure of taking a couple trips with Tristan and TAing together the best class I never had (ph136). He has become one of those comrades I have grown to count on for helping out when it really matters.

My "workout" buddies have been awesome. For the past two years Chris, Laura Book, Kevin Setter, Camilla Berretta and I have been meeting at some crazy early time every day. Not only have I trimmed off some of that Ph.D. fat, but having the routine of the days begun interesting discussions about science, life, or the things that are happening in our lives has helped make the last two years really a lot of fun. There is true appreciation for these friends who I have seen *almost*

every day for so long both in and out of the gym.

Though infrequent, the awesome dinners with Tod Kurt accompanied by long walks and amazing discussions of all sorts of technological gadgetry have always been engaging and, just as importantly, a breath of fresh air. This is not to mention the few key times that he provided scripting help (and plenty of patience) for me during some hectic periods. Similarly, my times with the John Joseph Carrasco, a fellow "mildly-later-life" Ph.D. physicist, have been rewarding. Having a similar appreciation of physics and a general optimism, his conversations have always been full of life, thought, and fun. He and Tod have been a highlight of my time at Caltech.

There have been many other good friends along the way: Ketan Vyas who among the many great conversations, mediocre to pathetic games of tennis, and dinners eating cheap vegetarian food, also saved me oodles of money by letting me use one of his parking spaces. Without decreasing the value I give to the other points I would really like to emphasize this final point: seeing as I had sworn to never pay Caltech a dime for parking, and that street parking is a royal pain to find, I am particularly indebted to Ketan for offering me this honorable way to not eat crow (i.e., without having to walk in from San Marino like I almost have to do now). Thanks Ketan! (and of course, thanks for all of the good times too ⌣̈). It was on my first day at Caltech when I met Jenny Roizen. She was Chicagoan. . . to the bone. She had spark, chutzpah, and a developed sense of sarcasm, three things that have ensured that our friendship has endured. Her smile and friendship have made these years so much more enjoyable. Anusha Narayan has been another one of those wonderfully positive influences on my Caltech time. Many a wonderful lunch was passed with Anusha discussing various aspects of life, what we want out of it, and what we get out of it. Her clarity, thoughtfulness, and perspective are always a welcome addition to any conversation.

Of course, if discussing those people who have helped make my life better during my tenure in Los Angeles, I cannot leave out my non-Caltech buddies. Kris and John Shenk, Jon Ilaw and Patricia Wuillaume, and Karla Rourke (a.k.a. Chickaroo) have always been up for evening dinners, discussions about life, concerts and general revelry. Their friendships have been a highlight of Lucia and my time in Los Angeles.

Probably one of the greatest things in my life for which I am the most fortunate, is my incredible family. I am so lucky to have parents who have always been there for me, supporting me, and, by example, teaching me how to be a better person. The good fortune to always have a place which *feels* like home is both ineffable and beyond value. The support which they have given me during the challenging times, and the love that they have invariably given, have provided a foundation for the confidence necessary to be who I am and pursue the dreams that I have. This is the same confidence needed to intelligently pursue happiness, and to similarly surround myself with good people who make my life better. These acknowledgements are a tribute to this point. Like anybody deserving of deep gratitude, it is not the result of single acts, but of a lifetime of acts. Of course, I can only

attribute their greatness to great genes ...

In a section dedicated to people who have helped me complete this thesis happily, it goes without saying that my wonderful wife deserves a significant amount of credit. It is to our homes that we carry our greatest burdens, and I am grateful beyond words to have in my life someone in that home who helps me see those burdens in their contexts, and helps provide me the strength to carry them on until resolution. And, similarly, it is wonderful to have that same person always there to celebrate with me when those burdens are successfully resolved. The past ten years have been particularly wonderful, due in large part to the presence of Lucia in my life. My love and respect only grow.

These are some of the friends, family, and colleagues who have helped make this odyssey exceed all expectations. These years have been some of the most challenging of my life, while at the same time being equally rewarding and enjoyable. I have never been happier. I extend my gratitude to all of these people who have helped make this journey as special to me as it has been. Thank you!

Abstract

Over the past ten years it has become increasingly clear that most, if not all, galaxies have supermassive black holes lurking in their cores. The implications for this are large as they not only have significant effects on the host galaxies, far beyond what would have been naïvely expected, but would provide several significant gravitational wave sources to the *Laser Interferometer Space Antenna* (LISA). This thesis is primarily concerned with these gravitational wave sources and the possible electromagnetic counterparts. In particular, when two galaxies merge, it leads to the ultimate merger of their individual SMBHs. If gas is present near the time of merger a circumbinary disk forms around the binary. By assuming the disk is pressureless, and looking at the limits of this approximation, in Chapter 2 we develop an analytic theory of the reaction of such a gaseous disk to the gravitational wave mass loss and recoil kicks which occur during a SMBH merger. However, to understand the effects of finite pressure, in Chapter 3 we develop a one-dimensional hydrodynamic code. The efficiency of the code and the power of the analytic solution allow us to explain the entirety of possible reactions. These results are also favorably compared with far more complicated 3D relativistic magneto-hydrodynamics simulations. LISA will not see only the mergers of two SMBHs, it would also see the inspirals of stellar-mass objects into a SMBH. In Chapter 4 we discuss a new channel of formation of these extreme mass ratio inspirals (EMRIs). This new channel of EMRI formation is rich physically and, in particular, almost always requiring either the Kozai mechanism or an as-of-yet unnoticed phenomenon which we dub the reverse Kozai mechanism. We find that this channel of EMRI formation produces modest numbers of EMRIs when compared to the primary channel of EMRI formation, which, under optimistic detection scenarios for the most recent LISA design, results in the plausible detection of several. Finally, an unrelated project that considers solving the self-similar Type-II strong-shock problem in slightly asymmetric media is given in Chapter 5. We show that the results can even be applied to explosions along weak discontinuities in the density.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The Excitement

Gravitational waves (GWs) are one of the last frontiers of known fundamental physics. Their discovery will be not only a discovery of new, never-before-seen (heard) objects, but will be a detection of a theorized, but still only indirectly confirmed, consequence of general relativity (GR). Indeed, there is probably no other known field of physics where an expected, but still elusive, feature of the theory could provide such an incredible wealth of information about the "invisible" world on scales from kilometers to the size of the observable universe. Gravitational wave physics is not just a new field of physics, and it is not just a new type of telescope, it is a doorway to see parts of the universe which are impossible to probe with any other form of known physics.

1.2 Gravitational Wave Detectors and Sources

1.2.1 Gravitational Wave Experiments

Ground-Based: What we can learn from GWs depends on how we try to detect them. At present there are two primary types of detectors: bars and interferometers. However, bars are slowly being phased out for a variety of reasons and the primary emphasis in today's GW community is on interferometric GW detectors, though there are currently two spherical resonant-mass detectors in operation: MiniGrail and Mario Schenberg (named after the Brazilian physicist also known as Mario Schönberg). Five large-scale interferometric experiments have run, and are currently being upgraded to improved sensitivities (3 LIGO interferometers, 1 VIRGO interferometer, and the GEO600 interferometer), and several are planned for the future (AIGO in Australia, LCGT in Japan, and ET in Europe, among others). The current experiments' sensitivity ranges are all roughly the same, with a $\sim 10\,\mathrm{Hz}$ wall on the low end of the frequency range due to seismic and gravity gradient noise and a gradual high frequency rise beyond $\sim 200\,\mathrm{Hz}$.

These first (bars) and second (interferometers) generation of gravitational wave detectors have been designed with the primary goal of just detecting gravitational waves, though interferometers can say a significant amount about the sources. However, most sources lie in lower frequencies than observable by these Earth-based gravitational wave detectors.

Space-Based: To reach the majority of plentiful and strong GW sources it is necessary to go to space to escape the seismic and gravity gradient wall. The current leading design plan for a space-based detector is the *Laser Interferometer Space Antenna* (LISA). Interestingly it is the first GW experiment to be designed with the intent of doing astrophysics, and the only one which also has known sources it can detect. Currently LISA is undergoing redesign due to budget constraints, which may or may not lead to significant changes. The design of LISA adopted in this thesis is that of prior to the March 2011 exit of NASA from the joint NASA-ESA LISA project.

As LISA is the GW detector relevant to the physics presented in this thesis, its sensitivity curve is plotted (solid black) in Fig. 1.1 along with several of the key LISA sources: the final year of the merger of two equal mass black holes is plotted for four different cases (labeled in plot). We will discuss these sources in Sec. 1.2.2.

It should also be noted that there are other proposed, but more futuristic, space-based detectors, called the Big Bang Observer and DECIGO. These detectors would have even greater sensitivities at the 0.1–10 Hz range, reach even lower frequencies, and would have the capacity to see GW fluctuations direct from the inflationary beginnings of the universe, when the universe was completely opaque to light and particles.

1.2.2 Gravitational Wave Sources

There are several key characteristics of most GW sources: a characteristic mass, with a characteristic separation, a characteristic frequency of rotation, and the distance to the source. For the majority of sources relevant to this thesis, the paradigm is even simpler since the characteristic separation and characteristic frequency are related, and because the sources are made up of what can be assumed to be two massive point particles. That is, we are concerned primarily with the merger of two compact objects. Compact objects are a class of objects which include black holes, neutron stars (NSs), and white dwarf (WDs).

To understand the dependencies let's compare the strength of the GWs from a circular binary to the LISA noise curve. The strength of the gravitational wave is measured by its strain $h(t)$ which is defined to be the maximum fluctuation of some length divided by the length:

$$h(t) \equiv \frac{(\Delta L)_{\max}}{L} \,. \tag{1.1}$$

However, the LISA sensitivity is a function of frequency f, not time, so to compute the signal-to-noise

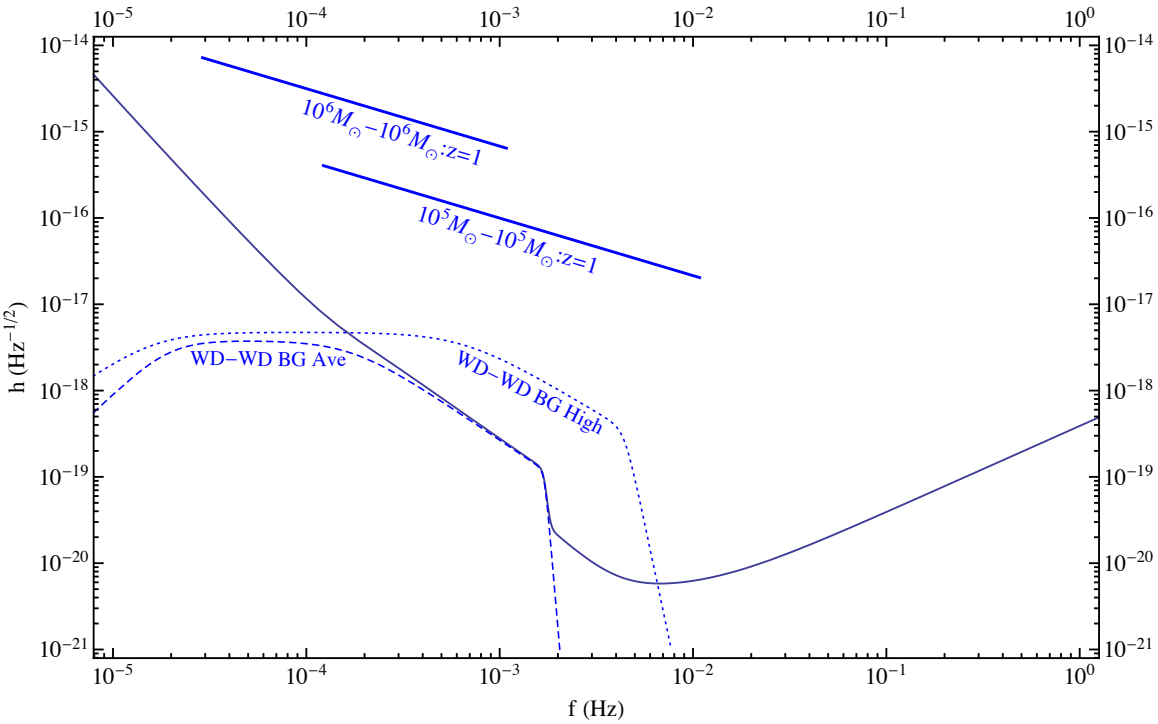


Figure 1.1: We plot the sensitivity curve for LISA with several sources relevant to this thesis (and several which aren't). The ordinate is the strain in units of $\mathrm{Hz}^{-1/2}$, and the abscissa is frequency in Hz. The average white dwarf binary background is plotted (dashed blue curve) along with a high estimate (dotted blue) [1]. Also shown are the Galactic binaries known to be in the LISA band [2]. One of the primary sources considered in this thesis are the mergers of SMBHs. We plot the signals from the final year of inspiral for equal mass mergers with $10^5\,M_\odot$ and $10^6\,M_\odot$ SMBHs at a redshift of 1 (solid blue curves). (Code to produce the figure was graciously provided by Michele Vallisneri.)

ratio (SNR) of the two it is easiest to use the Fourier transform of h, $\tilde{h}$:

$$(\mathrm{SNR})^2 = \int \frac{f\tilde{h}^2(f)}{S_n(f)} d\ln f\,. \tag{1.2}$$

The integrand is then the SNR^2 in a logarithmic band of f, so that

$$\mathrm{SNR} = \frac{\tilde{h}\sqrt{\Delta f}}{\sqrt{S_n(f)}}\,. \tag{1.3}$$

Thus, to compare signal from a source to the sensitivity (or, more pejoratively, noise) of LISA we plot the $\sqrt{S_n(f)}$ curve in plots such as Fig. 1.1 (plotted in solid black) along with the signal strength from the given source $\tilde{h}\sqrt{f}$. On a log-log plot, the difference between these two curves is then the SNR.

[3] calculated $\tilde{h}$ for circular binaries

$$\tilde{h}(f)\sqrt{f} \propto \frac{c}{D_L(z)}\left(\frac{G\mathcal{M}}{c^3}\right)^{5/6} f^{-7/6}\sqrt{f} = \frac{c}{D_L(z)}\left(\frac{G\mathcal{M}}{c^3}\right)^{5/6} f^{-2/3}\,, \tag{1.4}$$

where $D_L(z)$ is the luminosity distance to the source and $\mathcal{M}$ is the redshifted chirp mass given by

$$\mathcal{M} = (1+z)\nu^{3/5}(M_1+M_2)\,, \tag{1.5}$$

with z being the redshift of the source, $M_{1,2}$ being the masses of the compact objects, and $\nu = M_1M_2/(M_1+M_2)^2$ being the symmetric mass ratio.

Thus we see that the signal is proportional to the mass to 5/6 power, the frequency to the $-2/3$ power, and the inverse distance to the source, as long as the orbits in the binary are circular. That is, the signal $S(f)$ obeys

$$S(f) \propto \nu^{1/2}\frac{(M_1+M_2)^{5/6}}{D_L(z)f^{2/3}}(1+z)^{5/6}\,, \tag{1.6}$$

where ν ranges from $q \equiv M_2/M_1 \leq 1$ in the extreme mass ratio limit to 1/4 in the equal mass limit.

In general, in the Newtonian limit the orbital frequency at some separation d is given by

$$\Omega = \sqrt{\frac{G(M_1+M_2)}{d^3}}\,. \tag{1.7}$$

Using this to approximate the gravitational wave frequency from the merger of a binary of compact objects at the final stage of the circular inspiral we write

$$f_{\rm final} = \frac{2\Omega}{2\pi} = \frac{1}{\pi}\sqrt{\frac{G(M_1+M_2)}{d^3_{\rm final}}}\,, \tag{1.8}$$

where the extra factor of 2 is due to the fact that the gravitational wave frequency is twice the orbital frequency.

SMBH-SMBH Mergers – Cosmological Sources: Super-massive black holes (SMBHs) are defined as black holes with masses larger than $\sim 10^5\,M_\odot$ and lie at the centers of most galaxies. Observational evidence and other specific details for these large but compact objects will be discussed in more detail in Sec. 1.3. These are the objects of the greatest relevance to this work.

An object inspiraling on a quasi-circular orbit to a black hole can remain circular only until the innermost stable circular orbit (ISCO), which in the case of Schwarzschild black holes is given by $6\,GM/c^2$. Thus, the final frequency for the merger of two black holes would be given by approximately

$$f^{\rm SMBH}_{\rm final} \approx \frac{1}{\pi 6^{3/2}}\frac{1}{G(M_1+M_2)/c^3} \approx 2\times 10^{-2}\left(\frac{M_{\rm tot}}{2\times 10^5 M_\odot}\right)^{-1}\,{\rm Hz}\,. \tag{1.9}$$

Given that the documented locations of SMBHs are the centers of galaxies, the one sure way to

have a SMBH merger is following a merger of galaxies. Such mergers occur primarily at cosmological distances and so we use as our fiducial value $z = 1$ ($D_L \approx 6.6$ Gpc). In Fig. 1.1 we plot Eqn. 1.4 for an equal mass merger of two $10^6\,M_\odot$ and two $10^5\,M_\odot$ SMBHs (solid blue). The final frequency of the circular inspiral (just prior to merger) of the $10^5\,M_\odot$ merger is close to the peak sensitivity of LISA. In this case, with the $1/r$ scaling of the signal (Eqn. 1.6) these mergers can be detected by LISA virtually anywhere in the observable universe where such a merger would be expected to be.

EMRI Mergers – Marginally Cosmological Sources: Looking at the signal for a $10^5\,M_\odot$ equal mass merger in Fig. 1.1 and examining Eqn. 1.6 it is clear that the merger is still detectable even if we decrease the mass of one of the SMBHs. Indeed, at a redshift of $z = 1$, the secondary black hole could still be easily detected if it were as small as $1\,M_\odot$, as long as the secondary is on a circular orbit. In fact, in this case the secondary need not be a black hole, but could also be a neutron star or white dwarf, since both of these compact objects also have masses of order $1\,M_\odot$ (discussed below). In general, a merger where the ratio of the secondary to the primary is $\ll 10^{-3}$ is referred to as an an extreme mass ratio inspiral (EMRI).

Unfortunately, most mechanisms of EMRI formation produce EMRIs which initially do not have circular orbits but instead highly eccentric orbits (see Sec. 4.1 for a brief discussion and the rest of that chapter for an in-depth discussion of one example). This provides several key challenges in that the signal is significantly more complicated than that of a circular orbit. On the other hand, more information can be extracted from their signals. It is currently believed that SBH-SMBH mergers can be seen by LISA to a redshift of $z \approx 1$ (assuming a $10\,M_\odot$ SBH), while WD-SMBH mergers should be detectable to $z \approx 0.2$ (assuming a $0.6\,M_\odot$ WD) and NS-SMBH mergers to a similar distance [4].

It is also interesting to note that EMRIs have final frequencies which are larger than those for equal mass SMBH mergers by only a factor of 2 (Eqn. 1.9), so the EMRI signal still lies in the LISA band.

SBH-SBH Mergers - Local Extra-Galactic Sources: Stellar mass black holes are the final remnant of only the most massive stars, though their formation mechanisms and masses are not yet well understood [5]. Their final masses are expected to be approximately $8\,M_\odot$ but can vary between $\approx$ 3–30 $M_\odot$ [6]. Though their masses tend to be greater than most stellar objects and stellar remnants, their number densities are far lower, though not necessarily much different than NSs. Along with the less probable occurrence that two SBHs are born in a binary tight enough to merge in 10^{10} yr, this makes their mergers far less probable, and pushes their rates into a region where it is unlikely that we would have a SBH-SBH merger in the Milky Way over a decade of observing. The most likely detection would come from the nearby Virgo cluster, which sits at approximately 14 Mpc from the Milky Way.

However, taking $8\,M_\odot$ as the fiducial mass of a SBH, the final frequency before merger is Eqn. 1.9

$$f_{\rm final}^{\rm SBH} \approx 2 \times 10^2\,{\rm Hz}\,. \tag{1.10}$$

This actually places these sources outside of the LISA band (and actually in the LIGO band).

SBH-NS-WD Mergers – Galactic Sources: Of course, there is no requirement that such a merger of stellar mass compact remnants contains only SBHs. They may also contain neutron stars and white dwarfs. Both neutron stars and white dwarfs are expected to have masses of approximately $1\,M_\odot$ [7].

White dwarfs are by far the most numerous of the compact remnants. So numerous in fact that WD-WD binaries should actually be a source of confusion noise for LISA at certain frequencies and strains (shown in Fig. 1.1).

However, there are several known "verification" binaries which have either WDs or NSs which should be detectable to LISA and therefore provide an almost guaranteed detection of gravitational waves. These are shown in Fig. 1.1.

1.3 Super-Massive Black Holes

1.3.1 SMBH Existence

There is a growing body of evidence that large SMBHs lurk at the center of most galaxies. The most compelling case is Sgr A* at the center of our own Galaxy. Measurements of the orbits of the "S-stars", main sequence B-stars with mass $\sim 10\,M_\odot$ and lifetimes few $\times\,10^7$ yr in the central 0.5" ($\approx 1\,$pc at $8\,$kpc), show that they are orbiting an object coincident with the radio source Sgr A* which has mass $(4.1 \pm 0.6) \times 10^6\,M_\odot$ [10]. In addition, the eccentric orbit of the star S2 at periapsis constrains the size to be less than 45 a.u. or 600 Schwarzschild radii of a $4 \times 10^6\,M_\odot$ black hole [11]. The remarkable orbit orbit of S2 about Sgr A* is shown in Fig. 1.2.

However, the first evidence of the existence of SMBHs goes back to the discovery of active galactic nuclei (AGN). These were the extremely luminous galaxy cores, which outshone even the host galaxy, that were visible to very high redshift. Even today it is hard to explain these extreme luminosities for such long durations without invoking the existence of SMBHs [12].

1.3.2 Active Galactic Nuclei

The best explanation for these ultra-luminous galaxy cores was gas accreting onto central SMBHs. The unified model of AGN (see Fig. 1.3) states that the many different types of AGN that have been observed are all very similar objects with the primary distinction being the direction. Fig. 1.3 crudely outlines the basic principles of the theory.

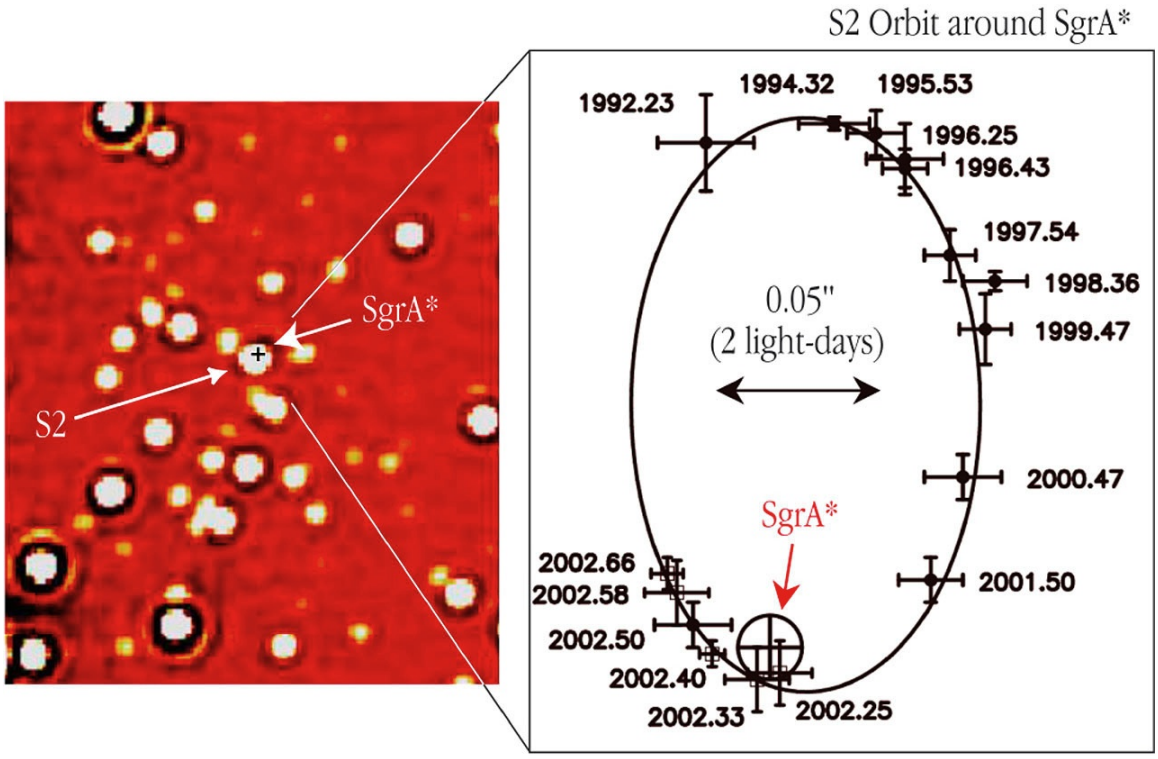


Figure 1.2: Convincing evidence for the existence of SMBHs: The star S2 is tracked over time using adaptive optics. The orbit of S2 passes within 600 Schwarzschild radii of a $4 \times 10^6\, M_\odot$ SMBH. S2 has also been tracked separately by [8] and the two results agree well [9]. (Released by ESO website under Creative Commons Attribution.)

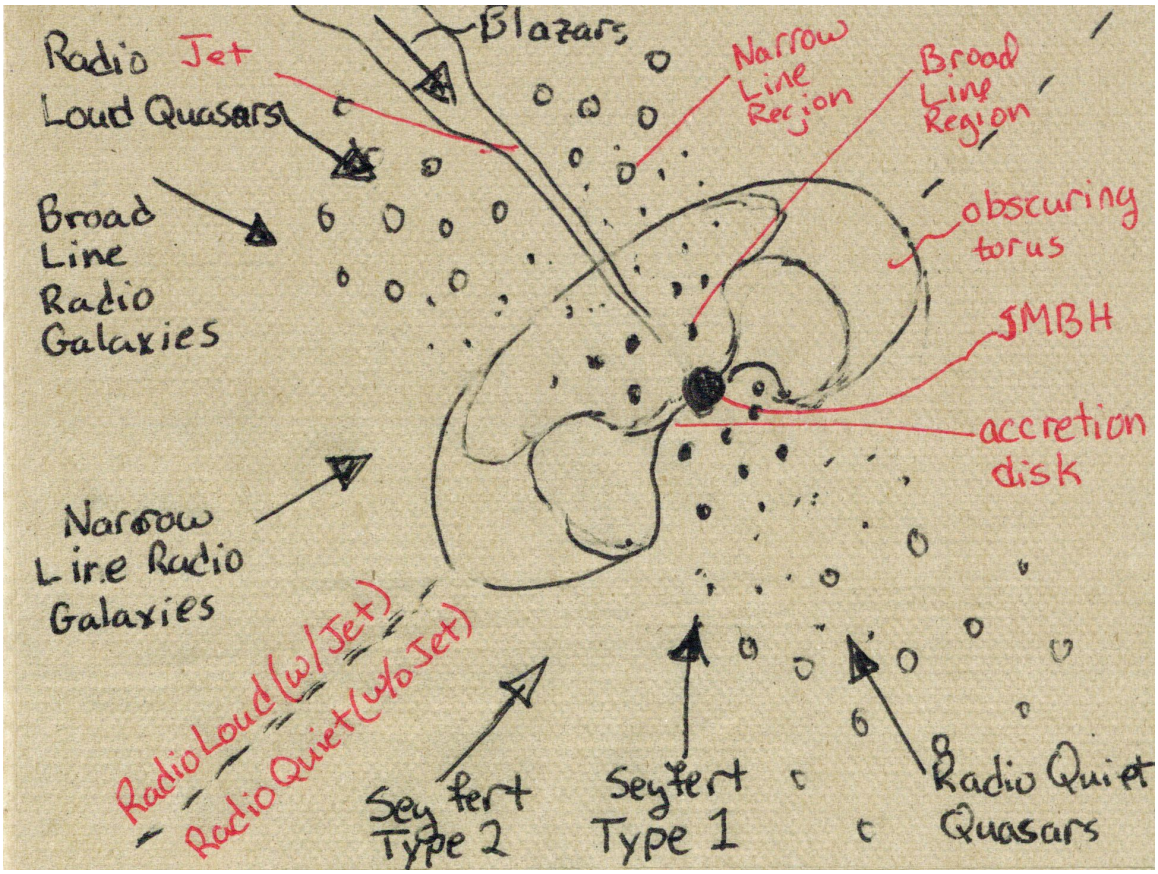


Figure 1.3: AGN unification theory: The principle of the AGN unification theory is that many of the objects AGN which were thought to be different because of their very different features are all gas being accreted onto a SMBH with the differences coming from the viewing angle and whether there exists a jet in the system. The presence of the jet distinguishes between whether the AGN is radio loud or radio quiet. As gas accretes onto the SMBH it heats up and radiates at a significant fraction of the Eddington luminosity. Gas close to the SMBH, but not in the accretion disk, is traveling at high velocities, and when illuminated by this bright accretion disk absorbs and reradiates this energy in doppler shifted lines. Farther out clumps of gas are traveling more slowly and thus re-emit the energy sourced from the accretion disk in more narrow lines than the faster moving gas closer to the SMBH. In some cases, the accretion produces strong relativistic jets in one or two directions, while in others this does not happen. In this picture the final distinguishing characteristic between different types of AGN is the viewing angle.

As gas accretes onto the SMBH it heats up and radiates. This radiation then exerts a pressure on the material in the disk falling inwards, and thus slows it down. Given this basic feedback mechanism there is a maximum steady-state rate at which gas can fall into the SMBH which is calculated by equating the radiation force on the gas to the gravitational force on the gas [there are a multitude of sources for similar discussions, but we follow 13]. In the case of spherical optically thick accretion the radiation force per unit volume is just given by

$$f_{\rm rad} = \sigma_{\rm T} n_e \int \frac{L_\nu}{4\pi c r^2} d\nu = \frac{\sigma_{\rm T} n_e L}{4\pi c r^2}\,, \tag{1.11}$$

where $\sigma_{\rm T}$ is the Thompson scattering cross section, n_e is the electron number density, and L is the bolometric luminosity. On the other hand the gravitational force per unit volume acting on the gas is

$$f_{\rm grav} = \frac{G M_\bullet \mu_e n_e}{r^2}\,, \tag{1.12}$$

where $M_\bullet$ is the mass of the SMBH, and μ_e is the mean mass per electron. Equating the gravitational force to the radiation pressure force and solving for L gives:

$$L_{\rm Edd} = \frac{4\pi c G M_\bullet \mu_e}{\sigma_{\rm T}} = 1.5 \times 10^{43} \frac{M_\bullet}{10^5 M_\odot}\,{\rm erg\,s^{-1}}\,. \tag{1.13}$$

This is the maximum luminosity which can be radiated in spherical steady state accretion, modulo radiation bubbles. Note that for a modest $10^5\,M_\odot$ SMBH this is a whopping $10^{43}\,{\rm erg\,s^{-1}}$, a luminosity comparable to that of a large galaxy with 10^{10} stars. For a $4 \times 10^6\,M_\odot$ SMBH like Sag A* at even a fraction of the Eddington luminosity it would outshine the Milky Way. It is this reason that explains the super-bright AGN.

It is interesting to note that the Eddington luminosity also corresponds to a maximum accretion rate for steady state spherical accretion. If we assume that mass falling in to the SMBH is converted to radiation with an efficiency η, then

$$\dot{M}_{\rm Edd} = \frac{L_{\rm Edd}}{\eta c^2}\,. \tag{1.14}$$

In reality, the flow will not likely be accreting at the Eddington rate, but instead will have some lower luminosity but with a similar η:

$$\dot{M} = \frac{L}{\eta c^2} = \frac{4\pi G \mu_e}{\eta \sigma_{\rm T} c} \frac{L}{L_{\rm Edd}} M\,. \tag{1.15}$$

That is, if $L/L_{\rm Edd}$ remains constant over time we have exponential growth with characteristic time (Salpeter time)

$$t_{\rm growth} = \frac{\eta \sigma_{\rm T} c}{4\pi G \mu_e} \frac{L_{\rm Edd}}{L} = 3.7 \times 10^8 \eta \frac{L_{\rm Edd}}{L}\,{\rm yr}\,. \tag{1.16}$$

Thus, assuming maximal accretion and the fiducial value of $\eta = 0.1$, a SMBH will double its mass every 3×10^7 yr. However, more realistic luminosities would be less.

1.4 SMBH Characteristics

There are several surprising relationships between the SMBHs in galaxy centers and large scale characteristics of the galaxy. These are surprising because, although the black holes have seemingly large masses, their sphere of gravitational influence given by

$$r_{\rm inf} = \frac{GM_\bullet}{\sigma^2} \tag{1.17}$$

is typically $1 - 10$ pc and contains only $2\,M_\bullet$ of the bulge mass which is on average $\sim 1000\,M_\bullet$[14].

Probably the most important relationship is that between the SMBH mass and the velocity dispersion of the bulge of the galaxy. Dubbed the M–σ relation, it says that the velocity dispersion σ and the SMBH mass follow [15]

$$\frac{M_\bullet}{10^8 M_\odot} = (1.32 \pm 0.27) \left(\frac{\sigma}{200 \mathrm{km\,s^{-1}}}\right)^{4.24\pm0.41} . \tag{1.18}$$

The Faber-Jackson relation relates the bulge luminosities and velocity dispersions, thus giving a relation between the SMBH mass and the bulge luminosity referred to as the $M_\bullet$–L relation. This is quite fortunate because dispersions are hard to measure, while bulge luminosities are quite straightforward. [15] provide the most recent values for the relation:

$$\frac{M_\bullet}{10^8 M_\odot} = (8.91 \pm 2.57) \left(\frac{L_V}{10^{11} L_{\odot,V}}\right)^{1.11\pm0.18} . \tag{1.19}$$

Because there is a relationship between the SMBH mass and the bulge luminosity, and also a relationship between the bulge luminosity and its mass, one may relate the mass of the SMBH to the mass of the bulge.

Of these relations the M–σ relation has the lowest scatter, and so is likely the fundamental relation [14].

1.5 SMBH Densities

The $M_\bullet$–L relation is particularly useful, since the luminosities of elliptical galaxies are easy to measure. This allows for the possibility of measuring the SMBH densities as a function of redshift using mechanized galaxy surveys. [16] do a similar process where they determine the velocity dispersion for a large sample of galaxies by using Hubble-type specific luminosity functions to relate the local

luminosity to the velocity dispersion. Using this method they find that the density of SMBHs is given by

$$\rho_\bullet = (2.4 \pm 0.8) h_{71}^2 \times 10^5 M_\odot \, \mathrm{Mpc}^{-3} \,, \tag{1.20}$$

where $h_{71} = H_0/71$ km s^{-1} Mpc^{-1} and H_0 is the Hubble constant [17].

This density is fundamental in determining the rates of GW sources from SMBHs detectable for LISA and has been used in [18] and in Chapter 5 of this text.

1.6 SMBHs and This Work

One of the primary sources for LISA are SMBH-SMBH mergers. The luminosities of these mergers in gravitational waves can well exceed the total electro-magnetic luminosity of the observable universe for a brief period of time. This makes them observable, virtually anywhere in the universe they can form.

However, LISA will not be able to localize the host galaxy of the event to much better than a square degree [19]. Having an electromagnetic counterpart significantly increases the astrophysical usefulness of the system. Relationships between the host galaxy and the galaxy merger which carried in the second SMBH could be better discerned. And, cosmologically, electromagnetic counterparts would allow for accurate determination of the redshift. This together with the luminosity distance measured by LISA from the gravitational wave signal, would give a physically clean and well understood *independent* method for determining the luminosity distance-redshift relation, and hence cosmological models and parameters.

To have an electromagnetic counterpart to a SMBH merger, there must be gas present in the system. And if there is a significant amount of gas present it is likely to be in the form of an accretion disk. However, unlike in the case of AGN, as discussed in Sec. 1.3.2, in the case of the equal mass binary accretion is expected to be minimal [20]. In the case of corotating disks coplanar with the SMBH binary, the binary exerts torque on the disk at the outer Lindblad resonances counteracting viscous losses in the disk which otherwise would drive mass inwards to ultimate accretion. This effectively clears a region directly surrounding the SMBHs and would significantly decrease the luminosity of the system.

In the case of two $10^5\ M_\odot$ SMBHs it has been found numerically [21] and analytically [22] that the binary clears a region equal to approximately twice the binary separation and causes a concomitant over-density in the inner disk of approximately an order of magnitude [22].

During the final period of the inspiral when the time to merger becomes comparable to the viscous time increasingly rapid gravitational wave shrinking will decouple the binary from the disk, and the disk will begin viscously-driven free fall. This viscous motion is slow compared to the merger time and the disk will effectively stall during the rest of the merger (see Sec. 2.2 for more details).

When the time to merger is of the order of the dynamical time of the inner disk, any change to the potential will be seen as impulsive (Sec. 2.3.1). There are two major effects: mass-energy loss due to gravitational wave radiation (Sec. 2.3.2) and recoil kicks given to the final SMBH due to asymmetric gravitational wave radiation (Sec. 2.3.3). One can compare the relative importance of these two effects on the inner disk by comparing the maximum radial velocities they induce. For all regions of the probed parameter space mass loss was dominant, though kicks became marginally important in certain regions (Secs. 2.3.2 and 2.3.3). Farther out in the disk, however, kicks always are the most relevant.

Thus, if one wishes to consider rapid emissions from the inner regions of the disk, one may largely neglect the effect of kicks. In this case the problem becomes axisymmetric. By doing the canonical vertical averaging of the disk, the problem can be reduced to a one-dimensional one.

In this case, it is straightforward to solve for the reaction of the surface density of the disk to the mass-energy loss in the limit of a pressureless disk (Sec. 2.4). The problem can be easily generalized to the case with kicks, if one is interested in effects farther out in the disk (Sec. 2.4.2). The analytic solution is surprisingly simple, and provides significant insight into the problem. In particular, one can derive the approximate time that a real gaseous flow (with pressure) will deviate from this analytic solution, which is also where in the disk the first shocks will form (Sec. 2.5). The strength of the shocks can be well approximated, offering an order of magnitude tracer for the jumps in temperature across the shock (Sec. 2.5).

The disk reactions can also be computed numerically, a necessary complication since a fully analytic treatment of disks with finite pressure and temperature is not possible. A finite differencing Lagrangian leap-frog scheme was used to evolve the fluid equations in 1D (Sec. 3.2). Because of the numerical efficiency of a 1D code, the entirety of parameter space can easily be probed (Sec. 3.4) and analyzed (Sec. 3.3.2). Over the past several years many groups have simulated circumbinary disks to varying degrees of accuracy (see Sec. 3.1 for a discussion of these many simulations). The 1D theory and simulations can be compared to the particular simulations of [23] (which consider 2.5–3 dimensions, general relativity, and in some simulations magneto-hydrodynamics), showing excellent agreement in the regions of parameter space simulated by [23].

Even after more than 30 years of observation of AGN, and theoretical effort to understand those observations, there still remain significant discrepancies. As such, there is still not universal agreement on initial conditions of the circumbinary accretion disk. Hence, there is great power in analytic (Chapter 2) and 'simple' numerical simulations (Chapter 3) that elucidate the key physical ingredients and physics in a well understood manner.

However, as noted in Sec. 1.2.2 there are many other types of inspirals to SMBHs, such as those by compact objects (COs) such as SBHs, NSs, and WDs. These extreme mass ratio inspirals would be challenging for LISA to see, but would provide a significant amount of information. There are

several known formation channels for EMRIs, including COs scattered by other stars leading to a slow random walk of their angular momentum until some become highly eccentric, and gravitational radiation emitted as they pass close to the SMBH leads to their inspiral and ultimate merger with the SMBH. Other less common channels are discussed briefly in Sec. 4.1.

In this thesis we describe a new channel of EMRI formation: perturbation by a secondary SMBH (Chapter 5). When two galaxies merge, the SMBH of the smaller galaxy, likely hosting a less massive SMBH (which we refer to as the secondary SMBH, or just secondary) inspirals due to dynamical friction until the stellar mass interior to it becomes of order its own mass. At this point the inspiral stalls.

But just as the secondary approaches its stalling radius it is entering the densest and most relaxed regions of the galaxy, and is scattering stars near the primary SMBH at an unprecedented rate (see Sec. 4.2 for a more detailed description of the process). Some of these have strong interactions with the primary and can inspiral to merger. However, most do not. Indeed, the most likely method of formation is through the secular effect of the Kozai mechanism (Sec. 4.3). We found that a previously unnoticed secular effect occurs when the relativistic precession rate of the Lenz vector becomes high (Sec. 4.4). This creates a new channel for formation of EMRIs. Ultimately, although we find the rate of EMRIs during this phase could be significantly enhanced, the short duration of this phase results in a total event rate approximately 10^{-3} times less common than EMRIs formed through the standard star-star scattering picture (Sec. 4.7).

Chapter 2

Time Evolution of Circumbinary Disks Following Super-Massive Black Hole Mergers

N.B.: This work will be published with the following authors: Sterl Phinney and Nate Bode in this order.

Abstract

This is the first chapter of a two-chapter series discussing the time evolution, of circumbinary disks following the merger of a super-massive black hole (SMBH) binary in the limit of a thin disk. Here, in Chapter 2, in the limit of a noninteracting disk, we provide an analytic solution to the circumbinary disk's evolution taking into account both mass-energy loss during the final inspiral and the recoil of resulting SMBH. We also show that given black hole spins aligned or anti-aligned with the orbital angular momentum, mass-energy loss is the dominant effect on the disk. Finally, we describe where the noninteracting assumption breaks down, approximate the strength of the shocks which inevitably form on this boundary, and describe the final density distribution, thus fully characterizing the dynamics of the disk relevant to observability. In Chapter 3 we use simulations to both demonstrate the relevance of the solutions given here, but also to survey the entirety of the relevant parameter space. These results are especially useful in the context of electromagnetic counterparts to SMBH mergers in the context of the Laser Interferometer Space Antenna (LISA).

2.1 Introduction

Over the past several years there has been a growing interest in understanding electromagnetic (EM) counterparts to gravitational wave (GW) signals. The reason is simple: they extend the scientific

reach of current (LIGO, VIRGO) and upcoming (LISA, DECIGO, Einstein Telescope) gravitational wave detectors. Because GW detectors have relatively poor angular resolution they cannot pinpoint their sources, making the primary importance of the EM counterpart to locate the host galaxy. Once the host galaxy is known, it may be followed with many different telescopes to watch for relationships between the host's EM signatures and the GW signal.

In this two-chapter series, we consider EM counterparts to GW signals from merging supermassive black holes (SMBHs), a signal relevant to LISA. The importance of EM counterparts in the LISA context has been well documented for Astro2010 [24–26]. There are many possible mechanisms [27–35], but we concern ourselves with a thin circumbinary disk reacting to the sudden reduction in the SMBH's gravitational mass due to mass-energy carried away by gravitational radiation (hereafter referred to in shorthand as "mass-energy loss"), and to the recoil kick due to asymmetric emission of GWs (hereafter referred to as "kick"). There has already been a great deal of work on such disks [21, 23, 36–39], but only a modest amount of analytic treatment [22, 40, 41].

These disks are a possible outcome of the SMBH binary formation process. Following the merger of two galaxies, the SMBHs, which are harbored in the cores of most galaxies [42, 43], will inspiral due to dynamical friction until GW radiation dominates, takes over, and propels the binary to its final merger [44–46]. The merger of galaxies is also expected to cause gas to migrate towards the galactic center [47]. When this happens a circumbinary disk should form, while quadrupolar torques from the binary and viscous torques within the disk will act to align the orbital angular momenta of the disk and binary [48]. The inner regions of the disk become evacuated as the binary exerts torques on co-rotating disks at Linblad resonances [21]. While the viscous time of the inner disk is shorter than the inspiral time of the binary, the disk will follow the binary inwards, while afterwards the binary will inspiral quickly, leaving the circumbinary disk in a pseudo-stationary state.

Here, in Chapter 2, we present an analytic solution to the evolution of a circumbinary disk with time, breaking the disk into three regions: a pre-shock non-interacting region, the shock(s) region, and the final post-shock region. We provide an analytic result for the surface density as a function of time in the regions preceding and following the shocks and give both the approximate locations of the shocks and a strong lower bound to their strength. In Chapter 3, we follow with an illustration of the solution presented here, in the case of pure mass-energy loss, using a 1D hydrodynamical simulation. The entirety of the relevant parameter space is explored and described, and a straightforward method is provided for using the solution given here to determine the type of flow produced given the merger and disk parameters.

Our chapter is presented as follows: (a) we first describe the circumbinary disk and the relevant distance and time scales in Sec. 2.2. (b) In Sec. 2.3 we describe why one can consider the mass-energy loss during merger in the impulse approximation (Sec. 2.3.1), and summarize the predictions of numerical relativity for mass-energy loss (Sec. 2.3.2) and recoil kicks, and show that recoils are at

most marginally relevant to the inner disk for fiducial disk parameters (Sec. 2.3.3). (c) The analytic solution to the non-interacting region preceding the shocks is then presented in the special case of the axisymmetric mass-energy loss (Sec. 2.4.1), and then in the general case of mass-energy loss with kicks (Sec. 2.4.2). (d) The solutions for the case of pure mass-energy loss form a two-parameter family, defined by the fractional mass-energy loss $\delta M/M_i$ and the disk aspect ratio h/r. However, it is more insightful to consider the parameter space defined by h/r and a new parameter, the epicyclic Mach number, M_e. We define M_e (Sec. 2.5.2) and relate it and the h/r to the time it takes for the flow to deviate from the preshock solution (Sec. 2.5.3). (e) Finally, we provide a strong lower bound for the Mach number of the shock (Sec. 2.5.5), give an analytic expression for the final post-shock quasi-steady-state surface density (Sec. 2.5.7), and generalize Secs. 2.5.2–2.5.7 to disks with kicks (Sec. 2.5.8).

For the reader in a hurry, the primary results of this chapter are summarized as follows: Figs. 2.5 and 2.6 outline the relative importance of mass-energy loss and kicks in the case of non-spinning SMBH and SMBHs with spins aligned and anti-aligned with their orbital angular momentum. The relevant preshock equations are Eqns. 2.30–2.31 for negligible kicks and Eqns. 2.62–2.63 when kicks are imparted, which give the time evolution of the circumbinary disk in Lagrangian coordinates. In Eqns. 2.104 and 2.120 the deviation/shock-formation time is again given for the cases of pure mass-energy loss and both mass-energy loss and kicks, while for the same two cases the shock Mach numbers are given in Eqns. 2.109 and 2.121. For reference, a table of notation in the chapter is given in the conclusion.

2.2 The Circumbinary Disk

The circumbinary disk is substantially different from a standard accretion disk, since the binary exerts torques on the disk at the Lindblad resonances. While these torques inject angular momentum into the disk and effectively excrete material from the region of the binary (thus their alternative name: excretion disks), the disk is also undergoing viscous losses due to viscous torques. Together these effects produce a region of possibly significantly enhanced density at a radius $r_{\rm inner}$ inside of which the disk is effectively evacuated. For an equal mass binary [21] found $r_{\rm inner} \approx 2r_{\rm bin}$ where $r_{\rm bin}$ is the binary separation.

However, the disk is not in steady state, since the binary is inspiraling. Instead the disk has several dynamically distinct regions:

The first is an inner region with viscous timescale less than the merger time which is in steady state and following the binary inwards. The radial position where the viscous timescale equals the time to merger, denoted $r_{\rm steady}$, separates this region from outer regions. The viscous timescale is

given by

$$t_{\rm visc} = \frac{2}{3}\frac{r^2}{\nu(r)} = \frac{1}{\alpha}\frac{1}{(h/r)^2} t_{\rm dyn}\,, \tag{2.1}$$

where $\nu(r)$ is the kinematic viscosity and we have assumed a standard α-accretion disk prescription [49] in the final equality. On the other hand, for a quasi-circular inspiral, the time to merger is given by

$$t_{\rm merger} = \frac{5}{256}\frac{M}{M_2}\left(r_{\rm bin}^4 - r_{\rm ISCO}^4\right)\,, \tag{2.2}$$

where t is in units of GM/c^3, $r_{\rm bin}$ is in units of GM/c^2, $M = M_1 + M_2$, $M_2 \leq M_1$, and $r_{\rm ISCO}$ is the position of the innermost stable orbit. Solving Eqn. 2.1 equal to Eqn. 2.2 for r gives $r_{\rm steady}$. However, $r_{\rm steady}$ depends strongly on the viscosity parameterization and disk parameters.

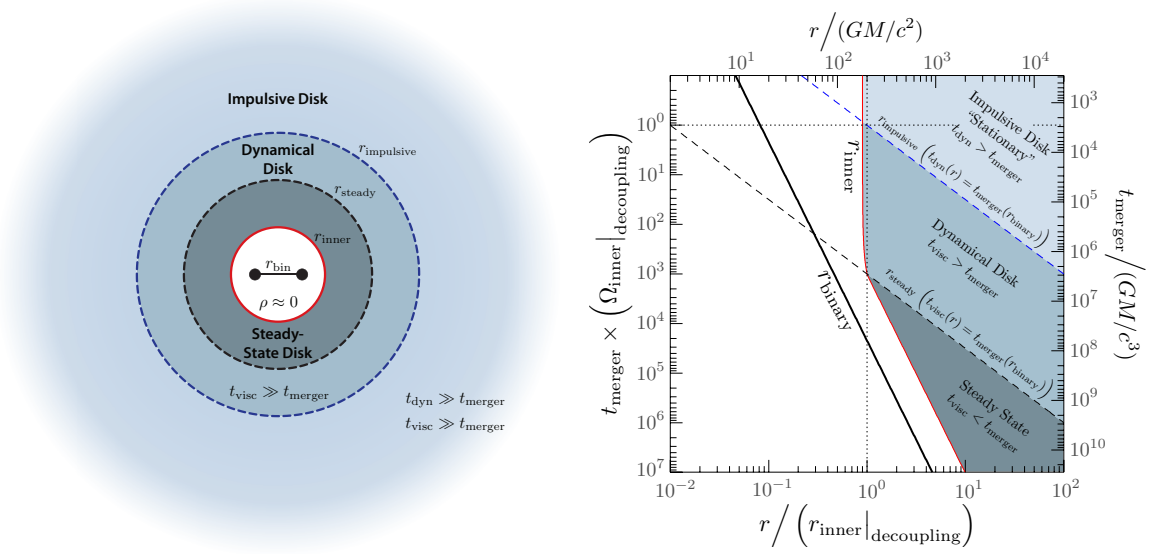


Figure 2.1: Left: This cartoon of the circumbinary disk illustrates the important regions of the disk: for $r < r_{\rm inner}$ the density is approximately 0 and gas has little effect on any dynamics of the disk, for $r_{\rm inner} < r < r_{\rm steady}$ the disk follows the binary adiabatically and retains essentially the same shape as in the steady state solution, for $r > r_{\rm steady}$ the disk has decoupled from the binary and will begin a process of changing its structure to that of a standard accretion disk over a timescale comparable to $t_{\rm visc}$, and for $r > r_{\rm imp}$ all changes to the potential will be essentially impulsive. Right: In this sketch we give an example of how a disk would evolve, due to the dependence of $r_{\rm bin}$, $r_{\rm inner}$, $r_{\rm steady}$, and $r_{\rm imp}$ on $t_{\rm merger}$. On the abscissa we plot time to merger (currently in wrong units) while the ordinate is distance in units of the $r_{\rm inner}$ at decoupling. The vertical dotted line represents a possible time when the cartoon on the left would be appropriate. At later times the steady state region no longer exists, and when $t_{\rm merger}$ equals $t_{\rm dyn}$ of the inner disk, the entirety of the disk is in the impulsive region and all subsequent changes in the potential will be perceived as instantaneous by the disk.

The second dynamical region is that where all subsequent changes in the potential are seen as

impulsive and is defined to be when $t_{\rm dyn} > t_{\rm merger}$. Since

$$t_{\rm dyn} = \frac{1}{\Omega(r)} = \sqrt{\frac{r^3}{GM}}\,, \tag{2.3}$$

we can use to Eqn. 2.2 to find $r_{\rm imp}$, the inner boundary of the impulsive region in units of GM/c^2:

$$r_{\rm imp} \sim \left(\frac{5}{256}\frac{M}{M_2}\right)^{2/3} \left(r_{\rm bin}^4 - r_{\rm ISCO}^4\right)^{2/3}\,. \tag{2.4}$$

Some values of $r_{\rm imp}$ and $qt_{\rm merger}$ are tabulated for specific values of $r_{\rm bin}$ in units of GM_i/c^2 and GM_i/c^3 in Table 2.2.

The final intermediate region where material lies between $r_{\rm steady}$ and $r_{\rm imp}$, and is neither in steady state nor effectively "unmoving". Depending on how one parameterizes the scale height of the disk there may exist times when this region does not exist, because $r_{\rm steady} > r_{\rm imp}$.

Note that because $t_{\rm merger}$ is constantly changing, and because $t_{\rm dyn}$ and $t_{\rm visc}$ scale differently with r, the regions are in constant flux. For a schematic of the disk structure and evolution see Fig. 2.1.

For the purposes of many figures and illustrative calculations we use a binary mass ratio $q = 1$, which has been shown to decouple from the binary at a binary separation of $r_{\rm bin} = 100\,GM/c^2$, for a fiducial disk [36]. Based on fluid simulations [21] have found that inner radius of the disk is $r_{\rm inner} \approx 2$. We scale most distances to be in units of the length scale of the disk, $r_{\rm inner}$, and, whenever it is necessary to perform a calculation, we assume the fiducial value of

$$r_{\rm inner} = 200\frac{GM}{c^2}\,. \tag{2.5}$$

The relevant dynamical time to the disk is then just $1/\Omega_{\rm inner}$, where $\Omega_{\rm inner}$ is the orbital frequency at $r_{\rm inner}$. Thus, we use

$$\Omega_{\rm inner} = \sqrt{GM/r_{\rm inner}^3} = 2^{3/2}10^3\frac{GM}{c^3}\,, \tag{2.6}$$

when calculations are needed.

2.3 Changes in the Potential

The evacuated central region of the disk suggests a natural timescale for the disk: the dynamical timescale at $r_{\rm inner}$. Any change in the potential which occurs on timescales shorter than this the disk will perceive as impulsive. In the case of SMBH binaries there are two such changes: the mass-energy loss due to gravitational wave emission and the final recoil given to the subsequent SMBH due to asymmetric GW radiation. Here, in the following subsections, we first clarify why the impulse approximation is valid for mass-energy loss. We then give a brief survey of the possible mass-energy

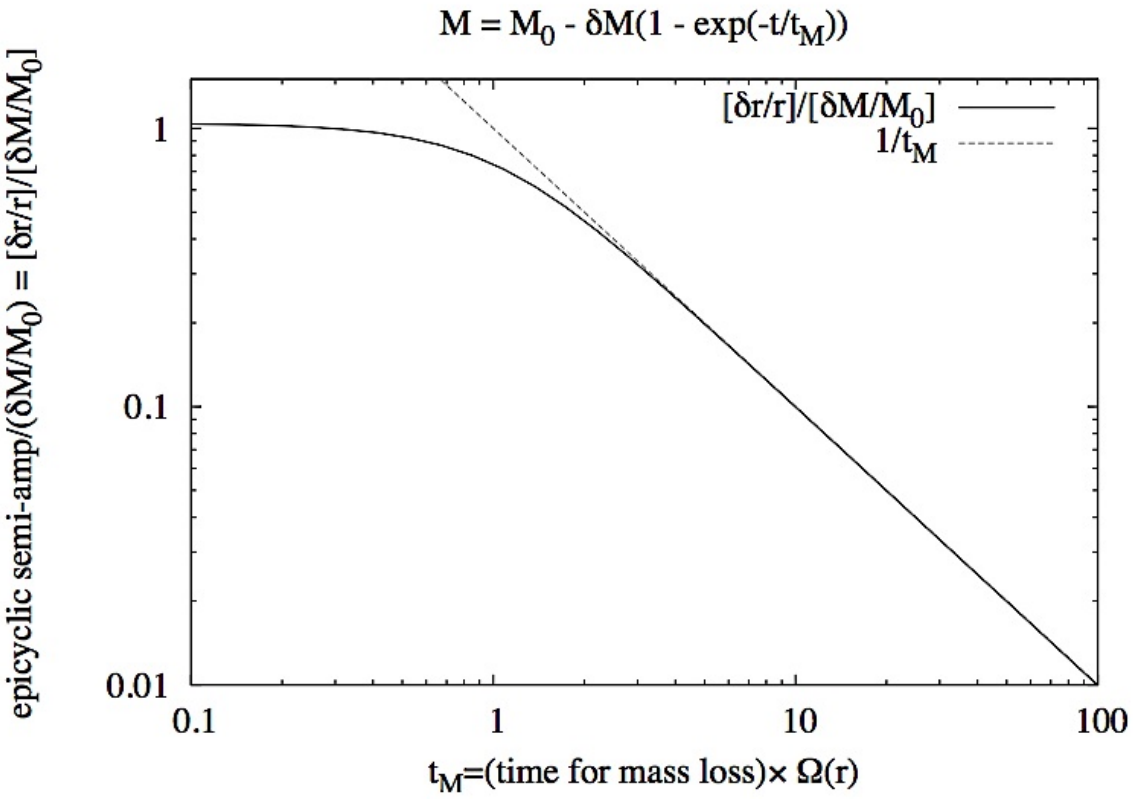


Figure 2.2: Solid curve: amplitude δr of radial epicyclic motion induced in previously circular orbits of orbital frequency Ω and radius r, by mass-energy loss δM on a timescale $\tilde{t}_M$. When the mass-energy loss occurs on a timescale short compared to the orbital time ($\Omega\tilde{t}_M \ll 1$), $\delta r/r = \delta M/M$. When the mass-energy loss occurs on a long timescale compared to the orbital time ($\Omega\tilde{t}_M \gg 1$), the amplitude is given by the asymptotic result $\delta r/r \to (\delta M/M)(1/\Omega\tilde{t}_M)$ shown by the dashed line.

loss, taking into account both the inspiral and the final several orbits. And finally we discuss kicks and their importance to these systems.

2.3.1 The Impulse Approximation

To understand how the mass-energy loss affects a particle in the circumbinary disk we present a simple toy model. Consider a particle in the disk on a circular orbit located a distance r from the binary's center of mass. When the timescale for mass-energy loss in gravitational radiation, $\tilde{t}_M$, is large compared to the dynamical timescale of the gas, $1/\Omega(r)$, then the orbits of the gas will preserve adiabatic invariants: the orbits remain circular while they slowly expand as $r \propto 1/M(t)$ [50]. However, if the mass-energy loss timescale is small, $\Omega\tilde{t}_M \ll 1$, then an eccentricity (epicyclic motion) will be induced, with radial semi-amplitude of $\delta r/r = \delta M/M$.

We demonstrate this with a toy model $M = M_i - \delta M\{1 - \exp(-t/\tilde{t}_M)\}$ with $\delta M < 0$. This

toy form of mass-energy loss is initially linear in time, a case which can be treated analytically [51]. In Fig. 2.2 we plot the ratio of the semi-amplitude of the induced epicyclic motion to $\delta M/M_i$ when $t \to \infty$ as a function of $\tilde{t}_M\Omega$, where Ω is the Keplerian orbital frequency for some particle of interest. The limiting cases are easy to understand. When $\tilde{t}_M\Omega \ll 1$ the mass-energy loss is essentially instantaneous and the impulse approximation is relevant and gives that

$$\frac{\delta r}{r} \sim \frac{\delta M}{M} \equiv \epsilon_{\rm m}\,, \tag{2.7}$$

as shown in Sec. 2.4.1 and given in Eqn. 2.29. When $\tilde{t}_M\Omega \gg 1$ the limiting behavior becomes transparent if we notice that the epicyclic motion will be induced for the maximum mass-energy loss, ΔM_Ω, which occurs on a timescale comparable to $1/\Omega$:

$$\frac{\delta r}{r} = \frac{\Delta M_\Omega}{M} \sim \frac{1}{M}\frac{\dot{M}_\Omega}{\Omega} \approx \frac{\delta M}{M}\frac{1}{\tilde{t}_M\Omega}\,, \tag{2.8}$$

where $\dot{M}_\Omega$ is the maximum rate of mass-energy loss on a timescale $1/\Omega$. The first equality follows since ΔM_Ω is the largest mass such that Eqn. 2.7 loosely holds. In the second relation we use that $\Delta M_\Omega \sim \dot{M}_\Omega/\Omega$, while in the last relation we merely solve for $\dot{M}_\Omega$ by differentiating $M(t)$ with respect to time.

This can also be shown in a real world problem similar to the one considered here, a particle reacting to the mass-energy being radiated during an extreme mass ratio inspiral (EMRI). In this case there is an analytic solution to the mass radiated during inspiral [52, 53]. Using Eqn. 4.18 of [53] we integrate the path of a particle originally orbiting the primary SMBH on a circular orbit with semimajor axis $r_0 = 200\,GM/c^2$ as the secondary inspirals and ultimately merges. The resulting evolution is plotted in Fig. 2.3. If one considers only the mass lost on timescales shorter than $1/\Omega(r_0)$, denoted $\Delta M_{\rm effective}$, then the resulting low eccentricity orbit is well described by epicyclic motion with radial semi-amplitude given by $\delta r/r_0 \approx \Delta M_{\rm effective}/M_i$.

2.3.2 Mass-Energy Loss

A significant fraction of a SMBH binary's total initial mass-energy is lost in GWs throughout the process of merger. Calculating the mass-energy loss naturally breaks up into two parts: the inspiral phase which is accessible analytically using post-Newtonian theory for quasi-circular orbits, and the merger phase where simulations must be used to solve the highly nonlinear last few orbits and final plunge:

$$\frac{\delta M}{M_i} = \frac{E_{\rm inspiral}}{M_i} + \frac{E_{\rm merger}}{M_i} \tag{2.9}$$

where M_i is the initial total mass, $E_{\rm merger}$ is the mass energy radiated during the final several orbits, the merger, and the subsequent ringdown, and $E_{\rm inspiral}$ is the mass-energy radiated prior to

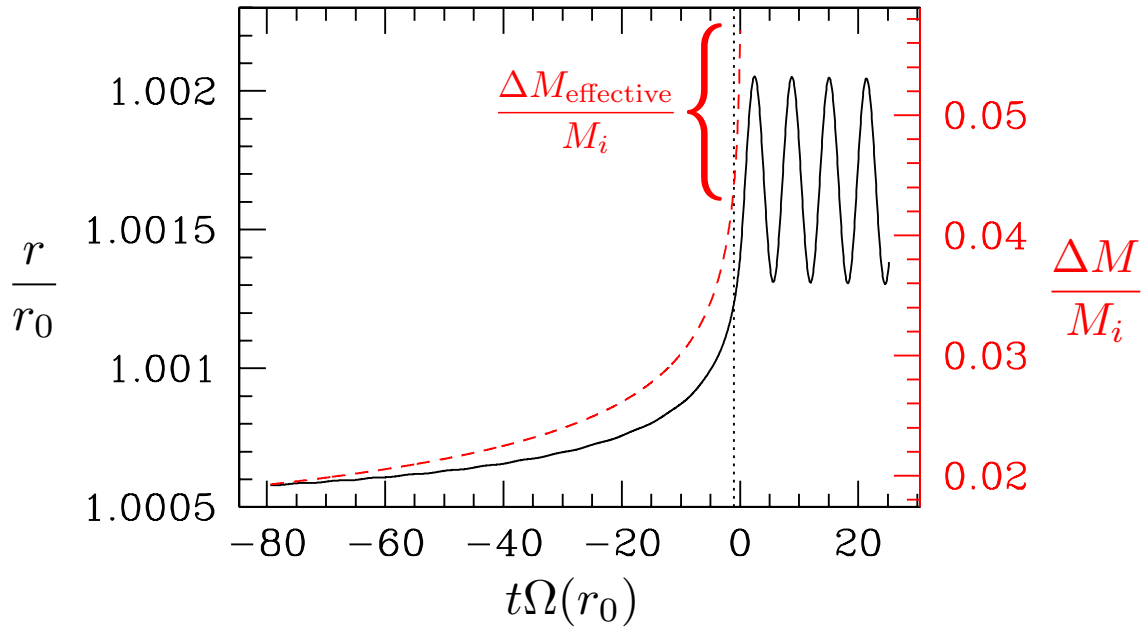


Figure 2.3: Solid (black) curve: We plot the radial position of a particle initially on a circular orbit reacting to mass-energy loss during the inspiral of a secondary SMBH with mass ratio $q = 0.03$. The particle begins at $r_0 = 200\,GM/c^2$ and is plotted in units scaled by that orbit. Dashed (red) curve: The fractional mass-energy radiated during inspiral is plotted as a function of time. Discussion: while the mass radiated occurs on timescales long compared to the dynamical timescale of the particle, the particle's orbit will preserve adiabatic invariants and remain circular while slowly expanding as $r \propto 1/M(t)$. When the mass radiated occurs on timescales short compared to the dynamical time of the particle, epicyclic motion is induced with radial semi-amplitude $\delta r/r = \delta M/M$. Here we mark one dynamical time prior to merger with a vertical dotted (black) line. Mass radiated after this point, denoted $\Delta M_{\rm effective}$ will be seen as impulsive by the particle. From the simulation $\Delta M_{\rm effective}/M_i \approx 3.7 \times 10^{-4}$, while $\delta r/r_0 \approx 3.7 \times 10^{-4}$, as expected.

the final several orbits. Several of the published simulations with both black holes of comparable masses do not provide the initial $M_{\rm ADM}$, from which can easily be derived the binding energy. So, to approximate $E_{\rm inspiral}$ we must rely on using post-Newtonian analysis to determine $E_{\rm inspiral}$ prior to reaching the ISCO, a value only well defined for when the mass ratio $q \equiv M_2/M_1 < 1$ is small.

For the merger of equal mass non-rotating black holes, [54] have done the highest precision simulation with the most (16) orbits to date. They also include both inspiral and merger in their calculation of the total mass lost and find the total mass radiated during inspiral to be

$$\frac{\delta M(q=1,\vec{a}=0)}{M_i} = 0.05\ . \tag{2.10}$$

These results are consistent with earlier numerical simulations finding $E_{\rm merger} = 0.04$, [55–59], when combined with the 3PN result ($\omega_{\rm static} = 0$) of [60] for the binding energy

$$\begin{aligned}\frac{E_{\rm inspiral}}{M_i} &= -\frac{\eta}{2}x\Bigg[1+\left(-\frac{3}{4}-\frac{\eta}{12}\right)x + \\ &\quad \left(-\frac{27}{8}+\frac{19}{8}\eta-\frac{\eta^2}{24}\right)x^2 + \\ &\quad \left\{-\frac{675}{64}+\left(\frac{34445}{576}-\frac{205}{96}\pi^2\right)\eta - \right. \\ &\quad \left.\frac{155}{96}\eta^2-\frac{35}{5184}\eta^3\right\}x^3\Bigg]\ .\end{aligned} \tag{2.11}$$

Here $x = (M_i\Omega_{\rm ISCO})^{2/3}$ and $\eta = q/(1+q)^2$ is the symmetric mass ratio (note $0 < \eta < 0.25$). Plugging in $\eta = 0.25$ and $M_i\Omega_{\rm ISCO} = 0.129$ [from Table XI of 59] one finds $E_{\rm inspiral} = 0.01M_i$.

For non-spinning mergers of unequal mass black holes, with mass ratio $q = M_2/M_1 < 1$, the energy radiated in the final plunge after the black hole orbits come within the effective innermost stable circular orbit is well fitted to the numerical simulations by $E_{\rm merger}/M_i = 0.0363(4\eta)^2$ [Eqn. 3.13 of 59]. We used a least squares linear fit to the 3PN $M_i\Omega_{\rm ISCO}$ values from Table XI of [59], and using Eqn. 2.11 find that the radiated energy in the inspiral is well fitted by $E_{\rm inspiral} = 0.061\eta(1+\eta)M_i$. Therefore the total radiated energy from the merger of Schwarzschild BHs is approximately

$$\left.\frac{\delta M(\eta)}{M_i}\right|_{\vec{a}=0} = 0.06\eta(1+10.7\eta)\,, \tag{2.12}$$

where $\vec{a}$ is the dimensionless spin parameter. Note that this produces mass perturbations approximately 10% larger than those predicted by [54] for a symmetric mass ratio of $\eta = 0.25$.

Simulations of mergers of equal mass rotating black holes with $a_1 = a_2 = 0.75$ and $a_1 = a_2 = -0.75$ show that the fractional mass-energy radiated in the final orbit is larger than for non-spinning black holes (0.07 vs. 0.04) when both spins are aligned with the orbital angular momentum, and

smaller (0.02) when both are anti-aligned to the orbit [61]. To these should again be added the intial binding energies: $M_i\Omega = 0.05$. For Kerr BHs with spins aligned or anti-aligned to the orbital angular momentum, and with $s \equiv (S_1 + S_2)/M_i^2$, the binding energy to 2PN is given by [62]

$$\frac{E_{\rm inspiral}}{M_i} = -\frac{\eta}{2}x\Bigg\{1 - \frac{37}{48}x + \frac{7}{6}sx^{3/2} - \left(\frac{1069}{384} + \frac{1}{4}s^2\right)x^2\Bigg\}\,, \tag{2.13}$$

and gives $0.016\,M_i$ and $0.013\,M_i$ for $E_{\rm inspiral}$ in the aligned and anti-aligned cases, respectively. This results in estimates for the total radiated energy of an equal-mass equal-spin BH binary

$$\left.\frac{\delta M(a{=}0.757)}{M_i}\right|_{q=1} = 0.08 \tag{2.14}$$

and

$$\left.\frac{\delta M(a{=}-0.757)}{M_i}\right|_{q=1} = 0.03\,, \tag{2.15}$$

where $a = a_1 = a_2$.

To understand mergers of varying spin magnitude, [63] runs four simulations of equal-mass equal-and-opposite-spin mergers. For their four cases $a = 0.2,\ 0.4,\ 0.6,\ 0.8$ (aligned and anti-aligned with the orbital angular momentum) they give $E_{\rm inspiral} = 0.016,\ 0.015,\ 0.015,\ 0.015$ and $E_{\rm merger} = 0.032,\ 0.033,\ 0.033,\ 0.033$. It is interesting that this yields the same total mass-energy loss for equal-mass equal-and-opposite-spin mergers in all four simulations:

$$\left.\frac{\delta M(a_1{=}-a_2)}{M_i}\right|_{q=1} = 0.05\,. \tag{2.16}$$

A series of simulations by [64] consider equal mass Kerr binaries with the spin of one BH aligned with the orbital angular momentum, $a_2 = 0.584$, and the spin of the other BH varying from $a_1 = 0.584$ to $a_1 = -0.584$. They present both $E_{\rm inspiral}$ and $E_{\rm merger}$ and we summarize these results in Table 2.1. The total mass-energy loss can be fit well linearly:

$$\left.\frac{\delta M(a_1/a_2)}{M_i}\right|_{q=1} = 0.011\frac{a_1}{a_2} + 0.0576\,. \tag{2.17}$$

This is consistent with the results of [63], summarized in Eqn. 2.16.

For mergers of Kerr BHs with mass ratio $q \ll 1$ the mass-energy loss can be dealt with analytically. [65] show $E_{\rm inspiral} = f_b M_2 = f_b q M_i$, where f_b is a function of the dimensionless spin parameter a_1. In Fig. 2.4 we plot f_b as a function of a_1 for prograde and retrograde orbits along with f_b averaged over all orbit inclinations for randomly-oriented quasi-circular orbits using the fitting formula of Eqn. 9 of [65]. In the limit of non-rotating black holes $f_b = 0.057$, while for a maximally spinning

Table 2.1: Total Mass-Energy Loss for [64]

a/M	$E_{\rm inspiral}/M_i$	$E_{\rm merger}/M_i$	$\delta M/M_i$
-0.5840	-0.0144	-0.0325	-0.0469
-0.4380	-0.0145	-0.0353	-0.0498
-0.2920	-0.0144	-0.0380	-0.0524
-0.1460	-0.0143	-0.0402	-0.0545
-0.0000	-0.0141	-0.0426	-0.0567
0.1460	-0.0138	-0.0456	-0.0594
0.2920	-0.0136	-0.0495	-0.0631
0.4380	-0.0133	-0.0530	-0.0663
0.5840	-0.0128	-0.0564	-0.0692

Table 2.2: Inspiral into Schwarzschild Black Hole.

$\frac{r}{M}$	$\frac{(E_b(r)-E_b(6M))}{\mu}$ [a]	$\frac{\mu}{M}\frac{t_{\rm insp}}{M}$ [b]	$\left(\frac{\mu}{M}\right)^{2/3} r_{\rm ad}$ [c]
6	0	0	0
10	0.0134	9.2×10^1	20
15	0.0261	7.3×10^2	81
20	0.0334	2.6×10^3	190
25	0.0379	6.7×10^3	360
30	0.0410	1.4×10^4	590
60	0.0490	2.4×10^5	3,900
100	0.0522	1.9×10^6	15,000

[a] Mass radiated as gravitational waves in inspiral from r to the last stable orbit at $r = 6\,M$ is the tabulated value times μ, the mass of the smaller black hole, assumed $\ll M$, the mass of the larger black hole.
[b] Time to inspiral from radius r is the tabulated entry times M/μ, in GM/c^3 units.
[c] Radius $r_{\rm ad}$ at which orbital angular frequency $\Omega(R)$ equals the inspiral time from radius r is the tabulated entry times $(M/\mu)^{2/3}$, in GM/c^2 units. At $R \gg R_{ad}$, disk particles respond adiabatically to mass-energy loss from inspiral. At $R \ll R_{ad}$, the mass-energy loss from inspiral is effectively impulsive, and epicyclic oscillations are excited. See text.

black hole $f_b = 0.42$, and $f_b = 0.038$ for prograde and retrograde inspirals, respectively. In contrast to the equal mass case, the energy radiated during the final plunge, merger, and ringdown is negligible: only a fraction $\sim q^2$ of the total mass. Therefore, with these values of f_b we have to first order in q

$$\left.\frac{\delta M(a_1)}{M_i}\right|_{q\ll 1} = q \cdot f_b(a_1)\ . \tag{2.18}$$

Event rates for SMBH mergers are strongly dependent on the chosen model's prescription for seed population, and can vary by orders of magnitude [66–68].

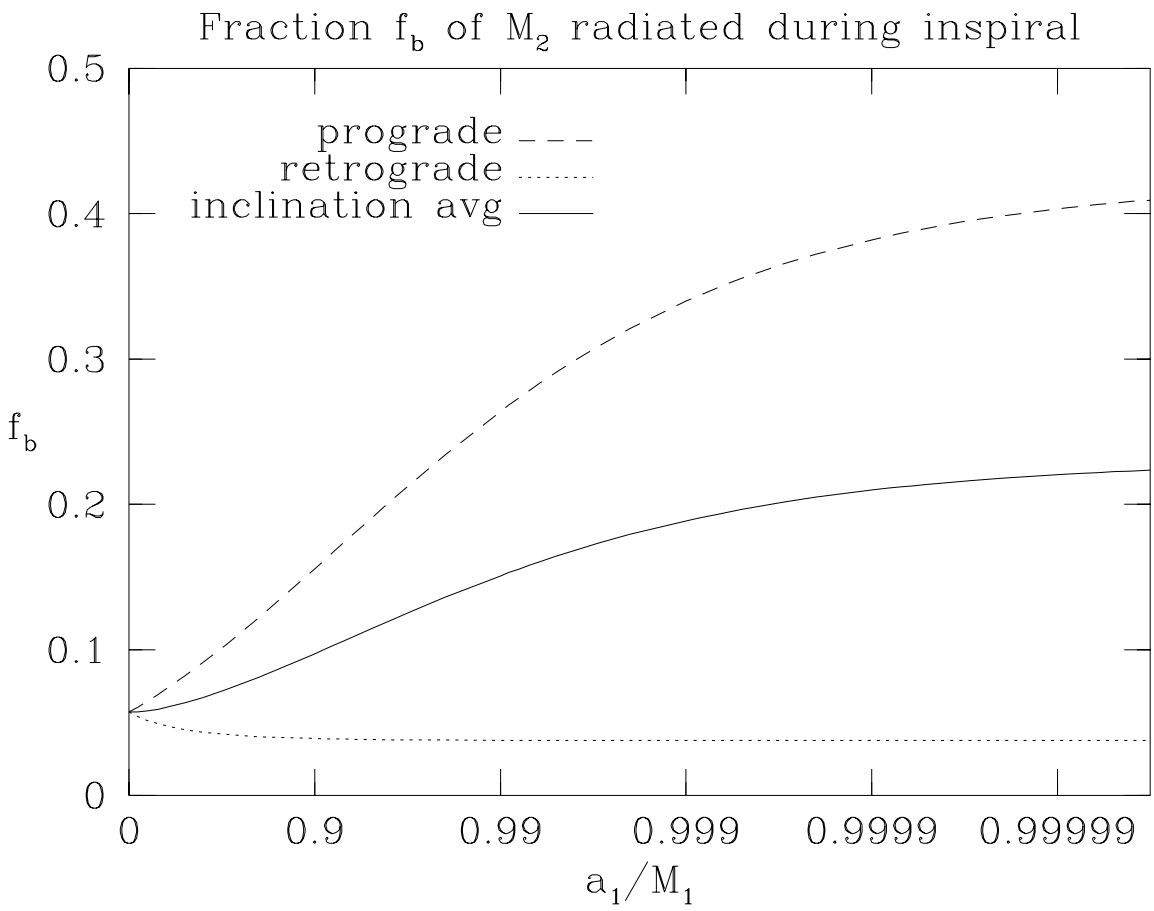


Figure 2.4: Total mass-energy, as a fraction f_b of M_2 radiated in gravitational radiation in the inspiral and merger of a mass M_2 with a black hole of mass $M_1 \gg M_2$ and dimensionless spin parameter a_1. Upper curve is for a prograde quasi-circular equatorial orbit, lower curve for retrograde, and the middle curve is the average over all orbit inclinations for randomly-oriented quasi-circular orbits.

2.3.3 Mergers with Kicks

Asymmetric GW radiation during the merger of two black holes imparts an impulsive kick to the final black hole whose magnitude is a function of q and vectorial spin, $\vec{a}$. In systems with significant quantities of gas, accretion acts to align the spins of Kerr black holes with the orbital angular momentum [69]. In this orientation, with the spins orthogonal to the orbital plane (or with the spins 0), symmetry about the orbital plane dictates that mergers give kicks in the orbital plane. This makes the problem at most two-dimensional.

We now consider the case of Kerr mergers with spins orthogonal to the orbital plane and Schwarzschild mergers. Our analysis collects the recoil velocities of the simulations discussed in Sec. 2.3.2 to determine the relevance of mass-energy loss and kicks to the dynamics of a circumbinary disk.

In the case of the simulations of equal-mass non-spinning mergers [55–59] there is no kick because of symmetry. Therefore the 6% mass-energy loss described in Sec. 2.3.2 is the sole perturber of the post-merger disk.

Recently, numerical simulations of Schwarzschild black hole mergers [70–72] have extended early analytic work [73, 74] considering kick magnitudes for high-mass-ratio non-spinning BH mergers, to equal mass binaries. These simulations have produced a modified Fitchett fitting formula:

$$|v_{\rm kick}| = A\eta^2 \frac{1-q}{1+q}(1+B\eta) \ {\rm km/\,sec}\ , \tag{2.19}$$

where $\eta(q)$ is the symmetric mass ratio and A and B are free parameters to be fit from simulations. [70] systematically performed 31 simulations of black hole mergers for q ranging from 0.25 to 1.0 and, using a least-squares fit, found $A = 1.2 \times 10^4$ and $B = -0.93$. Using these values for this range of q we plot Eqn. 2.19 in Fig. 2.5 (solid blue) and plot its continuation to include $q = 0$ (black dash-dot). To compare the effects between the mass-energy loss discussed in Sec. 2.3.2 and black hole recoil discussed in this section, we also plot in solid black the maximum radial velocity, $v_{\rm epicycle}$, of a particle located at $r_{\rm inner} = 200\,M_i$:

$$v_{\rm epicycle} = v_{\phi_0}\epsilon_{\rm m}\ , \tag{2.20}$$

as given by differentiating Eqn. 2.29 with respect to time and using Eqn. 2.12 to determine $\epsilon_{\rm m}$. Here v_{ϕ_0} refers to the initial angular velocity at some radius. In red (dashed) we plot the ratio of the epicyclic velocity to the kick velocity, which, as described in Sec. 2.4.2, is the appropriate quantity to determine which effect dominates. When $q = 0.14$ this ratio approaches its minimum of 3.2, indicating that kicks are moderately important to the dynamics of the disk. As the binary approaches being equal mass BHs the plot shows that kicks become irrelevant (since $v_{\rm kick}$ goes to

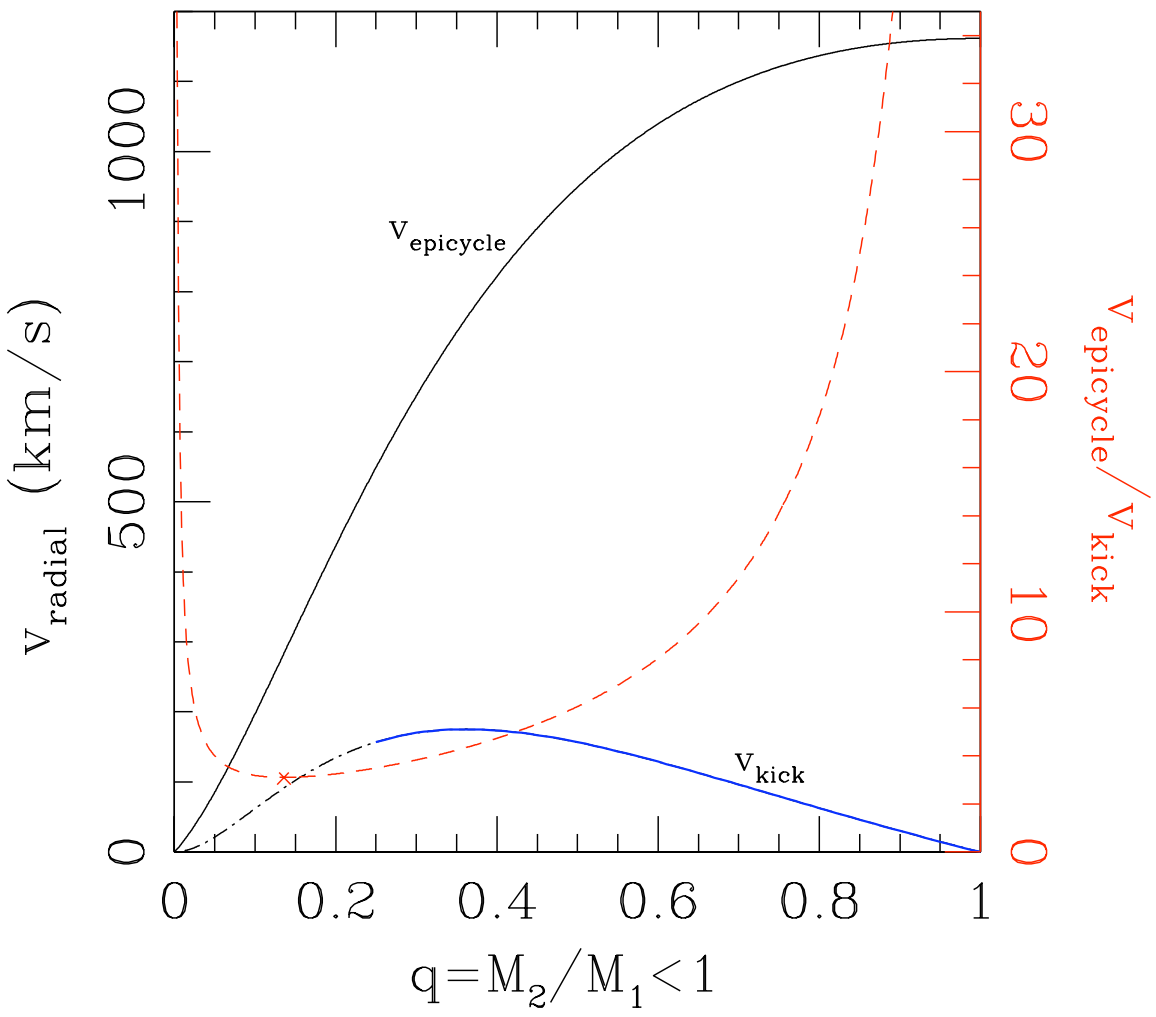


Figure 2.5: We plot figure 2 of [70] showing the kick velocity for non-spinning black holes as a function of mass ratio (solid blue). Also plotted is the maximum radial epicyclic velocity at $r_{\rm inner} = 200\,M_i$ due to mass-energy loss as a function of mass ratio (thick solid) and the ratio of the maximum radial epicyclic motion due to mass-energy loss to the SMBH kick velocity (dashed red plotted with ordinate on right). The latter is a measure of the relative importance of mass-energy loss when compared to kicks as is most clearly seen in Eqn. 2.63. The minimum is marked by an "X" and a 2D density plot of the density distribution for this point is shown in Fig. 2.11. Because $v_{\rm epicycle} \sim \tilde{r}_0^{-1/2}$ the importance of kicks will grow as $\tilde{r}_0^{1/2}$.

0) and the one-dimensional analysis of Sec. 2.4.1 is sufficient to understand the disk's outcome. In interpreting Fig. 2.5 it is important to remember that $v_{\rm epicycle} \sim \tilde{r}^{-1/2}$ meaning that farther out that $200\,M_i$ the effect of kicks will be scaled by a factor of $\tilde{r}^{1/2}$. For instance, at $2000\,M_i$ kicks and mass-energy loss will have comparable contributions to the reaction of the inner disk.

[63] considers an equal mass inspiral with BHs of equal and opposite spin parallel to the orbital angular momentum and changes the spin magnitude between simulations. As initial conditions they modify the initial position, linear momentum, and puncture masses, so as to keep the initial total angular momentum the same for all of their simulations. They find that the kicks are well fit by

$$|v_{\rm kick}| = 475 \text{ km s}^{-1}\text{a}\,, \tag{2.21}$$

where a is the dimensionless spin parameter which ranges, in their simulations, from 0.2 to 0.8. We plot $v_{\rm epicycle}(r_{\rm inner})/|v_{\rm kick}|$ as a function of a in Fig. 2.6 where $v_{\rm epicycle} = \frac{\delta M}{M_i} v_0$ is the maximum radial velocity, calculable from Eqn. 2.29, and $\frac{\delta M}{M_i}$ is given by Eqn. 2.16. For low spins the merger is dominated by mass-energy loss, since as the the spin goes to 0 so does the kick magnitude while the mass-energy loss remains finite. However, as the black holes become maximally spinning the effects of the spin on kicks are being maximized where for mass-energy loss they are being minimized.

In Fig. 2.6 we also plot $v_{\rm epicycle}(r_{\rm inner})|v_{\rm kick}|$ for the simulations of [64] described in Sec. 2.3.2. For $|v_{\rm kick}|$ we use their fitting formula for the resultant kicks:

$$\begin{aligned} |v_{\rm kick}| &= |a_2| \times \\ &\left\{ 109.3 - 132.5 \left(\frac{a_1}{a_2}\right) + 23.1 \left(\frac{a_1}{a_2}\right)^2 \right\} \text{km/ s}\,. \end{aligned} \tag{2.22}$$

2.4 Analytic Solutions of Collisionless Disk Following Merger

Just following merger a fluid particle in the disk will initially follow a collisionless path; the path of a non-interacting particle. It is only when pressure gradients significantly differ from the steady state pressure gradients that the fluid particle will deviate from this non-interacting path. We solve for the surface density as a function of time in the region where the flow is still non-interacting under the assumption that the gas is initially on circular orbits and that the changes to the potential are instantaneous (an assumption validated in Sec. 2.3.1). A discussion of what occurs on the boundary of this non-interacting region and the final density profile is given in Sec. 2.5.

For clarity we first consider the axially symmetric problem of a merger which undergoes only mass-energy loss in Sec. 2.4.1, and then generalize this solution to mergers with recoils in Sec. 2.3.3.

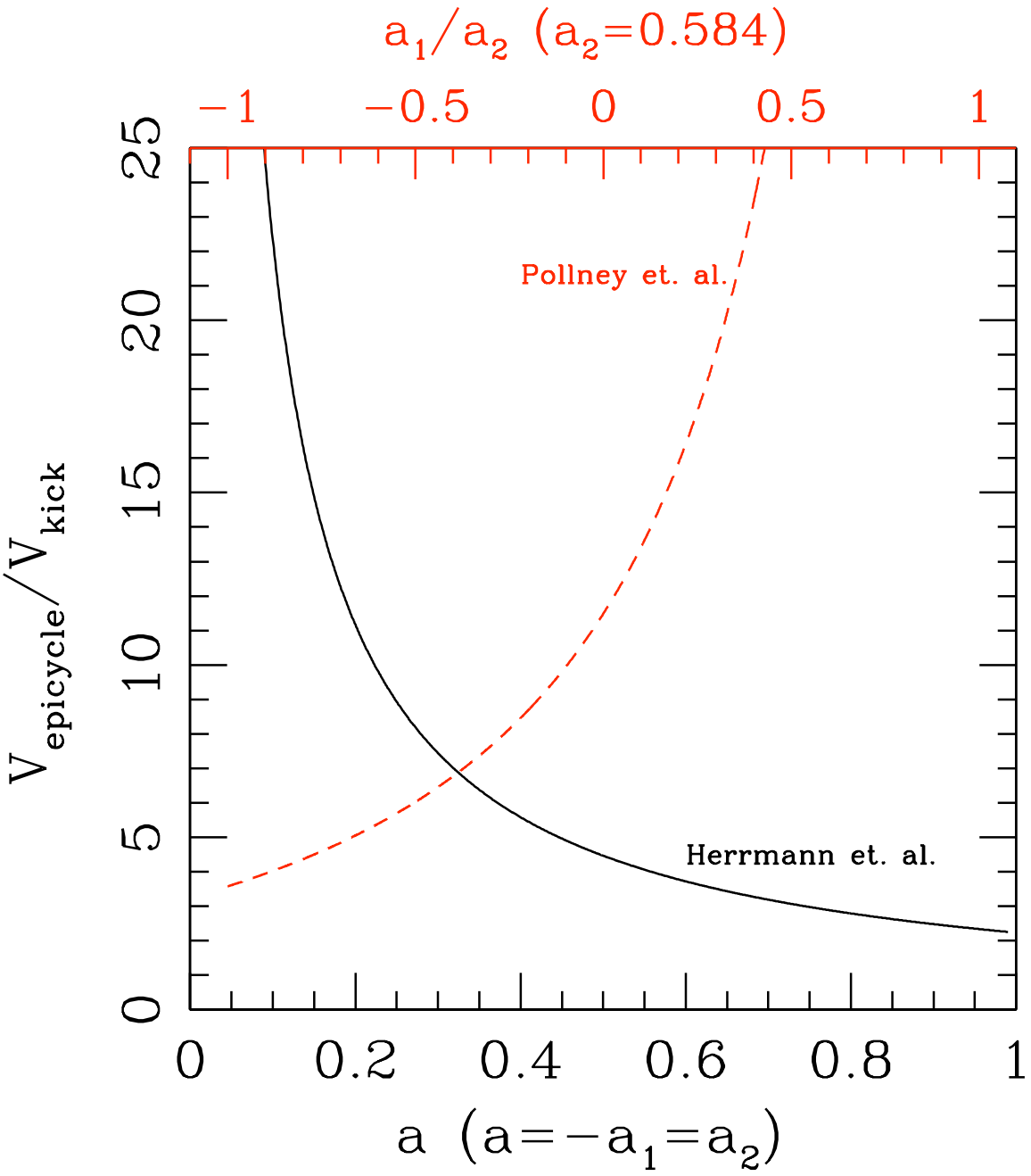


Figure 2.6: The ratio of the maximum radial velocity at $r_{\rm inner} = 200\,M_i$ due to mass-energy loss to the kick velocity is plotted for an equal mass merger with equal-and-opposite spins parallel to the orbital angular momentum as given by [63] (solid black plotted with bottom axis). The same ratio is plotted for an equal mass merger with one spin $a_2 = 0.584$ and the other spin ranging from $a_1 = -0.584$ to $a_1 = 0.584$ based on simulations by [64] (dashed red plotted with upper axis). Because of the radial dependence of the epicyclic velocity these curves are modified by a factor of $\tilde{r}_0^{-1/2}$ when considering other regions of the disk. Where the ratio is of order 1 the kicks are important to the dynamics of the inner circumbinary disk, while in regions where it greatly exceeds 1 the dynamics of the inner disk are dominated by the mass-energy loss and the problem reduces to the self-similar one-dimensional solution of Sec. 2.4.1.

2.4.1 Axisymmetric Case — Mass-Energy Loss

Consider a particle in the collisionless disk. Just before the merger it was in a circular orbit at radius r_0 about a central mass M_i, with tangential velocity $v_0 = \sqrt{GM_i/r_0}$ and zero radial velocity. Just after the merger, it is still at r_0, but now orbiting a central mass $M_f = M_i - \delta M \equiv M_i(1-\epsilon_{\rm m})$, where

$$\epsilon_{\rm m} \equiv \frac{\delta M}{M_i}\,. \tag{2.23}$$

Its velocity v_0 is too large for a circular orbit at r_0, so it is at the pericenter of a new slightly eccentric orbit. This new orbit is most simply described as retrograde epicyclic motion of amplitude X about a guiding center located a distance X beyond r_0, as shown in Fig. 2.7. The equation for a Keplerian orbit is

$$r = a - ae\cos\phi_t + O(e^2)\,, \tag{2.24}$$

where e is the orbital eccentricity, a is the semi-major axis, and ϕ_t is the mean anomaly. Thus $X = ae$. Immediately after the mass-energy loss, the *vis-viva* equation gives $v^2 = GM_i/r_0 = GM_f(2/r_0 - 1/a)$, and hence $a = r_0(1-\epsilon_{\rm m})/(1-2\epsilon_{\rm m})$. Since at this moment, the particle is at its new pericenter, $r_0 = a(1-e)$, we may read off, to first order in e (or $\epsilon_{\rm m}$),

$$e = \epsilon_{\rm m}\,, \tag{2.25}$$

$$a = r_0(1+\epsilon_{\rm m})\,, \tag{2.26}$$

$$X = ae = r_0\epsilon_{\rm m}\,. \tag{2.27}$$

To the same order the mean anomaly is given by

$$\phi(r_0,t) = \sqrt{\frac{GM_f}{a^3}}\;t = (1-2\epsilon_{\rm m})\Omega_0 t \equiv \phi_t\;, \tag{2.28}$$

where $\Omega_0 \equiv \sqrt{GM_i/{r_0}^3}$ is the particle's angular frequency prior to merger. Combining Eqn. 2.24 and Eqns. 2.25–2.27, the particle's radial motion after merger is then given by

$$r(r_0,t) = r_0 + r_0\epsilon_{\rm m}\left\{1-\cos\phi_t\right\} + O\left(\epsilon_{\rm m}\right)^2\;. \tag{2.29}$$

Immediately after the merger, particles at all radii are moving outwards. However, because the epicyclic frequency, given by Eqn. 2.28, is higher at smaller radii, as time progresses particles with small radial separations become out of phase, creating rarefactions, compressions, and ultimately caustics. In Fig. 2.8 this is demonstrated by evolving evenly spaced particles in a space-time diagram. Patterns form and move outwards rapidly and at late times orbit crossings occur (caustics). The surface density at the location of a given particle oscillates in time with increasing amplitude as the

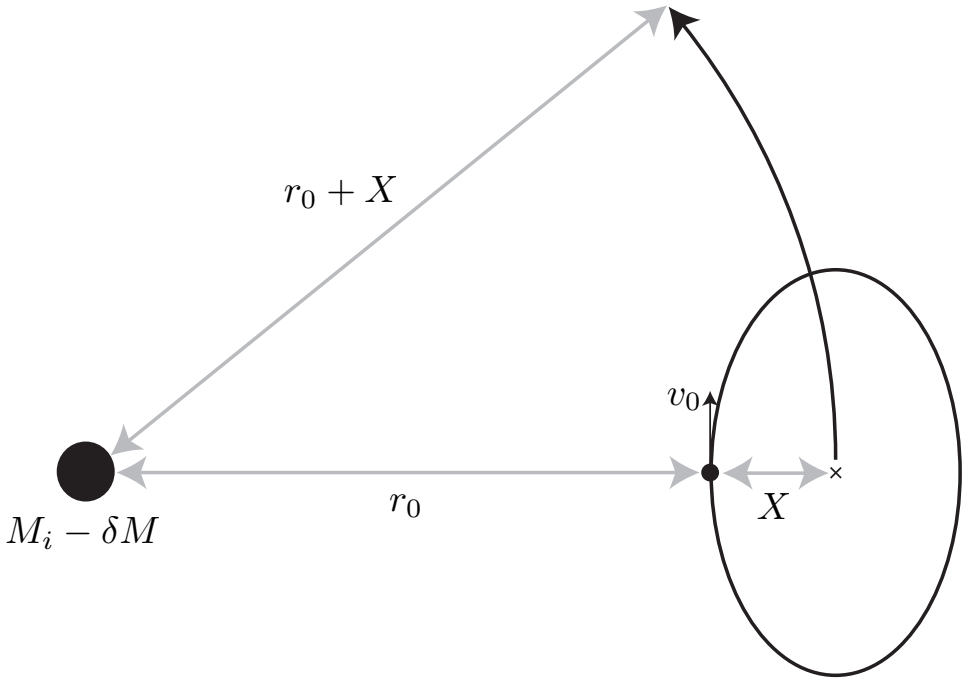


Figure 2.7: After the central binary black holes' mass-energy loss due to gravitational radiation, a disk particle's pre-merger circular velocity v_0 is now too fast for a circular orbit about the newly lowered central mass. So the disk particle begins epicyclic motion with radial amplitude $X = r_0\epsilon_{\rm m}$ about a guiding center at $r = r_0 + X$ in uniform rotation about the merged black hole.

particle becomes more out of phase with neighboring particles.

If the surface density of a fluid element immediately before the merger is $\Sigma_0(r_0)$, where r_0 is the fluid element's initial position, the surface density at the fluid element's location, $r(r_0)$, at times after the merger is

$$\Sigma(r) = \sum_{r_0:r(r_0)=r} \frac{\Sigma_0(r_0)}{\frac{r}{r_0}\left|\frac{\partial r}{\partial r_0}\right|} \equiv \sum_{r_0:r(r_0)=r} \frac{\Sigma_0(r_0)}{S(r_0,t)} \ , \tag{2.30}$$

where the summation is being taken over all particles at position $r(t)$. Using Eqn. 2.30 and differentiating Eqn. 2.29 with respect to r_0, we compute the denominator of Eqn. 2.30, keeping terms to first order in $\epsilon_{\rm m}$:

$$\begin{aligned} S(r_0,t) &= \frac{r}{r_0}\left|\frac{\partial r}{\partial r_0}\right| \\ &= \left|1 + 2\epsilon_{\rm m}\left\{1-\cos\phi_t\right\} - \frac{3}{2}\epsilon_{\rm m}\Omega_0 t\sin\phi_t\right| \ , \end{aligned} \tag{2.31}$$

where the mean anomaly, ϕ_t, is given in Eqn. 2.28.

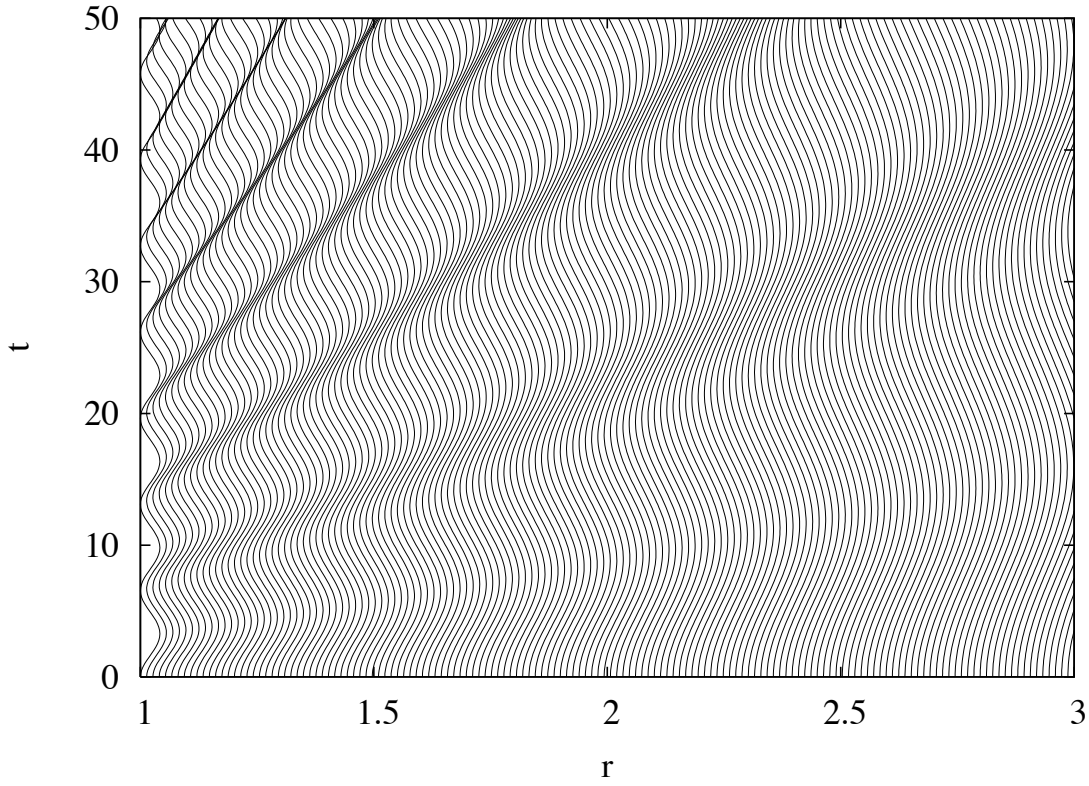


Figure 2.8: A space-time diagram of particles initially evenly spaced at time $t = 0$, but evolving according to Eqn. 2.29 due to a 2% mass perturbation. Patterns form and move outwards through the disk quickly, and near the time $t = 30$ the orbits of adjacent particles in the inner disk cross, forming the first caustic. We solve for the surface density of a particle (i.e., in Lagrangian coordinates) as a function of time in Eqn. 2.30.

The summation in Eqn. 2.30 deserves clarification. It is only relevant where orbit crossings (caustics) have occurred and multiple particles occupy the same spatial point. The relationship between orbit crossings and caustics is demonstrated in Fig. 2.9, where the surface density is plotted in Eulerian coordinates (assuming $\Sigma_0 = 1$) along with the initial positions of particles as a function of their final positions for a mass perturbation of $\epsilon_{\rm m} = 0.08$ and at a time $t = 10\,\Omega(1)$. There are multiple methods to produce such a figure. In one dimension it is most straightforward to first split your domain into a grid of initial positions of particles, propagate them forward to the desired time using Eqn. 2.29, find which neighboring pairs bound the position, $\bar{r}$, where you are finding the surface density, and then use one of the points from each pair as an initial guess to find the roots of $r(r_0) = \bar{r}$. However, this is not entirely necessary for practical application. Because a particle never differs from its original position by more than $\epsilon_{\rm m} r_0$, considering the flow in Lagrangian coordinates does not incur significant errors, remains analytically tractable, and therefore provides greater insight. For this reason (and for clarity), the approximations presented in this chapter are done using Lagrangian coordinates as a function of the particles' initial positions.

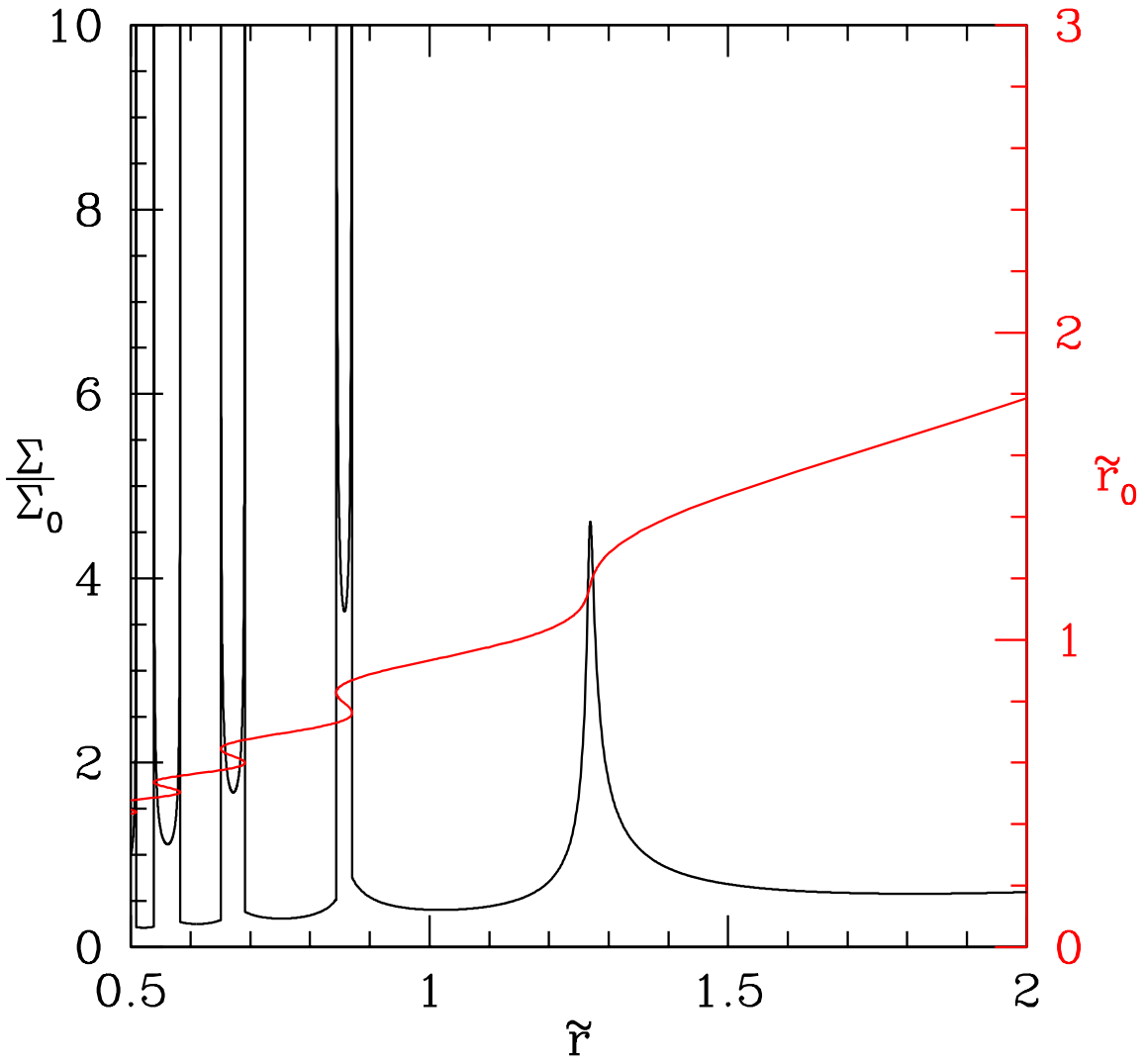


Figure 2.9: We plot the disk response, Σ/Σ_0, at time $\tilde{t} = 10\Omega(1)$ due to an instantaneous mass perturbation of $\epsilon_{\rm m} = 0.08$ using the left ordinate. Overlaid in red using the right ordinate is the disk's *initial* position, $\tilde{r}_0 \equiv r/r_{\rm inner}$, as a function of its *final* position, $\tilde{r}$. Where this curve becomes steep an initially constant density disk now has overdensities. Where the curve is multivalued several initially distinct particles have arrived at the same position, producing caustics. Each caustic is marked by a characteristic double peak where the surface density becomes infinite. Inside these peaks it is necessary to sum over the surface densities for each particle which occupies the point in space. This is analogous to caustic formation in geometric optics where initially separate rays converge to the same point or line (caustic) and the total intensity is the sum of the intensity from the distinct paths. Just as in this case, the caustic formation is an indication that the relevant approximations break down and the analysis fails. In our case, as particles approach the caustic and the density increases, the non-interacting approximation is no longer valid. Therefore the solution *can* only be relevant prior to the first caustic. Note that because the solution is self-similar there is no dependence on the value of $r_{\rm inner}$.

As an aside, it is interesting to note that since the mass-energy loss is axisymmetric and there are no special length scales relevant to the physics, Eqn. 2.31 is self-similar in the similarity variable $\Omega_0 t$.

However, it is more transparent to work with the natural dimensionless variables

$$\tilde{r} \equiv r/r_{\rm inner} \qquad \tilde{t} \equiv \Omega_{\rm inner} t \,, \tag{2.32}$$

which we will prefer.

Using Eqn. 2.31 we see that the extrema of S are defined by

$$\Omega_0 t \approx \frac{(n-1/2)\pi}{1-2\epsilon_{\rm m}} \approx (n-1/2)\pi \equiv C_n \,, \tag{2.33}$$

where n is an integer. Therefore at a time $\tilde{t}$ the particles located at the extrema were originally located at

$$\tilde{r}_{0_n}^{\rm extr} = \left(\tilde{t}/C_n\right)^{2/3} \tag{2.34}$$

and so travel at a rate

$$\dot{\tilde{r}}_{0_n}^{\rm extr} = \frac{2}{3} C_n^{-2/3} \tilde{t}^{-1/3} = \frac{2}{3}\frac{v_{\rm K}}{C_n} \,, \tag{2.35}$$

where the time derivative is with respect to $\tilde{t}$. It is interesting that though amplitude of the density perturbations go to 0 with the mass perturbation, the mass perturbation has only a weak effect on the speed of the overdensities. This is because the overdensity speeds are determined by the relative phases of the orbital frequencies of neighboring particles, something which is nonzero and only perturbed by the mass-energy loss.

Because S is a function of $\tilde{r}_0$, the initial position, Eqns. 2.33–2.35 refer to the *initial* positions, $\tilde{r}_0$ (or $\Omega_0 t$), of the particle which at time $\tilde{t}$ will lie at the n-th extrema. To find the *final* position of these particles at time $\tilde{t}$, and therefore the true positions of the extrema, one must use Eqn. 2.29. However, as 0th order approximations to the true locations of the extrema, Eqns. 2.33–2.35 are useful to gain a quick insight into the motion of the flow.

The minima (maxima) of S occur at odd (even) n and, at later times, lie approximately on the line

$$S_{\substack{\rm max \\ \rm min}} = 1 + 2\epsilon_{\rm m} \pm \frac{3}{2}\epsilon_{\rm m}\tilde{t}\tilde{r}^{-3/2} \approx 1 \pm \frac{3}{2}\epsilon_{\rm m}\tilde{t}\tilde{r}^{-3/2} \,. \tag{2.36}$$

Caustics (orbit-crossings, where the disk surface density goes to infinity) occur where $S(r_0, t) = 0$. Therefore, the first caustic crosses the path of a particle initially at a radius r_0 near the time $t_{\rm c1}$ given by

$$\Omega_0 t_{\rm c1} = \frac{2}{3}\frac{1}{\epsilon_{\rm m}} + \frac{4}{3} \approx \frac{2}{3}\frac{1}{\epsilon_{\rm m}} \,. \tag{2.37}$$

It is useful to use Eqn. 2.36 and Eqn. 2.37 to find the curves on which the peaks and troughs of

Σ/Σ_0 lie:

$$\left(\frac{\Sigma}{\Sigma_0}\right)_{\substack{\max\\\min}} \approx \frac{1}{1 \mp \tilde{t}/\tilde{t}_{\rm c1}}\,. \tag{2.38}$$

Interestingly, the maxima and minima of the density distribution are constants as a function of time. This is readily seen by plugging Eqn. 2.33 into Eqn. 2.36 to find

$$\left(\frac{\Sigma}{\Sigma_0}\right)_{\rm (extr)} \approx \begin{cases} \frac{1}{1-\frac{3}{2}\epsilon_{\rm m}\pi(2p-3/2)} & \text{for } p\text{th peak} \\ \\ \frac{1}{1+\frac{3}{2}\epsilon_{\rm m}\pi(2m-1/2)} & \text{for } m\text{th minima} \end{cases}. \tag{2.39}$$

Using Eqn. 2.33 along with Eqn. 2.37 one can show that the caustic is $p_{\rm c1}$-th *peak* in the density distribution, where

$$p_{\rm c1} = \left\lceil \frac{\Omega_0 t_{\rm c1}}{2\pi} + \frac{3}{4} \right\rceil . \tag{2.40}$$

Here, $\Omega_0 t_{\rm c1}$ is a constant given by Eqn. 2.37 and $\lceil x \rceil$ is the least integer greater than x (ceiling of x). In the extreme case of a 10% perturbation only a single density peak will precede the first caustic.

To determine the position of the first caustic, $\tilde{r}_{\rm c1}$, one may solve Eqn. 2.37 for the position as a function of the time:

$$\tilde{r}_{\rm c1}(\tilde{t}) \approx \left(\frac{\tilde{t}}{\frac{2}{3\epsilon_{\rm m}} + 4/3} \right)^{2/3} . \tag{2.41}$$

However, one may determine a more accurate approximation for $\tilde{r}_{\rm c1}$ by plugging $2p_{\rm c1} - 1$ (given in Eqn. 2.40) in for n in Eqn. 2.34. This yields

$$\tilde{r}_{\rm c1}(\tilde{t}) \approx \left(\frac{\tilde{t}}{(2p_{\rm c1} - 3/2)\pi} \right)^{2/3} . \tag{2.42}$$

Again, $\tilde{r}_{\rm c1}$ is the original position of the particle (which is within $\epsilon_{\rm m}\tilde{r}_{\rm c1}$ of the final location). To find the final position one must merely apply Eqn. 2.29 to $\tilde{r}_{\rm c1}$.

As an example, in Fig. 2.10 we consider a circumbinary disk with a merger releasing 2% of the initial mass energy and plot the ratio of the surface density to the original surface density as a function of $\tilde{r}$ for four different times $\tilde{t} = 25$, 35, 45, 55. The approximate position of the first caustic as given by Eqn. 2.37 is shown by the short thick red vertical line segment. Caustics can be easily distinguished from finite overdensities by their characteristic split peak as is most clearly visible for the final two caustics of the $\tilde{t} = 55$ plot. Note that the split peak of the first caustic is not visible due to the narrowness of the caustic and the size of the domain. As given by Eqn. 2.40 there are 5 finite peaks preceding the first caustic.

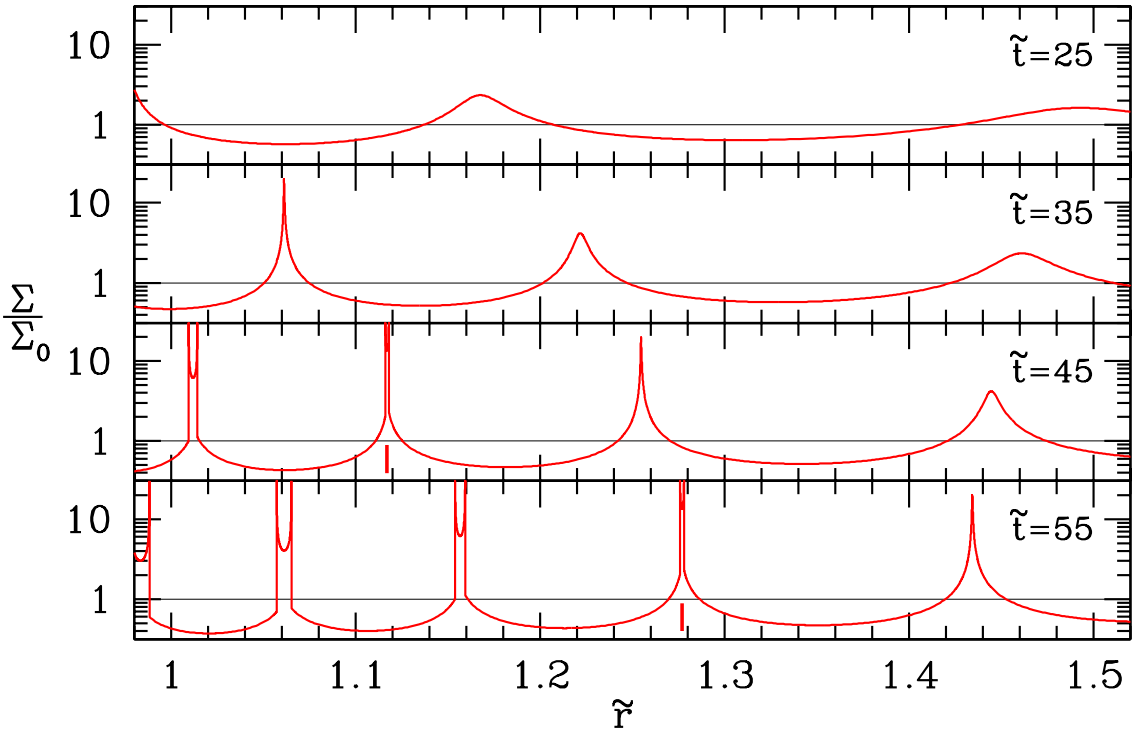


Figure 2.10: For a non-interacting disk and a perturbation of $\frac{\delta M}{M_i} = 0.02$, Σ/Σ_0 is plotted as a function of $\tilde{r}$ for four values of $\tilde{t}$. The thin horizontal black line demarcates the initial condition $\Sigma/\Sigma_0 = 1$ and the approximate position of the first caustic, given by Eqn. 2.37, is marked by the short thick red vertical line segment. Indeed, the peak above this short vertical line segment is the first caustic, though the characteristic double peak exhibited in the following caustics is not visible within the given domain.

2.4.2 2D Disk Evolution — SMBH Recoil

We now generalize the solution of Sec. 2.4.1 and consider the mergers of SMBHs with spins aligned or anti-aligned with the orbital angular momentum. In this case the kicks are in the plane of the disk, and the problem is two-dimensional.

We repeat the same process as in the axisymmetric case (Sec. 2.4.1), but now with a two scale expansion in the mass perturbation, $\epsilon_{\rm m} \equiv \delta M/M_i$ (as in the axisymmetric case), and the kick perturbation,

$$\epsilon_{\rm v} \equiv v_{\rm kick}/v_0(r_0) = \frac{v_{\rm kick}}{v_{\rm inner}}\sqrt{\tilde{r}_0}\,, \tag{2.43}$$

where $\tilde{r}_0 \equiv r_0/r_{\rm inner}$. As will become clear shortly, the spatial dependence of $\epsilon_{\rm v}$ has the added characteristic of breaking the self-similarity present in the axisymmetric solution for pure mass-energy loss, since it introduces a characteristic length scale when $\epsilon_{\rm v} = 1$ (and the perturbative solution being discussed is no longer relevant).

We begin by considering the particles' motions from the frame of the moving black hole which

has a kick in the $\theta = \pi$ direction in cylindrical coordinates and assume the disk co-rotates with the binary in the positive θ direction. The *vis-viva* equation then becomes

$$\frac{GM_i}{r_0}\left(1 - 2\epsilon_{\rm v}\sin\theta_0 + {\epsilon_{\rm v}}^2\right) = v^2 = GM_f\left(\frac{2}{r_0} - \frac{1}{a}\right) \,, \tag{2.44}$$

where M_i is the initial mass of the SMBH binary, M_f is the mass of the final SMBH, r_0 is the initial radial position of the particle, and θ_0 is the initial angular position measured in a counter-clockwise fashion. Solving for a we have

$$\begin{aligned} a &= r_0 \frac{1-\epsilon_{\rm m}}{1 - 2\epsilon_{\rm m} + 2\epsilon_{\rm v}\sin\theta_0 - {\epsilon_{\rm v}}^2} && (2.45)\\ &= r_0\left(1 + \epsilon_{\rm m} - 2\epsilon_{\rm v}\sin\theta_0\right) + O(\epsilon_{\rm m}, \epsilon_{\rm v})^2 && (2.46)\\ &= r_0(1-\epsilon) + O(\epsilon_{\rm m}, \epsilon_{\rm v})^2 \,, && (2.47) \end{aligned}$$

where the final equality defines the parameter ϵ.

In the case of no kick (only mass-energy loss) the particle was located at periapsis at the moment of merger. However in this case the kick breaks the symmetry and requires the eccentricity and mean anomaly to both have angular dependence and therefore depend on both r_0 and θ_0. The mean anomaly, ϕ, is the sum of a time-dependent term given by the angular frequency at a, plus a constant term ϕ_0 which indicates where in the epicycle the particle begins:

$$\phi(r_0, \theta_0, t) = \sqrt{\frac{GM_f}{a^3}}t + \phi_0(r_0, \theta_0) \equiv \phi_t + \phi_0 \,. \tag{2.48}$$

After using Eqn. 2.47 and expanding to first order in $\epsilon_{\rm m}$ and $\epsilon_{\rm v}$, we may write

$$\begin{aligned} \phi_t &= \frac{(1 - 2\epsilon_{\rm m} + 2\epsilon_{\rm v}\sin\theta_0 - {\epsilon_{\rm v}}^2)^{3/2}}{1-\epsilon_{\rm m}}\Omega_0 t && (2.49)\\ &= (1 - 2\epsilon_{\rm m} + 3\epsilon_{\rm v}\sin\theta_0)\Omega_0 t \,. && (2.50) \end{aligned}$$

To remove e and ϕ_0 from the evolution equations we solve the initial position and radial velocity equations for a particle:

$$\begin{aligned} r_0 &= a(1 - e\cos\phi_0) && (2.51)\\ v_{\rm kick}\cos\theta_0 &= \dot{r} = a\Omega(a)e\sin\phi_0 \,, && (2.52) \end{aligned}$$

to get

$$\begin{aligned} e\sin\phi_0 &= \frac{v_{\rm kick}\cos\theta_0}{a\Omega(a)} \approx \epsilon_{\rm v}\cos\theta_0 \,, && (2.53)\\ e\cos\phi_0 &= 1 - r_0/a \approx -\epsilon \,. && (2.54) \end{aligned}$$

When $\epsilon_{\rm v} = 0$ we have the correct limit that $\phi_0 = 0$.

Expanding $\cos\phi$ and applying approximations 2.53 and 2.54 we can rewrite $r = a(1 - e\cos\phi)$ as

$$r(r_0, \theta_0, t) = r_0(1 - \epsilon(1 - \cos\phi_t) + \epsilon_{\rm v}\cos\theta_0 \sin\phi_t) \ , \tag{2.55}$$

to first order.

We can find $\theta(r_0, \theta_0, t)$ via a similar process. The theory of epicycles gives that the magnitude of the angular epicyclic oscillation is

$$\begin{aligned} \delta\theta_{\rm e} &= 2e\sin(\phi) & (2.56)\\ &= 2e(\sin\phi_t\cos\phi_0 + \sin\phi_0\cos\phi_t) & (2.57)\\ &= -2\epsilon\sin\phi_t + 2\epsilon_{\rm v}\cos\theta_0\cos\phi_t \ . & (2.58) \end{aligned}$$

Then the angular position is given by the sum of the position of the guiding center plus the angular epicyclic oscillations:

$$\theta(r_0, \theta_0, t) = \theta_{\rm GC0} + \theta_{\rm GC} + \delta\theta_{\rm e} \ , \tag{2.59}$$

where $\theta_{\rm GC} = \phi_t$ is the angular distance traveled by the guiding center and $\theta_{\rm GC0} = \theta_0 - 2\epsilon_{\rm v}\cos\theta_0$ is given by the initial conditions. Together, this gives us,

$$\begin{aligned} &\theta(r_0, \theta_0, t) = & (2.60)\\ &\quad \theta_0 - 2\epsilon_{\rm v}\cos\theta_0 + \phi_t - 2\epsilon\sin\phi_t + 2\epsilon_{\rm v}\cos\theta_0\cos\phi_t \end{aligned}$$

to first order in $\epsilon_{\rm m}$ and $\epsilon_{\rm v}$.

Using Eqns. 2.55 and 2.60 and performing the differentiation, noting that $\epsilon_{\rm v}$ has an implicit ${r_0}^{1/2}$ dependence, we can write down

$$\begin{aligned} \Sigma(\vec{r'}) &= \sum_{\vec{r}_0:\vec{r}(r_0,\theta_0)=\vec{r'}} \frac{\Sigma_0(r_0,\theta_0)}{\frac{r}{r_0}\left|\frac{\partial(r,\theta)}{\partial(r_0,\theta_0)}\right|} & (2.61)\\ &\equiv \sum_{r_0:r(r_0)=r} \frac{\Sigma_0(r_0)}{S(r_0,t)} \ , & (2.62) \end{aligned}$$

where vectors denote coordinate pairs (r, θ), and S is given by

$$\begin{aligned} S &\equiv \frac{r}{r_0}\left|\frac{\partial(r,\theta)}{\partial(r_0,\theta_0)}\right| & (2.63)\\ &= 1 + \epsilon_{\rm m}\left\{2 - 2\cos\phi_t - \frac{3}{2}\phi_t\sin\phi_t\right\} - \frac{3}{4}\epsilon_{\rm v}\Big\{4\sin\theta_0 - \\ &\quad 2\phi_t\cos(\phi_t - \theta_0) + 3\sin(\phi_t - \theta_0) - \sin(\phi_t + \theta_0)\Big\} . \end{aligned}$$

So, the surface density at a point (r', θ') at a time t is equal to the sum of the surface densities of all of the particles who end up at this point at that time.

At later times, Eqn. 2.60 becomes an increasingly poor approximation: while the total error is small compared to the total angular distance traveled, it can grow to become large compared to a single orbit, i.e., $\sim 2\pi$. With negligible additional computational cost this can be accounted for to much greater accuracy by using the exact angular velocity ϕ_t at the approximate radius of the guiding center, a_0, as given by Eqn. 2.49.

Not all terms in Eqn. 2.63 are of equal importance: terms linear in ϕ_t grow linearly with time and $\epsilon_{\rm v}$ has a $\sqrt{r_0}$ dependence that makes it increasingly more important farther out in the disk. The same is the case for the second-order terms (unlisted). For this reason, when

$$\phi_t \sim \frac{1}{\epsilon_{\rm v}} + \frac{1}{\epsilon_{\rm m}} \tag{2.64}$$

the second-order terms proportional to ϕ_t can become first order. Writing this out explicitly tells us that Eqn. 2.63 is accurate until

$$\tilde{t} \sim \tilde{r}_0 \left(\frac{v_{\rm inner}}{v_{\rm kick}} + \sqrt{\tilde{r}_0} \frac{1}{\delta M / M_i} \right) . \tag{2.65}$$

To keep the first-order accuracy of Eqn. 2.63 at later times one should add to it the second-order terms proportional to ϕ_t:

$$\begin{aligned}
(\text{terms} \propto {\epsilon_{\rm v}}^2 \phi_t) &= \frac{3}{8} {\epsilon_{\rm v}}^2 \phi_t \big[14 \sin(\phi_t) + 3 \sin(2\phi_t) - \\
&\quad 29 \sin(2\theta_0) + 3 \sin(2\theta_0 + \phi_t) + \\
&\quad 5 \sin(2\theta_0 - \phi_t) + \sin\{2(\theta_0 - \phi)\} \big] \qquad (2.66) \\
(\text{terms} \propto \epsilon_{\rm m} \epsilon_{\rm v} \phi_t) &= -\frac{9}{2} \epsilon_{\rm m} \epsilon_{\rm v} \phi_t \big\{ \cos\theta_0 (\cos\phi_t - 3) + \\
&\quad \sin\theta_0 \sin\phi_t (1 + \cos\phi_t) \big\} \qquad (2.67) \\
(\text{terms} \propto {\epsilon_{\rm m}}^2 \phi_t) &= \frac{3}{2} {\epsilon_{\rm m}}^2 \phi_t (1 + \cos\phi_t) \sin\phi_t \, . \qquad (2.68)
\end{aligned}$$

At late times Eqn. 2.63 becomes

$$S = 1 + \frac{3}{2} \phi_t \left\{ -\epsilon_{\rm m} \sin\phi_t + \epsilon_{\rm v} \cos(\phi_t - \theta_0) \right\} . \tag{2.69}$$

At late times S reaches its extremal values earliest along the curve of particles initially located at $\theta_0 = 3\pi/2$. For these particles at late times

$$S = 1 - \frac{3}{2} \phi_t \sin\phi_t \left\{ \epsilon_{\rm m} + \epsilon_{\rm v} \right\} . \tag{2.70}$$

Therefore, for a given $\tilde{r}_0$,

$$S_{\min}^{\max} = 1 \pm \frac{3}{2}\phi_t \left\{\epsilon_{\rm m} + \epsilon_{\rm v}\right\} , \tag{2.71}$$

where the mth maximum occurs when

$$\phi_t = (2m - 1/2)\pi \, , \tag{2.72}$$

and the mth minimum occurs when

$$\phi_t = (2m + 1/2)\pi \, . \tag{2.73}$$

To solve for the positions of the extrema we find

$$\Omega_0\left(\theta_0 = \frac{3\pi}{2}\right) t \approx \frac{(n-1/2)\pi}{1 - 2\epsilon_{\rm m} - 3\epsilon_{\rm v}} \approx (n-1/2)\pi \equiv C_n \, , \tag{2.74}$$

which is approximately the same position as in the axisymmetric case of pure mass-energy loss (see Eqn. 2.33), except for the $\tilde{r}$ dependence of $\epsilon_{\rm v}$. Before $\epsilon_{\rm v}$ ceases to be just a small perturbation (as it is in the inner disk), this will cause the density peaks to slowly move off of an axisymmetric solution. However, for simple calculations, and within the relevance of our approximations, the positions of the extrema with kicks and without kicks are approximately the same. Therefore,

$$\tilde{r}_{0_n}^{\rm extr,kick} = \left(\tilde{t}/C_n\right)^{2/3} \, , \tag{2.75}$$

and the speeds of the peaks are given by

$$\dot{\tilde{r}}_{0_n}^{\rm extr,kick} = \frac{2}{3} C_n^{-2/3} \tilde{t}^{-1/3} = \frac{2}{3} \frac{1}{C_n \left(\tilde{r}_{0_n}^{\rm extr,kick}\right)^{1/2}} \, . \tag{2.76}$$

It is important to understand that a line of particles initially all at the same θ_0 will quickly wind up around the center, as particles closer in have significantly quicker orbital periods ($\propto \tilde{r}^{3/2}$). The maxima on the line of $\theta_0 = 3\pi/2$ represent points where a spiral arm of the surface density intersects this line, *and* is also where this spiral arm reaches its maximum values. Thus, we expect to see in a solution a spiral arm which oscillates in its maximum value. This can be seen in Fig. 2.11 where we plot $\Sigma(\tilde{r}, \theta, \tilde{t}{=}90)/\Sigma_0$ for a perturbation $\epsilon_{\rm v} = 0.0043\sqrt{\tilde{r}}$ and $\epsilon_{\rm m} = 0.013$ in units of $G = M_i = 1$. This is approximately the case where the kick imparted during the merger of two Schwarzschild black holes has its maximum effect (see Fig. 2.5), which occurs for a mass ratio of roughly 0.14. While the inner disk approaches the one-dimensional solution discussed in Sec. 2.4.1, the outer disk clearly shows the asymmetry introduced by the kick as a spiral structure.

From Eqn. 2.71 it's seen that the first caustic passes a radius $\tilde{r}$ near the time

$$\tilde{t}_{\rm c1}^{\rm (kick)} = \frac{2}{3}\frac{1}{\epsilon_{\rm m}+\epsilon_{\rm v}}\tilde{r}^{3/2}\,. \tag{2.77}$$

Using this to rewrite Eqn. 2.71 we have

$$S_{\substack{\rm max\\ \rm min}} = 1 \pm \tilde{t}/\tilde{t}_{\rm c1}^{\rm (kick)}\,. \tag{2.78}$$

And, therefore, it is also true for disks with kicks that the extrema lie on the curves

$$\left(\frac{\Sigma}{\Sigma_0}\right)_{\substack{\rm max\\ \rm min}} \approx \frac{1}{1\mp\tilde{t}/\tilde{t}_{\rm c1}^{\rm (kick)}}\,. \tag{2.79}$$

It should be noted that $\tilde{t}_{\rm c1}^{\rm (kick)}$ is less than the constraint given by Eqn. 2.64, indicating that along the curve of $\theta_0 = \pi/2$ Eqns. 2.66–2.68 offer little extra accuracy. However, for other θ_0 the first caustic will not arrive until later and it is in these cases that the extra terms provide extra accuracy.

2.5 Gaseous Disks

All particles in a disk will initially follow a non-interacting trajectory and only when pressure forces become significant will they deviate from this course. It is in this region prior to the pressure forces becoming relevant that the analytic solution of Sec. 2.4 is applicable. We now turn to understanding where the boundary of significant pressure forces is and what occurs beyond it.

We first discuss the simple disk model used. Then after introducing the epicyclic Mach number which characterizes the flow, we find the approximate time the flow deviates from the non-interacting solution. Shocks invariably form on this boundary and we give a strong lower bound for the strength of these shocks. Finally we solve for the surface density of the circularized regions of the disk following the shocks.

For clarity we first consider the axisymmetric case of pure mass-energy loss and then generalize the results to the complete problem including both mass-energy loss and in-plane kicks.

2.5.1 Disk Model

We assume a thin disk, initially in steady state, satisfying initial vertical hydrodynamic equilibrium. Then, the vertically integrated pressure is:

$$P(r) = \frac{GM_i}{r}\left\{\frac{h(r)}{r}\right\}^2\Sigma(r)\,, \tag{2.80}$$

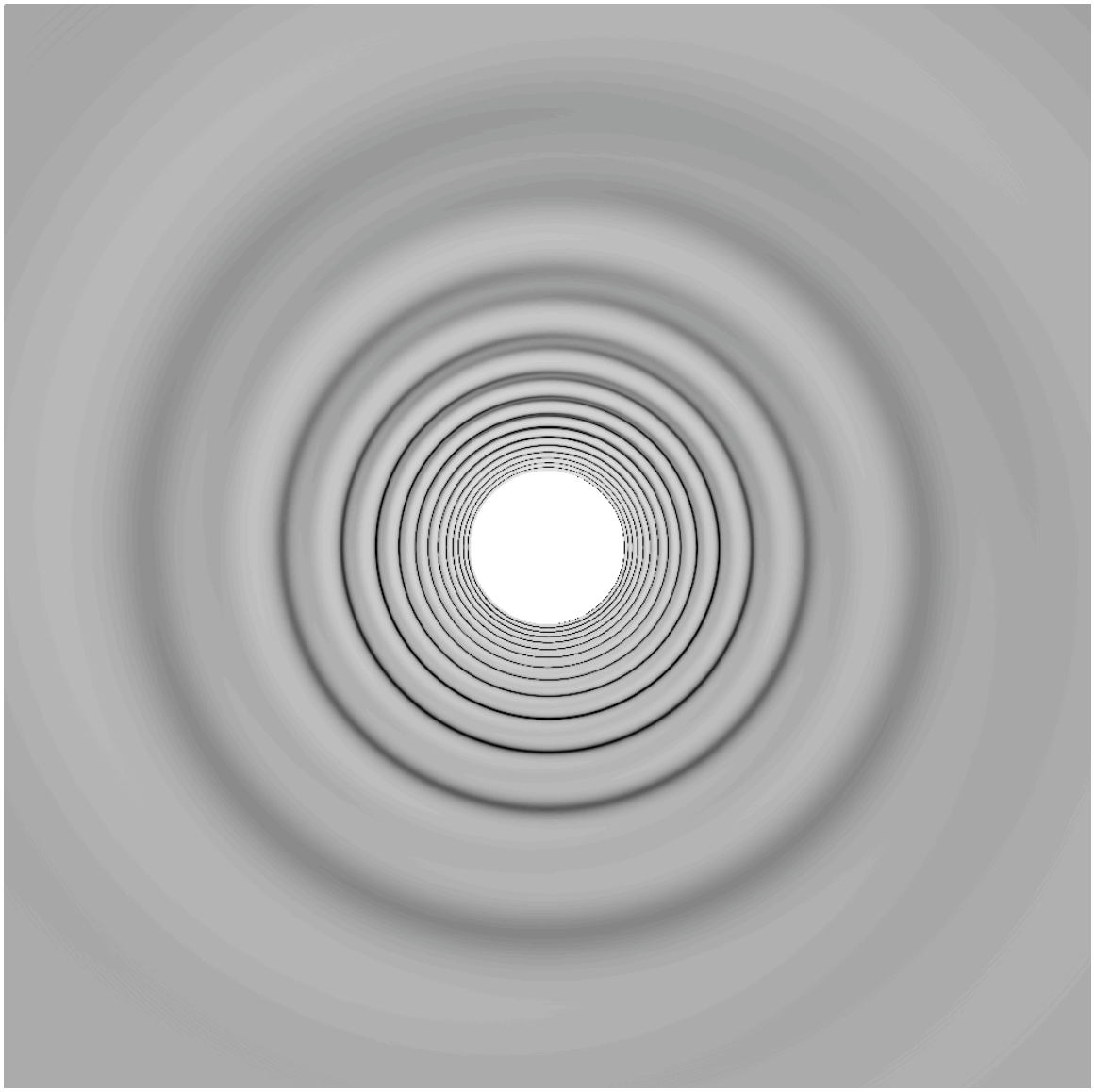


Figure 2.11: In units of $G = M_i = 1$ we plot $\Sigma(\tilde{t} = 90)/\Sigma_0$ for the merger of two Schwarzschild SMBHs chosen such that the kick is oriented to the left and the kick has maximal effect, as given by [70] (see Sec. 2.3.2). This point is marked on the dashed line of Fig. 2.5. In this case $\epsilon_{\rm m} \approx 0.013$ and $\epsilon_{\rm v} \approx 0.0043\sqrt{\tilde{r}}$. In the outer disk the perturbations become clearly two-dimensional while in the inner disk the perturbations approach an axisymmetric solution.

where $h(r)$ and $\Sigma(r)$ are the disk scale height and surface density at radius r, and the sound speed is given by

$$c_s^2 = v_{\rm K}^2 (h/r)^2 \,, \tag{2.81}$$

implying that $c_s \propto h/r^{3/2}$.

2.5.2 Epicyclic Mach Number

To understand the dynamics of the flow we begin with the radial Euler equation which governs the radial flow:

$$\frac{\partial v_r}{\partial t} + v_r \frac{\partial v_r}{\partial r} = -\frac{1}{\Sigma}\frac{\partial P}{\partial r} - \frac{\partial \Phi}{\partial r} + \frac{v_\phi^2}{r}\,, \tag{2.82}$$

where Σ is the surface density, Φ is the gravitational potential, P is given by Eqn. 2.80, and v_r and v_ϕ are the radial and azimuthal velocities, respectively. We define the epicyclic Mach number as the order of magnitude ratio of the kinetic term with the pressure term (analogous to the Mach number but specific to radial flow):

$$M_e \equiv \frac{v_{\rm epicycle}}{c_s} \sim \frac{v_{\rm orbital}\epsilon_{\rm m}}{v_{\rm orbital}\frac{h}{r}\left(\frac{\Sigma}{\Sigma_0}\right)^{\frac{\gamma-1}{2}}} \approx \frac{\epsilon_{\rm m}}{h/r}\,, \tag{2.83}$$

where c_s is given by Eqn. 2.81, $v_{\rm epicycle}$ is the maximum radial velocity and given by Eqn. 2.25 and Eqn. 2.29, and $(\Sigma/\Sigma_0)^{(\gamma-1)/2} \sim 1$. When $M_e \gg 1$ the sound speed is small compared to the radial velocity of the fluid. In this limit the kinetic term in the radial Euler equation dominates the pressure term and the flow is, therefore, indistinguishable from the non-interacting case described in Sec. 2.4 until near the passing of the first caustic.

On the other hand, when $M_e \ll 1$ the pressure term dominates the kinetic term and perturbations should evolve acoustically. It is important to remember that the non-interacting motions that initially gave rise to the perturbation continue to act and ensure that any perturbation continues traveling at the pattern speed, and that a rarefaction follows. It is only when $M_e \sim 1$ that shocks will form quickly (compared to $\tilde{t}_{\rm c1}$) from perturbations, since the kinetic and pressure terms in Euler's equations are comparable and maximize the nonlinear effects which turn acoustic waves into shocks.

2.5.3 Deviation Time

The time frame for the flow to deviate from the pressureless solution is merely the timeframe for a fluid particle to deviate from its epicyclic motion. To see when this occurs it is clearest to consider the radial Euler equation (Eqn. 2.82) in Lagrangian coordinates:

$$\frac{\rm D}{\rm Dt} v_r = -\left(1-\epsilon_{\rm m}\right)\frac{GM_i}{r^2} - \frac{\nabla P}{\Sigma} + \frac{v_\phi^2}{r}\,. \tag{2.84}$$

Here, the convective derivative is denoted $\frac{\rm D}{\rm Dt}$ and is the time derivative moving with a fluid particle. Initially, when the disk is in steady state this equation gives

$$0 = -\frac{GM_i}{r_0^2} + \frac{v_{\phi_0}^2}{r_0} - \frac{\nabla P_0}{\Sigma_0}\,. \tag{2.85}$$

Perturbing all quantities of Eqn. 2.84 except for the pressure and applying Eqn. 2.85 we get to first order in $\delta r/r_0$, $\delta v_\phi/v_{\phi_0}$, and $\epsilon_{\rm m}$:

$$\begin{aligned}\frac{\rm D}{\rm Dt}v_r &= \frac{GM_i}{r_0^2}\left(\epsilon_{\rm m}+2\frac{\delta r}{r_0}\right)+\frac{v_{\phi_0}^2}{r_0}\left(2\frac{\delta v_\phi}{v_{\phi_0}}-\frac{\delta r}{r_0}\right)\\ &\quad -\left(\frac{\nabla P}{\Sigma}-\frac{\nabla P_0}{\Sigma_0}\right). \end{aligned} \tag{2.86}$$

We cannot perturb the pressure because there is no reason why the pressure force itself remains small; it is only that the difference between the steady state pressure and the perturbed pressure must remain small. The final term on the right-hand side reflects this point, while the first two terms on the right-hand side determine the pressureless motion. When the pressure term is of order the pressureless term the flow will cease to follow the pressureless evolution.

To make this comparison we can employ the analytic solution found in Sec. 2.4.1:

$$\frac{\rm D}{\rm Dt}v_r = r_0\Omega_0^2\epsilon_{\rm m}\cos\phi_t\,, \tag{2.87}$$

$$\frac{\delta r}{r_0} = \epsilon_{\rm m}(1-\cos\phi_t)\,. \tag{2.88}$$

Conservation of angular momentum relates initial angular velocity to later values: $rv_\phi = r_0 v_{\phi_0}$. Therefore,

$$\frac{\delta v_\phi}{v_{\phi_0}} = -\epsilon_{\rm m}(1-\cos\phi_t)\,. \tag{2.89}$$

Plugging Eqns. 2.87–2.89 into Eqn. 2.86 and simplifying yields:

$$\begin{aligned}0 &= \epsilon_{\rm m}\frac{v_{\phi_0}^2}{r_0}\left(1-\frac{GM_i/r_0}{v_{\phi_0}^2}\right)(3-2\cos\phi_t)+\\ &\quad \left(\frac{\nabla P}{\Sigma}-\frac{\nabla P_0}{\Sigma_0}\right). \end{aligned} \tag{2.90}$$

There is a subtlety in this equation due to the fact that, as implied by Eqn. 2.85, when particles interact the circular velocity does not equal the Keplerian circular velocity and the first term on the right-hand side would be proportional to $\nabla P_0/\Sigma_0$. However, we do not want to compare the pressure term to the "non-interacting" terms traveling on orbits modified by pressure, we want to compare the pressure term to the non-interacting terms following the non-interacting motion. Therefore the circular velocity is the Keplerian circular velocity and the first term on the right-hand side of Eqn. 2.90 is 0 to first order. What does it mean to be 0 to first order? It means that the value of the term is smaller than the first-order terms. Therefore, deviation from the non-interacting solution occurs when

$$O(1) > b = \frac{r_0}{\epsilon_{\rm m}v_{\rm K}^2}\left(\frac{\nabla P}{\Sigma}-\frac{\nabla P_0}{\Sigma_0}\right), \tag{2.91}$$

where b is some constant smaller than unity and $v_{\rm K}$ is the Keplerian circular velocity. It is expected that an arbitrary constant appears here, since there is an arbitrariness to how one determines when one flow has deviated from another.

By noting that the vertically integrated pressure is such that $\nabla P = \nabla\left(\sqrt{\pi/\gamma}P_c(r)h(r)\right)$, where P_c is the central pressure, we find

$$\begin{aligned}
\nabla P &= \sqrt{\frac{\pi}{\gamma}}\left\{h(r)\nabla P_c(r) + P_c(r)\nabla h(r)\right\}, && (2.92)\\
&\sim \sqrt{\frac{\pi}{\gamma}}\left\{h(r)\frac{P_{cp}-P_{ct}}{\ell_g} + \frac{P_p}{r}\right\}, && (2.93)\\
&= \frac{P_p - P_t}{\ell_g} + \frac{P_p}{r} && (2.94)\\
&= \frac{P_p}{r}\left\{\lambda\left(1-\frac{P_t}{P_p}\right)+1\right\} && (2.95)\\
&= \frac{P_p}{r}\left[\lambda\left\{1-\left(\frac{1-\tilde{t}/\tilde{t}_{\rm c1}}{1+\tilde{t}/\tilde{t}_{\rm c1}}\right)^{\gamma}\right\}+1\right] && (2.96)
\end{aligned}$$

where the p and t subscripts denote whether a value is taken at the peak of a density perturbation or at the trough which precedes it, ℓ_g is the length scale relevant to the gradient (half the reduced wavelength is given approximately by the distance between peak and trough, divided by π) and $\lambda \equiv r/\ell_g$. In the final equality we used the adiabatic equation of state, $P \propto \Sigma^{\gamma}$ where γ is the adiabatic index, and Eqn. 2.38 to evaluate P_p and P_t. Using Eqn. 2.38 is equivalent to assuming a continuum of density peaks, an assumption which necessarily fails at early times when peaks are well separated. The effects of this are discussed in Fig. 2.12 and surrounding text.

Using the final equality above along with $\nabla P_0 \sim P_0/r$ we have

$$\begin{aligned}
b &= \frac{P_0/\Sigma_0}{\epsilon_{\rm m} v_{\rm K}^2}\cdot \\
&\quad \left[\frac{P}{P_0}\frac{\Sigma_0}{\Sigma}\left\{\lambda\left(1-\left(\frac{1-\tilde{t}/\tilde{t}_{\rm c1}}{1+\tilde{t}/\tilde{t}_{\rm c1}}\right)^{\gamma}\right)+1\right\}-1\right] && (2.97)\\
&= \frac{v_{\rm K}^2(h/r)^2}{\epsilon_{\rm m} v_{\rm K}^2}\cdot \\
&\quad \left[\left(\frac{\Sigma}{\Sigma_0}\right)^{\gamma-1}\left\{\lambda\left(1-\left(\frac{1-\tilde{t}/\tilde{t}_{\rm c1}}{1+\tilde{t}/\tilde{t}_{\rm c1}}\right)^{\gamma}\right)+1\right\}-1\right] && (2.98)\\
&= \frac{(h/r)^2}{\epsilon_{\rm m}}\cdot\left[\left(\frac{1}{1-\tilde{t}/\tilde{t}_{\rm c1}}\right)^{\gamma-1}\cdot \right.\\
&\qquad \left.\left\{\lambda\left(1-\left(\frac{1-\tilde{t}/\tilde{t}_{\rm c1}}{1+\tilde{t}/\tilde{t}_{\rm c1}}\right)^{\gamma}\right)+1\right\}-1\right]. && (2.99)
\end{aligned}$$

In the limit of small $\tilde{t}/\tilde{t}_{\rm c1}$ this reduces to

$$b = \frac{(h/r)^2}{\epsilon_{\rm m}} \left\{ 2\lambda \frac{\tilde{t}}{\tilde{t}_{\rm c1}} + \frac{\gamma - 1}{\gamma} \frac{\tilde{t}}{\tilde{t}_{\rm c1}} \right\} . \tag{2.100}$$

Using Eqn. 2.34 one can find ℓ_g by solving for the separation between the mth trough and the following peak. Then, making the approximation that $4m \gg 1$ (accurate to 5% for $m = 1$) one finds that

$$\lambda \approx \frac{\tilde{t}/\tilde{t}_{\rm c1}}{\epsilon_{\rm m}} . \tag{2.101}$$

Plugging this into Eqn. 2.100 we get

$$b = \frac{(h/r)^2}{\epsilon_{\rm m}} \left\{ \frac{2}{\epsilon_{\rm m}} \left(\frac{\tilde{t}}{\tilde{t}_{\rm c1}} \right)^2 + \frac{\gamma - 1}{\gamma} \frac{\tilde{t}}{\tilde{t}_{\rm c1}} \right\} . \tag{2.102}$$

When $\tilde{t}/\tilde{t}_{\rm c1} > \epsilon_{\rm m}$ the first term in the brackets dominates, while when $\tilde{t}/\tilde{t}_{\rm c1} < \epsilon_{\rm m}$ the second term dominates. However, the earliest we would expect the flow to deviate is when the first peak crosses a given fluid element. But the first peak crosses a given position at time $\tilde{t}_{p1} = \frac{5}{2}\pi \tilde{r}_{p1}^{3/2}$ (Eqn. 2.34), while the first caustic passes the same position at time $\tilde{t}_{\rm c1} = \frac{2}{3}\frac{1}{\epsilon_{\rm m}}$ (Eqn. 2.37). Then the earliest time to expect deviation is $\tilde{t}/\tilde{t}_{\rm c1} \approx 2\pi\epsilon_{\rm m}$. Therefore we can neglect the second term of Eqn. 2.102 and solve for $\tilde{t}/\tilde{t}_{\rm c1}$:

$$\frac{\tilde{t}_{\rm dev}}{\tilde{t}_{\rm c1}} \approx \sqrt{\frac{b}{2}} M_e . \tag{2.103}$$

Solving for $\tilde{t}_{\rm dev}$:

$$\tilde{t}_{\rm dev}^{\rm (cont)} = \sqrt{\frac{b}{2}} M_e \tilde{t}_{\rm c1} = \sqrt{\frac{b}{2}} \frac{2}{3} \frac{M_e}{\epsilon_{\rm m}} \tilde{r}^{3/2} \equiv \sqrt{\frac{2b}{9}} \frac{\tilde{r}^{3/2}}{h/r} , \tag{2.104}$$

where the "cont" indicates that we are in the continuous peak limit. For the remainder of the chapter we assume that $b = 1/2$, a value consistent with simulations.

When $M_e < 1$, Eqn. 2.104 implies that the speed with which a flow deviates from the non-interacting solution depends solely on the scale height. When $h/r \sim 0.1$ we expect either the first or second density peak to deviate from the non-interacting solution and when $h/r \ll 0.1$ we expect several density peaks following the non-interacting solution prior to deviation.

As mentioned earlier, the continuous peak limit used in Eqn. 2.96 to derive Eqn. 2.104 fails when the deviation time is smaller than or around the time for the third peak to arrive ($\tilde{t}_{\rm dev} \lesssim \frac{5\pi}{2}\tilde{r}^{3/2}$), since until then the peaks are well separated (see Fig. 2.12 for an example). In this case a better

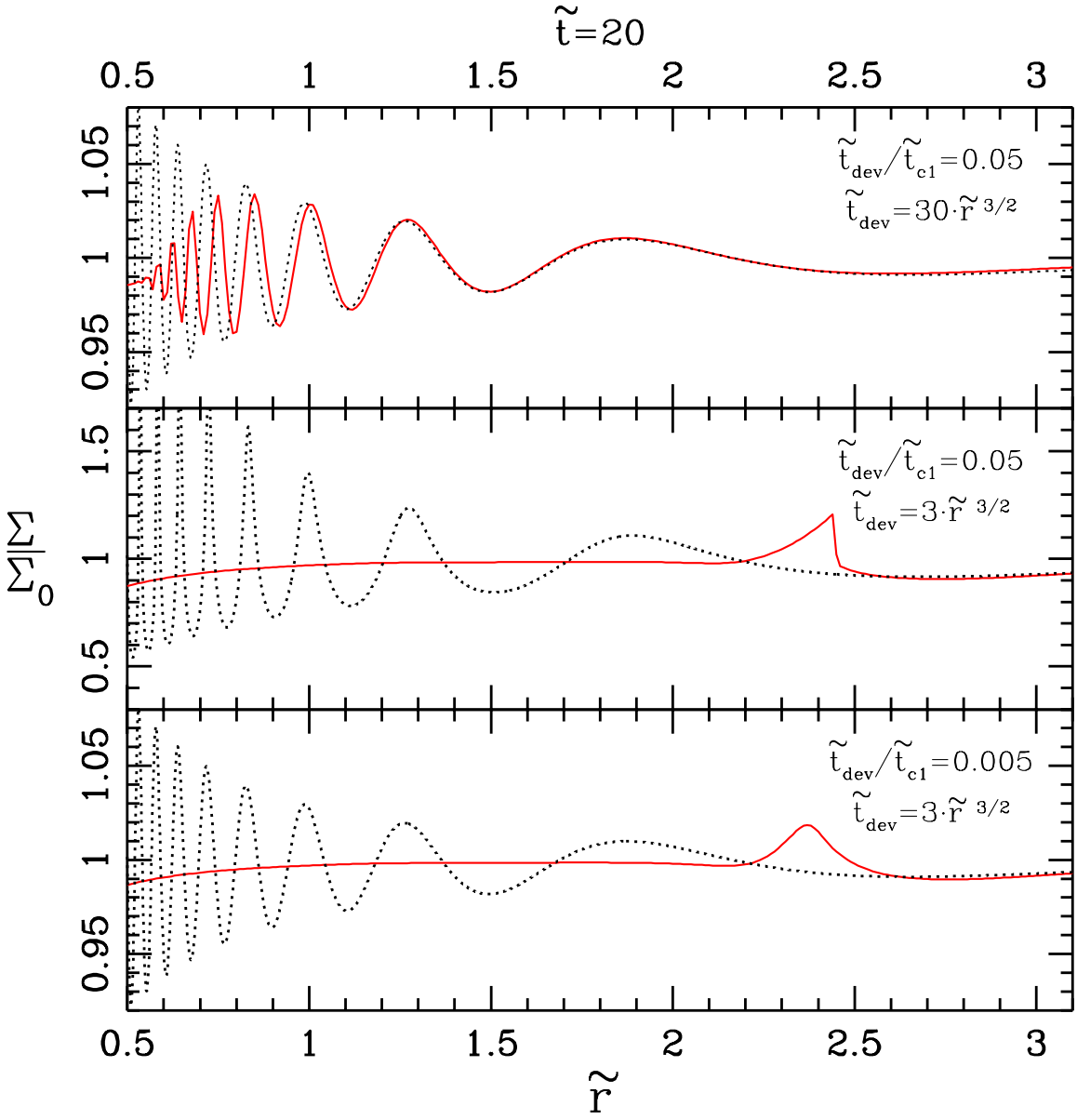


Figure 2.12: The disk response, Σ/Σ_0, in Eulerian coordinates for three example flows with small M_e all plotted at time $\tilde{t} = 20$. In each window the analytic solution given by Eqn. 2.30 is plotted in solid red and a 1D hydrodynamic simulation (discussed in Chapter 3) is plotted in dotted black. The parameters used for each window, from top to bottom, are $\{\epsilon_{\rm m} = 0.001, h/r = 0.001\}$, $\{\epsilon_{\rm m} = 0.01, h/r = 0.01\}$, and $\{\epsilon_{\rm m} = 0.001, h/r = 0.01\}$. The top two plots have the same $\tilde{t}_{\rm dev}/\tilde{t}_{\rm c1}$ (see Eqn. 2.103), while the bottom two plots have the same $\tilde{t}_{\rm dev}$ (see Eqn. 2.104). For the values of $\tilde{t}_{\rm dev}/\tilde{t}_{\rm c1}$ and $\tilde{t}_{\rm dev}$ labeled in each window we have assumed $b = 1/2$. It is important to note that there is an implicit spatial dependence in $\tilde{t}_{\rm dev}/\tilde{t}_{\rm c1}$ since $\tilde{t}_{\rm c1}$ depends on position (Eqn. 2.37). Considering the lower window, if we evaluate the deviation time at position $\tilde{r} = 2.5$ (approximately where the flow is deviating from the non-interacting solution) using Eqn. 2.104 we find $\tilde{t}_{\rm dev} \approx 13$. However, if we consider the discrete character of the peaks we find that the deviation time lies in the range: $\tilde{t}_{\rm dev} \in [19, 31]$, which is correct.

approximation to the deviation time is that it lies in the range:

$$\pi\tilde{r}^{3/2}\left(2p-\frac{5}{2}\right) \lesssim \tilde{t}_{\rm dev} \lesssim \pi\tilde{r}^{3/2}\left(2p-\frac{3}{2}\right), \tag{2.105}$$

where

$$p = \left\lceil \frac{1}{3h/r}\frac{1}{2\pi}+\frac{3}{4}\right\rceil . \tag{2.106}$$

Physically p is such that the pth peak of the non-interacting solution is the next to pass $\tilde{r}$ after time $\tilde{t}_{\rm dev}^{\rm (cont)}$. The upper (lower) bounds of Eqn. 2.105 are then just the times of the passing of the pth peak (the $(p-1)$th trough) over position $\tilde{r}$.

2.5.4 Physical Limitation of 1D Assumption

The 1D assumption will cease to be relevant when the density peak separations become comparable to the scale height, since the timescale for the disk to adapt in the vertical direction becomes comparable to the timescale for one peak to thermally interact with neighboring ones. Therefore the 1D approximation is only valid when

$$1 > \frac{h}{\ell_g} = \frac{h}{r}\lambda = \frac{h/r}{\epsilon_{\rm m}}\frac{\tilde{t}}{\tilde{t}_{\rm c1}}, \tag{2.107}$$

where we have applied the definition of λ in the first equality and Eqn. 2.101 in the second equality. This produces a limit just slightly weaker than the deviation time given by Eqn. 2.103:

$$\frac{\tilde{t}_{\rm 1D}}{\tilde{t}_{\rm c1}} < M_e , \tag{2.108}$$

where $\tilde{t}_{\rm 1D}$ is the range of times when the 1D approximation should be correct.

2.5.5 Shock Mach Number

Shocks form on the non-interacting overdensities as pressure forces become relevant, which occurs at approximately the time $\tilde{t}_{\rm dev}(\tilde{r})$ given in Sec. 2.5.3. One might expect that once a shock forms it could evolve independently of the peak on which it first formed, but this is not entirely true. The reason is that the flow preceding it continues to evolve as though the shock weren't there, meaning that the contraction which initially produced the shock will begin to form again if the the shock begins to recede. This effectively injects energy into the shock and ensures that the shock goes no slower than the non-interacting peak on which it formed. However, there is no reason the shock cannot move faster than the peak. Therefore we can put a clear lower bound on the shock speed. And, because in many cases the speed of the non-interacting peaks greatly exceeds the sound speed (Eqn. 2.35), this is a strong lower bound. Using Eqn. 2.35 and Eqn. 2.81, this allows us to determine

a lower bound for the Mach number of a shock having formed on the pth peak:

$$M_{\rm shock} \gtrsim \frac{\dot{r}}{c_s} = \frac{2}{3C_{2p-1}} \frac{v_{\rm K}}{v_{\rm K} h/r} = \frac{2}{3C_{2p-1}} \frac{1}{h/r} \,. \tag{2.109}$$

If $\tilde{t}_{\rm dev}/\tilde{t}_{\rm c1} \ll 1$ then one may use p as given by Eqn. 2.106. Otherwise, since $\tilde{t}_{\rm dev} \sim \tilde{t}_{\rm c1}$, the best lower bound for the shock strength would be to choose the number of the peak of the first caustic, given by $p_{\rm c1}$ (see Eqn. 2.40). In either case, it is clear that thin disks with strong mass perturbations will produce the strongest shocks.

2.5.6 Local Response Time

To determine the profiles at later times it is necessary to know whether the epicyclic patterns will affect an area of the disk before shocks will, if shocks have formed. Since shocks travel at $M_{\rm shock} c$, where $M_{\rm shock}$ is the shock's Mach number, when shocks form in the inner disk they will arrive at an outer radius r after a time

$$t_{\rm shock} \sim \frac{r}{M_{\rm shock} c_0} \,. \tag{2.110}$$

The time, $t_{\rm od1}$, for the peak of the first overdensity to reach a given radius r is attained from Eqn. 2.33:

$$t_{\rm od1} = \frac{\pi r}{2\ v_\phi(r)} = \frac{\pi}{2} \Omega(r)^{-1} \,.$$

If the shocks formed quickly then the ratio of the speeds gives the ratio of times that it will take for each to affect a given radius in the disk.

To understand whether the first overdensity will affect some region of the outer disk before shocks which have formed in the inner regions of the disk do, we define the dimensionless local response time as

$$\begin{aligned} L_r &\equiv \frac{t_{\rm od1}}{t_{\rm shock}} \sim \frac{r/v_\phi}{r/M_{\rm shock} c_0} \\ &= M_{\rm shock} \frac{h}{r} \left(\frac{\Sigma}{\Sigma_0} \right)^{\frac{\gamma-1}{2}} \sim \frac{h}{r} M_{\rm shock} \,. \end{aligned} \tag{2.111}$$

When $\tilde{t}_{\rm dev} \ll \tilde{t}_{\rm c1}$ then $L_r = \frac{2}{3C_{2p_{\rm c1}+1}}$, where $p_{\rm c1}$ is given by Eqn. 2.40. Because n_p cannot realistically be less than 1, then in the case of quick forming shocks $L_r \lesssim 1/20$, since one would not expect Eqn. 2.109 to be low by more than a factor of 2.

It should be noted that M_e and L_r have radial dependence which can allow the dynamics to change as the perturbations move outwards through the disk.

2.5.7 Final Surface Density

Let us first consider a fluid element whose center is located at r_0 directly following a merger. Initially its sides will attempt to follow the non-interacting path defined by Eqn. 2.29, but due to pressure forces will be damped and will eventually fall onto a circular orbit with its new orbital radius dictated by conservation of angular momentum: $r_0/(1-\epsilon_{\rm m})$. Because the fluid element is now slightly farther out than it was originally, the final volume it encompasses, and therefore the final density, is also perturbed.

If r_{a0} and r_{b0} are the initial positions of the sides of the fluid element, conservation of angular momentum also ensures that they circularize to the final positions of $r_{a0}/\left(1-\epsilon_{\rm m}\right)$ and $r_{b0}/\left(1-\epsilon_{\rm m}\right)$, respectively. Then the final surface density will be

$$\Sigma_{\rm final} = \frac{\rm mass}{2\pi r \Delta r} = \frac{m_{\rm ab}}{2\pi \frac{(r_{\rm a0}+r_{\rm b0})}{2}(r_{\rm a0}-r_{\rm b0})\frac{1}{(1-\epsilon_{\rm m})^2}} \tag{2.112}$$

$$= \Sigma_0 \left(1-\epsilon_{\rm m}\right)^2 \tag{2.113}$$

$$= \Sigma_0 \left(1-2\epsilon_{\rm m}\right) + \mathcal{O}(\epsilon_{\rm m})^2 \tag{2.114}$$

where $m_{\rm ab}$ is the mass of the fluid element. So, a perturbation in the central mass causes, following circularization, a perturbative decrease in the surface density of the disk of twice the magnitude of the mass perturbation.

2.5.8 Disks with Kicks

When considering the general case including kicks, the analysis of L_r remains unchanged, since to first order the motion of the overdensities is a function only of the original orbital frequency. The derivation for M_e is analogous to that of Sec. 2.5.2: find the maximum radial velocity at some $\tilde{r}$ determined by Eqn. 2.55, $v_\phi(\tilde{r}_0)(\epsilon_{\rm m}+\epsilon_{\rm v})$. The sound speed is still given by Eqn. 2.81. This gives the generalization of Eqn. 2.83 for M_e in the case of kicks in the plane of the disk:

$$M_e = \frac{\epsilon_{\rm m}+\epsilon_{\rm v}}{h/r}\,. \tag{2.115}$$

It is important to note that because this is the maximum radial velocity at a given $\tilde{r}$ (located at $\theta_0 = 3\pi/2$), it is also the maximum M_e at that radius.

The analysis to determine the time before the flow deviates at a given radius is identical up until Eqn. 2.86. From that point we may still proceed similarly and use the non-interacting solution with kicks (Eqn. 2.55 and Eqn. 2.60) to show that the non-interacting terms cancel out, leaving us with the equation:

$$O(1) > b^{(\rm kick)} = \frac{r_0}{(\epsilon_{\rm m}+2\epsilon_{\rm v})v_{\rm K}^2}\left(\frac{\nabla P}{\Sigma} - \frac{\nabla P_0}{\Sigma_0}\right)\,. \tag{2.116}$$

Because Eqn. 2.78 is of exactly the same form as Eqn. 2.38, ∇P is still given by Eqn. 2.96, except with $\tilde{t}_{\rm c1}^{\rm (kick)}$ instead of $\tilde{t}_{\rm c1}$. In the limit of $\tilde{t}/\tilde{t}_{\rm c1}^{\rm (kick)} \ll 1$ this yields the analogue of Eqn. 2.100:

$$b^{\rm (kick)} = \frac{(h/r)^2}{\epsilon_{\rm m} + 2\epsilon_{\rm v}} \left\{ 2\lambda^{\rm (kick)} \frac{\tilde{t}}{\tilde{t}_{\rm c1}^{\rm (kick)}} + \frac{\gamma - 1}{\gamma} \frac{\tilde{t}}{\tilde{t}_{\rm c1}^{\rm (kick)}} \right\} . \tag{2.117}$$

At $\theta_0 = 3\pi/2$ the peaks and troughs of the density distribution are located at the same $\tilde{r}$ as in the axisymmetric case of pure mass-energy loss, so taking into account that $\tilde{t}_{\rm c1}^{\rm (kick)}$ has $\epsilon_{\rm m} + \epsilon_{\rm v}$ instead of just $\epsilon_{\rm m}$, as in the case of pure mass-energy loss, we find

$$\lambda^{\rm (kick)} \approx \frac{\tilde{t}/\tilde{t}_{\rm c1}^{\rm (kick)}}{\epsilon_{\rm m} + \epsilon_{\rm v}} . \tag{2.118}$$

From there the calculation is straightforward and we find

$$\frac{\tilde{t}_{\rm dev}^{\rm (kick)}}{\tilde{t}_{\rm c1}^{\rm (kick)}} \approx \sqrt{\frac{b^{\rm (kick)}}{2} \frac{\epsilon_{\rm m} + 2\epsilon_{\rm v}}{\epsilon_{\rm m} + \epsilon_{\rm v}}} M_e . \tag{2.119}$$

This is not quite the same equation as for the case of pure mass-energy loss (Eqn. 2.103), though it differs by no more than a factor $\sqrt{2}$. The reason for this difference is the introduction of the prefactor of 2 for $\epsilon_{\rm v}$ in Eqn. 2.116. Solving Eqn. 2.119 for $\tilde{t}_{\rm dev}^{\rm (kick)}$ gives

$$\tilde{t}_{\rm dev}^{\rm (kick)} \approx \sqrt{\frac{2b^{\rm (kick)}}{9} \frac{\epsilon_{\rm m} + 2\epsilon_{\rm v}}{\epsilon_{\rm m} + \epsilon_{\rm v}}} \frac{\tilde{r}^{3/2}}{h/r} . \tag{2.120}$$

This is approximately the time when shocks will form at a position $\tilde{r}$. Just as in the axisymmetric case of pure mass-energy loss, these shocks should "ride" the face of the non-interacting peak from from which they formed. Thus, an approximate Mach number of these shocks can be calculated by taking the ratio of the velocity of the non-interacting peak crossing a position $\tilde{r}$ just following the time $\tilde{t}_{\rm dev}^{\rm (kick)}$:

$$M_{\rm shock} \gtrsim \frac{\dot{r}}{c_s} = \frac{\frac{2}{3}\frac{v_{\rm K}}{C_{2p-1}}}{v_{\rm K} h/r} = \frac{2}{3C_{2p-1}} \frac{1}{h/r} . \tag{2.121}$$

This is only true when $\tilde{t}_{\rm dev}^{\rm (kick)} \ll \tilde{t}_{\rm c1}^{\rm (kick)}$. Just as was the case with shocks resulting from pure mass-energy loss, if $\tilde{t}_{\rm dev}^{\rm (kick)} \sim \tilde{t}_{\rm c1}^{\rm (kick)}$ the Mach number can be bounded by using $p_{\rm c1}$ given by Eqn. 2.40, but with $\tilde{t}_{\rm c1}$ replaced by $\tilde{t}_{\rm c1}^{\rm (kick)}$. Again, in both of these cases, thin disks with strong perturbations (and, therefore, quickly forming caustics) will produce the strongest shocks.

Table 2.3: Notation: Equations and Descriptions of Parameters and Variables Listed in Alphabetical Order

Parameter	Description	Eqn. #
G	gravitational constant	n/a
M, M_i	initial total mass-energy of SMBHs	Eqn. 2.9
M_e	epicyclic Mach number	Sec. 2.5.2, Eqn. 2.83
M_f	final total mass-energy of SMBHs	text preceding Eqn. 2.23
r	radial position of particle in disk at some time following merger	Eqn. 2.24
$\tilde{r}$	$r/r_{\rm inner}$	Eqn. 2.32
r_0	initial radial position of particle in disk	text preceding Eqn. 2.23
$r_{\rm steady}$	radius inside of which disk follows binary	text preceding Eqn. 2.1
$r_{\rm bin}$	binary separation	text preceding Eqn. 2.5
$r_{\rm imp}$	radius outside of which all subsequent changes to the potential are seen as instantaneous	Eqn. 2.4
$r_{\rm inner}$	radius of inner disk ($200\,GM/c^2$ for calculations)	Eqn. 2.5
$r_{\rm ISCO}$	innermost stable circular orbit	Eqn. 2.2
$\tilde{t}$	$t\Omega_{\rm inner}$	Eqn. 2.32
$t_{\rm merger}$	time to merger	Eqn. 2.2
$t_{\rm visc}$	viscous timescale	Eqn. 2.1
$v_{\rm epicycle}$	maximum radial velocity during epicyclic orbit: $\epsilon_{\rm m} v_{\rm K}$	Eqn. 2.20
v_{ϕ_0}	angular velocity	Eqn. 2.20
$v_{\rm K}$	Keplerian circular velocity: $\sqrt{GM_i/r}$	Eqn. 2.91
$v_{\rm kick}$	recoil kick velocity received at merger	Sec. 2.3.3
v_r	radial velocity	Eqn. 2.82
$\epsilon_{\rm m}$	$(M_i - M_f)/M_i$	Eqn. 2.23
$\epsilon_{\rm v}$	$v_{\rm kick}/v_{\phi_0}$	Eqn. 2.43
η	$1/(1+q)^2$	Eqn. 2.11
ϕ_t	mean anomaly	Eqn. 2.28
$\Omega_{\rm inner}$	orbital frequency at $r_{\rm inner}$	Eqn. 2.6
θ_0	initial angular position of particle in disk	Eqn. 2.44

2.6 Conclusion

Circumbinary disks are one of the likely candidates for a prompt electromagnetic counterpart to a gravitational wave (GW) signal from super-massive black hole (SMBH) mergers. These electromagnetic events will allow for LISA to be used to independently determine the cosmological dark energy parameters, but also be astrophysically interesting events in themselves, offering, among other things, insights to circumbinary disks and their structure which have yet to be observed. We provide an analytic solution to the evolution of the surface density of these disks following merger in the limit of no pressure (Eqns. 2.30 and 2.62). Our solution provides not only an intuitive approach to understanding the evolution of a thin circumbinary disk, but it also provides an attractive test for simulations.

The two primary sources of dynamics in the post-merger disk are mass-energy loss due to GW radiation and recoil kicks due to asymmetric GW radiation during merger. In all cases the mass-energy loss is constrained to be $\lesssim 10\%$. When gas is present during the inspiral of two SMBHs it is possible that interactions with the gas act to align the spins of the SMBHs with their orbital angular momentum. In this aligned-spin configuration the kicks are in the plane and are constrained to have effects on the inner disk which are at most modest when compared to the effects of mass-energy loss (Figs. 2.5 and 2.6). The preshock flow for disks reacting to pure mass-energy loss is self-similar (brief discussion surrounding Eqn. 2.32), while the introduction of a relevant length scale in the general solution removes this special characteristic (Eqn. 2.43 and surrounding text).

To understand when the analytic solution breaks down and shocks begin to form we introduce the deviation time (Sec. 2.5.3). This time measures at what point the first density peak which will turn into a shock will pass a given position. In the limit that it is much smaller than the caustic formation time the caustic formation time is inversely proportional to h/r (mass-energy loss: Eqn. 2.104; with mass-energy loss and kicks: Eqn. 2.120). There is a strong lower bound for the strength of the shocks which invariably form as the flow deviates from the non-interacting solution, and it too goes as the inverse of h/r (mass-energy loss: Eqn. 2.109; with mass-energy loss and kicks: Eqn. 2.121). After the shocks pass the flow returns to a quasi-steady state. In the case of pure mass-energy loss, the surface density in this final disk returns to a value dependent only on the initial density and the mass perturbation (Eqn. 2.114). In this way the distinct regions of circumbinary disks reacting to the merger of SMBHs can be well explained.

In conclusion, we provide 1) an analytic solution in the limit of a pressureless disk to the evolution of the surface density of a circumbinary disk following the merger of super-massive black holes with their spins aligned or anti-aligned with the orbital angular momentum, 2) an expression for the shock formation time relevant throughout a significant fraction of parameter space, 3) a strong lower bound for the strength of these shocks, and 4) the final density profile after the disk returns to

a quasi-steady state. In general, the solution presented offers a method to predict and understand the evolution of all regions of circumbinary disks.

Chapter 3

1-D Simulations of Circumbinary Disks Following Massive Black Hole Mergers

N.B.: This work will be published with the following authors: Nate Bode and Sterl Phinney in this order.

Abstract

When two galaxies merge, the supermassive black holes (SMBHs) they may harbor at their centers inspiral, driven initially by interactions with stars and gas. If these interactions bring the SMBHs close enough, the radiation of energy and angular momentum in gravitational waves can complete the inspiral, ending with their ultimate merger. During this inspiral and final plunge, a significant fraction of the mass-energy of the binary system is released in the form of gravitational waves on timescales short compared to the dynamical timescales in the disk, leading to an impulsive reduction in the total black hole mass and a recoil "kick". In Chapter 2 we provided an analytic solution to the reaction of a non-interacting disk of particles, considering both the mass loss and kicks in the plane of the disk. We also showed that over a significant region of parameter space, mass loss dominates the reaction of the inner disk. Here we consider only the mass decrement. We use a 1-D hydrodynamic code to explore the full parameter space of gaseous (hydrodynamic) disks, and expose the diversity of possible responses to the dominant effect of mass loss due to gravitational waves. We show how the qualitative and quantitative features of these responses can be understood from the analytic solution from Chapter 2. Finally, this simple analysis is shown to reproduce well the majority of results of the 3D hydrodynamic code and 2.5D magnetohydrodynamic code of [23].

3.1 Introduction

Interest in electromagnetic counterparts (EMCs) to gravitational wave (GW) signals is increasing as we learn more about the number of ways they can be used to constrain cosmological and astrophysical parameters. EMCs corresponding to GW signals from supermassive black hole (SMBH) mergers would allow one to determine the $D_L - z$ relation with comparable accuracy to current and proposed Type Ia supernovae studies. This would be an independent and physics-based determination of the $D_L - z$ relation [75]. Though [75] found that the accuracy of using the EMC with GW signals is limited by weak lensing, [76] demonstrated that the error in the luminosity distance can be decreased by a factor of 2–3 using the non-Gaussianity of the magnification distribution.

In Chapter 2 we showed that in the case of a SMBH merger orbited by a gaseous disk, it should be possible to use the dynamics inferred from the EMC evolution to determine the disk thickness and sound speed as a function of radius in the disk, something which as of yet has eluded astronomers trying to study the structure and dynamics of SMBH accretion disks.

There have been several studies of circumbinary disks in the context of SMBH mergers: [36] discussed the turn-on of an AGN after viscosity refilled the circumbinary "hole" and gas finally reached the recently merged SMBH. [37] considered a non-interacting disk and showed that the kick of the final SMBH will produce spiral caustic structures in the disk. In Chapter 2 we extended their work by providing an analytic solution to the problem of a non-interacting disk reacting to both GW recoil kicks in the plane and GW mass decrement. We provided a way to relate this non-interacting solution to gaseous disks and to qualitatively and quantitatively understand what should happen with the gaseous flows. [38] simulated the reactions of a thick disk to both mass decrement and kicks. They included radiation transfer so as to have a more realistic understanding of the luminosity as a function of time. [41] presented a partially analytic study with simulations, considering both kicks and mass decrement. Though they concluded that mass decrement is of negligible importance to the dynamics of the disk, they did not take into account that the mass decrement and kicks are coupled: Schwarzschild black holes will only have modest kicks in the plane of the disk and will therefore be dominated by mass decrement (see Chapter 2), while when the SMBHs are spinning, the spins tend to align, significantly reducing the recoil velocities [77]. When the coupling between mass decrement and kicks is considered, throughout almost all parameter space where kicks remain in the plane of the disk, the mass decrement is found to be the dominant source of perturbations (Chapter 2). The work most comparable to that presented here, is that of [23], which considered a thin disk reacting to mass decrement. They ran two different types of code, one a 3D hydrodynamic code and the other a "2.5"D magnetohydrodynamic (MHD) code, with the central SMBH losing 10%, 5%, and 1% of its mass during merger. Of particular interest, they found that the MHD code and hydrodynamic code resulted in essentially the same flow, indicating that MHD effects were not

significant. Moreover, they noted that the most significant change in the luminosity was an initial *drop* as material moved outwards. This is largely corroborated by our results (Sec. 3.4.2).

Our primary focus in this chapter is to demonstrate the ease and accuracy with which the analytic non-interacting solution of Chapter 2 can be applied to understand almost the entire range of parameters for thin hydrodynamic disks reacting to a merger dominated by GW mass decrement. Because GW mass decrement is axisymmetric, and we are working in the thin disk approximation we can use a 1D code to elucidate the theory. After discussing the necessary assumptions and the code itself (Sec. 3.2), we present a survey of simulations spanning three orders of magnitude in both percentage mass decrement and disk thickness (Sec. 3.4). For each of the simulations we explain why we should expect what we see, given the tools presented in Chapter 2. This has the dual purpose of demonstrating both the wide variety of types of flows which can occur, and the utility of the analytic solution for understanding these vastly differing flows. One of the primary limitations of a 1D code is that it cannot account for turbulence or expansion in the vertical direction. For this reason we also use the analytic solution to describe and predict a variety of phenomena, including the shock speeds and luminosities, manifest in the simulations of [23] (Sec. 3.4.2). This demonstrates that the analytic solution gives insight into the behavior of a wide variety of disks, even those that are subject to non-axisymmetric perturbations and have entirely different density distributions. Sec. 3.5 gives our conclusions.

3.2 Simulation

3.2.1 Assumptions

Axisymmetric: As yet, there is no firm observational evidence for a circumbinary disk around a SMBH binary. Though certain aspects of the formation of these disks have been studied, a detailed understanding of their formation is currently beyond current numerical abilities. It is thus generally supposed that when gas is present, a gaseous disk will form based on roughly the same physics that would yield a standard accretion disk in the case of a single SMBH. In a binary, however, torques are exerted on the inner regions of the disk by the binary. [20], [36], [21], [22], and various other authors, have shown that when such a disk is co-rotating with the binary these torques inject angular momentum into the disk at resonant radii and effectively "push" material outwards. This process evacuates a region surrounding the binary whose size depends on the mass ratio of the binary. In the case of equal mass mergers [21] used simulations to model these circumbinary disks and showed that the inner radius of the disk, $r_{\rm inner}$, is about twice the binary separation. This was also recently shown analytically by [22]. When merger timescale of the binary becomes shorter than the viscous timescale of the disk, the disk effectively decouples from the binary, and diffuses inwards on a viscous timescale.

Following decoupling, the disk will evolve viscously in what will effectively be an axisymmetric gravitational potential. Then, by the time of merger, viscosity will have smoothed the disk to a state which is, to reasonable approximation, axially symmetric.

$r_{\rm inner} = 100\, r_{\rm S}$*:* Though we assume $r_{\rm inner} = 100\, r_{\rm S}$ in our simulations, as discussed in Chapter 2, the analytic solution is self similar and scales, along with all of the relevant positions and times in the flow, with wherever the peak of the density distribution is located. Thus, should the true inner radius of the disk be farther in or out, one need only scale our results be the appropriate factor such that the two inner radii match.

This point is made abundantly clear in Sec. 3.4.2 where our simulations using these dimensionless variables compare favorably, after scaling, with the 3D relativistic magneto-hydrodynamic simulations run by [23] with $r_{\rm inner} = 10\, r_{\rm S}$.

Our choice of $r_{\rm inner} = 100\, r_{\rm S}$ is appropriate for an equal-mass ratio merger [22, 36, 78]. However, as the mass ratio $q \leq 1$ decreases, $r_{\rm inner}$ will also decrease, until gas can eventually accrete onto the SMBHs [22]. For example, [22] find that, for $10^8\, M_\odot$ black hole, a gap will form in the disk for mass ratios greater than $q = 3 \times 10^{-3}$. In this case the binary decouples from the disk at binary separation $r_{\rm bin} \approx 15$. They consider several mass ratios, and find that the radius of decoupling changes smoothly with q: decoupling occurs at $r_{\rm bin} \approx 24$ for $q = 10^{-2}$ and $r_{\rm bin} \approx 56$ for $q = 10^{-1}$. Even in these extreme cases, the response of the disk exterior to the binary will undergo a response to what we describe here.

Note that as the mass ratio becomes more extremal, the mass decrement occurring in less than a dynamical time of the inner disk decreases for two reasons: The first is that as $r_{\rm inner}$ decreases, the dynamical time of the disk decreases ($\propto r_{\rm inner}^{3/2}$), so that only a smaller fraction of the mass loss occurs in less than a dynamical time. Moreover, as the mass ratio becomes more extremal the total mass-energy radiated during inspiral, merger and ring down also decreases. In fact, in the limit of $q \ll 1$ both of these points are accentuated, because, in contrast to comparable-mass mergers, only a negligible amount of the mass-energy $\propto q^2$ is radiated during the final plunge, merger, and ring down.

Newtonian: We do not consider relativistic effects in our simulations. According to [22], as long as $q \geq 10^{-2}$ this leaves $r_{\rm inner} \geq 12 r_{\rm S}$. Note that although these values are calculated for a SMBH primary mass of $10^8\, M_\odot$, [36] showed that $r_{\rm inner}$ has only a weak dependence on the primary's mass. Thus, general relativistic effects are negligible.

Thin Disk: In most circumstances, the disk is predicted [36] to be cool enough to be described using the well-developed theory of geometrically thin disks, where the scale height h, is less than the radial distance r.

We consider a disk in the central potential of a mass M_0, and model the disk as geometrically thin and vertically isothermal, with a scale height measured from the mid-plane (half-height) and

parameterized by

$$\tilde{h}(r) \equiv \frac{h}{r_{\rm inner}} = H\left(\frac{r}{r_{\rm inner}}\right)^N \equiv H\tilde{r}^N\,, \tag{3.1}$$

where $\tilde{h}$ and $\tilde{r}$ are, respectively, the scale height h and the radius r in units of $r_{\rm inner}$, and H and N are dimensionless positive constants.

Because the dependence on N is relatively weak, to simplify the analysis in this chapter we have chosen $N = 1$ for all of our simulations. This makes the effects of the varied parameters clearer. Moreover, it is then also the case that

$$\frac{h}{r} = \frac{\tilde{h}}{\tilde{r}} = H\,. \tag{3.2}$$

One-Dimensional: Because we are assuming an axially symmetric disk (see first assumption above), which is geometrically thin (previous assumption) the problem is reduced to one dimension. We ignore turbulence and MHD effects. In 1D, the waves and shocks in the disk necessarily propagate purely radially, with no vertical variation or refraction. This is unlikely to be accurate over many dynamical times, but as is corroborated in Sec. 3.4.2, it should be reasonable on the short timescales of interest to us.

Non-Self-Gravitating: We assume that the mass of the inner regions of the disk is significantly less than the mass of the SMBHs. This assumption is consistent with the results of [22] which find disk masses with masses around $10^3\,M_\odot$.

Impulsive: The disk responds adiabatically to mass-energy loss of the central binary over timescales longer than the dynamical time, slowly expanding but remaining circular to accommodate the change (see Chapter 2 for a more in-depth discussion of this). However, the mass-energy lost on timescales shorter than the dynamical time is perceived by the disk to be an impulsive change to the central potential and leads to the effects discussed in this chapter. Thus, we changed the central mass instantaneously at the beginning of every simulation with initial conditions appropriate for a steady-state disk orbiting the original mass.

Adiabatic: For simplicity, our simulations do not account for radiative losses, and thus strictly apply only to disks whose cooling time is longer than the timescales considered.

3.2.2 Code

We solve the 1-D problem using a standard Lagrangian finite-differencing leapfrog method with von Neumann-Richtmyer artificial viscosity [79]. Our code evolves the fluid-dynamical equations written in cylindrical coordinates (r, ϕ, and z), but simplified to 1-D using the assumptions discussed above (Sec. 3.2.1). This produces a geometric term in the radial Euler equation equal to u_ϕ^2/r where u_ϕ is the angular velocity.

Lagrangian codes are ideal for 1-D hydrodynamics because they naturally accommodate a flow

by populating denser regions with more grid points (and less dense regions with fewer points), also making them better than standard Eulerian finite-differencing methods in capturing shocks.

In a Lagrangian code it is beneficial to rewrite the derivatives in terms of $m(r)$, the mass inside the radius r, and the surface density $\Sigma(r)$. The derivatives with respect to r in the standard hydrodynamic equations are modified with the prescription:

$$\partial_r = 2\pi r \Sigma \partial_m \,. \tag{3.3}$$

Using this method we first compute, at the first half-time step, the radial velocity using the radial Euler equation:

$$\frac{\mathrm{D}}{\mathrm{Dt}} u_r = -(1-\epsilon_{\mathrm{m}}) \frac{GM_0}{r^2} - 2\pi r \partial_m (P+Q) + \frac{{u_\phi}^2}{r} \,. \tag{3.4}$$

Here u_r and u_ϕ are the radial and angular velocities of a fluid element, $\epsilon_{\mathrm{m}} \equiv |\delta M/M_0|$ is the fractional mass decrement of the initial SMBHs radiated "impulsively" in GWs during merger, and D/Dt is the convective derivative. The vertically integrated pressure is given by P, and Q is the von Neumann-Richtmyer effective viscous pressure described in more detail below (Eqn. 3.11).

The new radial position, a full time step ahead and at the same spatial position (in Lagrangian coordinates), is then determined using the definition of velocity:

$$\frac{\mathrm{D}}{\mathrm{Dt}} r \equiv u_r \,. \tag{3.5}$$

Because we have assumed that the disk is thin (no variation in the z direction) and axially symmetric (no variation in the ϕ direction), and because the fluid is effectively inviscid because we are considering timescales much shorter than the viscous timescale, the angular Euler equation reduces to conservation of the azimuthal angular momentum:

$$\frac{\mathrm{D}}{\mathrm{Dt}} (r u_\phi) = 0 \,. \tag{3.6}$$

This allows us to find u_ϕ at the same position and time as the radial position.

At the same time step at which r and u_ϕ are calculated, but a half-spatial step aside, we then calculate the thermodynamic quantities: the surface density Σ, pressure P, internal energy e, and temperature T. The surface density is calculated using Eqn. 3.3:

$$\Sigma(r) = \frac{\partial_r m}{2\pi r} \,. \tag{3.7}$$

It would be simple to compute the pressure and temperature using an adiabatic equation of state, but practically this introduces problems for two reasons. First, in shocks an adiabatic equation of state is not appropriate since shocks generate entropy. Artificial viscosity is a fairly effective kludge

used to deal with this. The second problem is that in our flows the temperatures transition between regimes where radiation pressure dominates and gas pressure dominates; e.g., in the case of strong shocks, the temperature may rise significantly across a shock, possibly converting a gas-pressure-dominated region into a radiation-pressure-dominated one. This we dealt with by using an iterative scheme to determine the pressure, energy, and temperature as follows.

The total integrated pressure is calculated as the vertically integrated sum of the gas and radiation pressures assuming temperature is independent of z:

$$P = \frac{k}{\mu m_p}\Sigma T + \frac{2}{3}a_R h(r) T^4 \,, \tag{3.8}$$

where k is Boltzmann's constant, μ is the mean molecular weight, m_p is the proton mass, and a_R is the radiation constant. The specific internal energy in a fluid cell is then given by

$$e = \frac{3}{2}\frac{k}{\mu m_p}T + 2a_R h(r)\frac{T^4}{\Sigma} \,, \tag{3.9}$$

where we have approximated the density as $\Sigma/h(r)$. It should be noted that the units of Eqns. 3.8 and 3.9 are removed by converting to code units ($G = M = r_{\rm inner} = 1$).

The temperature is then solved for iteratively using a form of Newton's method to solve for the roots of the conservation of energy equation

$$\frac{\mathrm{D}}{\mathrm{Dt}}e + (P+Q)\frac{\mathrm{D}}{\mathrm{Dt}}\frac{1}{\Sigma} = 0 \,, \tag{3.10}$$

solving for P and e at each step using Eqns. 3.8 and 3.9. When $\frac{\Delta T}{T} < 10^{-14}$ the process is terminated.

The final step is to calculate the artificial viscosity. We use an effective viscous pressure [79] (assuming a von Neumann-Richtmyer artificial viscosity) in cylindrical coordinates:

$$Q = \begin{cases} \frac{4}{3}\Sigma\ell^2\left(\frac{\partial u_r}{\partial r}\right)^2 & \text{if } \frac{\partial \Sigma}{\partial t} > 0 \\ \\ 0 & \text{else} \end{cases} \,. \tag{3.11}$$

Here ℓ is some small multiple of the particle separation and is usually written $\ell = k\Delta r$, where k is of order unity and Δr is the initial radial size of the fluid elements (thus ℓ is a constant).

3.2.3 Test Cases and Checks

There are three different types of flows which must be tested to ensure that the code works correctly in all the limits of interest: acoustic flows, flows with shocks, and non-interacting flows.

Acoustic Limit: To test the acoustic regime we considered a disk of constant surface density and sound speed ($N = 3/2$ where N is defined by Eqn. 3.1) and the inner boundary was oscillated to excite low amplitude acoustic waves (see Fig. 3.1). The waves followed the sound speed given by Eqn. 3.15 with accuracy limited by the determination of the peak position (the size of the fluid element).

Non-interacting Limit: In the non-interacting limit, the simulations provided in Sec. 3.4 are the best check available, since we have an analytic solution available to compare against. Though it is not strictly in the non-interacting limit, the late formation of the first caustic makes simulation 9, shown in Fig. 3.15, the best example of the code following the non-interacting solution for an extended period of time. The simulation is shown in solid (red) while the analytic solution to the non-interacting disk is shown in dashed (black). The two curves are almost indistinguishable, only becoming distinct at later times when the flow begins to interact with itself.

Shocks: Testing shocks is quite a bit more complicated since no exact analytic solutions exist for cylindrical coordinates. Therefore, the first test case considered was the SOD shock tube in Cartesian coordinates. We modify our code for cartesian coordinates following [79] and fix the adiabatic index to $\gamma = 5/3$.

For this test we follow [80] and use code units. The fluid velocity is initially set to 0 everywhere and the density and pressure are both set to 1 from $x = 0$ to $x = 0.5$, and to 1/8 elsewhere. We compare the density and pressure from the simulation to the analytic solution at time $t = 0.2$ in Figs. 3.2 and 3.3. There is excellent agreement. For instance, the magnitude of the density in each region is accurate to approximately 0.6%, except along the faces of the shock and contact discontinuity, where their finite width necessarily produces significant errors when compared to the discontinuous exact solution. We can see that the shock is traveling at the correct speed, because the discontinuity of the exact solution lies in the face of the shock as can be seen by the positive and negative error spikes at the shock face. Though this test case does not directly test the cylindrical equations it tests the fundamental ability of the algorithm to capture shocks and propagate them.

A final important check for the code discussed in Chapter 2, is that because of angular momentum conservation, any given fluid element located some distance from the inner boundary will eventually circularize to a new radius located at: $r_0(1 + \epsilon_{\rm m})$, where $\epsilon_{\rm m} \equiv \left|\frac{\Delta M}{M_0}\right|$, which we take to be positive. The slight motion outwards causes a geometric decrease in the surface density of the fluid element so that the final surface density can be easily determined (Eqn. 3.30). For each of our simulations we check this by averaging the value of $\Sigma_{\rm final}/\Sigma_0$ for 100th to 200th points in the simulation when this is possible (i.e., when the flow has completely circularized over this range). For simulations 6 and 9, this was not possible and we averaged over the 10th to 60th points in the domain. We do not start the averaging close to the inner boundary because of spurious variations due to a weak instability, discussed in more depth in Sec. 3.4.1. These values of $\Sigma_{\rm final}/\Sigma_0$ are shown in Table 3.2.

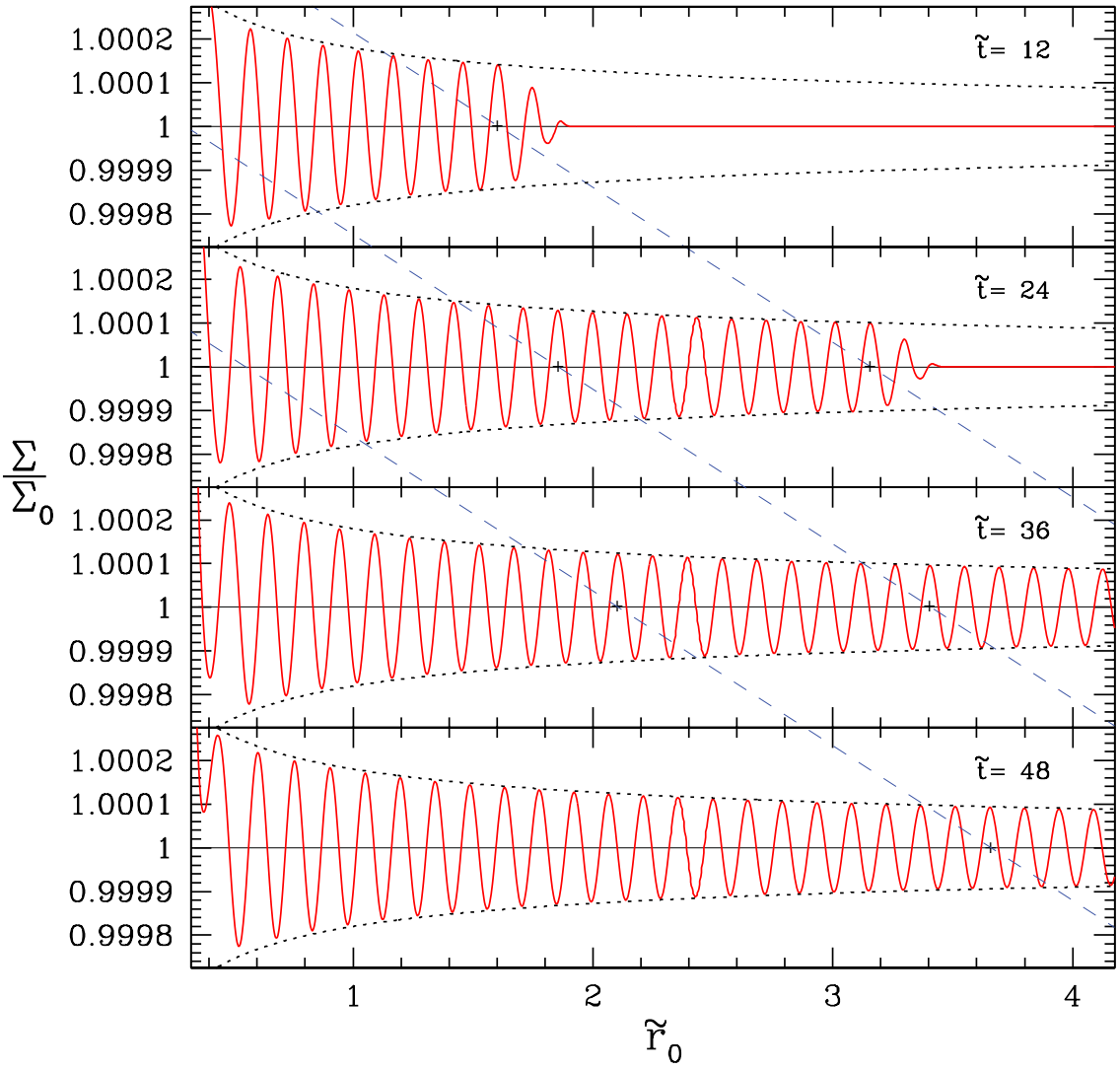


Figure 3.1: We test the ability of the code to deal with acoustic waves (Σ/Σ_0 shown in solid red) by oscillating the inner boundary to form low amplitude acoustic waves which propagate outwards through the disk. Parameters are chosen such that the sound speed is constant as a function of radius: We set $N = 3/2$ so that the sound speed has no explicit dependence on radius and the adiabatic index was forced to be 5/3 so that the implicit radial dependence of the sound speed is also removed. In the far field the amplitude of cylindrical acoustic waves is expected to decrease as $\tilde{r}^{-1/2}$ (shown in dotted black). Also shown in dashed blue are lines with slope appropriate for the sound speed as given by Eqn. 3.15. Peaks follow the sound speed to within relative errors of fractions of a percent, and limited by the determination of the location of the density peak.

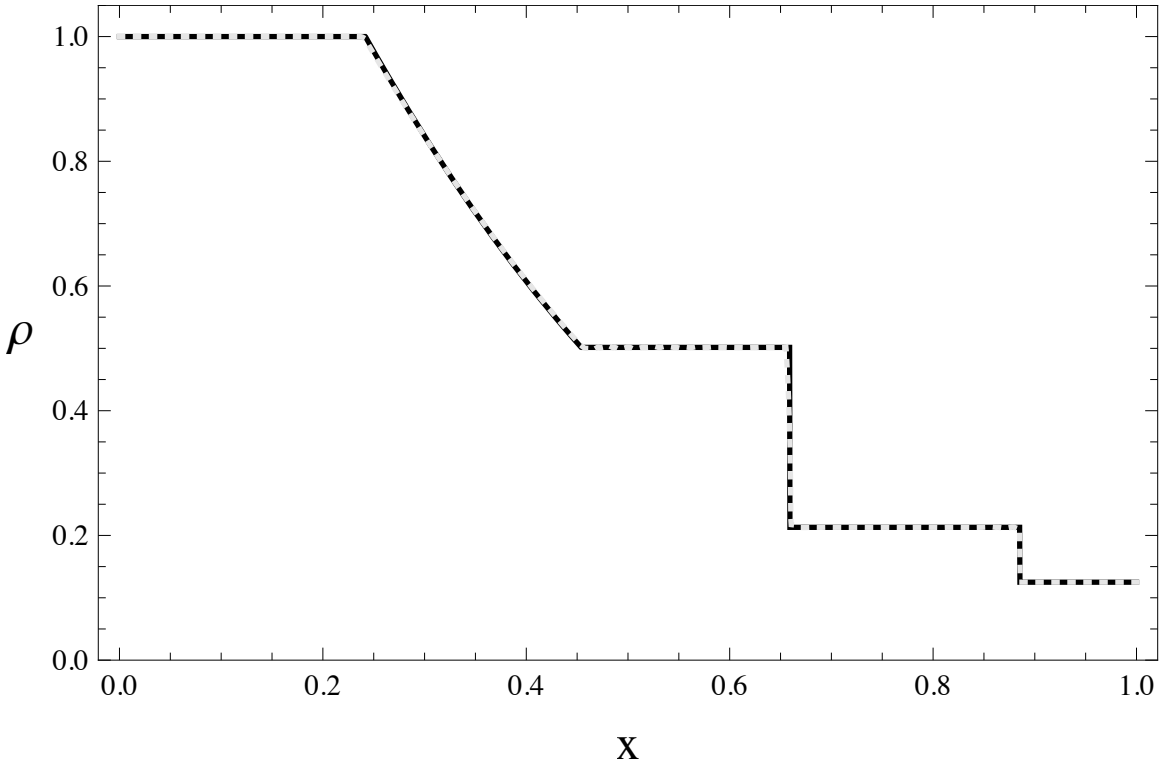


Figure 3.2: Sod shock tube results in code units. For $t = 0$ initial conditions specified in the text, at time 0.2 we compare the density of the analytic solution (solid black) to that of the simulation (dotted grey) as a function of position. There is excellent agreement in magnitude: excluding regions immediately surrounding the shock and contact discontinuity the analytic solution and simulation are in agreement to within half a percent. The errors are discussed more in Fig. 3.3.

3.2.4 Parameters and Initial Conditions

From here forward we specify radii in dimensionless units of the inner disk radius

$$\tilde{r} \equiv r/r_{\rm inner}\,, \tag{3.12}$$

where for $r_{\rm inner}$ we use the approximate radius of the *peak* of the initial density distribution (not the ill-defined inner edge of the initial disk; see Eqn. 3.16).

As described before (Eqn. 3.1), we parameterize the dimensionless scale height (in units of $r_{\rm inner}$) as a power law in $\tilde{r}$: $h(\tilde{r}) = H\tilde{r}^N$. In our simulations we keep N constant with a value of $N = 1$, and vary H from 10^{-3} to 10^{-1}. Due to the choice of N, $H = h/r$.

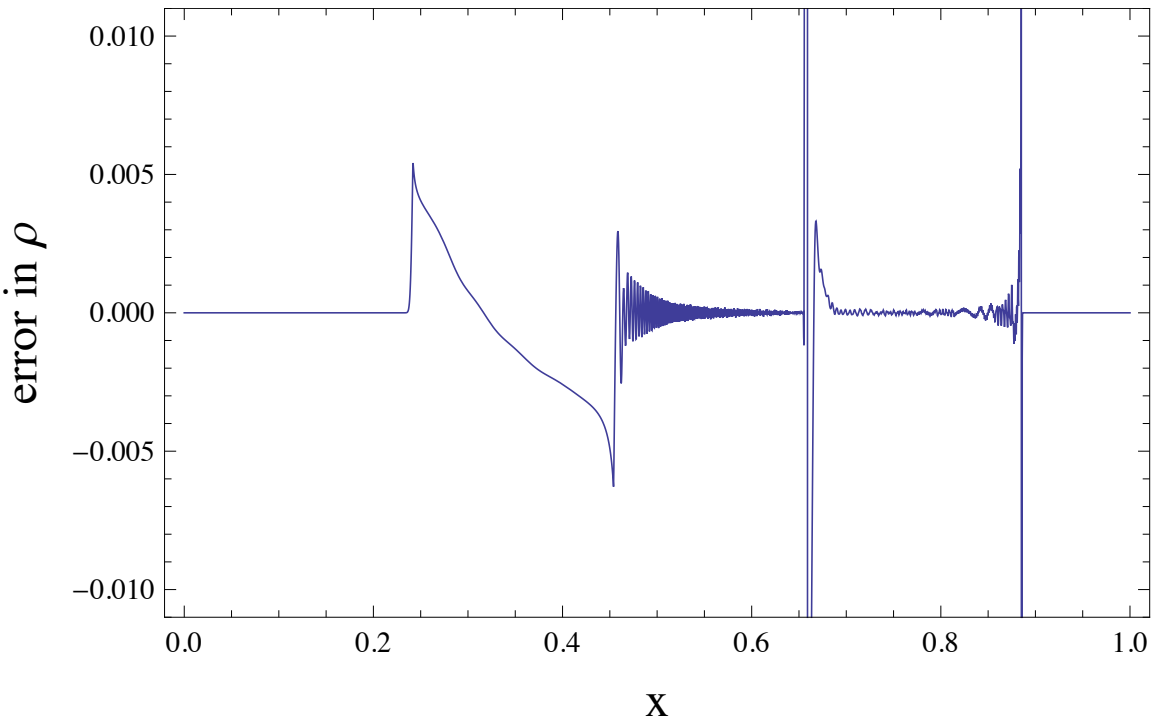


Figure 3.3: The fractional error of our simulation in the Sod shock tube case described in the text is plotted as a function of position (arbitrary units). Excepting regions close to the shock and contact discontinuity, the simulation produces a solution in excellent agreement with the analytic solution. The shock and contact discontinuity are traveling at the correct speed, to an accuracy limited by finite depth of the shock in our simulation. This can be seen by noting that the fractional error is both positive and negative across the faces of the shock and contact discontinuity; the face of the shock (and contact discontinuity) is spread across several cells, a necessary characteristic of finite differencing schemes. This makes the error spikes a necessary feature of any accurate finite differencing scheme, since the exact solution changes discontinuously.

In general, this produces a vertically integrated pressure:

$$P = \int_{-\infty}^{\infty} p(z)dz = -\int_{-\infty}^{\infty} z\frac{\partial p}{\partial z}dz = \Omega_K^2 \int_{-\infty}^{\infty} z^2 \rho dz \tag{3.13}$$

$$\equiv v_K^2 \left(\frac{h}{r}\right)^2 \Sigma\,. \tag{3.14}$$

Here Ω_K and v_K are the orbital frequency and angular velocity for a circular Keplerian orbit, respectively, and h, as before, is the standard Gaussian scale factor (therefore, measured from the mid-plane).

Then, the local sound speed is given by:

$$c_s = \sqrt{\gamma} v_{\rm K} \frac{h}{r}\,, \tag{3.15}$$

where $v_{\rm K}(\tilde{r})$ is the Keplerian circular velocity at radius $\tilde{r}$ and γ is the adiabatic index.

For all our simulations we take the initial density structure to be a flow similar to the final steady-state time-averaged surface density of the simulation done by [21] and parameterize it with a piecewise-defined function:

$$\Sigma_0(\tilde{r}) = \begin{cases} \sigma_0 \left(\frac{\tilde{r}}{\tilde{r}_{\rm m}}\right)^{-3} \exp\left\{-\left(\frac{\tilde{r}}{\tilde{r}_{\rm m}}\right)^{-2}\right\} & \text{for } \tilde{r} < \tilde{r}_{\rm p} \\ \\ \sigma_0 \frac{A\exp(-B\tilde{r}^C)}{(1+F\exp(-D(\tilde{r}-E)))} & \text{for } \tilde{r} > \tilde{r}_{\rm p} \end{cases} \tag{3.16}$$

Here, exterior to the peak of the distribution, $\tilde{r}_p$, the variables A–F are chosen (see below) to both approximate a digitized version of the final time-averaged profile of [21] and to match the inner region at $\tilde{r}_p$. The inner region is not chosen to be the exponential drop off seen in the profile of [21], but instead a slightly more modest power law which does not drop off quite as quickly since, in reality, by the time of merger the inner regions of the density profile will have diffused inwards and no longer be as steep as those of [21]. The fitted values are: $A \approx 0.730$, $B \approx 0.501$, $C \approx 0.714$, $D \approx 13.254$, $E \approx 0.797$, $F \approx 2.520$, $\tilde{r}_m \approx 1.397$, and $\tilde{r}_p \approx 1.140$. The disk is truncated approximately at $\tilde{r} = 0.25$.

Besides the density, the other variables solved for by the simulation are the radial velocity (u_r), the angular velocity (u_ϕ), pressure (P), the temperature (T), and the internal energy (e). To ensure particles are initially on circular orbits we initially set $u_r = 0$ and solve Eqn. 3.4 for u_ϕ. This is done by relating the vertically integrated pressure, P, to the surface density using Eqn. 3.14. The temperature is then just found by solving Eqn. 3.8 for T, while the internal energy is found using Eqn. 3.9, just as is the case at any other time step.

In summary, h/r and $\epsilon_{\rm m}$ are the input parameters to the simulation. The initial surface density is the same for all of the simulations and given by Eqn. 3.16, and P is then given by Eqn. 3.14.

Eqn. 3.8 then provides the initial T, which then gives e using Eqn. 3.8. Because the gas particles are assumed to be initially on circular orbits, $u_r = 0$ initially and then u_ϕ is found by solving Eqn. 3.4.

3.2.5 Boundary Conditions

The surface density at the inner boundary is 11 orders of magnitude smaller than the peak surface density, at $\tilde{r}_p$, and about 9 orders of magnitude smaller than anywhere in the main regions of the flow. This means that the gas at the inner boundary is so diffuse that as long as the simulation remains stable at or near the boundary it has no physical effect on the rest of the simulation. Therefore we can choose the inner boundary condition to ensure stability and not necessarily for strict physical realism. Finding a stable boundary condition proved to be particularly difficult because of the rapidly declining density distribution. Ultimately, a nonreflecting boundary condition consistently provided good results and is the chosen inner boundary condition for the simulations shown here. The nonreflecting boundary condition is described by

$$\frac{\partial u_r}{\partial t} = -c_s \frac{\partial u_r}{\partial \tilde{r}} = -2\pi \tilde{r} \Sigma c_s \frac{\partial u_r}{\partial m}\,, \tag{3.17}$$

where c_s is the local sound speed and given by Eqn. 3.15.

The outer boundary is dealt with using a linear extrapolation scheme. To check that these boundary conditions do not introduce significant errors, the energy flux through the boundary is monitored during the simulation. The total energy lost/gained through the boundaries always remains orders of magnitude smaller than the total energy in the flow.

3.3 Summary of Theory

3.3.1 Non-interacting Equations

For reference, we recapitulate the key points of Chapter 2 which we will use in this chapter. All flows will begin as though they are non-interacting and only when pressure forces are significant will they deviate from that motion. In Chapter 2 we presented the analytic solution for the non-interacting case. In particular, we showed that the surface density of a fluid element in Lagrangian coordinates as a function of time and initial position is:

$$\frac{\Sigma(\tilde{r}_0, t)}{\Sigma_0(\tilde{r}_0)} = \frac{1}{S(\tilde{r}_0, \tilde{t})}\,, \tag{3.18}$$

where $\tilde{t} \equiv \Omega_{\rm inner} t$, $\tilde{r}_0 \equiv r_0/r_{\rm inner}$, and

$$S(\tilde{r}_0, \tilde{t}) = \left| 1 + 2\epsilon_{\rm m} \left\{1 - \cos\phi_t\right\} - \frac{3}{2}\epsilon_{\rm m} \tilde{t} \tilde{r}^{-3/2} \sin\phi_t \right|\,, \tag{3.19}$$

$\epsilon_{\rm m} = \Delta M/M_0$ and ϕ_t is the first-order approximation to the mean anomaly:

$$\phi_t(\tilde{r}_0, \tilde{t}) = \frac{(1-2\epsilon_{\rm m})^{3/2}}{1-\epsilon_{\rm m}} \tilde{t}\tilde{r}_0^{-3/2} = (1-2\epsilon_{\rm m})\tilde{t}\tilde{r}_0^{-3/2} + O({\epsilon_{\rm m}}^2)\,. \tag{3.20}$$

Returning to the fluid equations, we showed where the non-interacting solution ceases to be valid. Because the non-interacting solution breaks down exactly when pressure forces become relevant to the flow, in the absence of dissipation this is also the region where shocks form. In fact, shocks invariably form at this boundary first (see Sec. 3.4.1), though they may not form until the outer regions of the disk (i.e., at $\tilde{r} \gg 1$).

Moreover, the analytic solution can be used to provide a strong lower bound on how strong the shock(s) can be and what occurs following the shocks. In general, the solution provides both quantitative and qualitative insights into the different regions of the flow.

To understand the different types of flows it was necessary to introduce the epicyclic Mach number:

$$M_e \equiv \frac{\delta M/M}{h/r} \equiv \frac{\epsilon_{\rm m}}{h/r}\,. \tag{3.21}$$

The physical importance of the epicyclic Mach number can be summarized as follows: when $M_e \ll 1$ the flow is in an "acoustic" regime where pressure forces dominate the flow and when $M_e \gg 1$ the flow is initially in a "pressureless" regime.

In terms of M_e we showed in Chapter 2 that the time it takes for the flow to deviate from the non-interacting solution is approximately:

$$\tilde{t}_{\rm dev} \sim \begin{cases} \frac{1}{3}\frac{\tilde{r}^{3/2}}{h/r} & M_e \leq 1 \\ \\ \tilde{t}_{\rm c1} & M_e \gtrsim 3 \end{cases}\,. \tag{3.22}$$

Here $\tilde{t}_{\rm c1}$, a function of $\tilde{r}$, is the caustic formation time at position $\tilde{r}$. Even a very cold flow ($M_e \gg 1$) will deviate from a pressureless one at $\tilde{t}_{\rm c1}$, since hydrodynamically forbidden intersecting flows occur in the non-interacting solution after the first caustic forms: so even in the cold hydrodynamic flows a shock will form at $\tilde{t}_{\rm c1}$. Hotter flows ($M_e \lesssim 1$) will deviate from the collisionless flow earlier than $\tilde{t}_{\rm c1}$ due to the effect of pressure forces in the perturbation (acoustic waves). $\tilde{t}_{\rm c1}$ is given by:

$$\tilde{t}_{\rm c1} \approx \frac{2}{3}\frac{1}{\epsilon_{\rm m}}\tilde{r}^{3/2}\,. \tag{3.23}$$

The caustic formation time plays a central role in the application of the analytic solution. This is

readily seen by solving Eqns. 3.18 and 3.19 for the curve on which the local extrema lie:

$$\left(\frac{\Sigma}{\Sigma_0}\right)_{\substack{\max\\\min}} \approx \frac{1}{1 \mp \tilde{t}/\tilde{t}_{\rm c1}} \,. \tag{3.24}$$

The surface density diverges as $\tilde{t}$ approaches $\tilde{t}_{\rm c1}$. Somewhere prior to the divergence, pressure forces will become relevant and the non-interacting solution will cease to be accurate (see Eqn. 3.22).

Interestingly, the maxima and minima of the normalized density distribution are constants as a function of time. This is readily seen by noting that the maxima of the surface density occur when

$$\Omega_0 t \approx \frac{(n-1/2)\pi}{1-2\epsilon_{\rm m}} \approx (n-1/2)\pi \equiv C_n \ . \tag{3.25}$$

Plugging Eqn. 3.25 into Eqns. 3.18 and 3.19 one finds

$$\left(\frac{\Sigma}{\Sigma_0}\right)_{\rm (extr)} \approx \begin{cases} \frac{1}{1-\frac{3}{2}\epsilon_{\rm m}\pi(2p-3/2)} & \text{for } p\text{th peak} \\ \\ \frac{1}{1+\frac{3}{2}\epsilon_{\rm m}\pi(2m-1/2)} & \text{for } m\text{th minimum} \end{cases} \,. \tag{3.26}$$

One can also invert Eqn. 3.23 to find the location of the first caustic as a function of time:

$$\tilde{r}_{\rm c1} \approx \left(\frac{3}{2}\epsilon_{\rm m}\tilde{t}\right)^{2/3} \,. \tag{3.27}$$

Similarly, one can find the deviation location, i.e., the location where the flow deviates from the non-interacting solution at some given time, by inverting Eqn. 3.22:

$$\tilde{r}_{\rm dev} \sim \begin{cases} \left(3\frac{h}{r}\tilde{t}\right)^{2/3} & M_e \leq 1 \\ \\ \left(\frac{3}{2}\epsilon_{\rm m}\tilde{t}\right)^{2/3} & M_e \gtrsim 3 \end{cases} \,. \tag{3.28}$$

At time t, the pressure forces begin to dominate the flow inside radius $\tilde{r}_{\rm dev}$, and it is at about $\tilde{r}_{\rm dev}$ that shocks will first start to form. The time it takes for a perturbation to evolve into a shock is flow dependent: in hot disks where pressure forces dominate the flow quickly (acoustic limit with $M_e \lesssim 1$) shocks form slowly and a perturbation does not form a shock until $\tilde{r} \gg 1$, while in cold disks (non-interacting limit with $M_e \gg 1$), where pressure forces are negligible until approximately the formation of the first caustic, shocks form quickly in the inner regions of the disk ($\tilde{r} \lesssim 1$). In both cases, the shock Mach numbers can be approximately determined analytically and are given

by

$$M_{\rm shock} \sim \begin{cases} 2 & M_e \leq 1 \\ & \\ M_e & M_e \gtrsim 3 \end{cases} . \tag{3.29}$$

Any given particle re-circularizes its orbit on a timescale comparable to its deviation time. As discussed in Sec. 3.2.3 (in the context of test cases for the code), because $t_{\rm visc} \gg \tilde{t}_{\rm dev}$, the angular momentum of the particle is approximately conserved, and the final surface density can be easily determined:

$$\frac{\Sigma_{\rm final}}{\Sigma_0} = (1-\epsilon_{\rm m})^2 . \tag{3.30}$$

In this way the analytic solution of Chapter 2 describes the regions preceding the shocks (Eqns. 3.18 and 3.19), where the flow will deviate from this solution and shocks will eventually form (Eqn. 3.22), how strong the shocks will be (Eqn. 3.29), and ultimately what surface density settles to after the shocks have passed (Eqn. 3.30).

3.3.2 Understanding Parameter Space

The plausible range of parameters covers a wide variety of disk response behaviors. And though there is some dependence on the rate at which $h(r)$ varies with r (the parameter N), over the relatively small inner region of the circumbinary disk that we consider, this dependence is weak. Thus, the variety of possible flows is almost entirely determined by the two other physical parameters in our model, h/r and $\epsilon_{\rm m}$. As discussed in Chapter 2 and reviewed in Sec. 3.3.1, the epicyclic Mach number M_e (Eqn. 3.21) separated parameter space into two regions: an "acoustic" region and a "non-interacting" region.

To better understand the range of disk responses we provide a "map" of parameter space in Fig. 3.4. The independent variables, $\epsilon_{\rm m}$ and h/r, are the x and y axes, respectively, while lines of constant M_e are dotted blue. Regions with $h/r > 0.3$ are marked in gray cross hatch and are not well described by our thin disk solution. Also, numerical relativity simulations of the final stages of merger, combined with post-Newtonian approximations of mass-energy loss during the early inspiral currently indicate an upper limit on $\epsilon_{\rm m}$ of about 10% the total initial mass. Regions outside the area of this limit are also marked by gray cross hatch. The epicyclic Mach number M_e splits parameter space into two regions: an "acoustic" region with $M_e \lesssim 1$ (shaded light blue) and a "non-interacting" region with $M_e \gg 1$ (varying shades of red). Analytic approximations to lines of constant deviation time (Eqn. 3.22) in the acoustic limit (dashed green) and the non-interacting limit (dashed red) are shown and connected to each other (dashed black). In general, we expect weak shocks with $M_{\rm shock} \sim 2$ in the acoustic region and strong shocks with $M_{\rm shock} \sim M_e$ in the non-interacting region (see Eqn. 3.29).

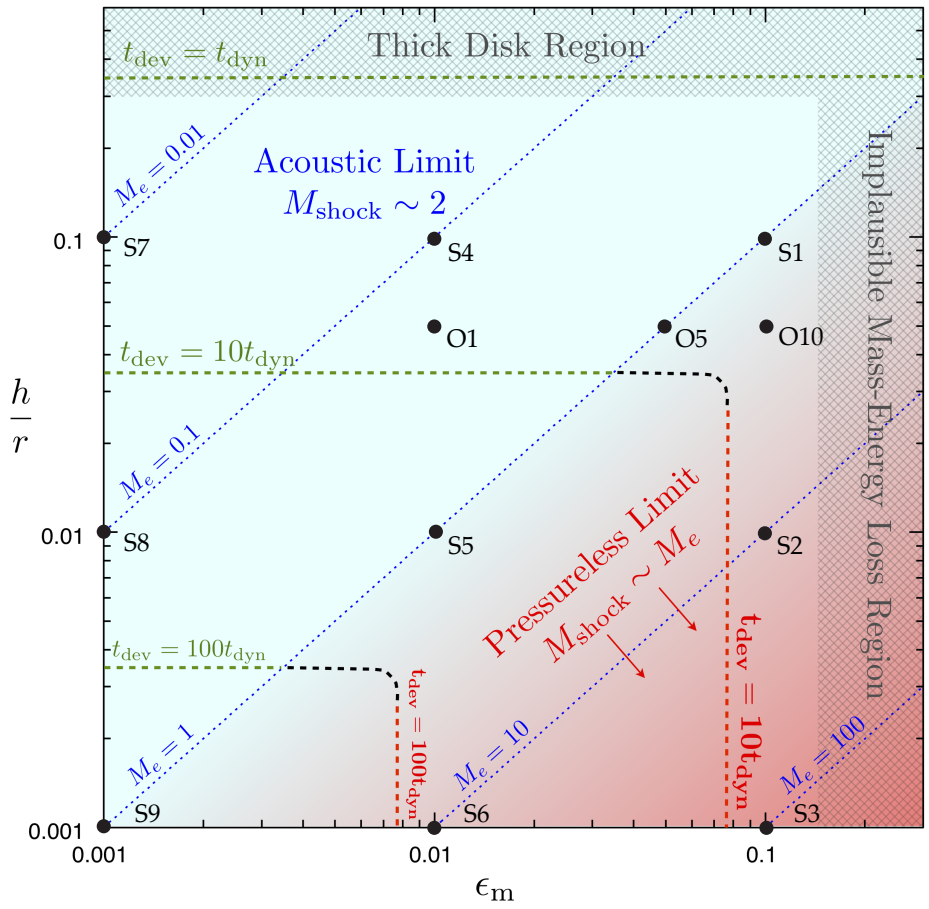


Figure 3.4: The parameter space of disk response as a function of the fractional SMBH mass decrement to gravitational waves, $|\epsilon_{\rm m}| \equiv |\delta M/M_0|$, and the local disk half scale height h (measured from the mid-plane) in units of the local disk radius r. Also shown are lines of constant M_e (defined in Eqn. 3.21). When $M_e \lesssim 1$ the flows lie in the weak shock region (blue), while when $M_e \gg 1$ the flows lie in the strong shock region (red) where the shock Mach numbers are of order the epicyclic Mach number. Gaseous flows deviate from the non-interacting solution around the deviation time, $\tilde{t}_{\rm dev}$ (Eqn. 3.22). Lines of constant deviation time, in units of the local dynamical time, are shown in dashed lines, though the curved (black) segments are connections between the two limits in which the deviation time is known (Eqn. 3.22). For reference we include the locations of the simulations presented in Sec. 3.4 numbered by their simulation number given in Table 3.1 ($S\#$), and give snapshots of these simulations in Fig. 3.5.

Table 3.1: Simulations, Their Parameters, and the Analytic Predictions

#	$\epsilon_{\rm m}$ [a]	$H = h/r$ [b]	N [c]	k [d]	$r_{\rm num}$ [e]	$t_{\rm num}$ [f]	M_e [g]	$\tilde{t}_{\rm c1}(\tilde{r}=1)$ [h]	$\tilde{t}_{\rm dev}(\tilde{r}=1)$ [i]	$\tilde{r}_{\rm dev}(\tilde{t}=30)$ [j]
1	0.1	0.1	1.0	1.25	5000	10^5	1	7	3	2.9
2	0.1	0.01	1.0	3.25	3000	10^5	10	7	7	2.4
3	0.1	0.001	1.0	1.25	3000	10^6	100	7	7	2.4
4	0.01	0.1	1.0	3.25	5000	10^5	0.1	70	3	2.9
5	0.01	0.01	1.0	1.25	5000	10^5	1	70	30	0.9
6	0.01	0.001	1.0	1.25	5000	10^5	10	70	70	0.6
7	0.001	0.1	1.0	8.25	5000	10^5	0.01	700	3	2.9
8	0.001	0.01	1.0	1.25	5000	10^5	0.1	700	30	0.9
9	0.001	0.001	1.0	1.25	10000	10^5	1	700	300	0.2

[a] The total mass decrement due to GW radiation occuring on timescales less than the dynamical time at $r_{\rm inner}$.
[b] Because $h = H(r/r_{\rm inner})^N$, and $N = 1$ for all simulations, $H = h/r$.
[c] Dimensionless parameter characterizing flaring of disk: $h = H(r/r_{\rm inner})^N$
[d] Dimensionless parameter characterizing strength of artificial viscosity (see Eqn. 3.11)
[e] Number of fluid cells used
[f] Number of time steps used
[g] The epicyclic Mach number (see Eqn. 3.21)
[h] The approximate time when the first caustic forms at $\tilde{r} = 1$ (see Eqn. 3.23)
[i] The approximate time that the flow deviates from the analytic solution at $\tilde{r} = 1$. (see Eqn. 3.22)
[j] The approximate radius where the gaseous flow will deviate from the analytic solution at time $\tilde{t} = 30$ (see Eqn. 3.28)

The nine simulations discussed in Sec. 3.4.1, which are the primary results of this chapter, are marked by their simulation number as given by Table 3.1 in the format $S\#$. For context and reference a snapshot of each of these simulations at time $\tilde{t} = 30$ is given in Fig. 3.5. The detailed results of these simulations are discussed in Sec. 3.4, but these snapshots are provided in Fig. 3.5 to illustrate the disk responses in the theoretically defined regions of Fig. 3.4, and are therefore laid out in the same form as Fig. 3.4. In each plot the simulation is shown in solid red and the analytic solution (Eqns. 3.18 and 3.19) is shown in dotted black. The analytic solution is truncated either where the first caustic forms (marked by a short thick vertical black line) or where it begins to obscure the results of the simulation. In each plot the theoretical deviation position (Eqn. 3.28) is given and when a shock is present the leading (largest) shock's Mach number in the simulation, $M_{\rm shock}$, is also given.

In Fig. 3.5 and in the figures of the simulations (Figs. 3.7–3.15) we plot regions of the disk dominated by gas pressure in green and regions dominated by radiation pressure in red. As expected from the non-interacting solution, morphologically there is little dependence of the surface density on the internal characteristics of the gas (e.g., radiation- or gas-pressure-dominated). However, whether the gas is gas-pressure-dominated or not has an important impact on the observability of these disks reacting to the merger of the SMBHs.

To gain a sense of the observability one must consider not only whether the flow is radiation- or gas-pressure dominated, but also the strength of the shocks, and where they form.

If strong shocks form in a flow which is gas-pressure-dominated both up and downstream of the shock, then the ratio of temperatures of the post-shock gas to the pre-shock gas is proportional to the shock's Mach number squared,

$$\frac{T_{\text{post-shock}}}{T_{\text{pre-shock}}} = \frac{2\gamma(\gamma-1)}{(\gamma+1)^2} M_{\text{shock}}^2 \sim M_{\text{shock}}^2 \,, \tag{3.31}$$

However, even if the gas upstream of the shock is gas-pressure dominated, a sufficiently strong shock will cause the post-shock gas to be radiation-pressure-dominated (e.g., simulation 3 shown in Fig. 3.9 with gas-pressure-dominated regions shown in green and radiation-pressure dominated regions shown in red), meaning that the temperature jump will be lower than in Eqn. 3.31. For a radiation dominated flow *down-stream* of the shock, but arbitrary gas up-stream, the temperature jump across a shock is given by

$$\frac{T_{\text{post-shock}}}{T_{\text{pre-shock}}} \sim \left(1 + \frac{P_{\text{gas},0}}{P_{\text{rad},0}}\right)^{1/4} M_{\text{shock}}^{1/2} \tag{3.32}$$

where $P_{\text{gas},0}$ and $P_{\text{rad},0}$ are the gas and radiation pressures just preceding the shock, respectively.

One may compute Eqns. 3.31 and 3.32 by using Eqns. 3.8 and 3.14 along with Eqn. 3.29 to determine which of these equations is appropriate, and determine the value of M_{shock} along with, if needed, $P_{\text{gas},0}$ and $P_{\text{rad},0}$.

If the shock does not form until after the perturbation leaves the hot inner regions of the disk, the bolometric luminosity may not change significantly. This can happen deep in the acoustic region (upper left of Fig. 3.4). By contrast, deep in the pressureless region (lower right of Fig. 3.4), strong shocks form quickly at small radii, and a large prompt change in the bolometric luminosity is expected.

Normally dissipation mechanisms keep weak perturbations from turning over into shocks, but as our simulations have no form of dissipation perturbations *invariably* turn into shocks. However, depending on the forms of dissipation in real disks when $h/r \gtrsim 0.1$ and $\epsilon_{\text{m}} \ll 1$ the weakest and slowest forming shocks may never actually form: For $h/r \gtrsim 0.1$ single weak shocks result. These weak shocks form by steepening of acoustic waves; and the more "acoustic" the flow is (the smaller M_e is and therefore the smaller ϵ_{m} is), the slower the perturbations evolve into shocks. This is readily visible from comparing figures 3.7 (S1), 3.10 (S4), and 3.13 (S7).

3.4 Results

In Chapter 2 the different types of possible flows are outlined. Here we choose a comprehensive set of initial conditions which exemplify the various types of flows. Table 3.1 gives the various parameters used for the different runs.

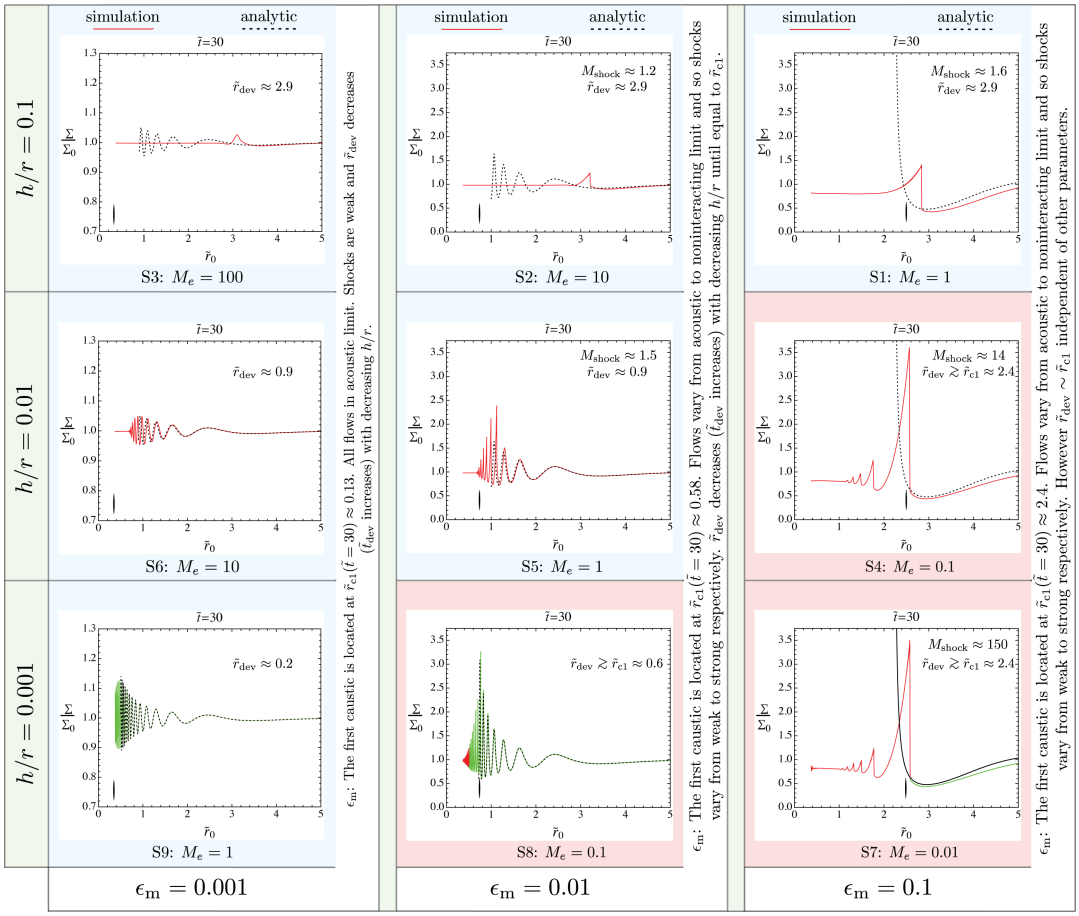


Figure 3.5: We plot the normalized surface density, Σ/Σ_0, at a time $\tilde{t} = 30$ as a function of the Lagrangian radial coordinate $\tilde{r}_0$ for the nine simulations discussed in Sec. 3.4, parameterized and labeled in Table 3.1, and shown in relation to parameter space in Fig. 3.4. The two parameters varied between the simulations are the mass decrement, $\epsilon_{\rm m}$, and disk aspect ratio, $h/r = H$ (see text following Eqn. 3.1). In each plot the simulation is plotted in solid red where the gas is radiation-pressure-dominated and in green where it is gas-pressure-dominated. The analytic solution discussed in Chapter 2 is plotted in dotted black from large radii down to the radius at which the first caustic forms or the analytic solution begins to obscure either itself or the simulation, whichever occurs farther out. Where shocks have formed the leading (largest) shock's Mach number is labeled. The theoretical deviation position (Eqn. 3.28) is also listed in each plot and the theoretical location of the first caustic is marked by a thick vertical line at the base of each plot region. A more in-depth discussion is given in the text. Also, compare with Fig. 3.4.

3.4.1 Simulations

We provide descriptions of the nine simulations outlined in Table 3.1. In all of the figures the simulation is plotted in solid (red), the non-interacting analytic solution (Eqn. 3.18) is plotted in dotted (black), and, where shown, the theoretical final value for Σ/Σ_0 (Sec. 3.2.3) is plotted in dash-dot (black). It might seem natural to plot $\Sigma(\tilde{r}_{(E)},t)/\Sigma_0(\tilde{r}_{(E)})$ in Eulerian coordinates, where both Σ and Σ_0 are calculated at a fixed position, $\tilde{r}_{(E)}$, in space. However, we have found that such plots are much harder to understand than plots of $\Sigma(\tilde{r}_0,t)/\Sigma_0(\tilde{r}_0)$ in Lagrangian coordinates, where both surface densities are calculated for a single fluid element as a function of its initial position $\tilde{r}_0$.

This is because, as described in Sec. 2.5.7 of Chapter 2, Lagrangian coordinates offer the advantage, and useful code test, of making the final position of the gas a known function of purely the mass perturbation and the initial position, whereas in Eulerian coordinates, the surface density ratio depends strongly on the (rather arbitrary) initial density profile. This is seen in simulation 1, shown in Fig. 3.7 using Lagrangian coordinates. Simulation 1 considers a mass perturbation $\epsilon_{\rm m} = 0.1$ and a disk thickness $h/r = 0.1$ (other relevant parameters are listed in Table 3.1). For this mass decrement, the theoretical final value of $\Sigma_{\rm final}/\Sigma_0 = 0.81$ (Eqn. 3.30 and listed in Table 3.2) as shown in dash-dot. There is an instability on the inner boundary at times $\tilde{t} = 21$ and $\tilde{t} = 28$ which will be discussed in the following paragraph. We re-plot simulation 1 in Fig. 3.6, using Eulerian coordinates. Note the final value of $\Sigma(\tilde{r}_{(E)})/\Sigma_0(\tilde{r}_{(E)})$ provides no insight to the dynamics of the flow or the reliability of the simulation, instead retaining a shape dependent largely on the initial density profile chosen. For this reason we find Lagrangian coordinates more illuminating and choose them over Eulerian coordinates for the plots of our simulations.

The instability on the inner boundary manifest in Fig. 3.7 is due to the rapidly changing surface density in the inner region. The surface density changes by about five orders of magnitude over a fraction of r_0. It is computationally challenging to populate this region with enough grid points that the assumption of small changes in the density between neighboring cells becomes valid. However, in these regions the surface density is orders of magnitudes smaller than everywhere else in the disk, and therefore carries only a minuscule fraction of the energy of a standard cell in the main regions of the flow. Therefore, there is little possibility for the instability to influence the main regions of the flow. To ensure that this is indeed the case, all simulations are monitored for significant deviations to the total energy in the flow; no such deviations were detected. Because this region has no physical relevance, in Figs. 3.8–3.15 we suppress the parts of the flow which have a surface density smaller than $10^{-5}\Sigma_{\rm max}$.

For the sake of clarity we have chosen, as much as possible, to keep the parameters other than h/r and $\epsilon_{\rm m}$ constant (see Table 3.1). As is visible in Table 3.1 two exceptions to this are the number of spatial points, $r_{\rm num}$, and the number of time steps, $t_{\rm num}$. For most simulations we chose $r_{\rm num} = 5000$ and $t_{\rm num} = 10^5$. This is far in excess of what was necessary to do the majority of

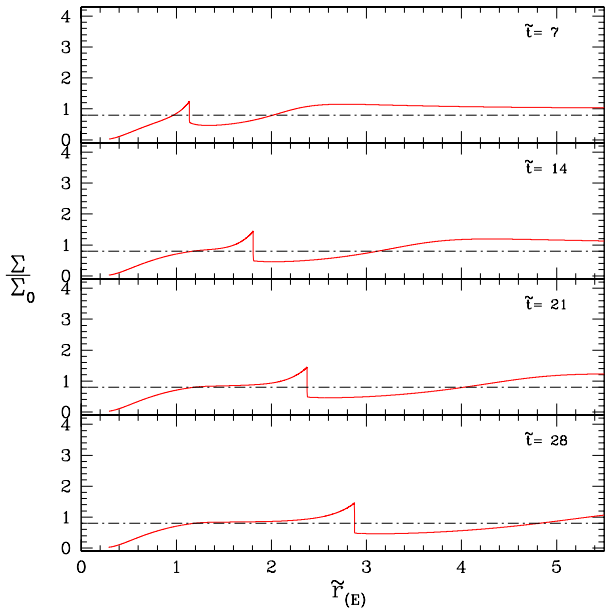


Figure 3.6: Simulation 1 (Eulerian coordinates): The same simulation as that shown in Fig. 3.7, except plotted in Eulerian coordinates, $\tilde{r}_{(E)}$. In Eulerian coordinates the region following the shock (left of the shock) does not have the simple analytic value it does in Lagrangian coordinates (see Fig. 3.7), and it depends strongly on the initial density profile. However, when plotted in Lagrangian coordinates as in Fig. 3.7, post-shock matter reaches a surface density which depends only on $\epsilon_{\rm m}$, and is entirely independent of the disk's initial density profile.

the simulations well, but close to what was necessary for the more challenging simulations. The number of spatial points was chosen with a balance between good capturing of the shocks and reasonable computational time. Similarly, the time step was chosen to ensure stability, and varies from simulation to simulation. Also, all simulations were run with multiple grid spacings to ensure convergence.

The k parameter, which controls the strength of the artificial viscosity (text surrounding Eqn. 3.11), is usually kept at a constant value of $k = 1.25$, though in several cases it was raised (see Table 3.1 for exact values) as a way to reduce the growth of the instability in the inner diffuse regions of the disk (see above discussion).

In the case of the N parameter, which is the exponent of the spatial dependence of $h(r)$ (Eqn. 3.1), we keep it at a constant value of $N = 1.0$ across all simulations. This produces a sound speed which changes as $\tilde{r}^{-1/2}$ (see Eqn. 3.15). Changing N does affect both shock velocity and shock formation speed, though only mildly inside the domain we consider. We have chosen to keep N constant to make clearer the dependencies of the flows on the primary parameters, h/r and $\epsilon_{\rm m}$.

The nine simulations which are presented come in three sets: the first presented is made up of simulations with a strong mass perturbation of $\epsilon_{\rm m} = 0.1$ (S1,S2,S3), the second set consists of simulations with a moderate mass perturbation of $\epsilon_{\rm m} = 0.01$ (S4,S5,S6), and the final set is made

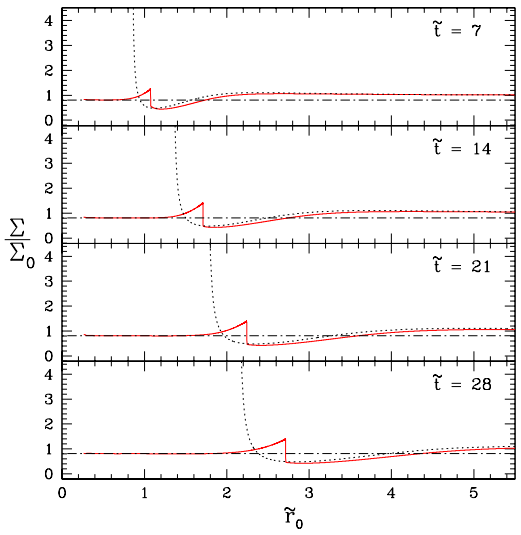


Figure 3.7: Simulation 1: Σ/Σ_0 is plotted as a function of Lagrangian radius $\tilde{r}_0$, for a mass decrement of $\epsilon_{\rm m} = 0.1$ and a disk of thickness $h/r = 0.1$. The simulation is plotted in red where the flow is dominated by radiation pressure and in green where dominated by gas pressure. Here the flow is entirely radiation-pressure-dominated. $\Sigma_{\rm final}/\Sigma_0$ is given in dash-dot (black), and the analytic solution for a non-interacting disk is plotted in dotted black at radii where no caustic has yet formed. Here $M_e = 1$: the simulation is in the acoustic region of Fig. 3.4. In this case a weak shock with Mach number $M_{\rm shock} \sim 2$ is expected (Eqn. 3.29) to form at time at $\tilde{r} = 1$ at time $tdev \approx 4$ (Eqn. 3.22). The simulation produces a shock with Mach number $M_{\rm shock} \sim 1.6$, when calculated from the pre-shock and post-shock densities (Eqn. 3.39). As expected (Eqn. 3.30) the surface density following the shock is $\Sigma_{\rm final}/\Sigma_0 = 0.81$. Moreover, the analytic solution predicts that a fluid element initially at position $\tilde{r} = 1$ will deviate around the time $\tilde{t}_{\rm dev} \approx 4$, while the simulation finds the particle deviates at $\tilde{t}_{\rm dev} = 5$.

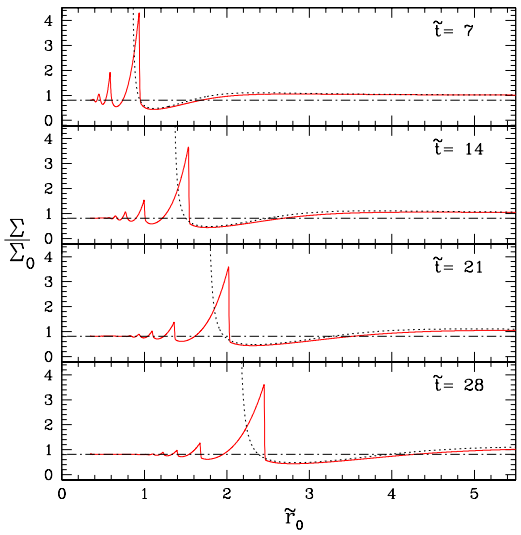


Figure 3.8: Simulation 2: Σ/Σ_0 is plotted as a function of Lagrangian radius $\tilde{r}_0$, for a mass decrement of $\epsilon_{\rm m} = 0.1$ and a disk of thickness $h/r = 0.01$. Lines are defined the same as in Fig. 3.7. Because $M_e = 10$ the flow should closely follow the non-interacting solution until close to the formation of the first caustic, since this indicates the flow is close to the pressureless limit. Therefore the flow should form its first shock on the face of the first caustic (as seen), and this shock should have Mach number $M_{\rm shock} \gtrsim M_e = 10$. In fact, the simulation finds the shock to have $M_{\rm shock} = 14$. Following the shocks, the flow circularizes to the final surface density $\Sigma_{\rm final}/\Sigma_0 = 0.80$, which is only marginally different from the expected value of 0.81.

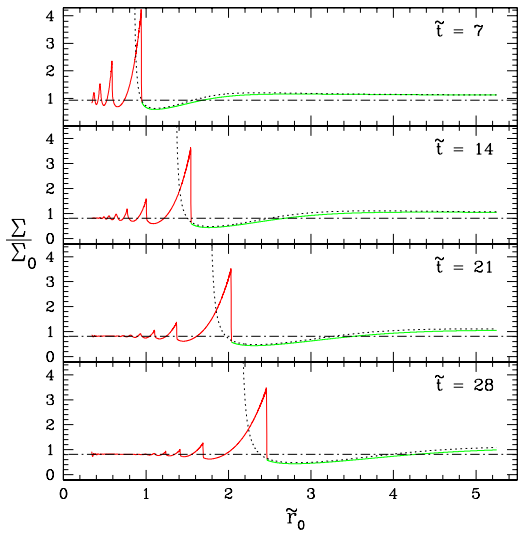


Figure 3.9: Simulation 3: Σ/Σ_0 is plotted as a function of Lagrangian radius $\tilde{r}_0$, for a mass decrement of $\epsilon_{\rm m} = 0.1$ and a disk of thickness $h/r = 0.001$. Lines are defined the same as in Fig. 3.7. In this case $M_e = 100$ and we see from Fig. 3.4 that this is merely an even more extreme case of simulation 2 (see Fig. 3.8). Therefore, we expect an almost identical flow, but with shocks that are an order of magnitude stronger. In line with these expectations, this flow deviates at the same time as simulation 2, the resulting Mach number of the first shock in the simulation is $M_{\rm shock} = 150$, and the final surface density is the expected value $\Sigma_{\rm final}/\Sigma_0 = 0.81$.

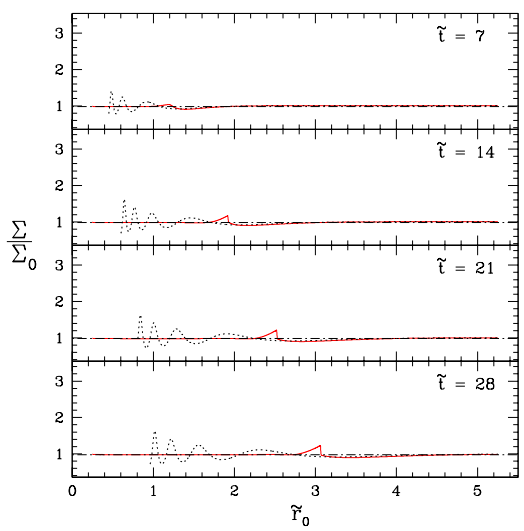


Figure 3.10: Simulation 4: Σ/Σ_0 is plotted as a function of Lagrangian radius $\tilde{r}_0$, for a mass decrement of $\epsilon_{\rm m} = 0.01$ and a disk of thickness $h/r = 0.1$. The simulation is plotted in red where the flow is dominated by radiation pressure and in green where dominated by gas pressure. Here the flow is entirely radiation-pressure-dominated. $\Sigma_{\rm final}/\Sigma_0$ is given in dash-dot (black), and the analytic solution for a non-interacting disk is plotted in dotted black at radii where no caustic has yet formed. Here $M_e = 0.1$, so we are in the acoustic regime, with an expected deviation time $\tilde{t}_{\rm dev} < 7$, shock strength $M_{\rm shock} \sim 2$, and final surface density $\Sigma_{\rm final}/\Sigma_0 = 0.98$. Indeed, from the simulation shown we find that at $\tilde{r} = 1$ the flow is expected to deviate at a time $\tilde{t}_{\rm dev} = 3.3$, while the simulation finds that $\tilde{t}_{\rm dev}(1) = 5$.

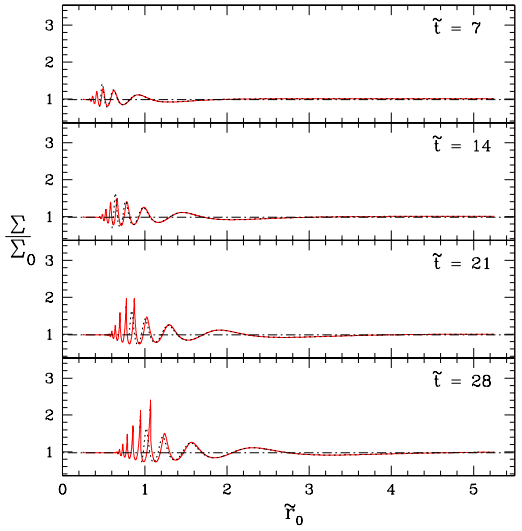


Figure 3.11: Simulation 5: Σ/Σ_0 is plotted as a function of Lagrangian radius $\tilde{r}_0$, for a mass decrement of $\epsilon_{\rm m} = 0.01$ and a disk aspect ratio $h/r = 0.01$. Lines are defined the same as in Fig. 3.10. Here $M_e = 1$ placing it on the boundary of the acoustic region (Fig. 3.4). In this simulation a particle initially at $\tilde{r} = 1$ deviates from the non-interacting flow in a time $\tilde{t}_{\rm dev} = 33$, which is also the expected theoretical deviation time (Eqn. 3.22). The Mach number of the first shock produced (located at approximately $\tilde{r} = 1.1$ at time $\tilde{t} = 28$) is 1.5 which is close to the expected Mach number 2 (Eqn. 3.29). Once the flow has circularized, it settles to the expected final value $\Sigma_{\rm final}/\Sigma_0 = 0.98$.

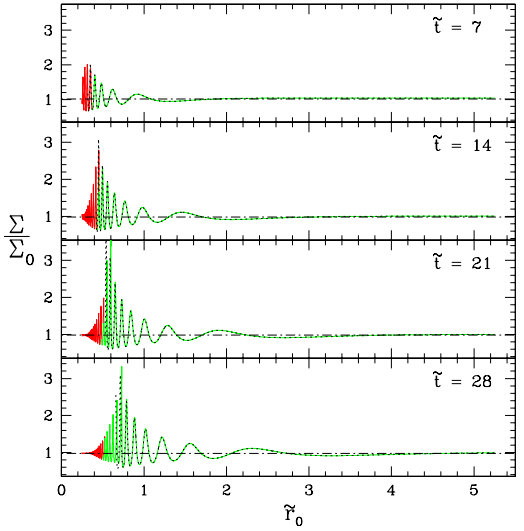


Figure 3.12: Simulation 6: Σ/Σ_0 is plotted as a function of Lagrangian radius $\tilde{r}_0$, for a mass decrement of $\epsilon_{\rm m} = 0.01$ and a disk of thickness $h/r = 0.001$. Lines are defined the same as in Fig. 3.10. This flow is in the non-interacting limit with $M_e = 10$: the flow closely follows the analytic solution (Eqns. 3.18 and 3.19) prior to the first shock, which should form close to the location of the first caustic and have a Mach number roughly given by $M_{\rm shock} = M_e = 10$. The location of the first caustic is given by $(3\epsilon_{\rm m}/2)^{2/3} = 0.6$, which is also the location of the first shock in the simulation as expected. This first shock has Mach number $M_{\rm shock} \approx 14 \sim M_e$ in accordance with expectations. The final surface density in the simulation, after the flow has reached a new steady state, is $\Sigma_{\rm final}/\Sigma_0 = 0.98$, as Eqn. 3.30 predicts.

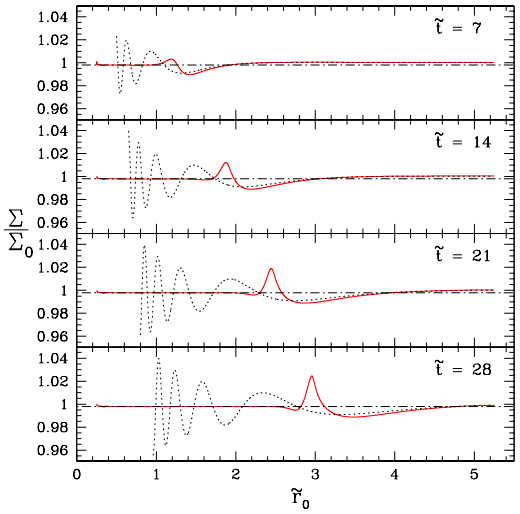


Figure 3.13: Simulation 7: Σ/Σ_0 is plotted as a function of Lagrangian radius $\tilde{r}_0$, for a mass decrement of $\epsilon_{\rm m} = 0.001$ and a disk of thickness $h/r = 0.1$. The simulation is plotted in red where the flow is dominated by radiation pressure and in green where dominated by gas pressure. $\Sigma_{\rm final}/\Sigma_0$ is given in dash-dot (black), and the analytic solution for a non-interacting disk is plotted in dotted black at radii where no caustic has yet formed. Because $M_e = 0.01$ this perturbation is evolving in the acoustic limit (see Fig. 3.4). A particle initially located at $\tilde{r} = 1$ should deviate from the non-interacting flow at a time $\tilde{t}_{\rm dev} = 3.3$, while in the simulation the particle deviates approximately at the time $\tilde{t}_{\rm dev} = 5$. Indeed, the rarefaction following the primary perturbation will evolve into a second shock, producing the "N" wave standard in cylindrical flows [81]. See text for a comparison between this figure and Figs. 3.7 and 3.10.

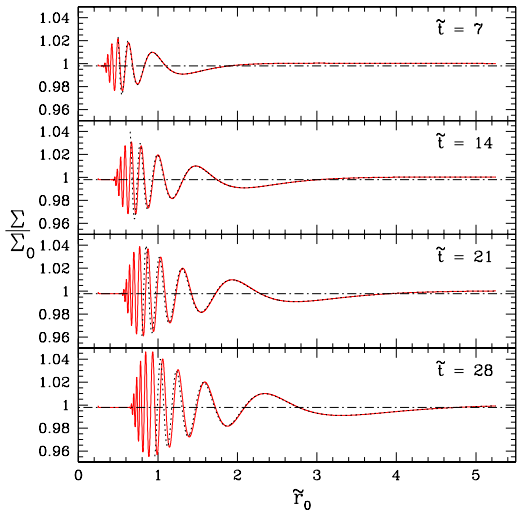


Figure 3.14: Simulation 8: Σ/Σ_0 is plotted as a function of Lagrangian radius $\tilde{r}_0$, for a mass decrement of $\epsilon_{\rm m} = 0.001$ and a disk of thickness $h/r = 0.01$. Lines are defined the same as in Fig. 3.13. The epicyclic Mach number for this flow is $M_e = 0.1$ placing the flow in the acoustic limit (Fig. 3.4). At time $\tilde{t} = 28$ the flow should deviate from the non-interacting solution around position $\tilde{r} = 0.9$ (Eqn. 3.22), which is approximately what is seen in the simulation.

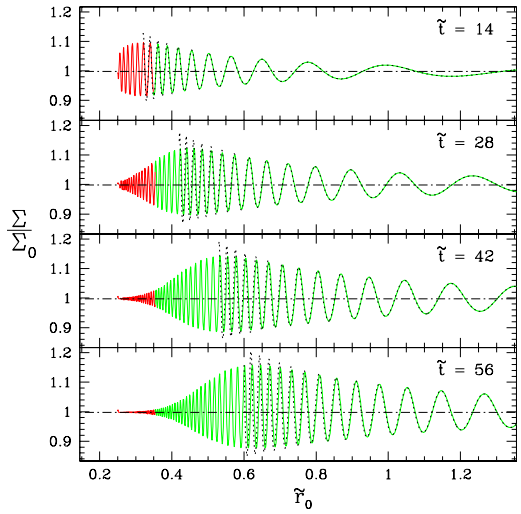


Figure 3.15: Simulation 9: Σ/Σ_0 is plotted as a function of Lagrangian radius $\tilde{r}_0$, for a mass decrement of $\epsilon_{\rm m} = 0.001$ and a disk of thickness $h/r = 0.001$. Lines are defined the same as in Fig. 3.13. Because $M_e = 1$, at a time $\tilde{t} = 28$ we expect the flow to deviate from the analytic solution at approximately $\tilde{r} = 0.2$ (Eqn. 3.22), approximately a factor of 2 off of the deviation position as given by the simulation.

Table 3.2: Expected Outcomes and Simulation Outcomes

#	$\epsilon_{\rm m}$ [a]	h/r [b]	M_e [c]	$\tilde{t}_{\rm dev}$ [d] pred	$\tilde{t}_{\rm dev}$ [e] simul	$M_{\rm shock}$ [f] pred	$M_{\rm shock}$ [g] simul	$\Sigma_{\rm final}/\Sigma_0$ [h] pred	$\Sigma_{\rm final}/\Sigma_0$ [i] simul
1	0.1	0.1	1	4	5	~ 2	1.6	0.810	0.816
2	0.1	0.01	10	7	8	~ 10	14	0.810	0.817
3	0.1	0.001	100	7	8	~ 100	150	0.810	0.819
4	0.01	0.1	0.1	3	5	~ 2	1.3	0.9801	0.9807
5	0.01	0.01	1	30	33	~ 2	1.5	0.9801	0.9801
6	0.01	0.001	10	70	50	~ 10	14	0.9801	0.9802
7	0.001	0.1	0.01	3	5–6	~ 2	n/a	0.998001	0.998007
8	0.001	0.01	0.1	30	46	~ 2	n/a	0.998001	0.998000
9	0.001	0.001	1	300	220	~ 2	n/a	0.998001	0.998002

[a] The total mass decrement due to GW radiation occuring on timescales less than the dynamical time at $r_{\rm inner}$.

[b] Since $h = H(r/r_{\rm inner})^N$, and $N = 1$ for all simulations, $h/r = H$.

[c] The epicyclic Mach number (Eqn. 3.21), which indicates the type of flow (Fig. 3.4), when ($\tilde{t}_{\rm dev}$) and where shocks form (Eqns. 3.22 and 3.28, respectively), and the strength of the shocks once they form ($M_{\rm shock}$, given by Eqn. 3.29).

[d] $\tilde{t}_{\rm dev}$ pred is the theoretical time a flow will deviate from the non-interacting solution at $\tilde{r} = 1$. It is given by Eqn. 3.22.

[e] $\tilde{t}_{\rm dev}$ sim is the deviation time at $\tilde{r} = 1$ calculated in the simulation.

[f] $M_{\rm shock}$ pred is the theoretical strength of the first (strongest) shock in the flow, once it has formed. It is given by Eqn. 3.29.

[g] $M_{\rm shock}$ sim is the shock Mach number of the first (strongest) shock in the flow, if a shock has formed. It is the average of the shock Mach numbers calculated using Eqns. 3.39 and 3.40.

[h] $\Sigma_{\rm final}$ pred is the theoretical value of the surface density of a fluid element after it has re-circularized (i.e., reached a new steady state). It is given by Eqn. 3.30.

[i] $\Sigma_{\rm final}$ sim is the final value of the surface density averaged over the 100th-200th cells (measured from the cell with smallest $\tilde{r}$) when these cells have circularized. In other cases (simulations 6 and 9) the average was done over the 10th-60th cells.

up of simulations with a weak mass perturbation of $\epsilon_{\rm m} = 0.001$ (S7,S8,S9). Each set consists of three simulations of varying disk thickness: a thick disk with $H = 0.1$, a moderately thin disk with $H = 0.01$, and then a very thin disk with $H = 0.001$.

Note that the non-interacting solution (plotted in dotted black) is a function of only the mass decrement and, in particular, has no dependence on the disk model. Therefore the location of the first caustic is the same for simulations in the same set (Eqn. 3.23).

On the other hand the deviation time depends on the epicyclic Mach number (Eqn. 3.22), which in turn is a function of both the mass perturbation and the disk thickness. So, in each set we expect all of the flows to initially be the same and merely deviate from the same non-interacting solution (Eqns. 3.18 and 3.19) at increasingly later times with increasingly thinner disks.

Once shocks have formed the flow will reach steady state when particles have circularized due to interactions. The speed with which this occurs can again be qualitatively understood using M_e: when $M_e \lesssim 1$ pressure forces quickly circularize particles following the first shock, while when $M_e \gg 1$ interactions only occur close to the locations of caustics in the non-interacting solution and it takes multiple contractions to completely circularize a flow. In general, the process of circularization occurs on a timescale comparable to the deviation time.

We present many of the key results for each of the simulations in Table 3.2. So, instead of

discussing each simulation in detail we will discuss only unusual features in each simulation. To give a clear understanding of where all of the values in Table 3.2 come from we will discuss one of the more feature rich simulations (Simulation 3) in detail, though in-depth discussions of the other simulations will be suppressed in favor of shorter discussions of characteristics unique to each.

Fig. 3.8 shows a simulation (simulation 2) of a moderately thin disk with $h/r = 0.01$ reacting to a strong mass perturbation of $\epsilon_{\rm m} = 0.1$. What is interesting is that multiple shocks form. Unlike the case of simulation 1 where the first shock completely extracted all radial motion of the particles' orbits, here the particles' radial motions are still significant following the first shock. Because the shock only decreases the amplitude of the epicyclic oscillation, but not the frequency, neighboring particles again come together with an attempted orbit crossing at the location of the second caustic in the non-interacting solution. Thus, a second shock forms at the location of the analytic solution's second caustic. Similarly, a third and even forth shock forms in the location of each subsequent caustic of the non-interacting solution, until the radial motion of the particles is completely extracted.

Again, simulation 3, shown in Fig. 3.9, considers a significant mass decrement of $\epsilon_{\rm m} = 0.1$, but affecting a very thin disk with $h/r = 0.001$. Simulation 3 is merely a more extreme case of simulation 2 and manifests the exact same structure, only with shocks of different strengths. In this case the epicyclic Mach number is $M_e = 100$ and we are placed deep into the non-interacting region of parameter space (see Fig. 3.4). We expect the shock to have a Mach number of order the epicyclic Mach number (Eqn. 3.29). Indeed, using the change in density and adiabatic index across the shock face we can solve for the shock Mach number. This calculation gives that in the case of the primary shock $M_{\rm shock} \approx 150 \sim M_e$. Moreover, because we are well into the non-interacting region we also expect that the first shock will form on the face of the first caustic. Here, the analytic solution is truncated after the first caustic, and we see that, as expected, the primary shock of the numerical solution rides the caustic's face. That the numerical flow circularizes to the expected final value of $\Sigma_{\rm final}/\Sigma_0 = 0.81$ indicates the accuracy to which the code obeys conservation of angular momentum (Eqn. 3.6).

The second set of simulations considers moderate mass perturbations of $\epsilon_{\rm m} = 0.01$. Again, the three simulations considered here all share the same non-interacting solution and, therefore, the same caustic formation time which is an order of magnitude longer than in the case of the strong mass perturbations of the first set of simulations. This allows for essentially two types of flows depending on the epicyclic Mach number: either a single slow-forming weak shock or a series of "non-interacting" density peaks followed by a series of shocks.

We consider a moderate mass perturbation of $\epsilon_{\rm m} = 0.01$ in simulation 4, shown in Fig. 3.10, acting on a thicker radiation-pressure-dominated disk with $h/r = 0.1$. One of the clear distinctions

between this simulation and simulations 1–3 is that the shock forms slowly. This reflects the acoustic nature of this flow, which has an epicyclic Mach number $M_e = 0.1$. The pressure term of the radial Euler equation dominates the kinetic term and, in the absence of strong dissipation, nonlinear forces slowly turn the perturbation into a shock. This is in contrast to simulations 2 and 3 where neighboring particles are being rapidly forced into one another upon the passing of the first caustic. Of observational interest, these slow forming shocks are unlikely to be detected promptly, because they form outside the hot inner ring of the disk which dominates the local disk luminosity. So, even if the shock were to significantly heat the disk, it forms far enough outside of the hot inner ring that it would only have a marginal effect. In this case however the shock is very weak with a shock Mach number $M_{\rm shock} \approx 1.2$ at time $\tilde{t} = 28$.

Simulation 5, shown in Fig. 3.11, considers a small mass perturbation of $\epsilon_{\rm m} = 0.01$ acting on a thin disk with $h/r = 0.01$. This is the first example of a flow which has density peaks prior to the first shock. This is because in this case $M_e = 1$ and our deviation location is a factor of $10^{2/3}$ closer to the SMBH than in simulation 4, placing it after several of the non-interacting density peaks. See Chapter 2 for a description for why this is the case. Having a higher epicyclic Mach number has the added benefit of having the shock form faster and ultimately heat the inner hot regions of the disk, unlike in the case of simulation 4.

Also with a mass perturbation of $\epsilon_{\rm m} = 0.01$ simulation 6, shown in Fig. 3.12, demonstrates the reaction of a very thin disk with $h/r = 0.001$. Upon first inspection one might find this simulation to be largely similar to simulation 5, but this is only true morphologically. In fact, this simulation is a good example of a flow in the non-interacting region of parameter space (see Fig. 3.4 and accompanying text) with a late-forming caustic. The flow only deviates from the analytic non-interacting solution close to when the first caustic passes. This is most readily apparent at time $\tilde{t} = 21$ when the first caustic lies at $\tilde{r} = 0.5$, while the primary shock lies at $\tilde{r} = 0.6$. In particular, as a flow which approaches the non-interacting limit the primary shock has a shock Mach number $M_{\rm shock} \approx 14$, which stands in stark contrast to the shock Mach number $M_{\rm shock} \approx 1.5$ of the primary shock in simulation 5. As mentioned in Chapter 2 the minimum shock speed is ultimately determined by $\epsilon_{\rm m}$, so this change in the shock Mach number between simulation 5 and simulation 6 is actually due, almost entirely, to the order of magnitude change in the sound speed.

Although mass perturbations of $\epsilon_{\rm m} = 0.001$ are likely not observable we consider them in the third set of simulations to demonstrate the theory. In this case the first caustic forms so late in the non-interacting solution ($\sim$ 100 orbital periods) that the gaseous flow will deviate from the non-interacting flow well before the passing of the first caustic.

In simulation 7 we consider a small perturbation with a thicker disk of $h/r = 0.1$. As is visible in Fig. 3.13 this flow is very similar to simulation 4, except that the shock is forming so slowly that it has not yet formed by $\tilde{t} = 28$, the final time plotted. Indeed, the same arguments describing simulation 4

are relevant here, except that now we are even farther into the acoustic regime with $M_e = 0.01$. In reality this perturbation is so small that in a real disk one would expect non-axisymmetric features of the disk to quickly destroy this structure.

A moderately thin disk with $h/r = 0.01$ reacting to a small mass change of $\epsilon_{\rm m} = 0.001$ (simulation 8) is considered in Fig. 3.14. Keeping in mind that the black dotted line is identical between this simulation in Fig. 3.14 and that of simulation 7 (Fig. 3.13), we see that this flow is identical to simulation 7 except that any given fluid particle will deviate from the non-interacting solution at a later time (Eqn. 3.22). Again, the perturbations which are deviating from the non-interacting solution will only slowly evolve into shocks, because we are again in the acoustic limit ($M_e = 0.1$).

Simulation 9, the final simulation presented and shown in Fig. 3.15, considers a very small mass perturbation of $\epsilon_{\rm m} = 0.001$ with a very thin disk with $h/r = 0.001$. Because $M_e = 1$, the deviation time still only depends on h/r which is now an order of magnitude smaller than in the case of simulation 8 and two orders of magnitude smaller than in the case of simulation 7. Therefore, because the non-interacting solution is the same as with both of these simulations, we expect an identical flow to both of these cases only with the deviation location even farther in in the disk. Noting that this is the only simulation with a different ξ_0 scale, this is easily seen in the figure.

3.4.2 Other Simulations

Several papers discussing electromagnetic counterparts due to mass decrement and kicks have been published recently, but the closest to our configuration is the paper by [23]. They simulated a disk which extends to approximately $10\,GM/c^2$ with $h/r = 0.05$, a constant. They use two different types of codes: a 3D hydrodynamical code on disks reacting to a mass decrement of $\epsilon_{\rm m} = 0.1$, 0.05, and 0.01, and a "2.5"D (r and z) magnetohydrodynamic (MHD) simulation for a disk reacting to mass decrements of $\epsilon_{\rm m} = 0.1$ and 0.01. For the rest of this section we scale the positions by $r_{\rm inner} = 10\,GM/c^2$ and the times by the orbital frequency at this radius.

To compare their simulation of a mass decrement of $\epsilon_{\rm m} = 0.01$ with ours, it is best to first consult a comparison diagram (see Fig. 3.16). We plot P/P_0 obtained in three different ways: we show the results of [23] taken digitally from their Fig. 1 (solid black), the results from our 1D simulation described in this chapter (dashed red), and the analytic solution for the normalized surface density discussed in Chapter 2 in Eulerian coordinates (dotted brown).

The plotting of the analytic surface density deserves a more detailed discussion. The analytic solution finds the surface density of a fluid particle as a function of time $\tilde{t}$, its *initial* position $\tilde{r}_0$, and the initial surface density at the particle $\Sigma_0(\tilde{r}_0)$. However, we wish to plot $\Sigma(\tilde{r},\tilde{t})/\Sigma_0(\tilde{r})$ (i.e., without the dependence on $\tilde{r}_0$). To plot this at a specified time we divvy up our domain into small intervals and follow a large number of particles to determine which interval they fall into. We then sum up the surface densities of all of the particles, whose initial positions $\tilde{r}_0$ we have been keeping track of, which

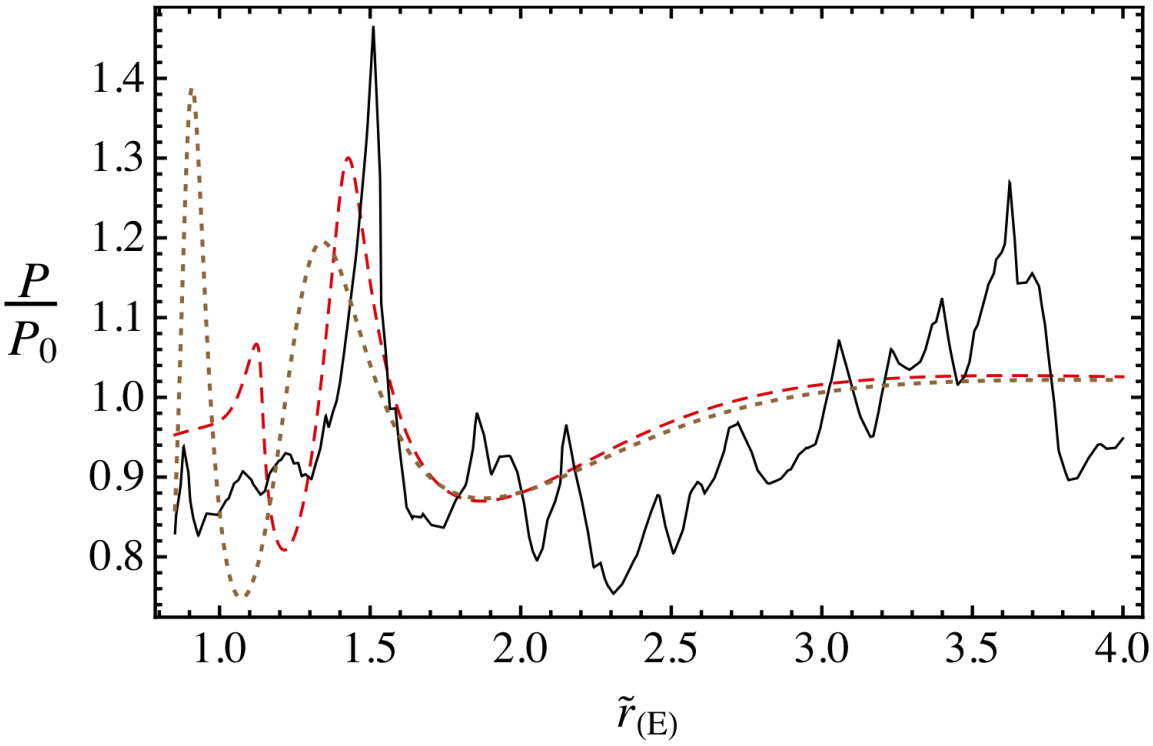


Figure 3.16: Comparison of the results of our code with the results from [23] for the case of a mass perturbation of $\epsilon_{\rm m} = 0.01$ affecting a moderately thick disk with a constant disk aspect ratio of $h/r = 0.05$. We plot the normalized pressure, P/P_0, as a function of r. Our 1D simulation is plotted in dashed red, while their 3D simulation is plotted in solid black. It should be noted that each group considers a fundamentally different initial density distribution. The similarities are significant, but three major differences are clear: the presence of turbulence in the O'Neill simulation, the presence of a shock-like structure (O'Neill) instead of a density perturbation, and the lack of the second density perturbation in the O'Neill simulation. The difference in turbulence is expected, as it can not be dealt with in a 1D code. The rapid formation of the shock is not well understood, but may be a reflection of the different density gradients used by each group, while the lack of the second density peak may very well be an outcome of the rapid formation of the shock.

have fallen into a given interval using Eqn. 3.18 and using the initial surface density distribution of our simulation for $\Sigma_0(\tilde{r}_0)$. This process is repeated for all intervals, ultimately producing the desired function $\Sigma(\tilde{r},\tilde{t})$. We then use an adiabatic equation of state to find the normalized pressure: $P/P_0 = (\Sigma(\tilde{r},\tilde{t})/\Sigma_0(\tilde{r}))^\gamma$ where $\gamma = 5/3$ is the adiabatic index used by [23].

By design our simulation implicitly determines the pressure, energy, and temperature independence of γ, effectively allowing γ to vary both in time and space if the thermodynamics requires this. For the purposes of comparison we remove this ability by setting the radiation constant $a_{\rm R} = 0$.

In all three examples in Fig. 3.16 the same trend is observed, where the flow initially has a small over-density (on the right), followed by a larger under-density, then followed by an even larger over-density. What follows this final over density is inconsequential in all of the examples, since the flow is expected to deviate at about this peak ($\tilde{r}_{\rm dev} = 1.5$).

There are a couple obvious differences between these curves: the presence of both turbulence and a shock-like structure in the [23] simulation which are absent in our simulation. The turbulence is an expected difference, since our 1D code clearly cannot take it into account, while the shock has merely formed faster in their simulation than in ours. This is likely due to differences in the initial conditions. It should be noted that in our simulation a shock is in the process of forming, as can be seen by our peak steepening and moving off of the non-interacting peak.

It is also important to note the power of the analytic solution. Prior to the shock the non-interacting solution loosely approximates the [23] simulation and expects a weak shock to form or have been formed at radius $\tilde{r} = 1.5$, approximately at the position of the shock (which is weak). As one would expect from inspection of Fig. 3.16, the analytic treatment can also reasonably well approximate the light curve of the O'Neill simulation.

To understand the fluctuations in the luminosity as plotted in [23] we must know the extrema of the analytic solution. These could be computed fairly easily using Eqns. 3.18 and 3.19, but a more straightforward method would be to just use Eqn. 3.26 instead. This is not only simple computationally, but it also highlights that for the nth extremum $(\Sigma/\Sigma_0)^{(n)}_{\rm max,min}$ is a constant.

Doing this in the case of $\epsilon_{\rm m} = 0.01$ gives

$$\left(\frac{\Sigma}{\Sigma_0}\right)^{(1)}_{\rm max} \approx 1.004\,, \left(\frac{\Sigma}{\Sigma_0}\right)^{(1)}_{\rm min} \approx 0.92\,, \left(\frac{\Sigma}{\Sigma_0}\right)^{(2)}_{\rm max} \approx 1.11\,, \tag{3.33}$$

where the subscript indicates whether the value is for a minimum or maximum and the superscript indicates the number of the maximum/minimum.

[23] uses bremsstrahlung emission as an upper limit to the emissions from the disk. For the sake of comparison we too compute it. If we consider just the region in some narrow annulus of the disk,

the relative bremsstrahlung emission can be approximated as

$$L = L_{\rm brem} \quad \sim \quad \rho^2 T^{1/2} V \,, \tag{3.34}$$

so that

$$L/L_0 \quad = \quad \left(\frac{\Sigma}{\Sigma_0}\right)^{3/2+\gamma/2} \frac{V}{V_0} \tag{3.35}$$

$$= \quad \left(\frac{\Sigma}{\Sigma_0}\right)^{\frac{3+\gamma}{2}} (1 + (N+2)\epsilon_{\rm m}) \,. \tag{3.36}$$

In this we have assumed an adiabatic equation of state, which should be accurate prior to the passing of the first caustic or shock. Taking $\gamma = 5/3$ as done in [23] we find that the first density peak would produce a variation in the luminosity such that

$$\left(\frac{L}{L_0}\right)^{(1)}_{\rm max} = 1.01 \,, \left(\frac{L}{L_0}\right)^{(1)}_{\rm min} = 0.81 \,, \left(\frac{L}{L_0}\right)^{(2)}_{\rm max} = 1.28 \,. \tag{3.37}$$

To relate this to the [23] it is best to compare with their Fig. 5 where they consider the luminosity of a small annulus of the disk restricted to the range of radii between 20 and $25\,GM/c^2$. There they have

$$\left(\frac{L}{L_0}\right)^{(1)}_{\rm max} = 1.03 \,, \left(\frac{L}{L_0}\right)^{(1)}_{\rm min} = 0.78 \,, \left(\frac{L}{L_0}\right)^{(2)}_{\rm max} = 1.28 \,. \tag{3.38}$$

Here there is remarkable agreement, especially considering that the luminosity is determined from the 1D non-interacting solution which takes into account no parameters of their 3D disk.

We now consider their simulations with a mass decrement of $\epsilon_{\rm m} = 0.05$ and begin by comparing the non-interacting solution (brown dotted), our 1D simulation (red dashed), and the 3D simulation of [23] (solid black) in Fig. 3.18. Again, there are several clear similarities. There is an initial over-density followed by a rarefaction which is then followed by a shock or, in the case of the non-interacting flow, a large density peak. The differences are also clear. In this case our 1D simulation produces a stronger shock than in O'Neill's 3D case, something which is expected both because we have an order of magnitude higher resolution and because we don't account for non-axisymmetric phenomena such as turbulence.

Just as in the case of the 1% mass perturbation, it is remarkable how well the non-interacting flow approximates both the 1D and 3D simulations. Here, just like before, the epicyclic Mach number is in the weak shock limit ($M_e \lesssim 1$), and, because h/r is the same as in the 1% case, we have the same deviation position, $\tilde{r}_{\rm dev} = 1.5$ (Eqn. 3.28). Prior to this position (farther out in the disk) the flow should be well approximated by the non-interacting solution, while at this position we expect a weak shock with a shock Mach number of order 2. In line with this prediction, our simulation

produces a shock approximately at $\tilde{r} = 1.5$ with shock Mach number $M_{\rm shock} = 2.2$, while the shock found in O'Neill's simulation is weaker with $M_{\rm shock} \approx 1.8$ and located at approximately the same position.

Here the shock Mach numbers have been computed using relationships between the pre- and post-shock densities, pressures, and Mach numbers. In particular,

$$\frac{\rho_2}{\rho_1} = \frac{(\gamma+1)M_1^2}{(\gamma-1)M_1^2+2} \tag{3.39}$$

and

$$\frac{P_2}{P_1} = \frac{2\gamma M_1^2 - (\gamma-1)}{\gamma+1}\,. \tag{3.40}$$

To determine the shock Mach numbers for our simulations we use both of the these equations to ensure consistency. However, because we only have the pressures from [23] we rely on the latter equation for an approximate Mach number. When computed in this way the shock Mach number is the ratio of the shock velocity *relative to the gas just preceding it* to the sound speed of that gas. Because the fluid preceding the shock may actually be moving quickly relative to the coordinate system the apparent shock velocity in that frame is not necessarily readily related to the Mach number.

The analytic expectations of the emissions can be found by proceeding as before. We first find the extrema of the surface density:

$$\left(\frac{\Sigma}{\Sigma_0}\right)^{(1)}_{\rm max} \approx 1.03\,, \qquad \left(\frac{\Sigma}{\Sigma_0}\right)^{(1)}_{\rm min} \approx 0.72\,. \tag{3.41}$$

This would produce the bremsstrahlung luminosities:

$$\left(\frac{L}{L_0}\right)^{(1)}_{\rm max} = 1.08\,, \qquad \left(\frac{L}{L_0}\right)^{(1)}_{\rm min} = 0.39\,. \tag{3.42}$$

There would then be a final rise which would likely be due to a weak shock yielding only a modest rise above the initial luminosity. [23] show in Fig. 5 the luminosities

$$\left(\frac{L}{L_0}\right)^{(1)}_{\rm max} = 1.04\,,\ \left(\frac{L}{L_0}\right)^{(1)}_{\rm min} = 0.28\,,\ \left(\frac{L}{L_0}\right)^{(2)}_{\rm max} = 1.37\,. \tag{3.43}$$

For the simulations with $\epsilon_{\rm m} = 0.1$ we plot the non-interacting solution (dotted brown), our 1D simulation (dashed red), and O'Neill's 3D simulation (solid black) in Fig. 3.18. The flows are qualitatively the same as before: a small peak followed by a low-density region, then followed by a shock or, in the case of the non-interacting solution, a strong peak. The first peak in the [23] simulation is somewhat different than the 1D versions, a discrepancy likely due to similar oscillations

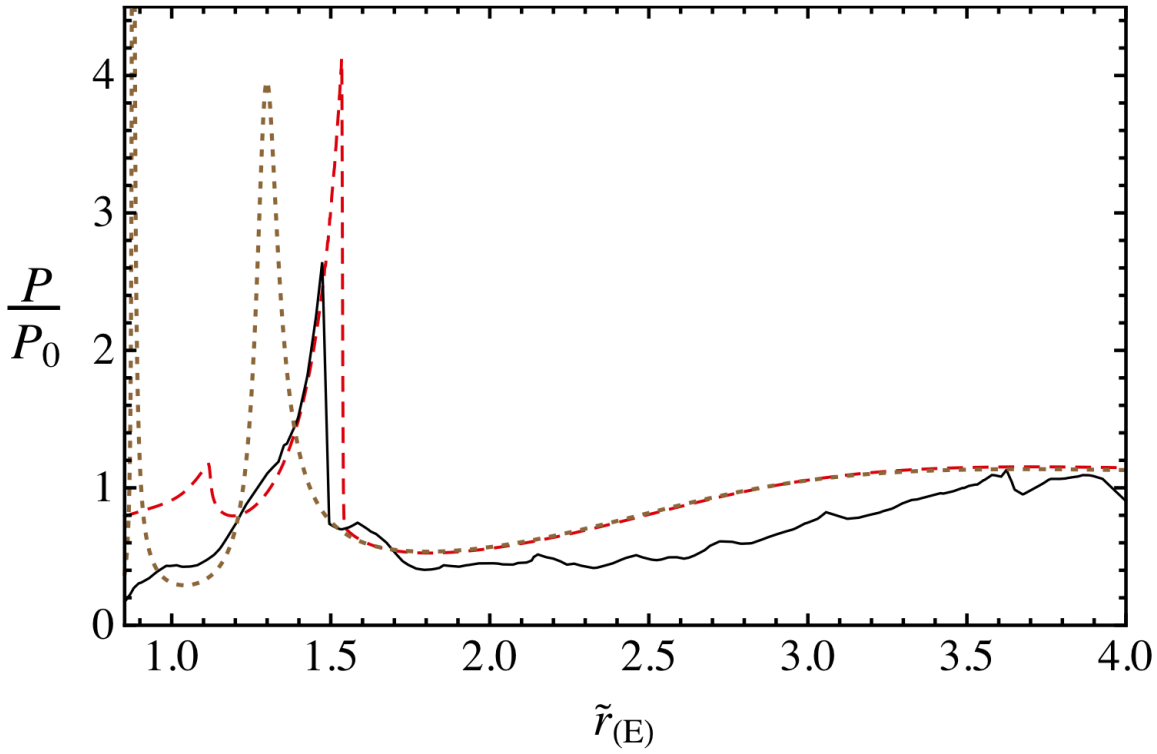


Figure 3.17: We plot the normalized pressure, P/P_0, for the analytic solution (dotted brown), our 1D simulation (dashed red), and the 3D simulations of [23] (solid black and taken digitally from their Fig. 1), all simulating the response of a disk with $h/r = 0.05$, a constant, reacting to a 5% mass perturbation. A description of the process undertaken to plot the analytic solution in Eulerian coordinates (which requires taking into account the correct initial density distribution) is described in the text. The analytic theory of Chapter 2 predicts a single shock located at approximately $\tilde{r}_{\rm dev} = 1.5$ (Eqn. 3.28) with $M_{\rm shock} \approx 2$, and that the gas preceding this shock can be approximated by the non-interacting solution. Indeed, both simulations produce a shock approximately located at $\tilde{r} = 1.5$; our 1D shock has $M_{\rm shock} = 2.2$, while the 3D shock has $M_{\rm shock} = 1.8$. Preceding the shock there is pretty good agreement. That the 3D simulation is more diffuse in the regions preceding the shock is likely due to oscillations out of the plane of the disk. Despite this marginal discrepancy, the 1D simulation and the analytic theory well capture the 3D vertically and azimuthally averaged flow.

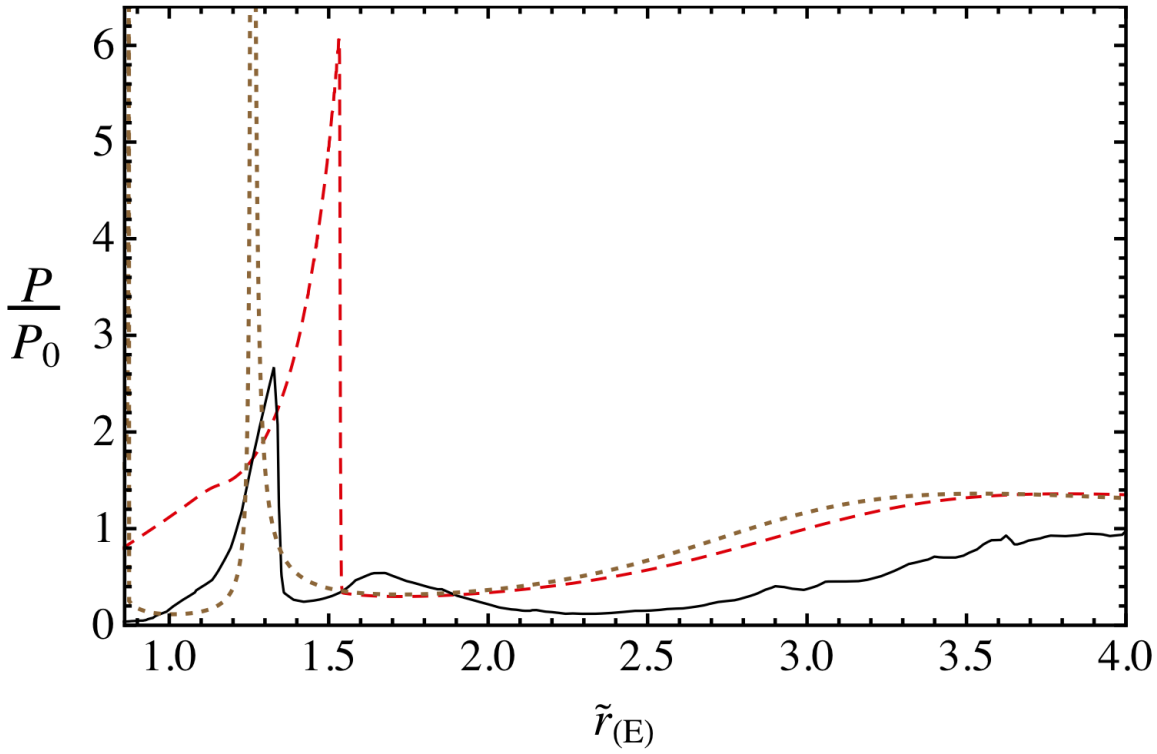


Figure 3.18: We plot the normalized pressure, P/P_0, for the analytic solution (dotted brown), our 1D simulation (dashed red), and the 3D simulations of [23] (solid black and taken digitally from their Fig. 1), all simulating the response of a disk with $h/r = 0.05$, a constant, reacting to a 10% mass perturbation. A description of the process undertaken to plot the analytic solution in Eulerian coordinates (which requires taking into account the correct initial density distribution) is described in the text. Upon examination of Fig. 3.4 we see that this flow should have approximately the same deviation location as for $\epsilon_{\rm m} = 0.05$ and $\epsilon_{\rm m} = 0.01$: $\tilde{r}_{\rm dev} = 1.5$ (also see Eqn. 3.28). In general, the analytic theory predicts a shock located at approximately $\tilde{r}_{\rm dev}$ (Eqn. 3.28), and given $M_e = 2$ we expect the shock to be mildly stronger than in the weak shock limit with $M_{\rm shock} \gtrsim 2$. Indeed, both simulations produce a shock approximately located at $\tilde{r} = 1.5$; our 1D shock has $M_{\rm shock} = 3.8$, while the 3D shock has $M_{\rm shock} = 2.5$. As can be seen in the figure, and as is expected in the analytic theory, the gas preceding this shock can be reasonably approximated by the non-interacting solution. That the 3D simulation is somewhat more diffuse in the regions preceding the shock is likely due to oscillations out of the plane of the disk. Despite this marginal discrepancy, the 1D simulation and the analytic theory well capture the primary features, both in time and space, of 3D vertically and azimuthally averaged flow.

in the vertical direction (out of the plane of the disk). A discussion of this vertical motion is available in [23].

The non-interacting solution again provides insight into almost every region of the flow. In this case, $M_e = 2$, which is loosely between the weak shock region and the non-interacting limit. But examination of lines of constant deviation time (and position) in Fig. 3.4 tells us that the deviation time (and position) should be the same as in the other simulations given in [23], $\tilde{t}_{\rm dev} = 1.5$. Thus we expect a shock at this position. Preceding this shock the flow is largely dictated by a purely

non-interacting disk, while following the shock we expect the flow to quickly circularize. The shock Mach number is expected to be of order 2, while in our 1D simulation we find $M_{\rm shock} = 3.8$ and in O'Neill's 3D simulation $M_{\rm shock} \approx 2.5$.

Following the same method as above, we can again examine the ability of the non-interacting solution to explain the emission pattern. The normalized surface density extrema are given by

$$\left(\frac{\Sigma}{\Sigma_0}\right)^{(1)}_{\rm max} \approx 1.10\,, \qquad \left(\frac{\Sigma}{\Sigma_0}\right)^{(1)}_{\rm min} \approx 0.48 \tag{3.44}$$

which would yield bremsstrahlung luminosities

$$\left(\frac{L}{L_0}\right)^{(1)}_{\rm max} = 1.39\,, \qquad \left(\frac{L}{L_0}\right)^{(1)}_{\rm min} = 0.20\,. \tag{3.45}$$

Their final luminosities are

$$\left(\frac{L}{L_0}\right)^{(1)}_{\rm max} = 1.19\,, \qquad \left(\frac{L}{L_0}\right)^{(1)}_{\rm min} = 0.05\,. \tag{3.46}$$

The analytically predicted changes in the luminosities ($\Delta L/L_0$) are systematically higher than those produced with the simulations by [23], but still within a factor of several and sometimes much better. This is indeed quite remarkable considering that within the analytic model we assume only that the disk is thin and use no other information. The differences are likely due to the inclusion of the z direction in their simulations, something which would certainly affect the flow on these timescales and possibly to these amounts.

3.5 Conclusion

We have used a Lagrangian 1D code to model the axisymmetric response of a thin disk to the merger of a super-massive black hole binary with mass ratio $1/3 \lesssim q \leq 1$. In particular, we compare our results to the analytic model presented in Chapter 2 and compare both the theory and our simulations to the results of [23] and show strong agreement in all cases.

In particular, we present a series of nine simulations which present the full range of possible flows, as a function of the disk scale height ratio (h/r) and the merger's gravitational mass decrement ($\epsilon_{\rm m} = \delta M/M_0$). Various aspects of these simulations are compared to the analytic solution of Chapter 2, including the regions preceding shocks, the shock formation region, the strength of the shocks, and the final steady-state surface density. Ultimately, the predictions of the theory are borne out by our simulations.

The 2.5D and 3D simulations by [23] have also been compared to both our simulation and the analytic solution. Our simulations approximate their results well, and, using the analytic machinery

of Chapter 2, the bremsstrahlung luminosity resulting from their simulations can be reasonably well described. It is notable that the analytic theory, which assumed one dimension, no particular disk model, and only that the disk was thin, could — with modest accuracy — model the output of their much more complicated simulations with a different disk structure.

In summary, it has been demonstrated that the reaction of a geometrically thin circumbinary disk to GW mass decrement can be largely understood using the analytic machinery of Chapter 2, and, where more accurate results are necessary, 1D codes can capture the majority of the physics. Together, the analytic and numerical work provide a comprehensive physical understanding of the range of possible disk responses to the merger of a binary black hole.

Chapter 4

Production of EMRIs in Super-Massive Black Hole Binaries

N.B.: This work will be published with the following authors: Nate Bode and Chris Wegg in this order.

Abstract

Extreme mass ratio inspirals (EMRIs) are one of the laser interferometer space antenna's (LISAs) sources of greatest astrophysical interest. The conventional source of EMRIs are stellar-mass black holes scattered onto orbits passing close to a super-massive black hole by multiple encounters with other stellar objects. However, scattering of a stellar-mass black hole by *any* other perturber could potentially produce an EMRI. We consider here compact objects in a relaxed stellar cusp with central super-massive black hole (SMBH) reacting to a secondary SMBH inspiraling through the cusp to its stalling radius. We find that: 1) this process produces approximately 10^{-3} $\mathrm{yr}^{-1}\mathrm{Gpc}^3$ EMRIs; 2) The formation of these EMRIs is largely due to a previously unnoticed mechanism. Normally apsidal precession due to a stellar potential (SP) and/or general relativity (GR) "detunes" the Kozai mechanism entirely. However, we find that sufficiently rapid precession produces, in the precessing frame, a perceived motion of the secondary. This induces a secular effect similar to the Kozai mechanism, though weaker and more chaotic, and tends to reduce the time required to produce EMRIs; 3) The chaotic aspect of the mechanism producing the EMRIs makes unreliable the simple secular approximations for the rates of EMRI production due to a secondary SMBH.

4.1 Introduction

One of the most interesting sources for the *Laser Interferometer Space Antenna* (LISA) is the capture of stellar mass compact objects (COs) by a super-massive black hole (SMBH). COs are the final state of stellar evolution and include stellar mass black holes, neutron stars, and white dwarfs. Due to the significant mass difference between the SMBH and the inspiraling CO, these sources are referred to as extreme mass-ratio inspirals (EMRIs).

Such sources for LISA enable many new and exciting physical measurements: 1) an accurate measurement of the spin and mass of the SMBH [82] along with a moderate determination of its location, 2) tests that the spin and mass are the only parameters characterizing the black hole's space-time [termed 'bothrodesy'; 83, 84], 3) information about the presence of a secondary (less massive) SMBH orbiting the primary (more massive) SMBH [85], 4) information about the presence of a gaseous disk in the system [86, 87], and, if the source is a white dwarf, 5) a possible electromagnetic counterpart to the LISA signal [88, 89]. Such an electromagnetic counterpart would provide an accurate localization of the host, allowing astronomers to, among other things, determine the luminosity distance-redshift relation to comparable accuracies as Type Ia supernova surveys, but using an independent and physical rather than empirical [75].

There are several plausible methods to produce EMRIs, a process which amounts to either forming or driving COs onto orbits whose gravitational wave (GW) inspiral time is shorter than the timescale for other orbital perturbations. The standard method of EMRI production [90] is that the transport of COs to the GW inspiral regime is via gravitational scattering with other stellar mass objects. The timescale for such scatterings must be short enough to drive a CO to the GW inspiral regime, but infrequent enough not to subsequently perturb the orbit either into the black hole, or out of the inspiral regime before the GW driven merger occurs. This is a challenging constraint, because as the CO's orbit becomes more eccentric, and the rate of orbital energy loss to GW emission increases, ever smaller kicks to its angular momentum may remove it from this orbit. The primary process that meets these limitations is star-CO scatterings, and this has been the focus of most previous work [see 4, for a good review].

Other possible EMRI formation mechanisms are that the COs may be formed in situ, via a massive self-gravitating accretion disk [91] like that which is believed to have existed in the Milky Way [92]. Alternatively, the CO can be carried to the SMBH in a stellar binary on a highly eccentric orbit which interacts strongly with the SMBH and ejects the CO's partner, while leaving the CO on a low eccentricity orbit with small semi-major axis [93].

We consider a different scattering method; one where a secondary SMBH is present and entering the final stage of its inspiral [44, 94] due to dynamical friction. In this scenario the scattering phase is short-lived, but the number of stars scattered to highly eccentric orbits is significantly increased.

Moreover, the secondary SMBH induces a secular effect on stars with semi-major axes smaller than its own, in which the eccentricity and inclination oscillate, referred to as the Lidov-Kozai mechanism (see Sec. 4.3.1 for a brief discussion of the mechanism). Because some of these "stars" are COs, when these COs pass close enough to the primary SMBH they can radiate some fraction of their orbital energy in gravitational waves. If the passage is close enough, and the orbit is sufficiently stable in the presence of the secondary to have many orbits, the CO may form into an EMRI. This is the process we consider here.

It is possible, using the standard Kozai formalism, to compute the shortest inspiral time of a star undergoing the Kozai mechanism analytically (Appendix A). However, we find that the EMRI production is not dominated by CO's on these standard Kozai orbits, but instead by CO's whose apsidal precession periods are comparable or faster than the orbital period of the the secondary SMBH. These exhibit much richer behavior, which we refer to as a 'reverse Kozai' mechanism (Secs. 4.3.2–4.3.3 and Sec. 4.4).

Thus, we approach the problem of understanding the effects of the secondary SMBH on the EMRI rates numerically and use a modified version of the simulation code used to study tidal disruptions in [18] (see Sec. 4.5 for differences). There we were interested in the possibility of observing multiple tidal disruptions from the same galaxy due to the presence of a secondary SMBH. The similarities to the problem considered here make this code particularly appropriate. In both cases, here and in [18], in our simulations we initially distribute stars isotropically according to an η-model [95] for a single-mass stellar distribution around the primary SMBH. The secondary is then spiraled inwards on a slightly eccentric orbit approximating the orbit of a SMBH evolving by dynamical friction and stellar ejection, until it reaches the stalling radius [96] where we smoothly stop the inward motion [see 18, for a more thorough discussion of our method]. The evolution of each star is then followed until the end of the simulation is reached, the star has entered the LISA band, or 10^{10} steps were taken. The probability that the stellar object was a CO is then determined from its initial semi-major axis using multi-mass distributions taking into account mass segregation [97].

This channel of EMRI formation is rich, manifesting several distinct physical effects. Almost all captured objects have undergone some form of Kozai oscillations, while only a few are formed by strong scatterings by the secondary. The results are presented in Sec. 4.6, rates are estimated in Sec. 4.7 and summarized along with future work in Sec. 4.8 and Sec. 4.9.

4.2 The Paradigm

Throughout this paper, we consider a system made up of three objects: 1) a primary super-massive black hole (SMBH) surrounded by 2) a stellar cusp of mass equal to twice the primaries mass and orbited by 3) a secondary SMBH. Holding the stellar potential centered on the primary SMBH and

not allowing it to evolve, we follow the paths of stars in the test particle limit. Their initial orbits are drawn from the self-consistent isotropic distribution given by the Jeans equations for an η-model potential-density pair, meaning that the initial stellar distribution is drawn from [95, with $\mu = 0.5$]

$$\rho(r) = \frac{\eta}{2\pi r_c^3} \frac{M}{\left(\frac{r}{r_c}\right)^{3-\eta}\left(1+\frac{r}{r_c}\right)^{1+\eta}}\,, \tag{4.1}$$

where r_c is the size of the cusp and M is the mass of the primary. The stellar mass interior to radius r is therefore given by:

$$M_{\mathrm{stellar},\eta}(r) = \frac{2Mr^\eta}{(r_c+r)^\eta}\,. \tag{4.2}$$

The potential around the primary due to the stellar distribution is then

$$\begin{aligned} \Psi(r) &= \frac{2GM}{\eta-1}\left[1 - \frac{r^{\eta-1}}{(r_c+r)^{\eta-1}}\right], && \eta \neq 1\,, \\ &= 2GM\ln(1+r_c/r)\,, && \eta = 1\,. \end{aligned} \tag{4.3}$$

Throughout this work we use $\eta = 1.25$ since this is the relaxed form of the distribution close to the SMBH [98]. Multi-mass models manifest both mass segregation, and that close to the black hole only the most massive objects have $\eta = 1.25$ [97]. We discuss the consequences of our assumption of a universal $\eta = 1.25$ when calculating the rates in Sec. 4.7.2. In addition we use total stellar mass twice the black hole's mass [i.e. $\mu = 0.5$ in 95]. This has the convenient property that when matching the central density to a power law, the radius at which the mass enclosed by the power law is $2\,M$ (the total stellar mass) is just r_c. This allows easy comparison to measurements.

In general, we write quantities relevant to the primary SMBH without subscripts, those relevant to the secondary SMBH with a subscripted large black "dot", and those relevant to the stars we write with a subscripted star symbol. In particular, the mass of the primary SMBH hole is written M, while those of the stars and secondary are written $m_\star$ and $m_\bullet$, respectively. Similarly the semi-major axes of the star and the secondary are written $a_\star$ and $a_\bullet$ (though sometimes the $\star$ subscript is suppressed when unambiguous). Quantities written with tildes (such as $\tilde{r}$) are written in units of $G = c = 1$, where G is the gravitational constant and c is the speed of light. The mass ratio of the two SMBHs is written $q \leq 1$. The simulations presented here adopt $M = 10^6\,M_\odot$ and $q = 0.3$ or 0.1, and masses of the stars and COs (relevant for gravitational wave radiation) are either $1\,M_\odot$ or $10\,M_\odot$. In each context, q and $m_\star$ will be labelled. Also, for succinctness, the primary and secondary SMBHs will be referred to as simply the primary and secondary.

We choose to focus on $M = 10^6\,M_\odot$ since this will result in EMRIs with frequencies best suited for detection by the current LISA design [4]. Throughout we use a fiducial cusp radius $r_c = 1.7\,\mathrm{pc}$. This is motivated by the fits from [100] to the inner regions of ACS Virgo Cluster galaxies [101].

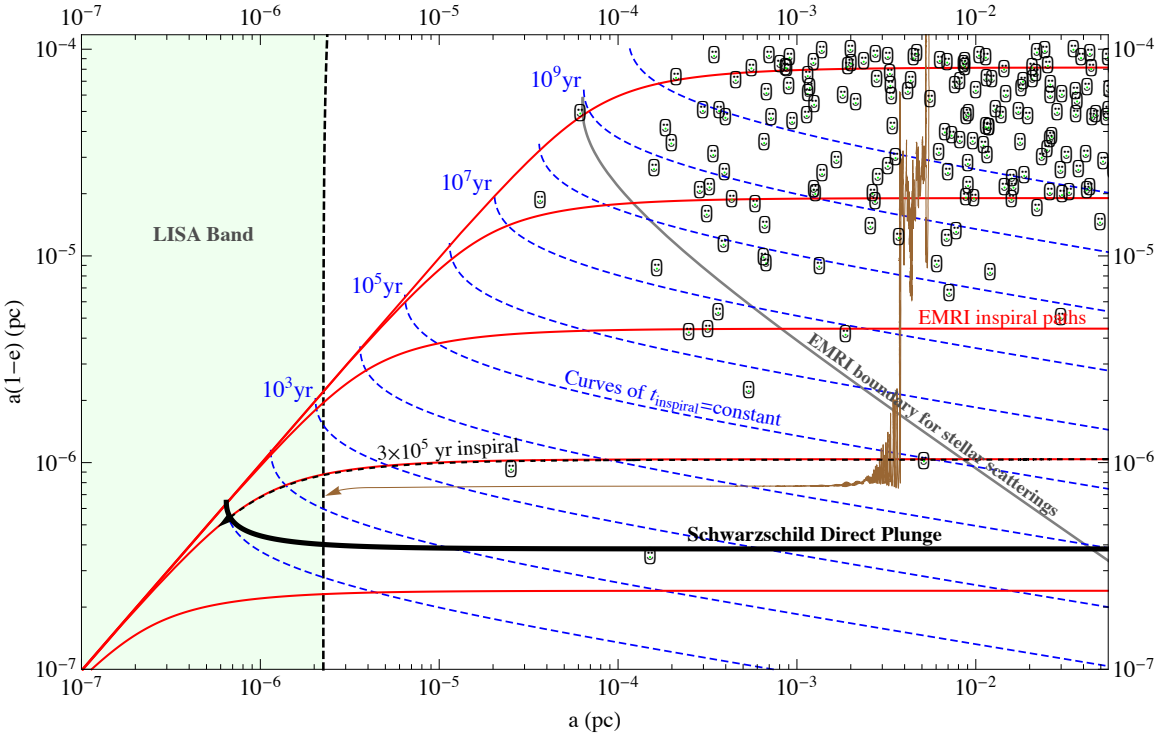


Figure 4.1: The standard EMRI picture with new addition: We plot the EMRI parameter space (which assumes a single central SMBH): $a(1-e)$ as a function of a. In the case of Keplerian orbits $a(1-e)$ is merely the periapsis distance, while in the case of EMRIs which have highly eccentric orbits and periapses well into the relativistic regions of the central SMBH $a(1-e)$ is more precisely a function of the angular momentum of the orbit (though it is still a crude approximation of the periapsis distance). We define a and e using Eqns. 4.5 and 4.6, respectively. A compact remnant (stellar-mass black hole, neutron star, white dwarf) inspirals due to GW radiation along paths shown in solid red. The timescale for inspiral is approximately given by the times on the dashed blue curves. However, if the initial a, $a(1-e)$ pair lies above the solid gray line, the star is unlikely to complete its inspiral before two-body stellar scatterings move the orbit to larger or smaller $a(1-e)$. The unstable circular orbit is plotted in thick solid black and represents the absolute minimum angular momentum a star may have without plunging into the SMBH. We plot the initial conditions of the 10^6 simulated stars whose initial orbital parameters lie in the plot region with smiley faces with green noses. In dashed black, we use our simulation discussed in Sec. 4.5 to follow a compact object's inspiral from the edge of the plotted domain until its final moments, assuming [99] gravitational wave energy losses. That the simulation follows the theoretical path (plotted in red) for such a long period is one of the many basic tests the simulation has undergone. We also plot the path of the star discussed in Figs. 4.4 and 4.5, which was simulated with the secondary SMBH and apsidal precession from both general relativity and the stellar potential. This has a path similar to many seen in our simulations, elucidating the new channel for a binary SMBH system to produce EMRIs.

For power-law galaxies these give [1]

$$r_{\rm inf} = 22\,(M/10^8 M_\odot)^{0.55}\,{\rm pc}\,, \tag{4.4}$$

where $r_{\rm inf}$ is defined such that the stellar mass interior to $r_{\rm inf}$ is $2\,M$ and M is the mass of the SMBH. Matching this to the η-model such that the central densities are equal gives $r_c = r_{\rm inf}$. Extrapolating to Sgr A* which has a mass of $\approx 4 \times 10^6\,M_\odot$ [10] gives $r_{\rm inf} = 3.8\,{\rm pc}$ which agrees well with the observations of $r_{\rm inf} \approx 4$ pc [102]. Using $M = 10^6\,M_\odot$ gives our fiducial $r_{\rm inf} = r_c = 1.7$ pc.

The secondary is moved inwards from the edge of the stellar cusp following a path roughly consistent with dynamical friction until it reaches its stalling radius, which we take from [96]. Details of the calculation of the secondaries orbit are given in Wegg and Bode [18]. During this inspiral the secondary SMBH strongly scatters significant numbers of stars, some of which come close to the primary. In reality these stars could be any stellar object including main sequence stars (MSSs), stellar mass black holes (SBHs), neutron stars (NSs), or white dwarfs (WDs), and accordingly we use the word 'star' to refer to any of these objects.

Normally one would describe the stars' orbits with Keplerian orbital elements, but here the stellar potential, the secondary, and general relativity (during close passages) all cause these orbits to be non-Keplerian. However, we use unambiguous analogous quantities to describe the orbits of our stars. For example, we write the semi-major axis as a function of the star's energy

$$a \equiv \frac{GM}{2E}\,, \tag{4.5}$$

and we define the eccentricity using the angular momentum and a:

$$e \equiv \sqrt{1 - \frac{L^2}{GMa}}\,, \tag{4.6}$$

where L is the angular momentum of the star. These are the quantities shown in the figures unless stated otherwise.

Since even in non-Keplerian potentials L and E are constant, then, provided the secondaries orbital period is long compared to the star's, on the star's orbital timescale both a and e defined as above are constant and well defined quantities. Practically, however, there are oscillations both on the binary's orbital timescale (shown in Fig. 4.3 and discussed in Sec. 4.3.2), and on the star's orbital timescale, due to numerical errors in the integration as discussed in Sec. 4.5.3.

The orbits which are interesting to us are those which approach the inner several Schwarzschild

[1] D. Merritt, personal communication. From fitting to Fig. 2 of Merritt et al. [100]

radii, meaning that they are highly eccentric orbits. In this case,

$$L \approx \sqrt{2GMa(1-e)} \propto \sqrt{a(1-e)}\,. \tag{4.7}$$

For a Keplerian orbit $a(1-e)$ would be the periapsis distance, however general relativistic effects cause deviations from this relation. We use $a(1-e)$ as one of our fiducial parameters because of its correspondence to both the angular momentum in these high eccentricity orbits, but also for its correspondence to the periapsis distance in the Keplerian limit, which provides useful insights to the problem (though we urge the reader not to assume equality with periapsis distance, especially close to the primary where general relativity causes strong departures).

When the periapsis of the stars is small enough, gravitational wave emission can affect the orbit. To first order in the periapsis distance $\tilde{r}_p \approx a(1-e)$, the energy radiated per orbit is given by [99]

$$\left\langle \frac{dE}{dt} \right\rangle = -\frac{\beta}{2} \frac{GMm_\star}{a^5(1-e^2)^{7/2}} \left(1 + \frac{73}{24}e^2 + \frac{37}{96}e^4\right)\,, \tag{4.8}$$

where

$$\beta = \frac{64}{5} \frac{G^3 M m_\star (M+m_\star)}{c^5}\,,$$

and e is the eccentricity of the star.

Essentially all of this energy is radiated during the periapsis passage and, in general, this is what we assume in our simulations (see Sec. 4.5.4). Radiating at this rate and assuming no interactions an inspiral takes approximately [99]

$$T(a_0, e_0) = \frac{12}{19} \frac{c_0^4}{\beta} \int_0^{e_0} \frac{e^{29/19} \left[1 + (121/304)e^2\right]^{1181/2299}}{(1-e^2)^{3/2}}\, de\,, \tag{4.9}$$

where c_0 is found using the initial condition that $a(e_0) = a_0$ and

$$a(e) = \frac{c_0 e^{12/19}}{1-e^2} \left[1 + \frac{121}{304}e^2\right]^{870/2299}\,. \tag{4.10}$$

As shown by the smiley faces in Fig. 4.1, an initial isotropic stellar distribution function puts very few stars on EMRI orbits. But stars put onto sufficiently eccentric orbits will, through gravitational radiation, find themselves on a one-way trip into the black hole. So what matters is the flux of stars whose orbits are perturbed into this EMRI region. For an isolated black hole rare 2-body interactions between stars in the cusp surrounding it are the primary perturbing force creating EMRIs. However, during a SMBH binary phase the secondary black hole will, for a short period, exact a much stronger perturbation, and these are our focus here. We must simultaneously consider both strong and (weak) secular perturbations, and these are sensitive to GW energy and angular

momentum loss, and to the relation between the secondary orbit and the apsidal precession due to the cusp potential and relativistic effects. These combine to produce a rich and intricate set of behaviors (see Sec. 4.4).

Many of these points are illustrated in Fig. 4.1 where we plot $a(1-e)$ as a function of a. A star lying in this region of parameter space will inspiral due to GW radiation along curves like the solid red lines, which are calculated using Eqn. 4.8 and would take approximately the period of time labeled along the blue dashed lines to complete the inspiral (assuming no other interactions). A star lying below the solid gray line would likely turn into an EMRI, while one above (even in the absence of some strong perturber like a secondary SMBH) would likely undergo some stellar scattering event which would increase its angular momentum (moving it upwards in the figure) and ultimately put it on a new trajectory with an unrealistic inspiral time. For reference we place those stars from our initial conditions of 10^6 stars drawn from the distribution of Eqn. 4.1 which lie in the plotted region. These are the initial conditions for a set of stars from one of our simulations. In the standard EMRI picture, weak stellar scatterings extract or add angular momentum from a star while keeping its energy approximately constant. This moves the star up or down in the figure. When the star scatters below the gray line, gravitational radiation can extract energy from the orbit faster than it is perturbed and the star will begin to follow one of the red lines. In the case presented in this paper, the picture is more complicated, but rich. Kozai oscillations with important effects from apsidal precession (both precession due to the stellar potential and that due to general relativistic effects) drive the stars on an orbital evolution which has the angular momentum rise and fall many times before a close passage occurs and enough gravitational wave radiation can be emitted so as to extract a significant fraction of the star's orbital energy. The path of an example star which we discuss in detail later in the paper (see Figs. 4.9 and 4.10) is shown in thin solid brown. This star's path manifests many of the important characteristics which affect this method of EMRI formation.

4.3 Kozai Mechanism

4.3.1 The Good — Historic Formalism

The Kozai mechanism is a secular effect on a body's motion around a central potential due to some periodic perturbation to the system. In our context, this is the situation of a star orbiting a primary SMBH being perturbed over long timescales by a secondary SMBH, also orbiting the primary. The original theory [103, 104] of the Kozai mechanism assumed (in the context of our problem) not only that the semi-major axis of the star is less than half that of the secondary SMBH, but also that the star is on what would otherwise be a Keplerian orbit in the absence of the secondary (i.e., general relativistic effects along with effects due to the stellar potential are ignored). Within this problem and these assumptions we initially discuss the Kozai mechanism.

Instead of giving a detailed description of the problem, something already done well by the original papers [103, 104] and much subsequent work [105–108] in various different contexts, we aim to provide in this section a brief description of the problem and give a conceptual but quantitative outline of the rich variety of different phenomena the Kozai mechanism produces.

Here, and throughout the rest of this paper, we will use the following conventions: a is the semi-major axis, e is the eccentricity, $\epsilon \equiv 1 - e^2$, i is the inclination to the orbital plane of the secondary SMBH, ω is the argument of periapsis, χ is the longitude of the ascending node, $T_\bullet$ is radial the period of the secondary, $T_\star$ is the radial period of the star, $q \leq 1$ is the ratio of masses of the SMBHs, and

$$T_{\rm Kozai} \equiv \frac{2}{3\pi q}\frac{T_\bullet}{T_\star}T_\bullet = \frac{4}{3q}\left(\frac{a_\star}{a_\bullet}\right)^{-3/2}\sqrt{\frac{a_\bullet^3}{GM}} \tag{4.11}$$

is the characteristic timescale on which the Kozai oscillations occur, which will be discussed shortly.

Starting from the exact equations of motion in the osculating elements, written with the true anomaly as the independent variable, and averaging over both the orbits of the star and the secondary SMBH, one may obtain an insightful set of differential equations governing the evolution of these elements [103] (written in the form of [107] but with $\epsilon = 1 - e^2$):

$$T_{\rm Kozai}\frac{da}{dt} = 0 \tag{4.12}$$

$$T_{\rm Kozai}\frac{de}{dt} = -\frac{5}{2}e\sqrt{\epsilon}\sin^2 i \sin 2\omega \tag{4.13}$$

$$T_{\rm Kozai}\frac{di}{dt} = -\frac{5}{4}\frac{e^2\sin(2i)\sin(2\omega)}{\sqrt{\epsilon}} \tag{4.14}$$

$$T_{\rm Kozai}\frac{d\omega}{dt} = \frac{2\epsilon + 5\sin^2(\omega)(\cos^2 i - \epsilon)}{\sqrt{\epsilon}} \tag{4.15}$$

$$T_{\rm Kozai}\frac{d\chi}{dt} = -\frac{\cos i}{\sqrt{\epsilon}}\left\{\epsilon + 5(1-\epsilon)\sin^2\omega\right\}\,. \tag{4.16}$$

The primary characteristic of the Kozai mechanism is that, as these equations imply, the star's orbital elements undergo an oscillatory motion which has a period given approximately by $T_{\rm Kozai}$, and which can be of significant magnitude. There are several key outcomes from the above equations:

1. The energy of the star's orbit remains constant (Eqn. 4.12).

2. The eccentricity (and inclination) reaches its extremal values only if $\omega = 0, \pm\pi/2, \, or \, \pi$ (solving Eqn. 4.13 equal to 0).

3. Setting $\dot{e} = \dot{i} = \dot{\omega} = 0$ we find a stationary solution when $\omega = \pm\pi/2$. In this case e and i are fixed and obey the relation

$$e^2 = \frac{5}{3}\sin^2 i - \frac{2}{3}\,, \tag{4.17}$$

while $\omega = \pm\pi/2$.

The second point implies that there can be two categories of behavior: solutions which have e and i extrema at $\omega = \pm\pi/2$ and *different* solutions which have e and i extrema at $\omega = 0$ or π. Indeed, one finds that there exist solutions which librate around $\omega = \pm\pi/2$ and have their extrema of e and i at $\omega = \pm\pi/2$, *and* there exist solutions for which ω evolves monotonically with extrema e and i at $\omega = 0, \pm\pi/2$ *and* π. We refer to these two possible solutions as librating and rotating solutions, respectively.

Besides a, equations 4.12-4.16 admit two further integrals of the motion [103, 104]:

$$\Theta = (1 - e^2)\cos^2 i = \epsilon \cos^2 i \tag{4.18}$$

$$Q = e^2 \left[5 \sin^2 i \sin^2 \omega - 2\right] , \tag{4.19}$$

which together tell us several things about the stars' evolution:

1. The z component of the angular momentum is conserved since $L_z = \sqrt{GMa\Theta}$ (Eqn. 4.18).

2. All solutions have $\epsilon \geq \Theta$ (Eqn. 4.18).

3. The eccentricity reaches its maximum (minimum) when the inclination reaches its minimum (maximum) (Eqn. 4.18).

4. Librating solutions have

$$e^{\max}_{\min} = \sqrt{\frac{5}{3}\sin^2 i^{\min}_{\max} - \frac{2}{3}} = \sqrt{1 - \frac{5}{3}\cos^2 i^{\min}_{\max}} , \tag{4.20}$$

   by solving

$$\Theta(e_{\max}, i_{\min}) = \Theta(e_{\min}, i_{\max}) \tag{4.21}$$

$$\text{and} \tag{4.22}$$

$$Q(e_{\max}, i_{\min}) = Q(e_{\min}, i_{\max}) \tag{4.23}$$

   for $e_{\max}(i_{\min})$.

5. The stationary solution obeying Eqn. 4.17 has

$$\epsilon = \sqrt{\frac{5}{3}\Theta}$$

   (joining the previous point with Eqn. 4.17).

Many of these points can be demonstrated graphically by plotting the possible orbital paths, calculated numerically from Eqns. 4.13–4.16, through parameter space for a given value of Θ. We

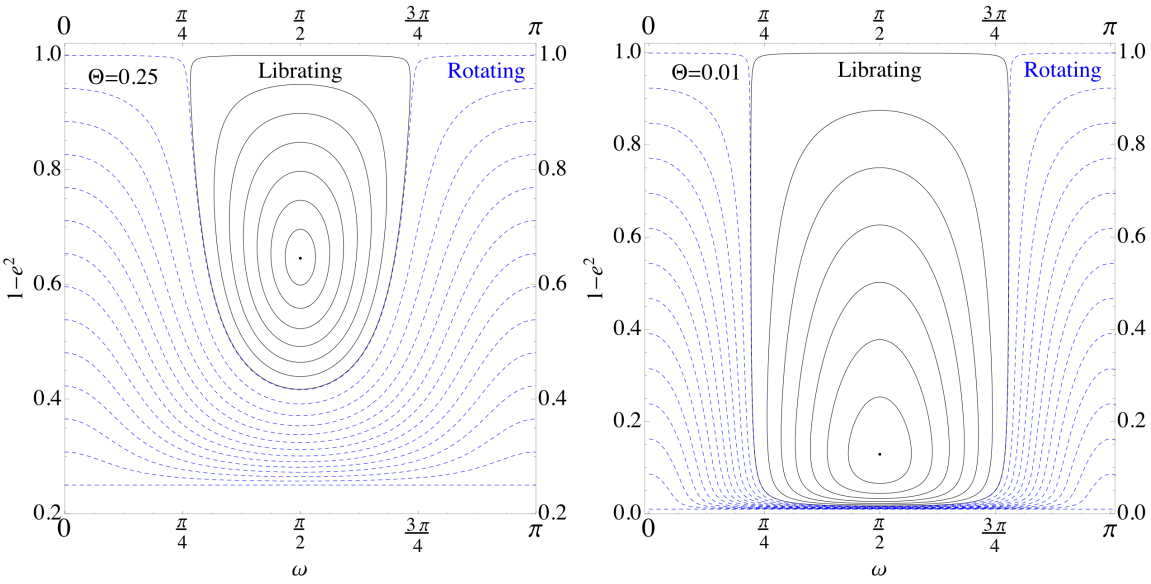


Figure 4.2: We plot the solutions to the Kozai differential Eqns. 4.12–4.16 for a series of orbits in a parametric plot of $1-e^2$ and the argument of pericenter ω as a function of time. Throughout Kozai oscillations a particle conserves the quantity $\Theta \equiv (1-e^2)\cos^2 i$. In the plot on the left all orbits have $\Theta = 0.25$, while all orbits in the figure on the right have $\Theta = 0.01$. The maximal attainable value of e for any orbit is $\sqrt{1-\Theta}$. When Θ is large only modest changes in eccentricity can occur, while all orbits with Θ close to 0 can be driven to modestly or extremely high eccentricities. There are two types of orbits shown: librating and rotating. In librating orbits ω oscillates around $\omega = \pi/2$ where the eccentricities and inclinations reach their extrema (visible in the figures). In rotating orbits have ω changes monotonically while the eccentricities reach their extrema at both $\omega = \pi/2$ and π.

do this in Fig. 4.2. For two different values of Θ we plot different orbits as a function of time in a parametric plot of ω and $\epsilon = 1 - e^2$. As was described earlier, and obvious from the figure, there are two different types of orbital evolution: the "rotating" solutions, whose argument of periapsis evolves monotonically (blue dashed lines), and the "librating" solutions, whose argument of periapsis oscillates around $\omega = \pm\pi/2$ (solid black lines). The region with $\omega \in [\pi, 2\pi]$ is just a reflection of the region plotted, and so is suppressed.

Orbits with small Θ are driven to high eccentricity, with the rotating solutions being driven to the highest eccentricities. Because $\Theta \propto L_z^2$, this also means that the stars which we are interested in, those which will become highly eccentric, must also have small L_z.

4.3.2 The Bad — Deviations from Standard Formalism

The picture which has been painted so far (Sec. 4.3.1) has been very clean: there has been no discussion of deviations from this analytic solution on the binary orbital timescale (this section) and there has been no discussion of the precession effects due to the stellar potential or GR (Sec. 4.3.3). These are both very important to our discussion.

Because the derivation of Eqns. 4.12–4.16 integrates over the orbit of the secondary SMBH, oscillations on the binary's orbital timescale are smoothed over. Throughout most of the orbit this has little effect, but close to the peaks of the eccentricity the oscillations on the orbital timescale can become significant. In the appendix of the paper by [107] it is shown that the magnitude of the variations in the angular momentum are given by

$$\Delta L = \frac{15}{8} \cos i_{\rm min} q \left(\frac{a_\star}{a_\bullet}\right)^2 \sqrt{G M a_\bullet}\,. \tag{4.24}$$

This produces a variation in $1-e$ given by

$$\Delta(1-e) = \frac{15}{8} \cos i_{\rm min} q \left(\frac{a_\star}{a_\bullet}\right)^{3/2} \sim q \left(\frac{a_\star}{a_\bullet}\right)^{3/2}\,. \tag{4.25}$$

Keeping in mind that for the cases of interest to us, $1-e$ is frequently of order 10^{-4}, and many times even lower, it is clear that these variations can become very significant close to the peak in eccentricity, as demonstrated in Fig. 4.3. There we plot $1-e$ as a function of time for a $10\,M_\odot$ star evolved without precession due to the stellar potential (abbreviated SP precession) and without general relativistic precession (abbreviated GR precession) using our simulation described in Sec. 4.5. In red is the approximate path predicted by the orbit averaged Kozai formalism, Eqns. 4.12–4.16, and in green is the predicted envelope (Eqn. 4.25). Each dot represents the calculated value of $1-e$ at apoapsis by our simulator. In most situations these oscillations would not be important, but as the Schwarzschild radius of the primary SMBH is approached small changes in the periapsis of the orbit can lead to significantly different evolutions. See Sec. 4.3.3 for a discussion of GR precession and Sec. 4.4 for a discussion of its dramatic effect on the orbital evolution of the star.

4.3.3 The Ugly — Other Forms of Precession

One can easily add precession to the Kozai equations (Eqns. 4.12–4.16) by adding a term with the relevant precession to the $d\omega/dt$ equation:

$$T_{\rm Kozai} \frac{d\omega}{dt} = \frac{2\epsilon + 5\sin^2(\omega)(\cos^2 i - \epsilon)}{\sqrt{\epsilon}} + \kappa g(e)\,, \tag{4.26}$$

where $\kappa g(e)$ is the precession over a time $T_{\rm Kozai}$ due to some other mechanism (e.g., SP precession or GR precession) and κ is independent of the eccentricity and $g(e)$ is a function of only the eccentricity.

For instance, in the case of the stellar potential, κ is given by [107]

$$\kappa_{\rm sp} = -\frac{2}{3\pi} K\left[\frac{M_{\rm st}(a)}{m_\star}\right]\,, \tag{4.27}$$

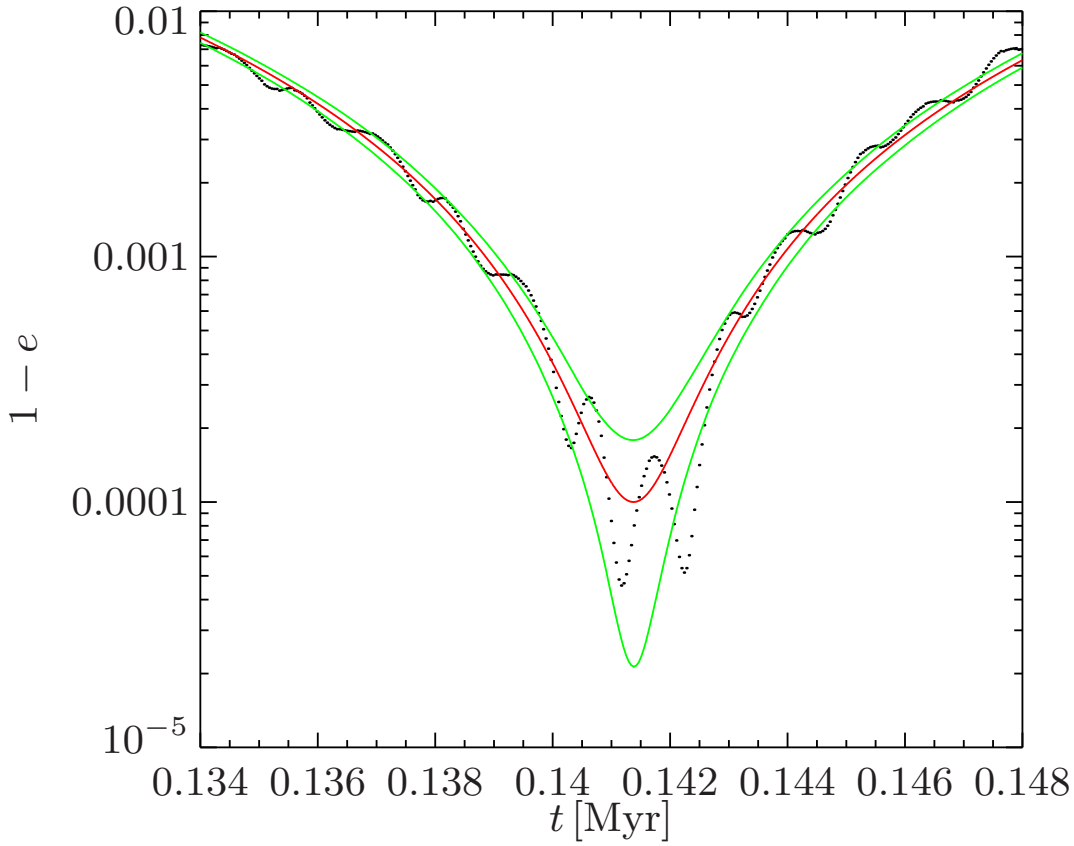


Figure 4.3: We plot $1-e$ as a function of time across the peak of the oscillation in eccentricity. In red is the approximate value predicted by the standard Kozai formalism (averaged over the secondary's orbit, Eqns. 4.12–4.16), while in green is the expected magnitude of the modulations in the eccentricity over the secondary's orbit given by Eqn. 4.25. Each dot represents the value of $1-e$ calculated by our simulation at apoapsis during the first maximum of the eccentricity in the simulation shown in the first column of Fig. 4.4 (in which both the stellar potential and the relativistic precession were turned off for clarity). The secondary's orbital modulations of the eccentricity can be significant in these regions, which also happen to be the regions of the Kozai oscillation where maximal relativistic effects will occur. Because of the strong dependence on position for these relativistic effects, it is not sufficient to use secular approximations to predict the outcome when the maximum eccentricity is reached.

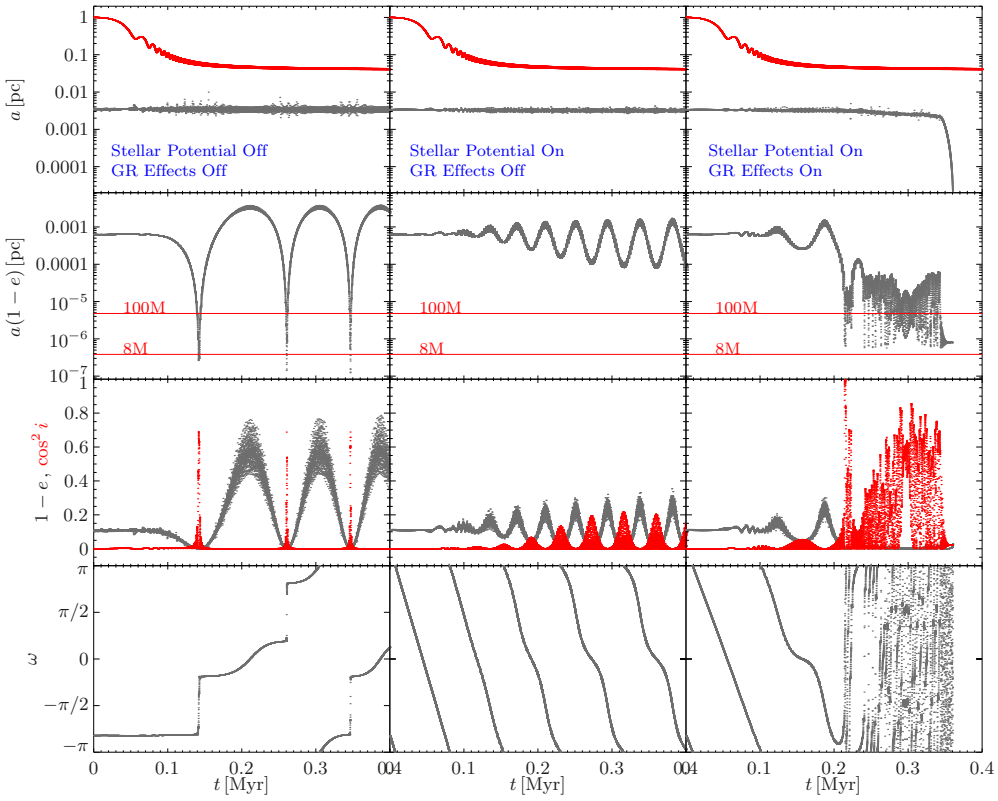


Figure 4.4: Comparison of Kozai inspirals with different forms of precession: in the first column we follow a key star without any form of precession (no SP or GR precession), in the second column we "turn on" SP precession and in the third column we finally add GR precession and GW energy loss. Describing the rows of the plot: In row 1 we plot in charcoal the semi-major axis of a $10\,M_\odot$ star (the mass is only relevant when GR effects are considered; final column) as a function of time along with the position of the secondary plotted in red. In row 2 we plot $a(1-e)$, an approximation for the square of the angular momentum, in charcoal as a function of time. Also plotted are the lines of constant $a(1-e) = 8\,GM/c^2$ and $100\,GM/c^2$, the latter being the arbitrary radius where we start calculating the energy loss for the orbit (see Sec. 4.5). In row 3 we plot $1-e$ in charcoal and $\cos^2 i$ in red as a function of time. The product of the two is L_z^2 which is a conserved quantity in the standard Kozai formalism. In the final row we plot the argument of periapsis as a function of time. From left to right there are clear changes in the star's orbit as new forms of precession are added: In the first column, with both stellar and relativistic precession turned off, the object undergoes traditional Kozai oscillations as described in Sec. 4.3. In the second column (with SP potential) three clear changes occur: the star begins retrograde precession (row 4), the Kozai period decreases (rows 2 and 3), and the magnitude of the oscillations decrease (rows 2 and 3). In the final column (with SP and GR precession) a very different phenomenon occurs: The relativistic precession is prograde and so weakens the effect of the stellar precession, which then allows for larger amplitude Kozai oscillations, which then drives the star to high enough eccentricities to reach a radius such that relativistic precession becomes significant. At this point the reverse Kozai mechanism takes over (Sec. 4.4) and an ostensibly chaotic orbit ensues. These regions where the relativistic precession dominate the orbit are shown in greater detail in Fig. 4.5.

where K is a factor depending on the stellar cusp (see [107]), and $g(e)$ is given by

$$g_{\rm sp}(e) = \sqrt{1-e^2}\,. \tag{4.28}$$

As $\kappa g(e)$ becomes significant the Kozai oscillations become "detuned", causing a decrease in both the amplitude and the Kozai period. This can be seen in Fig. 4.4 in the first two columns. In the first column the stellar potential is ignored, as is GR precession. In this case we get the classic orbital evolution of the star as predicted by the Kozai formalism, modulo the orbital oscillations discussed in Sec. 4.3.2 and illustrated in Fig. 4.3 (a zoomed-in image of the first major dip in $1-e$ in Fig. 4.4). In the second column SP precession is turned on, and the orbit evolves similarly to the standard Kozai oscillation without the stellar potential, but with the amplitude and period of the oscillations being smaller and shorter. In the third column GR precession is turned on along with the SP precession. Because the GR precession (prograde) and the SP precession (retrograde) are competing effects, the GR precession weakens the effect of the SP precession and allows the Kozai oscillations to have larger amplitudes and longer periods. However, once the periapsis of the star begins to approach the Schwarzschild radius of the primary SMBH, two important and related changes occur: 1) GR precession becomes dominant and 2) because of the strong radial dependence of the GR precession near the SMBH, the oscillations on the binary timescale discussed in Sec. 4.3.2 become important. Together these two phenomena cause the star to enter into a chaotic orbit which we illustrate in Fig. 4.5 and introduce and describe in greater detail in Sec. 4.4.

4.4 Reverse Kozai Mechanism

It is widely known that if a body, which would otherwise undergo a Kozai oscillation, also undergoes a rapid apsidal precession (compared to the Kozai apsidal precession), then the Kozai mechanism will no longer drive large oscillations in eccentricity. This situation is typically referred to as the Kozai mechanism being "de-tuned". It is commonly assumed that such precession kills *all* secular effects due to the presence of the secondary.

However, this is not quite the case. In fact, when the argument of periapsis precesses at a rate which greatly exceeds the orbital frequency of the secondary SMBH, the secondary effectively "sits still" while the orbit of the test particle executes a full 2π precession. In the frame of the precessing orbit (i.e., rotating with the argument of periapsis), the secondary executes a circular motion centered on and in a plane orthogonal to, the axis of the angular momentum. This perceived motion of the secondary, which is quite similar to the motion of the secondary in the standard Kozai geometry, again causes a secular effect which is analogous to the standard Kozai oscillation: The energy of the orbit remains constant as the eccentricity and inclination oscillate, while conserving

the z-component of the angular momentum.

There are some important and significant differences, though. The most important are the stability, speed, and strength of the effect. As one would expect, each of these depend on the source of the precession.

For instance, SP precession is relatively weak, and so does not completely neutralize the Kozai mechanism, but *can* instead cause the same stable oscillations with a significantly shorter Kozai period. See Sec. 4.3.2 for a qualitative discussion of this.

On the other hand, relativistic precession is more intricate. In this case the precession has a strong radial dependence which diverges as the periapsis of the particle approaches the radius of the maximum in the effective potential which defines the unstable circular orbit [UCO, defined as Eqns. 12-16 of 109, at $4M$ for $e \to 1$ non-rotating blackholes]. Thus very small changes in the periapsis, such as those due to weak non-secular effects of the secondary (Sec. 4.3.2), can produce vastly different precession rates. In particular, the position of the secondary in it's orbit can strongly modify the periapsis distance, as shown in Fig. 4.3 where the minimum periapsis (dots) is about half of the averaged Kozai value (red curve). This strong dependence of the periapsis on the phase of the secondary in turn creates a strong dependance on the GR precession and hence the reverse-Kozai period and amplitude. This gives rise to an apparently chaotic orbital evolution shown in the third column of Fig. 4.4 and in greater detail in Fig. 4.5.

4.4.1 Relativistic Precession

The requirement that the argument of periapsis precess at a rate faster than the orbital frequency of the secondary significantly constrains the region of parameter space where stars and compact objects can undergo the reverse-Kozai oscillation. For highly eccentric orbits, such as those we consider in this paper, and to first order in $\tilde{r}_{\rm p} \equiv \frac{r_{\rm p}}{GM/c^2}$, the relativistic precession per orbit is

$$\Delta\phi_{\rm rel} = \frac{3\pi}{\tilde{r}_{\rm p}}\,, \tag{4.29}$$

so that the apparent orbital frequency of the secondary (if it were not moving) would be

$$\Omega_\bullet^{\rm app} = \frac{3}{2}\Omega_\star \frac{1}{\tilde{r}_{\rm p}}\,. \tag{4.30}$$

This implies that

$$\frac{a_\star}{a_\bullet} \ll \left(\frac{3}{2}\frac{1}{\tilde{r}_{\rm p}}\right)^{2/3} \sim \left(\frac{1}{\tilde{r}_{\rm p}}\right)^{2/3}\,. \tag{4.31}$$

Because $\tilde{r}_p \geq 4r_{\rm S}$ for these highly eccentric orbits, these reverse Kozai oscillations can only occur due to relativistic precession when $a_\star \ll a_\bullet/3$.

This motion is almost identical to that of the Kozai mechanism, except that the secondary

appears to orbit an object located out of the plane of the orbit of the particle (but on the axis of the particle's angular momentum). Therefore, we expect the reverse-Kozai time to obey a similar relation to that of the Kozai mechanism, which is given by (see Eqn. 4.11)

$$T_{\rm Kozai} = \frac{2}{3\pi}\frac{1}{q}\frac{T_\bullet}{T_\star}T_\bullet\,, \tag{4.32}$$

where $T_\bullet$ is the secondary's orbital period, and $T_\star$ is the star's orbital period. Assuming the same relationship for the reverse-Kozai mechanism we have

$$T_{\rm RK} = \frac{2}{3\pi}\frac{1}{q}\frac{T_\bullet^{\rm app}}{T_\star}T_\bullet^{\rm app}\,, \tag{4.33}$$

where $T_{\rm RK}$ is the reverse Kozai time, and $T_\bullet^{\rm app}$ is the apparent orbital period of the secondary. Noting that Eqn. 4.30 implies that $T_\bullet^{\rm app} = \frac{2}{3}\tilde{r}_p T_\star$, we then have the expected reverse-Kozai timescale

$$T_{\rm RK} = \frac{16}{27}\frac{\tilde{r}_p^2}{q}\sqrt{\frac{a_\star^3}{GM}}\,. \tag{4.34}$$

Taking the ratio of Eqn. 4.34 to Eqn. 4.11 we find the fractional rate at which relativistic precession can speed up the Kozai timescale:

$$\frac{T_{\rm RK}}{T_{\rm Kozai}} = \frac{4}{9}\tilde{r}_{\rm p}^2\left(\frac{a_\star}{a_\bullet}\right)^3 \ll 1\,, \tag{4.35}$$

where we have used Eqn. 4.31 in the final relation. Thus, the reverse-Kozai period is much shorter than the Kozai period for objects with the same energy, and can even be shorter than the binary orbital timescale.

However, as noted by [110] and [108], in the case of the standard Kozai mechanism, as the relativistic precession rate becomes comparable to the Kozai precession rate, the oscillation is effectively detuned and the magnitude of the oscillations is decreased. A similar phenomenon occurs in the case of the reverse-Kozai mechanism, *except* that in this case this acts to speed the compact object to merger, as described below.

The typical scenario of an EMRI formed via the reverse-Kozai mechanism (see Figs. 4.4–4.5 for a multi-layered example) is that a star originally located at around $r \approx r_\bullet/10$ undergoes the standard Kozai oscillation, which drives the star to high eccentricity. As its periapsis approaches the unstable circular orbit [Eqns. 12–16 of 109], and is changing rapidly on the timescale of a single orbit, it has several close passages which produce a large apsidal precession. The standard Kozai oscillation immediately "breaks" and the star now begins a new orbit which is highly eccentric and rapidly precessing. Of course, for rapid EMRI formation it would be ideal if no change to the orbit were to occur, since it is in this highly eccentric phase that the compact object will most rapidly radiate its

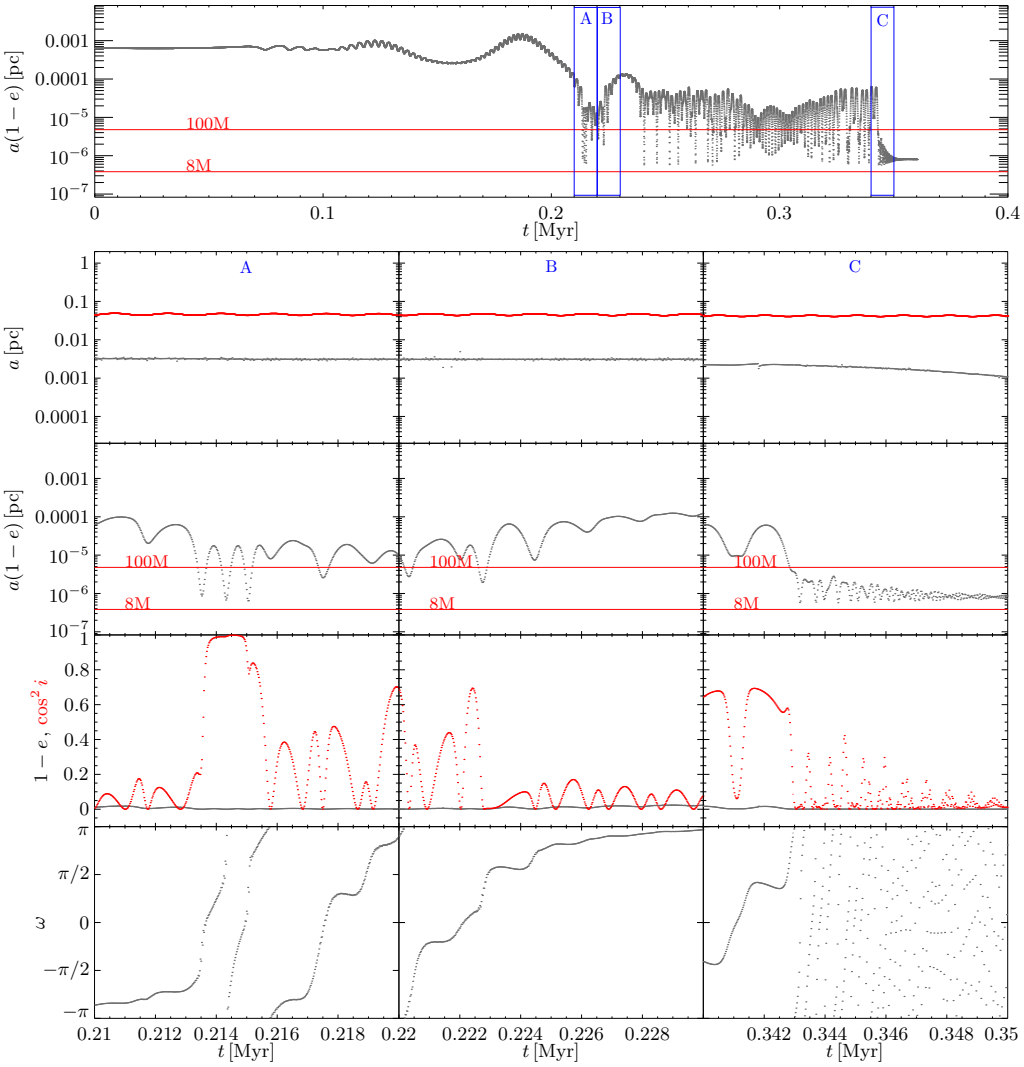


Figure 4.5: We plot $a(1-e)$ as a function of time for the entire simulation of the star shown in the right column of Fig. 4.4, but zoom in to show in more detail the regions where the relativistic precession has driven the star into the reverse-Kozai state. The three different boxed regions of the upper plot are shown in the lower plot in a separate column. In the first column is the first half of the region where the star first enters into the reverse Kozai state, while the second column is the second half of that region including where it exits the reverse Kozai state. The final column shows the final orbits which cause sufficient energy loss to reduce the semi-major axis and ultimately create an EMRI. Note that in regions where the oscillation in $a(1-e)$ is short, precession is higher. Slightly different initial conditions at the beginning of the first column produce qualitatively similar results on the star's orbital timescale, but diverge after just a few binary orbital periods. The complicated relationship between the small effects from the secondary which are important at high eccentricity, and the strong variations in the precession at high eccentricity is the likely source of the apparently chaotic nature of this star.

orbital energy in gravitational waves. Had the standard Kozai oscillation continued, the compact object would have only been in this configuration several times per Kozai cycle (see Appendix A) and the inspiral would be slow. In reality, the reverse Kozai mechanism begins and oscillations in periapsis distance have significantly shorter period and smaller amplitude, thus keeping the orbit at high eccentricity with small periapsis for longer (compare the left and right columns of Fig. 4.4).

4.5 The Simulation

To integrate the orbits of the test particles we utilize the symplectic integrator described in [111] and used in [112, 113]. This integrator has two particularly desirable properties for this work.

1. Its symplectic nature causes energy to be conserved up to round-off error. This is desirable since the spurious energy drifts found in many integrators would, over the many orbits simulated here, directly change the semi-major axis, so energy losses could lead to spurious EMRIs, or gains would destroy would-be EMRIs.

2. With an appropriate choice of step size (see Sec. 4.5.3) orbits in a Keplerian potential are reproduced exactly with only a phase error which is $\mathcal{O}\left(N^{-2}\right)$.

In particular, we use the version of the integrator used by [18], except extended to take into account both relativistic precession and the angular momentum and energy losses due to gravitational wave radiation. The latter must be done carefully to ensure that the energy-conserving quality of the symplectic integrator is preserved (Sec. 4.5.4).

For a detailed discussion of the integrator used in [18] see that work and its more detailed counterpart [114], and the works on which it was based [111–113]. We discuss here only how the integrator here differs from that of [18], focusing on how we deal with the effects of general relativity. In particular, we discuss how to incorporate these changes while still maintaining the separability of the Hamiltonian needed for the symplectic integrator. We introduce a new pseudo-Newtonian potential used to accurately reproduce the GR precession (Sec. 4.5.2). We discuss how the energy loss and angular momentum losses are taken into account to high accuracy using fitting functions to the Teukolsky equation for parabolic orbits (Sec. 4.5.4), and how EMRIs, direct plunges, and tidal disruptions (Sec. 4.5.5) are calculated in the simulation.

4.5.1 Determining GR Effects in Newtonian Code

One of the primary challenges is using our inherently Newtonian simulation to predict general relativistic effects on an orbit. While having many desirable properties, our integrator also requires that the Hamiltonian be separable, and therefore that the potential be a function of position only. Our approach is to use a pseudo-Newtonian potential that reproduces the correct precession rate

for the highly eccentric orbits from which EMRIs are produced. We discuss this potential below in Sec. 4.5.2.

This results however in positions and velocities close to the black holes that deviate from those which would be calculated from integration of the geodesic equation in a Schwarzschild space-time.

Therefore, instead of using position and velocity to calculate the instantaneous losses due to emission of gravitational radiation, we use the conserved energy and angular momentum in our pseudo-Newtonian potential to define an orbit. Then we calculate the orbit-averaged and energy and angular momentum loss and remove these in a single time step at each periapsis (Sec. 4.5.4).

4.5.2 Pseudo-Newtonian potential

We evolve test particles in the center of mass frame of the pseudo-Newtonian potential,

$$\begin{aligned} U(r) &= -\frac{GM}{r_1}\left(\frac{1}{1-\frac{5}{6}\frac{r_{\rm s,1}}{r_1}}+\frac{4}{3}\frac{r_{\rm s,1}}{r_1}\right) \\ &\quad -\frac{Gm_\bullet}{r_2}\left(\frac{1}{1-\frac{5}{6}\frac{r_{\rm s,2}}{r_2}}+\frac{4}{3}\frac{r_{\rm s,2}}{r_2}\right)-V(r_1)\,, \end{aligned} \tag{4.36}$$

where the numerical subscripts 1 and 2 are used to distinguish quantities measured with respect to the primary and secondary, respectively, r_i is the distance to the SMBH, $r_{\rm s,i}$ is the Schwarzschild radius of the SMBH, and $V(r_1)$ is the stellar potential given by Eqn. 4.3, which is assumed to follow the primary throughout the simulation. This expression is appropriate for parabolic orbits, which are the approximate orbits relevant to our method of EMRI formation and most studies of galactic dynamics in cores interacting with their (non-gaseous) surroundings.

Eqn. 4.36 agrees with the GR precession of periapsis to within 20% at all radii (see Fig. 4.6) for parabolic orbits, but is still conservative, and the resultant Hamiltonian is separable, allowing the use of the symplectic integrator of Preto and Tremaine [111] discussed in [18] and [114]. It also has the necessary feature of logarithmically diverging at the angular momentum appropriate for the UCO. As an example, in Fig. 4.7 we plot the trajectory of a star on a high eccentricity orbit with $1-e=10^{-4}$ for several orbits, along with the per-orbit apsidal precession compared to the proper precession predicted by general relativity.

Moreover, Eqn. 4.36 has the advantage that it is extremely quick to evaluate while still being accurate to better than 20% everywhere, and very accurate in the far field. While in principle one could derive a potential that exactly reproduces the correct precession for parabolic orbits as a function of angular momentum, the expression is unwieldy and involves elliptic functions, which in turn would significantly slow down the code since the potential and its derivative must be calculated at every step.

It is worthy to note that the commonly applied pseudo-Newtonian potential, the Paczynski-Wiita

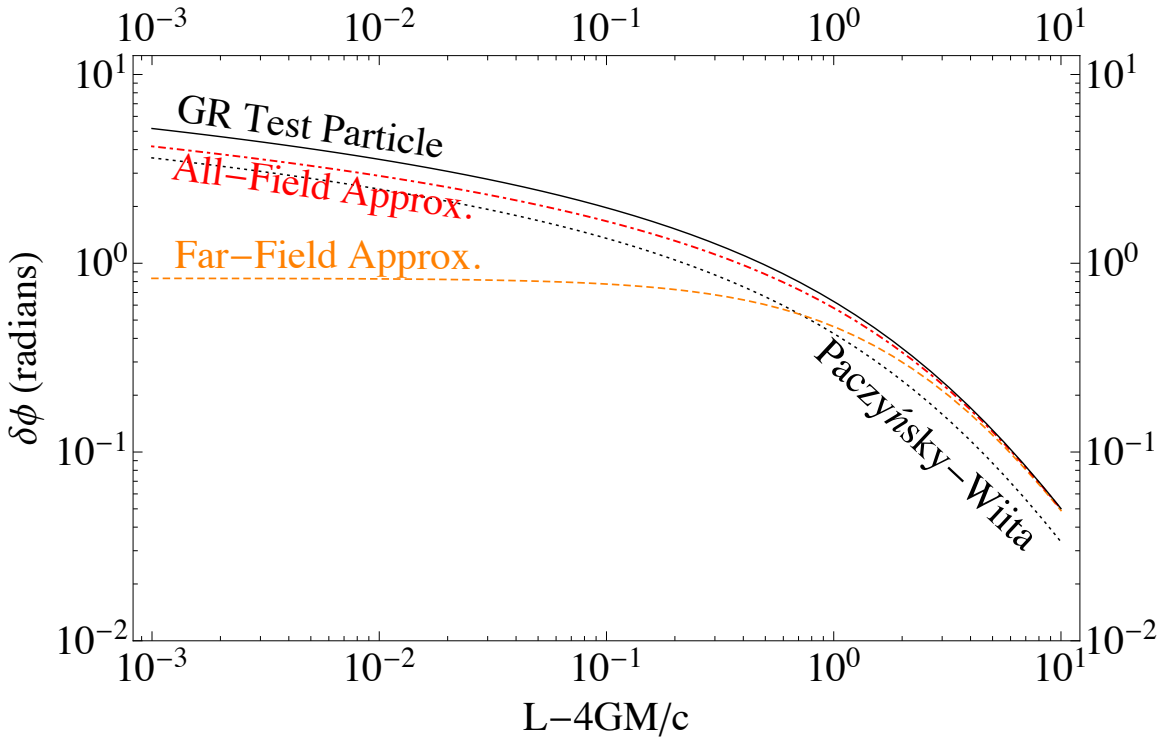
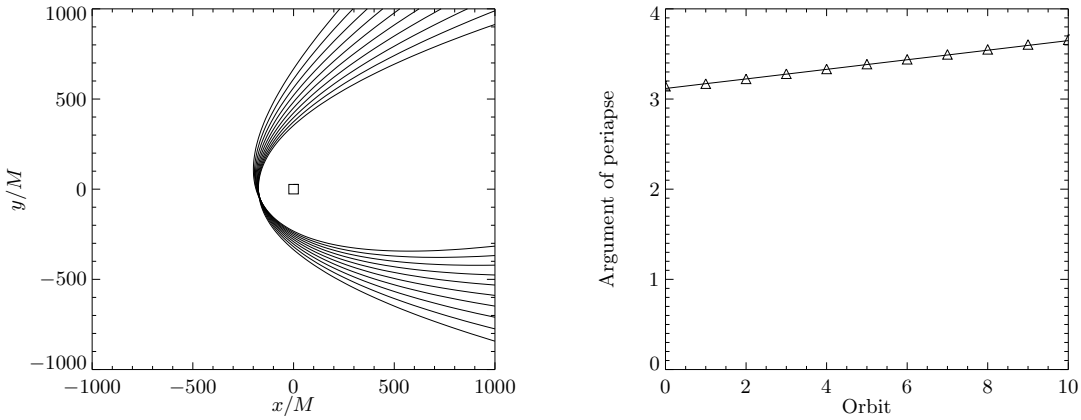


Figure 4.6: We plot the predicted apsidal precession as a function of the angular momentum close to the angular momentum of the innermost bound orbit (the unstable circular orbit for parabolic trajectories) for three pseudo-Newtonian potentials (intermittent lines) along with the apsidal precession for a test particle on a geodesic [115] in solid black. The commonly used Paczynski-Wiita pseudo-Newtonian potential, shown in dotted black, produces an error in GR precession greater than 30% at all radii. In dashed-orange is the far-field pseudo-Newtonian potential we introduce in the text (Eqn. 4.38) for computationally intensive tasks with $L \gtrsim 4GM/c$. In dash-dot red is the pseudo-Newtonian potential we use in our simulation (Eqn. 4.36) which produces the correct precession to within 20% for all angular momenta and to much higher accuracy in the far field.



(a) $1 - e = 10^{-4}, r_p = 100\,GM/c^2, x - y$ plane

(b) $1 - e = 10^{-4}, r_p = 100\,GM/c^2$ precession rate (triangles) and GR prediction (line)

Figure 4.7: Comparison of precession produced by code using the potential in Eqn. 4.36 with the rate predicted by GR for a highly eccentric orbit of a test particle

potential, produces an apsidal precession with no less than 30% error at all radii and in particular does not produce the correct precession rate in the far field limit (see Fig. 4.6). Because of the importance of apsidal precession to the Kozai mechanism and to many analyses of galaxy cores, correct precession is vital. See Secs. 4.3.2–4.3.3 and Sec. 4.4 for an in-depth discussion of these points. Because it requires only moderately more computations and provides the correct precession, it is best when working with highly eccentric orbits to use a potential like Eqn. 4.36.

A simpler potential accurate in the far field, but finite at the UCO, could also be used if one is limited significantly by computation. For reference, it is given by

$$U(r) = -\frac{GM}{r_1}\left(1 + \frac{3r_{s,1}}{2r_1}\right) \tag{4.37}$$

$$-\frac{Gm_\bullet}{r_2}\left(1 + \frac{3r_{s,2}}{2r_2}\right) - V(r_1)\,, \tag{4.38}$$

with parameters as in Eqn. 4.36, and is shown in Fig. 4.6.

For the stellar potential the self-consistent spherical η-models [116] with a central SMBH [95] were used. This potential is given by Eqn. 4.3

4.5.3 Step Size

Preto and Tremaine [111] show that for a Keplerian potential $U = \mu/r$ then using a step size $\propto 1/r$ reproduces Keplerian orbits exactly with only a phase error whose size is $\mathcal{O}\left(N^{-2}\right)$ where N is the number of steps per orbit. We therefore use a step size $\propto U$. The method of choosing the step size has not changed from [18], except that here we have chosen to have $20,000$ steps per orbit.

To check that we are not sensitive to step size we re-ran a simulation with $10,000$ steps per orbit. While individual stars evolved differently due to the chaotic nature of some orbits, the number of EMRIs and plunges was largely unchanged. In addition we re-ran key sections of Fig. 4.5 with $100,000$ steps per orbit and the qualitative behavior remained unchanged.

Close to either black hole numerical errors in our integrator manifest themselves as errors in the effective mass of the back hole [113]. Although these errors are small (typically the fractional error is $\delta M/M \leq 10^{-5}$), close to periapsis of highly eccentric orbits these errors result in a fluctuation in a. This is because the kinetic energy, $\frac{1}{2}v^2$, and potential energy, $\sim GM/r$, are both large and nearly cancel in the calculation of E. This, together with oscillations on a binary timescale result in the oscillations in a seen in Figs. 4.4 and 4.5.

4.5.4 Gravitational Wave Losses

When an object passes close to either SMBH, relativistic effects such as energy and angular momentum losses due to gravitational radiation become important. We incorporate these changes into the orbit by stepping out of the symplectic integrator at periapsis and calculating a new velocity vector using the energy and angular momentum loss due to gravitational radiation that were lost instantaneously at periapsis.

The energy and angular momentum loss are calculated assuming the orbit is parabolic. The assumption of a parabolic orbit is reasonable. For high eccentricity orbits the error is of order $1-e$, a number which is typically 10^{-4} or smaller for the orbits considered here.

Assuming a parabolic orbit, to calculate the energy and angular momentum loss we must first measure the angular momentum of the orbit. However, interactions with the binary affect the angular momentum far from the secondary, meaning that it is best to compute the angular momentum while close to the SMBH. This could be done at some arbitrary but close distance to the SMBH (such as $100\,GM/c^2$), but, despite the position and velocity of the object being somewhat unphysical close to the SMBH (see Sec. 4.5.1), the angular momentum will be conserved by the simulation (when far enough from the secondary) allowing us to calculate the angular momentum anywhere close to the SMBH. Then, for convenience, we calculate angular momentum at periapsis when we make the changes to the energy and angular momentum of the orbit.

To relate the angular momentum in the orbit to the energy and angular momentum lost during each periapsis pass we use the fitting functions from [117] in which they compute the fitting functions to the energy and angular momentum loss using the Teukolsky equation on parabolic orbits. For convenience we provide these fitting functions to the loss for parabolic orbits calculated from the

Table 4.1: Coefficients for Eqn. 4.39

	$n=0$	$n=1$	$n=2$
A_n^E	-0.318434	-5.08198	-185.48
B_n^E	0.458227	1645.79	8755.59
C_n^E	3.77465	-1293.27	-2453.55
A_n^L	-2.53212	-37.6027	-1268.49
B_n^L	0.671436	1755.51	9349.29
C_n^L	4.62465	-1351.44	-2899.02

Teukolsky equation here:

$$\begin{aligned}\frac{M}{m}\Delta X &= \cosh^{-1}\left[1+B_0^X\left(\frac{4}{\tilde{r}_{\rm p}}\right)^{N_X-1}\frac{1}{\tilde{r}_{\rm p}-4}\right] \\ &\quad\times\sum_{n=0}^{N}A_n^X\left(\frac{1}{\tilde{r}_{\rm p}}-\frac{4}{\tilde{r}_{\rm p}^2}\right)^n \\ &\quad+\frac{\tilde{r}_{\rm p}-4}{\tilde{r}_{\rm p}^{1+N_X/2}}\sum_{n=0}^{N}C_n^X\left(\frac{\tilde{r}_{\rm p}-4}{\tilde{r}_{\rm p}^2}\right)^n \\ &\quad+\frac{\tilde{r}_{\rm p}-4}{\tilde{r}_{\rm p}^{2+N_X/2}}\sum_{n=0}^{N-1}B_{n+1}^X\left(\frac{\tilde{r}_{\rm p}-4}{\tilde{r}_{\rm p}^2}\right)^n\,, \end{aligned} \tag{4.39}$$

where X is either the specific energy E/c^2 or the (scaled) specific angular momentum $cL/(GM)$, $\tilde{r}_{\rm p}=r_p/(GM/c^2)$, $N_E=7$, $N_L=4$, and the A_n^X, B_n^X, and C_n^X are coefficients given in Table 4.1. [117] note that $N=2$ is sufficient for better than 0.2% accuracy everywhere. This is the order used in our code and the coefficients from [117] are given in Table 4.1. Here, $\tilde{r}_{\rm p}$ is calculated based on the periapsis an orbit would have if it were parabolic and had the measured angular momentum (i.e., $\tilde{r}_{\rm p}$ is *not* calculated from the position of the object at its ostensible periapsis, since, as described in Sec. 4.5.1, this position is not reliable).

We subtract the energy and angular momentum loss given by Eqn. 4.39 at the step closest to periapsis. Introducing these changes at some other point other than periapsis introduces spurious precession.

At periapsis we calculate a new velocity, $\vec{v'}$, using the new specific energy, $E'=E+\Delta E$, and angular momentum, $L'=L+\Delta L$. Since the position is unchanged, the potential energy is unchanged and

$$v'^2=v^2+2\Delta E\,. \tag{4.40}$$

The orbital plane remains unchanged for a Schwarzschild black hole and therefore

$$\vec{L'}=\frac{L+\Delta L}{L}\vec{L}=\vec{r}\times\vec{v'}\,. \tag{4.41}$$

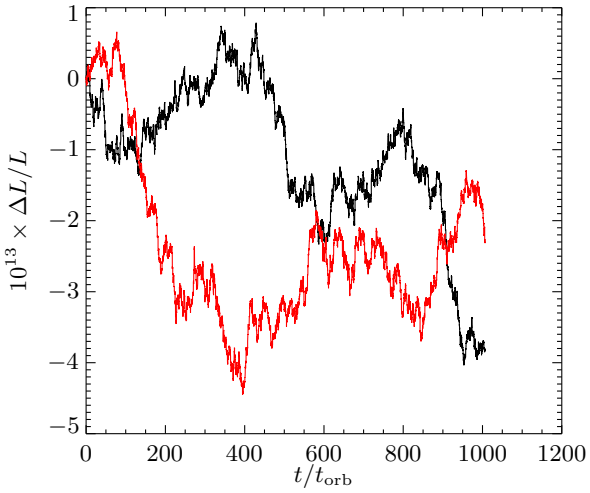


Figure 4.8: Plot showing the errors in conservation of reduced angular momentum, L over many orbital periods ($t_{\rm orb}$), of a high eccentricity ($e = 1 - 10^{-5}$) test particle. The red is without the procedure for calculating the change in $\vec{v}$ at periapsis, the black uses Eqn. 4.43 but with $\Delta L = 0$. The secondary has zero mass for both curves. The errors are still at the level $\Delta L/L \sim 10^{-12}$ indicating that the process of stepping in and out of the symplectic integrator does not inherently introduce significant errors.

Taking the dot product of this yields

$$\vec{r}.\vec{v'} = \sqrt{L'^2 - r^2 v'^2} \tag{4.42}$$

where we take the positive branch of $\vec{r}.\vec{v'}$ since this corresponds to the outgoing, post-periapsis solution. The cross-product $\vec{r} \times \vec{L'}$ yields

$$\vec{v'} = \frac{1}{r^2}\left[(\vec{r}.\vec{v'})\vec{r} - \vec{r} \times \vec{L'}\right] . \tag{4.43}$$

Eqn. 4.43 together with 4.40, 4.41, and 4.42 are used to calculate the new velocity $\vec{v'}$ following a periapsis passage.

In Fig. 4.8 we show the numerical accuracy of this procedure by considering whether L remains constant over many orbits. Despite the integrator no longer being truly symplectic due to the stepping out of the leapfrog scheme at periapsis, the errors remain small over many orbits.

4.5.5 Plunges

At each $< 100\,GM/c^2$ periapsis passage we check for plunges directly into the black hole. The marginally bound orbit is the appropriate limit since the particles lie on highly eccentric orbits. This has a reduced angular momentum around a Schwarzschild black hole of $L = 4\,GM/c$. On

Table 4.2: Number of EMRIs and Their Probabilities in Simulations

	Parameters				EMRI stars				
Sim #	q^a	$\frac{m_\star}{M_\odot}{}^b$	$N_\star{}^c$	Duration (Myr)^d	$N_{\rm K}{}^e$	$N_{\rm RK}{}^f$	$N_{\rm SS}{}^g$	$N_{\rm EMRI}{}^h$	$P_{\rm EMRI}{}^i$
1	0.3	10	10^6	1.5	14	13	0	18	1.8×10^{-5}
2	0.3	1	10^6	1.5	0	0	0	0	0
3	0.1	10	10^6	1.4	7	7	2	13	1.3×10^{-5}
4	0.1	1	10^6	1.4	1	1	0	2	0.2×10^{-5}

[a] $q = M_\bullet/M \leq 1$ is the ratio of the masses of the secondary SMBH and the primary.
[b] The assumed mass of the stars during this simulation. The mass of the stars is only relevant when the star passes within $100\,GM/c^2$ of the one of the SMBHs.
[c] $N_\star$ is the total number of stars simulated during the run.
[d] The duration of the simulation in megayears.
[e] The number of EMRIs whose formation required the Kozai mechanism.
[f] The number of EMRIs whose formation required the reverse-Kozai mechanism.
[g] The number of EMRIs whose formation required strong scatterings by the secondary.
[h] The total number of EMRIs formed during the simulation.
[i] The predicted probability of a CO of mass $m_\star$ becoming an EMRI.

periapsis passages, if $L \leq 4\,GM/c$, a direct plunge is assumed to result and the integration is halted.

4.5.6 EMRIs

At each apoapsis following a $< 100\,GM/c^2$ periapsis passage we additionally check whether the particle has entered the LISA band. We do this by checking if the semi-major axis is small enough that the test particle's orbital period is below 10^4 s, i.e.,

$$E \geq \frac{GM}{2a} = \frac{(GM)^{2/3}}{2}\left(\frac{2\pi}{10^4\text{ s}}\right)^{2/3}, \tag{4.44}$$

where E is the orbital energy. We then follow up with these objects to ensure that they would not have already been an EMRI had the secondary not been included, and to understand the different ways in which the secondary produces EMRIs.

4.6 Simulation Results

All simulations executed include 10^6 stars reacting to a binary with a mass ratio which is either $q = 0.3$ or $q = 0.1$. For each mass ratio two sets of simulations were run, one with $m_\star = 10\,M_\odot$ and one with $m_\star = 1\,M_\odot$. All runs have the same initial data, and are therefore not independent. This is significantly quicker computationally because for simulations with the same q we only need to reintegrate orbits that pass within $100\,GM/c^2$. Orbits that lie outside $100\,GM/c^2$ remain in the test particle limit and their evolution is unaffected by their mass. Only for orbits passing inside $100\,GM/c^2$, where gravitational radiation losses are significant, does the mass of the star matter.

Additionally it also provides a direct comparison between which stars from each simulation form EMRIs.

Throughout the stars' orbits their periapsis distances and semi-major axes were monitored. The simulation of a given star was stopped if the star either entered the LISA band (see Sec. 4.5.6 for definition), passed inside the unstable circular orbit ($L \leq 4\,GM/c$ for parabolic orbits), or if the number of steps required to complete the simulation exceeded 10^{10} steps. Those stars which were judged to be EMRIs were then re-examined more closely to judge the specific mechanisms which led to the EMRI, classifying them into Kozai inspirals, reverse-Kozai inspirals (which first rely on the Kozai mechanism), and strong scatterings where the secondary interacts strongly with the star and sends it close enough to the primary to result in an EMRI.

We provide the resultant number of EMRIs from these simulations in Table 4.2. Thus the numbers listed are appropriate if all of the stars being simulated were the appropriate species for the given phenomenon (compact object for plunges and main sequence star for tidal disruption).

Though Table 4.2 is the key output of the simulation, we may still learn a lot about the processes which deliver the stars to this region of parameter space. In Fig. 4.9 we plot the outcomes of all stars with initial positions between 10^{-3} and 10^{-1} pc by coloring the stars based on their final state. Stars are placed on the plot based on their initial angular momentum and initial position. Also plotted is the Kozai wedge; lines of constant azimuthal angular momentum at $|L_z| = 4, \sqrt{200}\,GM/c$. Since L_z is conserved in the Kozai mechanism, it is this region of low $|L_z|$ which is expected to have the highest probability of becoming EMRIs or directly merging (or being tidal disruptions in the case of non-compact objects). The black-lined gold stars represent stars which have ultimately become EMRIs, while the green dots demarcate stars which have ultimately merged via a plunge. Shades of blue indicate the closest distance that each star has gotten to the primary SMBH (see figure caption for more details). Most EMRI-forming stars originate from close to the Kozai wedge, with several becoming EMRIs after strong interactions with the secondary. This is quantified in Table 4.2 where $N_{\rm K}$, $N_{\rm RK}$, and $N_{\rm SS}$ refer to the number of EMRIs out of the original 10^6 stars that were formed using the Kozai mechanism, the reverse Kozai mechanism, and strong scattering, respectively. Note that these are not necessarily mutually exclusive groups.

For clarity we also give Fig. 4.10, which is essentially a re-make of Fig. 4.9 except with a non-normalized azimuthal angular momentum on the y-axis and the initial semi-major axis on the x-axis. The color scheme and lines are the same as in Fig. 4.9. There is a clear "sweet spot" around $r_{\rm stall}/10$ ($r_{\rm stall}$ being the secondary's stalling radius) where stars originally becoming EMRIs, direct plunges, or tidal disruptions originate.

To understand how many stars approach radii close to the primary SMBH we plot the normalized number of stars to have reached a radius r_p in a cumulative plot in Fig. 4.11. Also plotted is the same curve but for the secondary SMBH (solid red) and the same curve but for the same initial

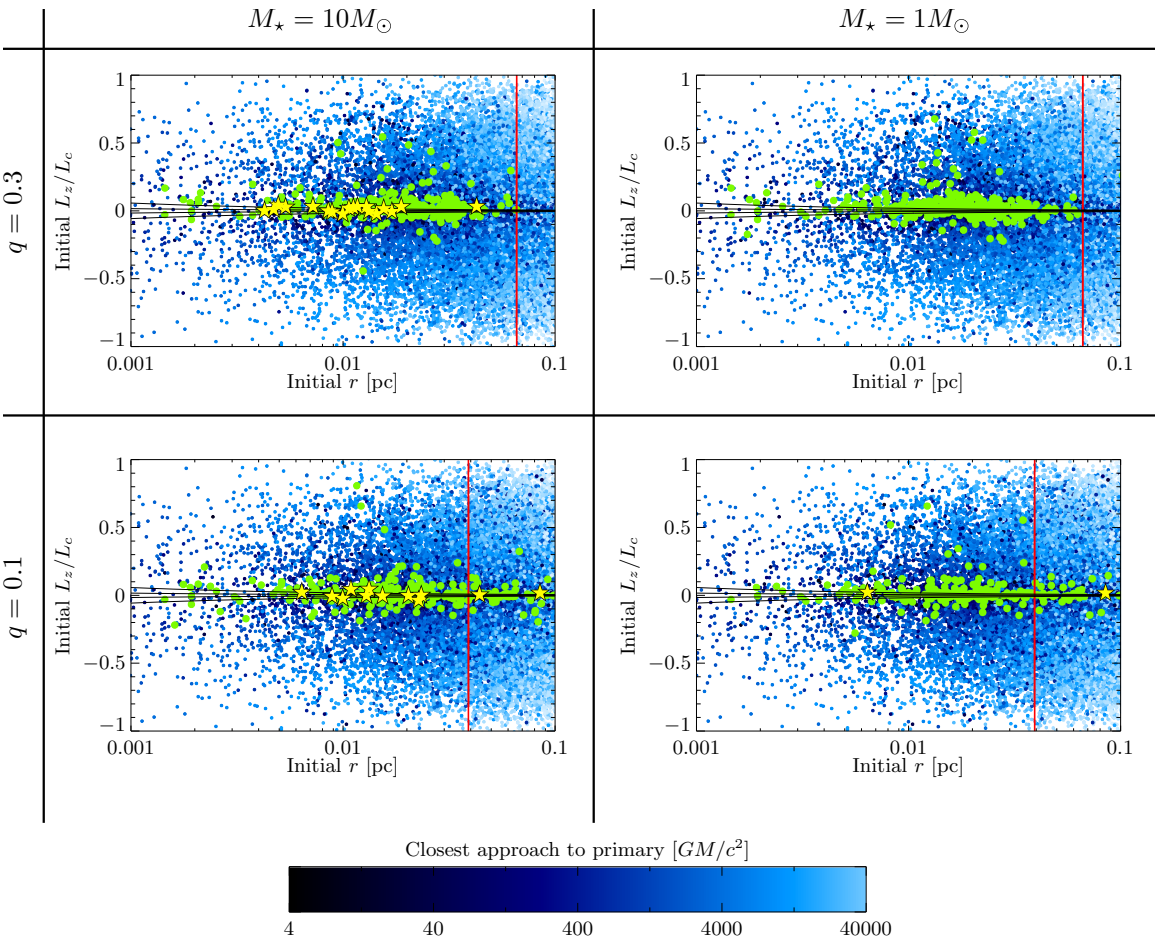


Figure 4.9: We plot the initial L_z/L_c as a function of the initial radial position for the stars of the four simulations done. The top and bottom rows are runs done with $q = 0.3$ and $q = 0.1$, respectively, while the first and second columns are runs done assuming stars of mass $10\,M_\odot$ and $1\,M_\odot$, respectively. The mass is only relevant if the star passes within 100 GM/c^2 of one of the SMBHs (see Sec. 4.5.4 for details). For reference, the final stalling radius of the secondary is shown as a vertical red line. By color we code how close the stars have come to the primary SMBH: gold stars represent stars which have turned into EMRIs (reached the LISA band; Sec. 4.5.6), green dots are stars which have by some method plunged into the primary SMBH (see Sec. 4.5.5 for definitioni), and all other stars are colored in shades of blue with the lightest versions being those which have never come close to the primary and the darker colors being those which have passed very close (see legend for exact definitions). Also plotted in solid black is the Kozai "wedge"; lines of $|L_z| = 4, \sqrt{200}\,GM/c$. Inside these lines the standard Kozai formalism predicts that it is possible (but dependent on initial conditions) that a star reaches 4, 100 GM/c^2, respectively. It is only inside these wedges that the standard Kozai formalism would predict possible EMRI formation. For a zoomed in plot of this region see Fig. 4.10. A significant number of stars out of the Kozai wedge eventually merge with (or are tidally disrupted by) the primary SMBH. This is due to either strong interactions with the secondary SMBH or the oscillations on the orbital timescale of the secondary SMBH (see Sec. 4.3.2).

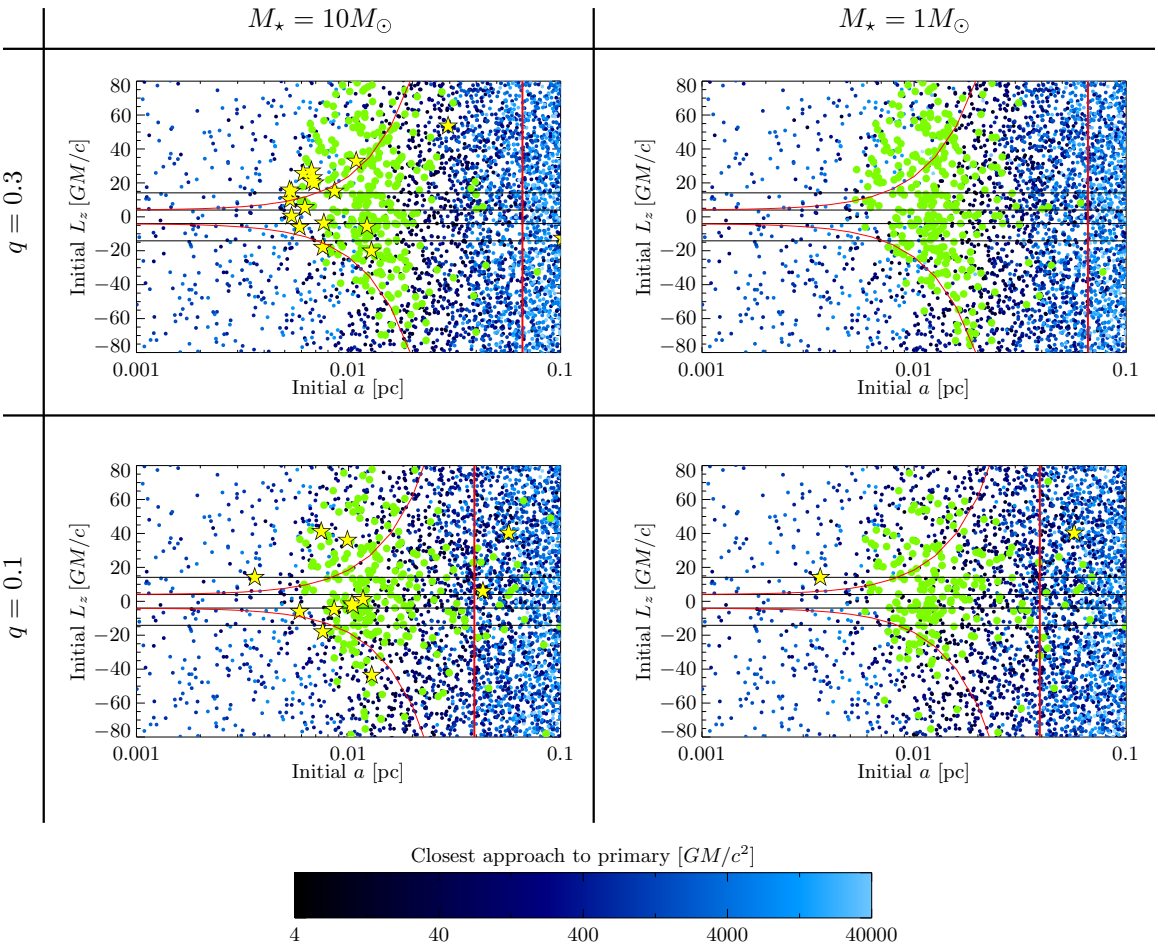


Figure 4.10: We plot L_z in units of GM/c as a function of the stars' initial semi-major axes for the four runs executed, each with the same 10^6 stars. The color scheme and data sets are the same as in Fig. 4.9. However, here we also plot the expected oscillations of L_z on the secondary's timescale (see Sec. 4.3.2 and Fig. 4.3). It is possible to see an asymmetry in these plots, indicating a preference for driving stars with positive L_z to high eccentricity. This is an expected outcome of the reverse-Kozai mechanism.

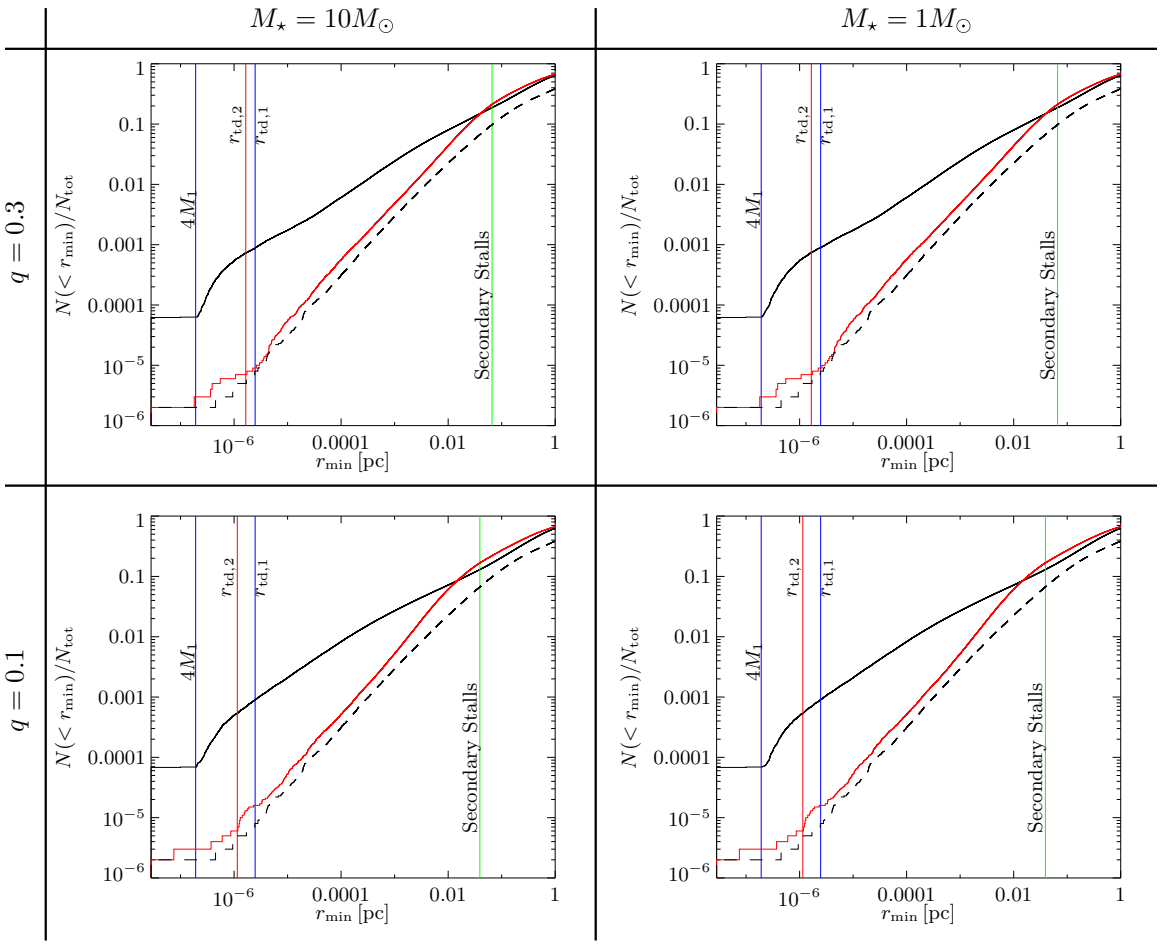


Figure 4.11: We plot the normalized number of stars to have reached a minimum radius less than r_p. Here r_p is calculated from the angular momentum of the star's orbit which is assumed to be parabolic. The ordering of the plots is the same as used previously: the upper and bottom rows are for simulations with $q = 0.3$ and $q = 0.1$, respectively, and the first and second columns are for simulations with $M_\star = 10\,M_\odot$ and $M_\star = 1\,M_\odot$, respectively. In solid black are the results for when r_p is calculated to be the distance to the primary, and in solid red when r_p is the distance to the secondary. In dashed black we plot the same as solid black, but have simulated the evolution of the stars without the presence of the secondary. In each plot the tidal disruption radii for each SMBH, $r_{\rm td,1}$ and $r_{\rm td,2}$, is demarcated and labelled, along with the stalling radius of the secondary. The unstable circular orbit for parabolic orbits, $r_{\rm UCO} = 4\,GM/c^2$, is labeled as $4M_1$.

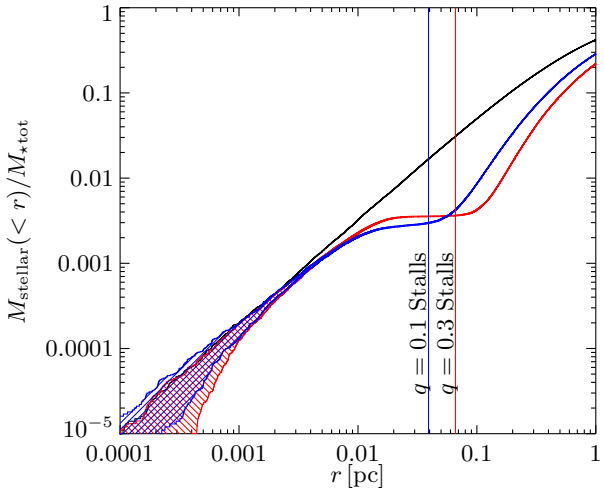


Figure 4.12: We plot the stellar mass interior to a given radius normalized to the total stellar mass, $M_{\rm stellar}(< r)/M_{\star{\rm tot}}$. In black we plot the initial cumulative stellar mass (c.f. Eqn. 4.2), while in blue we plot the cumulative stellar mass at the end of the simulation for the $q = 0.1$, $M_\star = 10\,M_\odot$ run, and in red we plot the cumulative stellar mass at the end of the simulation for the $q = 0.3$, $M_\star = 10\,M_\odot$ run. Note that the initial conditions for all runs are the same. The only difference between the outcomes of the runs with the same q but different stellar masses is due to those stars which go within $100\,GM/c^2$, only a small fraction of all stars. Thus, the cumulative stellar mass for the runs with $M_\star = 1\,M_\odot$ are virtually the same as the $10\,M_\odot$ counterparts with the same q. The hatched regions are stars that had not completed the full simulation within the preset limit of 10^{10} steps. For reference we also plot the stalling radii of the secondary as vertical lines.

stars evolving in the absence of a secondary SMBH. Labelled vertical lines mark the radius of the unstable circular orbit for parabolic orbits $r_p = 4\,GM/c^2$ and the tidal disruption radii for the primary and secondary. In our calculation of the number of tidal disruptions we assume that any star reaching a radius inside the tidal disruption radius is tidally disrupted. This is obviously an upper bound, though likely close to the correct number of tidal disruptions. Of some note, there are a non-zero number of tidal disruptions around the secondary. Of course, the true number would likely be significantly greater since the secondary will also carry with it a similar stellar cusp which will be scattered by the primary. However, even in the equal mass case, this could not increase the total number of tidal disruptions by greater than a factor of two.

One of the inconsistencies in our methodology is the assumption that the stellar potential does not evolve with time, though the stars' orbits do. This inconsistency is potentially significant. To better understand the effect of such an assumption we plot the mass interior to a given radius as a function of radius in Fig. 4.12. There the solid black line is the initial distribution given by Eqn. 4.2, the red solid line is the curve for the $q = 0.3$ and $m_\star = 10\,M_\odot$ simulation, and the blue solid curve is for the $q = 0.1$ and $m_\star = 10\,M_\odot$ simulation. The mass of the star has little to no effect on these curves, so the other two simulations look almost identical and are suppressed. The filled region

represents stars that required more than 10^{10} steps to complete the simulation and were therefore terminated.

Noting in Figs. 4.9 and 4.10 that most EMRIs originate from $\approx 10^{-2}$ pc, and noting here that the mass interior has not changed significantly at these radii, this assumption would not likely have a significant impact on our findings.

4.7 EMRI Rates

We now turn to discussing the predicted rates of EMRI mergers per cubic Gpc. The calculation is straightforward: Given a species X of CO or star (main sequence star, stellar mass black hole, neutron star, or white dwarf) the simulation simulates $N_\star$ test particles and outputs the number of EMRIs $N_{\rm EMRI}(X)$ assuming a mass M_X. Then

$$P_{\rm EMRI}(X) \equiv N_{\rm EMRI}(X)/N_\star \tag{4.45}$$

is the probability that a star of mass M_X eventually becomes an EMRI (Sec. 4.7.1). We then multiply by the expected number of stars in our model cusp with mass $M_\star$ assuming some initial mass function and some quantity of mass segregation (Sec. 4.7.2). This gives the approximate number of EMRI mergers for a given galaxy during the period of time that a secondary SMBH is just settling into its stalling radius. By determining the rate of the number density of galaxies undergoing a merger, $\dot{n}_{\rm merger}(M = 10^6\,M_\odot)$ (Sec. 4.7.3), we may produce the predicted EMRI rate density (Sec. 4.7.4). That is,

$$\mathcal{R}_{\rm EMRI}(X) = P_{\rm EMRI}(X) N_X \dot{n}_{\rm merger}(M = 10^6 M_\odot)\,. \tag{4.46}$$

4.7.1 EMRI Merger Probability

Each of our simulations contains 10^6 stars, with stars of mass $10\,M_\odot$ or $1\,M_\odot$, where we use the former to predict the rates of stellar mass black holes and the latter to predict the rates for both neutron stars and white dwarfs. We calculate the probability that an object in a given simulation will turn into an EMRI using Eqn. 4.45 (shown in Table 4.2) and will use the appropriate simulation to determine the rates of the distinct species in Sec. 4.7.

4.7.2 Species Number Density

There are several major unknowns in the calculation of rates, two of which culminate in a fairly significant uncertainty in the determination of the density of a given species of compact remnant in

Table 4.3: Approximate Mass and Number Densities of Species Deep ($< 0.05\,\mathrm{pc}$) in Stellar Cusp [97]

Species X	$\frac{M_X}{M_\odot}$ [a]	$\frac{\rho_X}{\rho_{\mathrm{MSS}}}$ [b]	$\frac{n_X}{n_{\mathrm{MSS}}}$ [c]	N_X [d]
Main Sequence Star	1	1	1	2×10^5
Stellar-Mass Black Hole	10	~ 10	~ 1	2×10^5
Neutron Star	1	~ 0.1	~ 0.1	2×10^4
White Dwarf	1	~ 0.3	~ 0.3	6×10^4

[a] The mass of species X is solar masses.

[b] The density of species X relative to the density of main-sequence stars in the region where EMRIs are sourced in our simulations.

[c] The number density of species X relative to the number density of main-sequence stars in the region where EMRIs are sourced in our simulations.

[d] The total number of species X in the entire stellar cusp, if the cusp were to have the same ratio of species X to main-sequences stars as in the region where the EMRIs are sourced in our simulation. See text for explanation.

the region of influence of the primary SMBH; both the stellar cusp profile and the quantity of mass segregation which has occurred directly effect the number density of each species of compact object.

In our simulations we use an η-model [95] of a spherical stellar cusp with a SMBH to establish the stellar distribution (see Sec. 4.2). Two parameters, the SMBH mass and a parameter η, which regulates the steepness of the stellar density, determine this self-consistent family of models of a *stable* stellar cusp. The η-model does not depend on the mass of the stars, and thus, in principle, leaves open the possibility of combining this model with a multi-mass model including mass segregation, such as has been done by [97].

However, for simplicity we do not choose this path, but instead simulate N_X stars of a characteristic mass M_X of a single species X. In line with the simplicity argument we then multiply $N_{\mathrm{EMRI}}(X)/N_X$ by the total number of compact objects of type X expected in the region from where most EMRIs are sourced.

Determining the number of compact objects of type X expected in some region is actually quite difficult, due to limitations in observations and the complex physics involved. However, [97] found, using multi-mass Monte-Carlo simulations of stellar remnants in galaxy cores, that mass segregation led to a significant over-density of stellar-mass black holes and an under-density of white dwarfs and neutron stars when compared to main sequence stars (see Table 4.3 for approximate values). We use these as our fiducial values.

It may appear concerning that the number of predicted SBHs in Table 4.3 implies an unrealistic number of SBHs in the cusp, but these numbers have been scaled to ensure the correct ratio of SBHs in the inner cusp, from where the EMRIs are sourced, given the constraint of enclosed mass and the [97] results. Thus, these numbers produce the correct number of SBHs in the relevant regions.

There are errors involved in using the [97] values which are appropriate for different values of η, especially in their application to determining the rates for white dwarfs and neutron stars which

are found to relax to a different value of η than we have considered here. However, those errors are small since we have normalized to the appropriate densities at 0.01 pc where the EMRIs originate.

In our simulations the total stellar mass assumed to be in the cusp is $2\,M = 2 \times 10^6\,M_\odot$, so that the number of any given species is approximately given by

$$N_X \approx 2 \times 10^5 \frac{n_X}{n_{\rm MSS}}\,. \tag{4.47}$$

4.7.3 Number Density of Mergers

An approximate form for the number density of mergers with SMBHs of mass $10^6\,M_\odot$ and mass ratios between 0.1 and 0.3 is merely to find the number density of SMBHs of mass $10^6\,M_\odot$ and assume one such merger per galaxy lifetime. In reality, the merger rate is far more complicated than this and depends strongly on the physics and the model But, crudely speaking, we don't expect many more significant mergers than 1 per Hubble time, since otherwise large disk galaxies like the Milky Way would not exist in the local universe, and we don't expect many fewer, since otherwise elliptical galaxies would not be plentiful today. Thus we use the naive order-of-magnitude estimate of 1 merger per galaxy per Hubble time.

[16] find that there are approximately 4×10^6 Gpc^{-3} SMBHs of mass $(10^{5.5}$–$10^{6.5})\,M_\odot$ assuming $H_0 = 71\,{\rm km\ s^{-1}Mpc^{-1}}$. Thus

$$\dot{n}_{\rm merger}(M = 10^6 M_\odot) \sim 3 \times 10^{-4}\,\frac{\rm mergers}{\rm Gpc^3\,yr}\,. \tag{4.48}$$

4.7.4 Final Rates

We can now calculate the rate of EMRI production per unit volume, $\mathcal{R}_{\rm EMRI}$, using Eqn. 4.46. We use the probability that each star in our simulations becomes an EMRI, $P_{\rm EMRI}$, from Table 4.2, the numbers of each species, N_X, from Table 4.3, and the SMBH merger rate, $\dot{n}_{\rm merger}$, from Eqn. 4.48. The rates for neutron stars (NSs) and white dwarfs (WDs) are uncertain since only 2 total EMRIs were produced across the 2×10^6 simulated $1M_\odot$ objects. Regardless of this uncertainty the rates from NSs and WDs from our production method are insignificant for the current LISA design. However, the rates for stellar-mass black holes (SBHs) are higher and significant for the current LISA design. The rates are given in Table 4.4.

4.8 Discussion

There are three key points to this paper: 1) SMBH binaries will likely produce EMRIs, 2) the formation of these EMRIs is intricate and not conducive to analytic secular approximations, and 3) to understand the formation phenomena which depend strongly on the GR precession, it is best to

Table 4.4: Final Rates of EMRIs Due To SMBH Binaries

q^a	$\mathcal{R}_{\rm EMRI}$ $(\mathrm{yr}^{-1}\mathrm{Gpc}^{-3})^b$		
	SBH	NS	WD
0.3	1×10^{-3}	0	0
0.1	8×10^{-4}	$\sim 1 \times 10^{-5}$	$\sim 4 \times 10^{-5}$

[a]The mass ratio of the secondary SMBH to the primary.
[b]Rate of EMRIs due to SMBH binaries for the cases of stellar-mass black holes (SBHs), neutron stars (NSs), and white dwarfs (WDs)

use a pseudo-Newtonian potential such as Eqn. 4.36, which reasonably well approximates the GR precession at all radii (to within $\approx 20\%$ at worst). We discuss each of these points in turn.

4.8.1 EMRI Rates

The EMRI rate densities predicted in Sec. 4.7.4 are high enough that, given optimistic numbers and under the current design of LISA, then of order several such EMRIs could be detected. This is particularly interesting in light of the recent work by [85, 87] noting that the EMRI waveforms carry in them information both about the gas present in the system and whether there is a secondary SMBH at distances appropriate for what we are discussing here. Because such a secondary would likely disrupt the inspirals expected in the standard picture of EMRI formation, and because the rates while the secondary is inspiraling are significantly higher than in the standard picture, EMRIs detected with a signal of a secondary in the waveform will have likely been formed by a mechanism described in this paper.

The numbers quoted in Sec. 4.7.4 have an important dependency: they will change linearly with the number density of compact objects at about 1/10 the stalling radius. If mass segregation has not yet had time to fully take place, the rates will decrease by an appropriate amount. If a different stellar cusp model is used, the rates should go as the rates presented here (Sec. 4.7.4) times the ratio of number densities at about 1/10 the stalling radius (and taking into account some model of mass segregation). This is in contrast to the standard picture of EMRI formation where the rates should change as the number density of COs times the number density of all objects.

Another interesting point is that in the standard picture of EMRI formation, stars must scatter to a state with low overall L, while in the Kozai picture stars must only have a low L_z (see Figs. 4.9 and 4.10). Thus, once the secondary SMBH has stalled and the timescales for inspiral become longer, star-star interactions become relevant and need only scatter stars to a state of low L_z to increase the probability of turning into an EMRI.

4.8.2 SMBH Binary EMRI Formation

Secular Predictions: EMRI formation in the presence of a secondary is surprisingly rich physically. The gentle balance between the Kozai mechanism, SP precession, and GR precession required to drive the stars to high eccentricity (Secs. 4.3 and 4.4) was unexpected, as was the importance of the reverse-Kozai mechanism (Sec. 4.4).

These intricacies make secular predictions of the EMRI rate through the Kozai mechanism using calculations such as in Appendix A essentially irrelevant without a more advanced formalism considering complications such as the reverse-Kozai mechanism.

Phase Errors: Our symplectic integrator is excellent for conserving energy along an orbit, and properly reproducing the path of the orbit. However, these integrators produce errors in the phase of the orbit [discussed in 18], possibly moving the star to an incorrect phase with respect to the secondary over the many thousands of orbits we integrate over. The phase relative to the secondary is particularly important to the maximum eccentricity reached during a Kozai cycle (see Sec. 4.3.2 and Fig. 4.3), which in turn has a significant impact on the reverse-Kozai mechanism (Sec. 4.4) required for the formation of many EMRIs (Table 4.2). We check for this possibility by running one of our runs with a factor of 2 smaller step size. That this change had no effect on the produced rates, while producing different phase errors indicates that this is not an issue.

Stellar Interactions: We have also not included stellar interactions for simplicity. This is not necessarily a good assumption since star-star relaxation occurs on similar timescales at 10^{-2}pc. In the near future we intend to address this issue by giving stars random weak kicks near apoapsis. It is the weak kicks that dominate the relaxation process in these mass segregated regions [97].

One important assumption that is potentially quite poor as one enters the diffuse regions (star-wise) near the primary SMBH, is that the stellar potential is spherical. When the number of stars interior to some radius becomes small the potential becomes highly non-spherical and dominated by discrete star-star potentials. We do not consider such complications due to lack of the appropriate resources, and leave it to future work.

4.8.3 Pseudo-Newtonian Potential

The Paczynski-Wiita pseudo-Newtonian potential provides reasonable accuracy regarding GR precession, but has greater than 30% error at all radii. For high eccentricity orbits this limitation can be important. For instance, that it under-produces far-field precession (see Fig. 4.6) may be seen as unimportant, until close examination of Fig. 4.4, where the significant effect of the stellar potential in the absence of GR precession is clear in the second column. However, *even in the far-field*, when the correct GR precession is used (third column), the evolution of the star's orbit is fundamentally different and, in particular, more extreme.

However, these inaccuracies are straightforward to overcome in the case of parabolic orbits with the potential given by Eqn. 4.36, though even this can be bettered as will be done in the final version of this paper.

4.8.4 Tidal Disruptions

In each of our simulations we found roughly 900 tidal disruptions, a number consistent with the results from our previous study on tidal disruptions [18]. This large rate of tidal disruptions during the inspiral [18, 118–120, and calculations herein] will likely mean that there is gas present in the vicinity of the primary SMBH, something which may be detectable in the EMRI waveform [87]. It is a nice feature of this system that it both predicts the existence of EMRIs and a gaseous environment, and can possibly validate that prediction in the EMRI detection.

4.9 Conclusion

We have used a symplectic integrator to integrate four sets of 10^6 stars orbiting a primary supermassive black hole (SMBH) of mass $10^6\,M_\odot$ with a secondary SMBH of varying masses inspiraling by dynamical friction to its stalling radius. When the stars get closer than $100\,GM/c^2$ from either SMBH, orbits which are locally nearly parabolic, we subtract the appropriate quantity of energy and angular momentum (due to gravitational wave radiation) at periapsis (Sec. 4.5.4). The energy and angular momentum loss is accurate for parabolic orbits to within a fraction of a percent. Moreover we use a new pseudo-Newtonian potential which produces the correct quantity of GR precession (to within 20% at the UCO and significantly more accurate farther away) all the way to the unstable circular orbit (UCO) for parabolic orbits (Sec. 4.5.2). Two quantities are followed throughout each star's evolution: its innermost periapsis distance and its semi-major axis. If the star's semi-major axis evolves such that it would be in the laser interferometer space antenna's (LISA's) band, we terminate the simulation and count the star as an extreme mass ratio inspiral (EMRI).

From this analysis there are several key conclusions which can be made

1. EMRIs can indeed be created by SMBH binaries and their rates are of order 10^3 times *smaller* than the optimistic published LISA rates.

2. A previously unnoticed phenomenon we refer to as the reverse Kozai mechanism is vital to the formation of a significant number of the EMRIs. This effect is due to rapid apsidal precession which leaves the secondary SMBH effectively sitting still while the orbit of the star executes full 2π precession. In the frame of the precessing orbit the secondary ostensibly executes a reverse circular motion around the orbital axis of the star, which then induces oscillations similar to those of the standard Kozai mechanism. These oscillations are much faster and

weaker. Because this only occurs at high eccentricity when periapsis distances are small, the weaker amplitude keeps the star at higher eccentricity and improves the rate at which the star radiates its orbital energy.

3. We introduce a new pseudo-potential appropriate for parabolic curves which produces the correct apsidal precession to high accuracy far from the SMBH and to within 20% as the UCO is approached.

4. Due to the important relationships between the Kozai mechanism, phase, oscillations in the eccentricity on the binary timescale, and GR precession, standard methods of determining the rates of EMRIs, or other similar phenomena with black holes, will not be accurate and a new formalism will need to be added or developed.

Because EMRI waveforms can be used to extract information about the secondary SMBH and gaseous disks present [85, 87], and therefore plausibly validate the existence of the secondary and the significant amounts of gas which should be present due to the highly increased rates of tidal disruptions, it is even more interesting that reasonable numbers of EMRIs can be produced from SMBH binary systems. This ability to produce EMRIs and then validate them along with their implied environments may make these systems particularly astrophysically rich.

Chapter 5

Slightly Two- or Three-Dimensional Self-Similar Solutions

N.B.: This work will be published with the following authors: Re'em Sari, unknown author, and Nate Bode in approximately this order.

Abstract

Self similarity allows for analytic or semi-analytic solutions to many hydrodynamics problems. Most of these solutions are one dimensional. Using linear perturbation theory, expanded around such a one-dimensional solution, we find self-similar hydrodynamic solutions that are two- or three-dimensional. Since the deviation from a one-dimensional solution is small, we call these slightly two-dimensional and slightly three-dimensional self-similar solutions, respectively. As an example, we treat strong spherical explosions of the second type. A strong explosion propagates into an ideal gas with negligible temperature and density profile of the form $\rho(r,\theta,\phi) = r^{-\omega}[1+\sigma F(\theta,\phi)]$, with $\omega > 3$, $\sigma \ll 1$. Analytical solutions are obtained by expanding the arbitrary function $F(\theta,\phi)$ in spherical harmonics.

5.1 Introduction

Astrophysics supplies ample examples of hydrodynamic problems that admit self-similar solutions. In supernovae explosions [121, 122] a shock wave is created by the release of an immense amount of energy during a short time in the center of an exploding star. When the shock wave propagates into the surrounding medium, the hydrodynamics is described by the the Sedov-Taylor solutions [123–126]. Gamma ray bursts, provide a relativistic analog of that [127–130]. If the external medium is

spherical, these are one-dimensional solutions. However, if the external density has angular dependence, it will cause the shape of the shock, and the flow behind it, to deviate from sphericity.

An inherently two-dimensional version of this problem is the explosion in half space. Here, space is assumed to be empty on one side of a plane, while the other side is filled with an ideal gas with constant density. A large amount of energy is then released at a point on the surface. This describes the propagation of shockwaves in the process of cratering caused by large impacts on a planetary surface. Qualitatively, this problem and its self-similar nature was described by [131], but a two-dimensional self-similar solution was not developed there.

Here, we obtain two-dimensional and three-dimensional self-similar solutions that deviate only slightly from some known one-dimensional solution. We show that when treating such solutions as perturbations, the analysis is analogous to the treatment of stability [132–136]. We call these solutions slightly 2D or slightly 3D self-similar solutions. As a working example, we analyze deviation from sphericity of the strong explosion problem with density falling as a power law of distance $\rho \propto r^{-\omega}$, $\omega > 3$. These are known to be self-similar solutions of the second type [126].

In Sec. 5.2 we briefly review the main features of the one-dimensional solution which serves as the unperturbed solution for our analysis. In Sec. 5.3, we discuss the perturbation formalism for this problem that allows the finding of slightly two-dimensional and slightly three-dimensional self-similar solutions.

5.2 The One-Dimensional Self-Similar Solution

Here we summarize the formalism leading to the one-dimensional self-similar solution [126]. The discussion here is taken from [135]. Consider the Strong Explosion Problem in which a large amount of energy is released at the center of a sphere of ideal gas with a density profile decreasing with the distance from the origin according to $\rho = Kr^{-\omega}$, forming a strong outgoing shock wave.

This problem was first investigated by Sedov (1946), Von-Neumann (1947), and Taylor (1950), who found the solutions for $\omega < 5$, known as the Sedov-Taylor solutions. [126] showed that the Sedov-Taylor solutions are valid only for $\omega < 3$, where the solutions are known as self-similar solutions of Type-I, and contain decelerating shock waves. New, Type-II, self-similar solutions for almost all the range $\omega > 3$ containing accelerating shock waves were constructed.

Here we briefly summarize the second type solutions for $\omega > 3$. The hydrodynamic equations for

an ideal gas with adiabatic index γ in spherical symmetry are given by:

$$\begin{aligned}(\partial_t + u\partial_r)\rho + \rho r^{-2}\partial_r(r^2 u) = 0 \ , \\ \rho(\partial_t + u\partial_r)u + \partial_r(\rho c^2/\gamma) = 0 \ , \\ (\partial_t + u\partial_r)(c^2\rho^{1-\gamma}/\gamma) = 0 \ ,\end{aligned} \tag{5.1}$$

where the dependent variables u, c, and ρ are the fluid velocity, sound velocity, and density, respectively. We now seek a self-similar solution to the hydrodynamic Eqn. 5.1 of the form:

$$\begin{aligned}u(r,t) = \dot{R}\xi U(\xi) \ , \ c(r,t) = \dot{R}\xi C(\xi) \ , \\ \rho(r,t) = BR^{\epsilon}G(\xi) \ , \ p(r,t) = BR^{\epsilon}\dot{R}^2 P(\xi) \ ,\end{aligned} \tag{5.2}$$

where $\xi = r/R(t)$ is the dimensionless spatial coordinate, and the length scale $R(t)$ (frequently abbreviated as simply R) is the shock radius and satisfies [126, 131]

$$\frac{\ddot{R}R}{\dot{R}^2} = \delta \quad \Rightarrow \quad \dot{R} \propto R^{\delta} \ . \tag{5.3}$$

The quantities G, C, U, and P, which are defined by Eqns. 5.2, give the spatial dependence of the hydrodynamic quantities. The diverging (exploding) solutions of Eqn. 5.3 are

$$R(t) = \begin{cases} A(t-t_0)^{\alpha}, & \delta < 1 \\ Ae^{t/\tau}, & \delta = 1 \\ A(t_0-t)^{\alpha}, & \delta > 1 \end{cases} \tag{5.4}$$

where $\alpha = 1/(1-\delta)$.

Solutions with $\delta < 1$ diverge in infinite time, and t_0 represents the time of the point explosion, which is usually taken to be $t_0 = 0$. For $\delta < 0$ the shock wave decelerates and for $0 < \delta < 1$ it accelerates. For $\delta > 1$ the shock wave accelerates so fast that it diverges in a finite time. In this case, t_0 represents the time of divergence rather than the explosion time. The transition between finite and infinite divergence occurs at $\delta = 1$ where we have exponential time dependance [137].

Substituting Eqn. 5.2 into the hydrodynamic equations (Eqns. 5.1) and using Eqn. 5.3, one gets regular differential equations for the similarity quantities U, C, and G (see for example Landau &

Lifshitz) with two free constants, the similarity parameters ϵ and δ:

$$\frac{dU}{d\log\xi} = \frac{\Delta_1(U,C)}{\Delta(U,C)} \ , \ \frac{dC}{d\log\xi} = \frac{\Delta_2(U,C)}{\Delta(U,C)} \tag{5.5}$$

and an explicit expression for the density G:

$$C^{-2}(1-U)^{\lambda} G^{\gamma-1+\lambda}\xi^{3\lambda-2} = \text{const} \ . \tag{5.6}$$

The functions Δ, Δ_1, and Δ_2 are given by:

$$\begin{aligned} \Delta &= C^2 - (1-U)^2 \ , \\ \Delta_1 &= U(1-U)(1-U-\delta) - 3UC^2 - 3C^2(\epsilon+2\delta)/\gamma \ , \\ \Delta_2 &= C(1-U)(1-U-\delta) - (\gamma-1)CU(1-U+\delta/2) - \\ &\quad - C^3 + \frac{2\delta-(\gamma-1)\epsilon}{2\gamma}\frac{C^3}{1-U} \ , \end{aligned} \tag{5.7}$$

and the parameter λ is

$$\lambda = \frac{2\delta-(\gamma-1)\epsilon}{3+\epsilon} \ . \tag{5.8}$$

The similarity parameter ϵ can be found from the boundary conditions at the strong shock, the Hugoniot jump conditions [138]. From these relations applied to strong shock one gets $\epsilon = -\omega$, and also

$$U(1) = \frac{2}{\gamma+1} \ , \ C(1) = \frac{\sqrt{2\gamma(\gamma-1)}}{\gamma+1} \ , \ G(1) = \frac{\gamma+1}{\gamma-1} \ . \tag{5.9}$$

The boundary conditions on the shock do not state any limits on the possible values of the similarity parameter δ. In order to determine the value of this parameter one should distinguish two kinds of similarity flows: Type-I and Type-II, defined first by Zel'dovich [131]. A solution of Type-I describes the flow in all space and therefore conservation laws must be obeyed by the self-similar solution. One can then deduce $\delta = (\omega-3)/2$, which gives the well-known Sedov-Taylor solutions. However, for $\omega > 3$ it is easy to see that the solution obtained with this value of δ contains an infinite amount of energy and therefore can not describe the flow over the whole space. Therefore, the flow must be of Type-II.

In Type-II solutions, there is a region, whose scale relative to the flow characteristic length $R(t)$ goes to zero with time, in which the similarity solution does not describe the physical system. Therefore, for this kind of solution the energy does not have to be conserved in the self-similar solution since this solution does not describe the whole flow. In order that the region which is not self-similar (located around the origin) does not influence the self-similar solution, the solution must

pass through the singular point defined by [126, 131]:

$$U + C = 1 \ . \tag{5.10}$$

From this singular point requirement, the dependence of δ upon the parameters ω and γ can be found. It was found [126] that for $\omega > \omega_g(\gamma) > 3$ there is a value of δ for which the solution passes through a singular point, and therefore a second type self-similar solution exists.

A fully analytic solution to Eqns. 5.5–5.8 exists for the case where $\omega = \omega_a(\gamma) \equiv 2(4\gamma-1)/(\gamma+1)$:

$$\begin{aligned} C(\xi) = \frac{\sqrt{2\gamma(\gamma-1)}}{\gamma+1}\xi^3 \ , \ U(\xi) = \frac{2}{\gamma+1} \ , \\ G(\xi) = \frac{\gamma+1}{\gamma-1}\xi^{-8} \ , \ P(\xi) = \frac{2}{\gamma+1} \ . \end{aligned} \tag{5.11}$$

For this analytical case the parameter δ is given by $\delta = (\gamma-1)/(\gamma+1)$.

5.3 Slightly Two- and Three-Dimensional Self-Similar Solutions

We shall use here the Eulerian perturbation approach. We define the perturbed quantities as the difference between the perturbed solution (i.e., the slightly two-dimensional self-similar solution) and the unperturbed one-dimensional solution at the same spatial point. The derivation of the perturbation equation is similar to the one given by [132], [134], and [135]. The perturbed hydrodynamic quantities are defined as

$$\begin{aligned} \delta\vec{v}(r,\theta,\phi,t) = \vec{v}(r,\theta,\phi,t) - v_0(r,t)\hat{r} \ , \\ \delta\rho(r,\theta,\phi,t) = \rho(r,\theta,\phi,t) - \rho_0(r,t) \ , \\ \delta p(r,\theta,\phi,t) = p(r,\theta,\phi,t) - p_0(r,t) \ , \end{aligned} \tag{5.12}$$

where $\vec{v}$, p, and ρ are the velocity, pressure, and density in the perturbed solution, while $v_0\hat{r}$, p_0, and ρ_0 are the same quantities as in the unperturbed solution. We consider perturbations that can be written in a separation of variables form [139]:

$$\begin{aligned} \delta\vec{v}(r,\theta,\phi,t) = \xi\dot{R}\left[\delta U_r(\xi)Y_{lm}(\theta,\phi)\hat{r} + \delta U_T(\xi)\nabla_T Y_{lm}(\theta,\phi)\right] f \ , \\ \delta\rho(r,\theta,\phi,t) = BR^{\epsilon}\delta G(\xi)Y_{lm}(\theta,\phi)f \ , \\ \delta p(r,\theta,\phi,t) = BR^{\epsilon}\dot{R}^2\delta P(\xi)Y_{lm}(\theta,\phi)f \ , \end{aligned} \tag{5.13}$$

where

$$\nabla_T \equiv \hat{\theta}\frac{\partial}{\partial\theta} + \hat{\phi}\frac{1}{\sin\theta}\,\frac{\partial}{\partial\phi} \tag{5.14}$$

are the tangential components of the gradient and $R(t)$ is the unperturbed shock radius which still satisfies Eqn. 5.3. The perturbed shock radius, $R(t,\theta,\phi)$, is given by

$$R(t,\theta,\phi) - R(t) \equiv \delta R(t,\theta,\phi) = Y_{l,m}(\theta,\phi)R(t)f\ . \tag{5.15}$$

Eqns. 5.13 and 5.15 define the quantities δU_r, δU_T, δP, δG, and f. The quantity f measures the fractional amplitude of the perturbation to the shock wave radius. Here we deviate from the standard treatment of stability. There, f is a function of time: if the function f increases with time then the solution is unstable, while if f decreases with time then the solution is stable. However, here, since we demand that the perturbed solution be self-similar, f has to be independent of time.

We linearize the hydrodynamic equation around the unperturbed self-similar solution to get a linear set of equations:

$$MY' = NY \tag{5.16}$$

where

$$Y = \begin{pmatrix} \delta G \\ \delta U_R \\ \delta U_T \\ \delta P \end{pmatrix}\ ,$$

$$M\ =\ \begin{pmatrix} \xi(U\text{–}1) & G\xi & 0 & 0 \\ 0 & (U\text{–}1)\xi^2 G & 0 & 1 \\ 0 & 0 & (U\text{–}1)\xi^2 G & 0 \\ -\dfrac{\gamma\xi(U\text{–}1)}{G} & 0 & 0 & \dfrac{\xi(U\text{–}1)}{P} \end{pmatrix}\ ,$$

$$N = \begin{pmatrix} \omega\text{–}3U\text{–}\xi U' & -\xi G'\text{–}3G & l(l+1)G & 0 \\ P'G^{-1} & (1\text{–}\delta\text{–}2U\text{–}\xi U')G\xi & 0 & 0 \\ 0 & 0 & (1\text{–}\delta\text{–}U)G\xi & -\xi^{-1} \\ -\dfrac{\xi\gamma(U\text{–}1)G'}{G^2} & -\xi\left(\dfrac{P'}{P} - \gamma\dfrac{G'}{G}\right) & 0 & \xi(U\text{–}1)\dfrac{P'}{P^2} \end{pmatrix}\ ,$$

and G, U, and P are defined by Eqn. 5.2.

Unlike the perturbation equations for stability, the equations above do not contain an unknown parameter. They are, in fact, a special case of the equations used in [135], but with the perturbation growth rate set to $q = 0$. In that sense they are similar to the equations of [140] for discretely self-similar solutions.

On the shock front, the linearized Hugoniot jump conditions are expressed as:

$$\begin{aligned} \delta G(1) &= \frac{\gamma+1}{\gamma-1}(d-\omega) - G' \quad , \quad \delta U_r(1) = -U' \\ \delta U_T(1) &= -\frac{2}{\gamma+1} \quad , \quad \delta P(1) = \frac{2}{\gamma+1}\,(2+d-\omega) - P' \ . \end{aligned} \tag{5.17}$$

Comparing with the stability treatment, two differences occur here. First, the parameter q which measures the perturbation growth rate in the stability analysis is set to zero. Instead, a new parameter d appears. It is the ratio of the amplitude of the external density perturbation σ to the amplitude of the perturbations in the shock: $d \equiv \sigma/f$.

For any value of the parameter d one can integrate Eqn. 5.16 beginning at the shock front using the shock boundary conditions. However, the singular point of the unperturbed solution, ξ_c, where $C+U=1$, is also a singular point of the perturbed solution. Therefore, in general, such integration will diverge at the ξ_c. Only for specific values of the parameter d, where an additional boundary condition at the singular point is satisfied, is the solution regular. These are the physical values for the parameter d.

Technically, solving these equations is easier than the equivalent perturbation case. The reason is that the unknown parameter d appears only in the shock boundary condition, and is absent from the differential equations. We can therefore solve these equations starting from the sonic point outward, and find the three independent solutions that are nonsingular at ξ_c. Then we can find a linear combination of these three solutions, and the value of d that can solve the four boundary conditions at the shock.

5.4 Results

For convenience we investigate the case $\gamma = 5/3$, $\omega = 17/4$, where the unperturbed solution is analytic. For $l = 1$ we obtain $d = -11.2$. This means that the fractional amplitude of perturbations in the shock wave position, f, are an order of magnitude smaller than the fractional amplitude of perturbations in the external density σ. The negative sign implies that at angles where the external density is higher, the shock wave position is retarded. This is expected intuitively. From the shock boundary conditions, we infer that the pressure at these angles is also lower. For $l = 2$ we find $d = -11.6$, and for $l = 3$ we find $d = -12.1$. A plot of d as function of l is given in Fig. 5.1.

5.5 Extension to Arbitrary Angular Dependence

The analysis above was limited to external density perturbations whose angular dependence was a spherical harmonic. This was necessary in order to obtain separation between the angular and

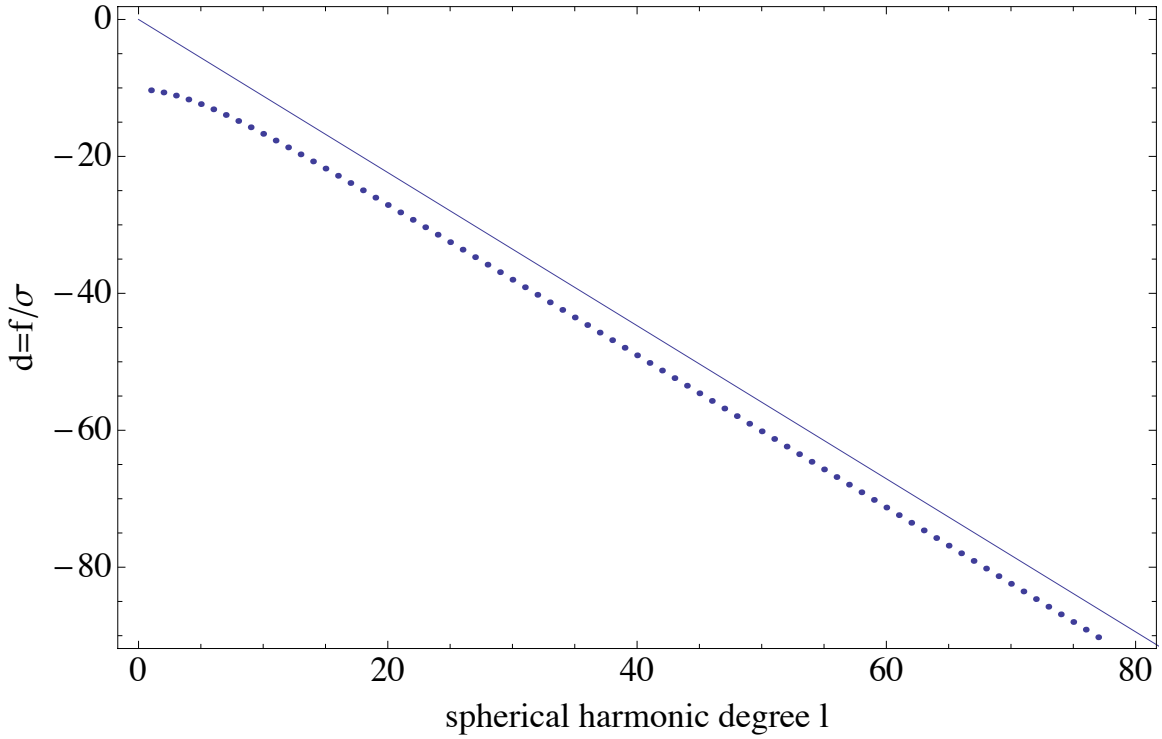


Figure 5.1: Dots show d as function of l as obtained by solving the differential equations for $\gamma = 5/3$ and $\omega = 17/4$. For small l we have $d \cong -10.$, while for large l, i.e., short wavelength, we obtain a linear relation: $d = -\sqrt{5/4}l$ (thin continuous line).

radial dependencies. However, since we are dealing with linear perturbations, any arbitrary angular dependence can be expanded into a sum of spherical harmonics, each of which could be solved in the method described in the previous section. Then, the solutions can be summed, leading to the perturbation solution for external density perturbations with arbitrary angular dependence.

As an example, we consider the following problem: A strong point-like explosion is launched into a surrounding which has a density on one side of a plane slightly different from the density on the other side of the plane. In our notation this is $\rho \propto r^{-17/4}(1 + \sigma H(\theta))$ where $H(\theta) = 1$ for $\theta < \pi/2$ and $H(\theta) = -1$ for $\theta > \pi/2$. The point explosion in half space could be thought of as an extreme version of this density profile with $\sigma = 1$. However, our solution formalism applies only for slightly two-dimensional cases where $\sigma \ll 1$.

Such a density profile could be expanded in spherical harmonics as

$$H(\theta) = \sum_{n=0}^{\infty} \frac{\pi\sqrt{4n+3}}{\Gamma(1/2-n)\Gamma(2+n)} Y_{2n+1,0}(\theta, 0) \ . \tag{5.18}$$

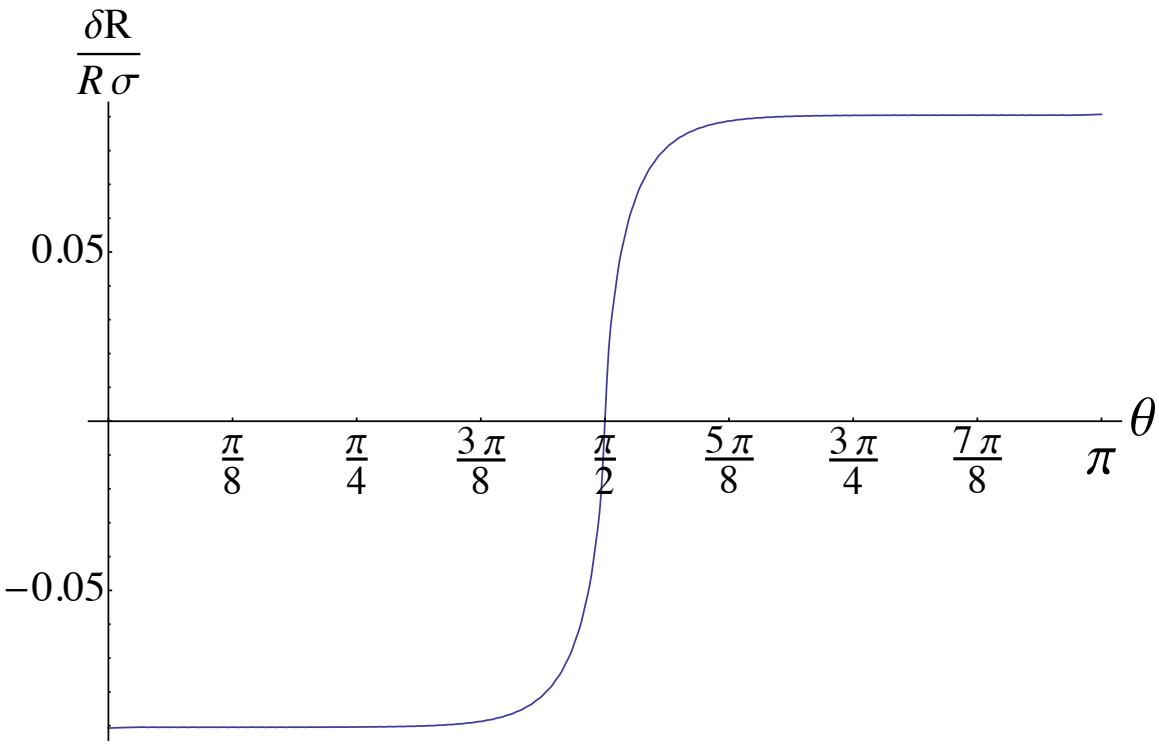


Figure 5.2: The fractional deviation of the shock position as function of θ for the heavyside density distribution. It is seen that the shock is composed of almost two hemispheres, connected smoothly over a short angular scale of about 0.2 radians.

The shape of the shock, $R + \delta R(\theta)$, deviates from its unperturbed value R by

$$\frac{\delta R(\theta)}{R} = \sigma \sum_{n=0}^{\infty} \frac{\pi\sqrt{4n+3}}{d(2n+1)\Gamma(1/2-n)\Gamma(2+n)} Y_{2n+1,0}(\theta, 0) \ . \tag{5.19}$$

This shape is plotted in Fig. 5.2 where the sum was taken from $n = 0$ to $n = 100$.

5.6 Short Wavelength Limit

Because the flow does not vary in the short wavelength limit, we may treat the matrices M and N as constants close to the shock front. By using the unperturbed values of the state variable at the shock, we find the four independent modes of the problem:

$$\lambda = -8, 3, \pm\sqrt{\frac{2\gamma}{\gamma+1}} l \, . \tag{5.20}$$

The first two are independent of l and indicate that the state functions, close to the shock, vary on the scale R, regardless of the wavelength of the perturbation. However the other two are linear in l meaning that close to the shock the state functions vary over small scales of order $\pm R/l$. Therefore,

for large l, the positive mode is growing inward very rapidly, and thus can not exist physically. For this reason we demand that the perturbation has no component along this mode on the shock front by requiring it to be written as a linear combination of the three eigenvectors associated with the other modes. This provides the extra boundary condition at the shock that allows us to determine d. Performing this calculation we find that for general ω and γ

$$d = -\sqrt{\frac{2\gamma}{\gamma+1}}l \tag{5.21}$$

in the limit $l \gg 1$.

5.7 Discussion

We have considered the problem of a strong shock propagating into a slightly aspherical medium made up of a density with a spherically symmetric radial power-law plus a perturbation of arbitrary angular dependence, and solved for the Type-II self-similar solution. Such an external medium has a density profile $\rho(r,\theta,\phi) = r^{-\omega}[1+\sigma F(\theta,\phi)]$, where $\omega > 3$, $\sigma \ll 1$, and F is an arbitrary function of θ and ϕ. Because the perturbations are small the hydrodynamic equations can be linearized around the unperturbed solution, which allows us to expand F as a series in spherical harmonics, and solve the problem term by term. In this way the general problem is reduced to one which includes only perturbations $F(\theta,\phi) \propto Y_{lm}(\theta,\phi)$.

The linearized self-similar equations are presented for this simpler case, $F(\theta,\phi) \propto Y_{lm}(\theta,\phi)$, along with the appropriate boundary conditions. There is a unique solution to these equations which depends on a single parameter d, which is determined by physical considerations only. That the only dependence on d is in the boundary conditions makes these equations particularly easy to solve.

We demonstrate this process on a specific example which deviates from spherical symmetry by a weak step function in the outside density across a plane containing the initial explosion. As expected, instead of the shock being spherical, it is composed of two hemispheres smoothly connected across the plane of the discontinuity.

5.8 Acknowledgments

We thank Yonatan Oren for helpful discussions.

Appendix A

Idealized Kozai Inspiral Time

We occupy ourselves with determining the inspiral time of a test particle orbiting the primary SMBH, while undergoing Kozai oscillations due to a secondary SMBH orbiting farther out, under the following assumptions:

1. We do not include SP precession.
2. We do not include GR precession.
3. We consider the Kozai formalism without consideration of the oscillations of the eccentricity on the timescale of the secondary [107].
4. We assume a librating solution where we can analytically calculate the maximum eccentricity in terms of the minimum inclination (Eqn. 4.20). Rotating Kozai orbits will always reach even higher eccentricities (see Fig. 4.2).

Given these assumptions it is straightforward to calculate a polite lower bound for the speed with which a CR undergoing the Kozai mechanism will inspiral to the SMBH. Ultimately, this is compared to both an object on a circular orbit with the same energy (and no secondary present) and an object on an eccentric Keplerian orbit (and no secondary present).

A.0.1 Analytic Calculation

There are three different phases and timescales to this inspiral: 1) orbital energy radiated away on an orbital timescale, 2) changes in eccentricity (but with constant semi-major axis) on a Kozai timescale that approaches $e \lesssim 1$, and 3) the inspiral timescale which will be many Kozai timescales. This allows us to calculate the inspiral time in a fairly straightforward fashion.

Orbital Timescale: During each orbit of the primary a CR radiates some of its orbital energy and angular momentum in gravitational waves. On the timescale of an orbit the eccentricity and semi-major axis of the orbit can vary only modestly if $t_{\rm orbit} \ll T_{\rm Kozai}$ and if the periapsis of the CR

is $\gtrsim 5r_S$ where r_S is the Schwarzschild radius. To find the energy lost in gravitational waves per orbit we use Eqn. 4.8 [99] which for simplicity we rewrite as

$$\left\langle \frac{dE}{dt} \right\rangle = -\frac{\beta}{2} a^{-5} f(e)\,. \tag{A.1}$$

Here e is an implicit function of t, since e is slowly evolving with time, β is a constant and the same as in Eqn. 4.8, and

$$f(e) = \frac{1 + \frac{73}{24}e^2 + \frac{37}{96}e^4}{(1-e^2)^{7/2}}\,. \tag{A.2}$$

So, to find the total energy radiated during one Kozai period we would integrate Eqn. A.1 over time for a full Kozai oscillation.

Kozai Timescale: The Kozai mechanism does not change the energy of the particle, and so during a Kozai cycle the particle's semi-major axis does not change. This allows us to write the integral as

$$(\Delta E)_{\rm Kozai} = \int_0^{T_{\rm Kozai}} \left\langle \frac{dE(e(t))}{dt} \right\rangle dt\,. \tag{A.3}$$

However, the energy radiated will be dominated by the orbits with highest eccentricity (see Fig. A.1). Thus, it is imperitive to know the maximum eccentricity $e_{\rm max}$ well. For inclinations greater than $i_{\rm min} = 39.2^{\rm o}$ the maximum eccentricity is given by Eqn. 4.20.

Knowing that we have an accurate expression for the maximum eccentricity, we now turn to computing the integral of Eqn. A.3.

For the orbits we're interested in (those with a small periapsis distance, but a large semi-major axis) $e \lesssim 1$. Moreover, because of the strong dependence in the numerator of $f(e)$ on e, it is only orbits with an eccentricity close to $e_{\rm max}$ which will have an effect on the integral of Eqn. A.3.

So, we write $e = 1 - \bar{\epsilon}$ where $\bar{\epsilon} \ll 1$. Plugging this in to Eqn. A.1 and expanding to highest order in $1/\bar{\epsilon}$ yields:

$$\left\langle \frac{dE}{dt} \right\rangle_{\text{near max}} = -\frac{\beta}{2} a^{-5} \left(\frac{425}{96 \cdot 2^{7/2}} \bar{\epsilon}^{-7/2} \right)\,. \tag{A.4}$$

The strong dependence on $\bar{\epsilon}$ means that an orbit with $\Delta\bar{\epsilon} \equiv \bar{\epsilon} - \bar{\epsilon}_{\rm min} = \bar{\epsilon}$, the energy radiated is already an order of magnitude smaller than the orbit with the maximum energy radiation.

By using an analytic approximation to $e(t)$ we now find an approximation for $\bar{\epsilon}$ near $e_{\rm max}$, which in turn lets us solve the integral of interest.

The differential equation which dictates the path of the eccentricity is given by [107]

$$\begin{aligned} T_{\rm Kozai} e \frac{de}{dt} \;=\; \pm \big[\big\{ \big[\big(3 + \kappa + Q - 5\Lambda^2\big)\, e^2 - (3+\kappa)e^4 - Q \big] \\ \times \big[Q + (2-\kappa)e^2 \big] \big\} \big]^{1/2}\,, \end{aligned} \tag{A.5}$$

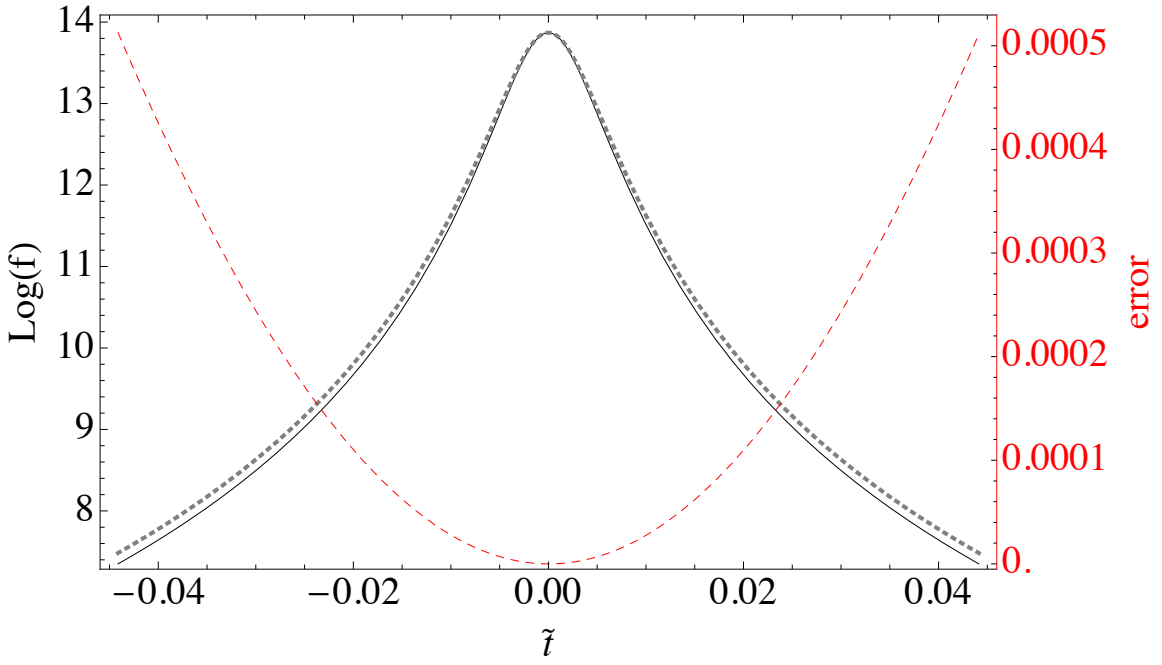


Figure A.1: On the left axis we plot $log[f(e)]$ as a function of t for $e_{\rm min} = 0.5$, $i_{\rm min} = 1.56$ (and therefore an $e_{\rm max} = 0.9999$) for the numerical solution to Eqn. A.5 (solid black). The function f peaks strongly when e approaches $e_{\rm max}$, meaning that to a good approximation we can expand the approximate analytic solution to Eqn. A.5 (Eqn. A.7) in $\bar{t} = t/T_{\rm Kozai}$ before integrating (see text). This approximate function of $f(e)$ is plotted in dotted gray. On the right axis we plot the relative error in this approximation (dashed red).

where

$$\Lambda = \sqrt{1-e^2}\cos(i) \text{ and } Q = e^2\left[5\sin^2(i)\sin^2(\omega) + \kappa - 2\right] \tag{A.6}$$

are constants of the motion, ω is the argument of pericenter, and κ is the apsidal precession over a Kozai oscillation due to the stellar potential. For simplicity we set $\kappa = 0$.

For the inclinations we are considering the solution to this equation is reasonably well approximated by a cos (see Fig. A.2):

$$e(\bar{t}) \approx \frac{1}{2}\left[(e_{\rm max} - e_{\rm min})\cos\bar{t} + (e_{\rm max} + e_{\rm min})\right], \tag{A.7}$$

where $\bar{t} \equiv t/T$, $e_{\rm min}$ is the minimum eccentricity,

$$T = \frac{\pi}{\left(\frac{2}{3}\right)^{5/2}\log\left[\sqrt{\frac{3}{5}}\frac{4}{e_{\rm min}}\right]}, \tag{A.8}$$

and we have assumed that $e(0) = e_{\rm max}$. Using this approximation near the peak we can expand in

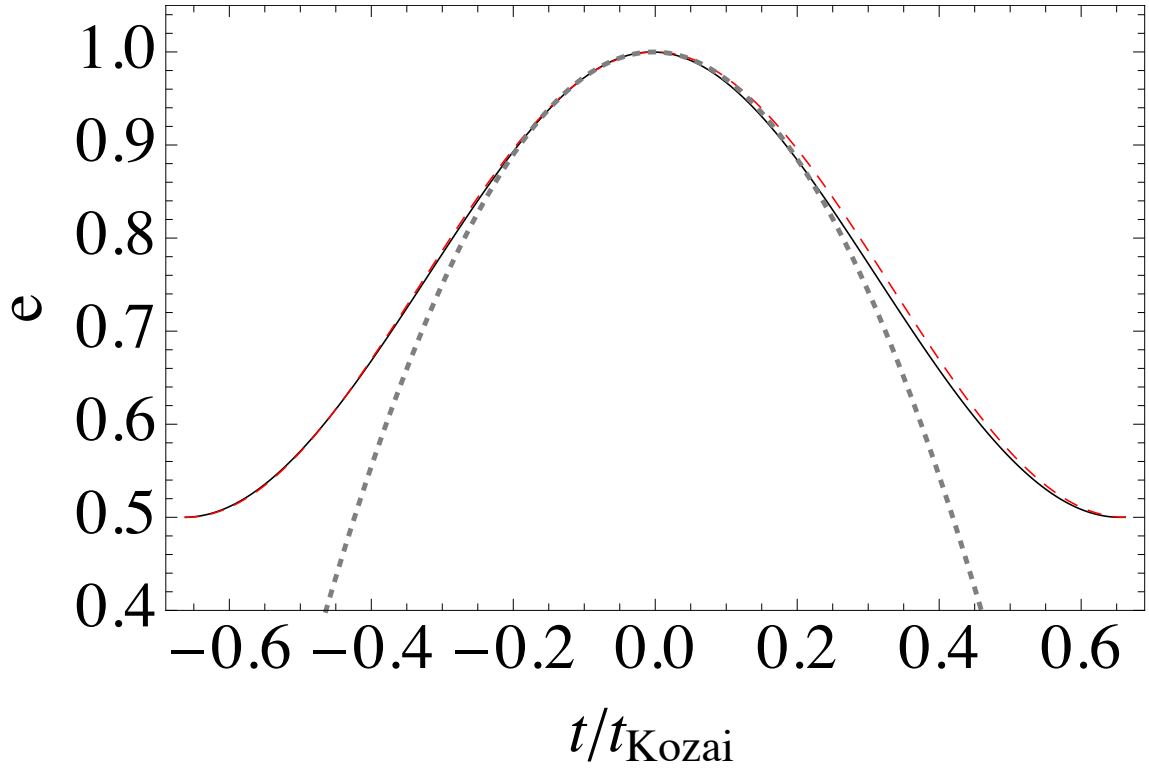


Figure A.2: We plot the eccentricity as a function of time in units of $T_{\rm Kozai}$ for both a numerical solution (black solid) and the analytic solution (red dashed). The region of interest is that where $e_{\rm max} - e(t) = \bar{\epsilon} = 1 - e_{\rm max}$. See text.

$\bar{t}$:

$$e(\bar{t}) = e_{\rm max} - \frac{e_{\rm max} - e_{\rm min}}{4}\bar{t}^2 + O\left(\bar{t}^4\right) , \tag{A.9}$$

so that

$$\bar{\epsilon}(\bar{t}) = 1 - e(\bar{t}) + O(\bar{t}^4) \tag{A.10}$$

$$= \epsilon_{\rm min} + \frac{e_{\rm max} - e_{\rm min}}{4}\bar{t}^2 + O(\bar{t}^4) . \tag{A.11}$$

Plugging this into Eqn. A.4, and that into Eqn. A.3, we find that the energy radiated per unit time is approximately

$$\dot{E} = -\frac{16}{15}\frac{\beta}{a^5}\frac{1}{\sqrt{e_{\rm max} - e_{\rm min}}\epsilon_{\rm min}^3} \tag{A.12}$$

$$= -\frac{16}{15}\frac{\beta}{a^5}\frac{1}{\sqrt{e_{\rm max} - e_{\rm min}}(1 - e_{\rm max})^3} . \tag{A.13}$$

Mathematically, this approximate solution to the integral of Eqn. A.3, produces only modest errors of several percent, while elucidating the primary dependencies. However, this approximation also

tacitly assumes that there are many orbits across the peak of the integrand of Eqn. A.3, in that we are assuming that the maximum eccentricity is in fact reached *while at periapsis*. This is not necessarily true and we discuss it further below.

Inspiral Timescale: $\dot{E}$ still retains a strong time dependence over the inspiral timescale, because as the orbit radiates energy in GWs its semi-major axis decreases. However, if we suppose that the minimum periapsis over each Kozai cycle, $\bar{r}_p$, roughly stays constant, then we may rewrite $\dot{E}$ as

$$\dot{E} = -\frac{16}{15}\frac{\beta}{\bar{r}_p^3}\frac{1}{\sqrt{e_{\max}-e_{\min}}}\frac{1}{a^2}\,. \tag{A.14}$$

In fact, both $e_{\max}$ and $e_{\min}$ must also evolve with time as the orbit circularizes, and as they approach the same value, $\dot{E}$ will grow oppressively small. However, as examination of Fig. 4.1 elucidates, once the CR has reached this point, gravitational wave radiation alone (without the Kozai mechanism to pump the eccentricity) will drive the CR to coalescence on a much shorter timescale than that required to deliver the CR to this orbit region. Thus, we may consider the factor containing $e_{\max}$ and $e_{\min}$ as being roughly ~ 1.

Finally, we may relate $\dot{a}$ to $\dot{E}$

$$\dot{a} = \frac{2a^2}{GMm_\star}\dot{E} \tag{A.15}$$

$$= -\frac{32}{15}\frac{\beta}{GMm_\star\bar{r}_p^3}\frac{1}{\sqrt{e_{\max}-e_{\min}}}\,, \tag{A.16}$$

which lets us ultimately determine the merger time:

$$t_{\rm merger} = \frac{15}{32}\frac{GMm_\star\bar{r}_p^3}{\beta}\sqrt{e_{\max}-e_{\min}}a_0 \tag{A.17}$$

$$= \frac{15}{32}\frac{GMm_\star}{\beta}\sqrt{e_{\max}-e_{\min}}(1-e_{\max})^3a_0^4 \tag{A.18}$$

$$= \frac{45}{2^{11}}\frac{c^5}{G^3}\frac{a_0^4\sqrt{e_{\max}-e_{\min}}}{Mm_\star(M+m_\star)}(1-e_{\max})^3\,. \tag{A.19}$$

The inspiral time for a quasicircular orbit is just

$$t_{\rm merger}^{\rm (circ)} = \frac{5}{256}\frac{c^5}{G^3}\frac{a^4}{Mm_\star(M+m_\star)}\,. \tag{A.20}$$

Thus we see that the Kozai mechanism increases the inspiral mechanism by up to

$$\frac{t_{\rm merger}^{\rm (Koz)}}{t_{\rm merger}^{\rm (circ)}} = \frac{9}{8}\sqrt{e_{\max}-e_{\min}}(1-e_{\max})^3\,, \tag{A.21}$$

when compared to a circular orbit of equal energy.

On the other hand, when compared to an eccentric Keplerian orbit whose eccentricity is the

maximum eccentricity of the star undergoing the Kozai oscillation and which has the same energy, the ratio of inspiral times is:

$$\frac{t_{\rm merger}^{\rm (Koz)}}{t_{\rm merger}^{\rm (e=e_{max})}} = \frac{t_{\rm merger}}{\frac{768}{425}(1-e_{\rm max}^2)^{7/2} t_{\rm merger}^{\rm (circ)}} \tag{A.22}$$

$$= \frac{1275}{2^{14}\sqrt{2}} \frac{\sqrt{e_{\rm max}-e_{\rm min}}}{\sqrt{1-e_{\rm max}}} \tag{A.23}$$

$$\sim \frac{10^{-2}}{\sqrt{1-e_{\rm max}}}\,. \tag{A.24}$$

Thus, in line with expectations, this upper bound on the merger time due to Kozai oscillations can be significantly slower than a Keplerian orbit of the same maximum eccentricity, while many times faster than a circular orbit.

A.0.2 Caveats

There are several important caveats to the above calculation which we clarify.

Apsidal Precession: The first is that we have ignored the SP and GR in this calculation, both of which cause the orbits of the CRs to not be Keplerian orbits. This is discussed in detail in Secs. 4.3.2–4.3.3 and Sec. 4.4.

Eccentric Oscillations on Binary Timescale: As discussed in Sec. 4.3.2, there are oscillations in the eccentricity on the timescale of the secondary's orbital period on top of the standard Kozai oscillation in the eccentricity. Though the magnitude in these oscillations is small, when the eccentricity becomes large and $1-e$ becomes comparable, these errors are significant. This is terribly important as one approaches $4\,GM/c^2$ and the energy lost per periapsis passage diverges, but far from $r_{\rm S}$ where the Peters equation (Eqn. 4.8) is most accurate, it will be important but not vital. This is also the case for the GRP (see Sec. 4.4).

Is $t_{orbit} \ll T_{Kozai}$?: As mentioned in the calculation of $t_{\rm merger}$ via the Kozai mechanism there is an implicit assumption that there are many orbits during the period of time that the orbit has an eccentricity near $e_{\rm max}$. This is not always true: the ratio of the orbital period to the Kozai period is roughly

$$\frac{1}{n_{\rm orbits}} = \bar{t}_{\rm orbit} \approx \frac{3\pi q}{2}\left(\frac{a\star}{a_\bullet}\right)^3 \sim 5q\bar{a}^3\,, \tag{A.25}$$

where $q \leq 1$ is the mass ratio of the SMBHs and $\bar{a} \equiv a_\star/a_\bullet$ and $\bar{t}_{\rm orbit} = t_{\rm orbit}/T_{\rm Kozai}$. The maximum plausible value of a for the Kozai mechanism to work is $a_\bullet/2$. In this work we consider a mass ratio of $q = 0.1$ so that there are at minimum some 16 orbits per Kozai period. However, $n_{\rm orbits}$ is a strong function of $\bar{a}$, so that an orbit with $\bar{a} = 8$ has $n_{\rm orbits} \approx 1028$.

The width of the peak in units of $T_{\rm Kozai}$ can also be determined. Defining the peak of $f(e)$ to be the region within a factor of 10 of its peak value, and using Eqns. A.3 and A.11, we find that the

width of the peak, in units of Kozai periods is:

$$\bar{t}_{\rm peak} = \frac{4}{\sqrt{e_{\rm max} - e_{\rm min}}}\sqrt{\epsilon_{\rm min}} \sim \sqrt{\epsilon_{\rm min}}\,. \tag{A.26}$$

To find the number of orbits per peak we take the ratio of Eqns. A.25 and A.26:

$$N_{\rm peak} = \frac{\sqrt{\epsilon_{\rm min}}}{5q\bar{a}^3}\,. \tag{A.27}$$

If we set the number of required orbits across the peak of $f(e)$ to be the somewhat arbitrary, but reasonable, value of 10, then we require

$$\bar{a} \leq \bar{a}_{\rm max} \equiv \frac{\epsilon_{\rm min}^{1/6}}{50q} = \frac{\left(1 - (5/3)\cos^2 i_{\rm min}\right)^{1/6}}{50q}\,. \tag{A.28}$$

For regions of the potential inspiral where the semi-major axes is greater than $\bar{a}_{\rm max}$ one must reduce $\dot{E}$ by a factor of order the ratio of the time spent near periapsis to the orbital period, a rough estimate for the probability that the CR will reach periapsis when the eccentricity is in the required bound. Roughly speaking this factor is

$$\frac{t_{\rm peri}}{t_{\rm orbit}} \sim \frac{3r_{\rm peri}}{c}\frac{\sqrt{GM}}{2\pi a_\star^{3/2}} = \frac{3}{2\sqrt{2}\pi}(1-e)\left(\frac{r_{\rm S}}{a_\star}\right)^{1/2}\,. \tag{A.29}$$

We see that the rate of inspiral is substantially reduced by both a factor of $1-e$ and a factor of $\sqrt{r_{\rm S}/a_\star}$.

While $\bar{a} \leq \bar{a}_{\rm max}$, $\dot{E}$ is independent of $T_{\rm Kozai}$. This is because the total number of orbits passing by the minimum periapsis remains constant; as the Kozai period changes so does $\bar{t}_{\rm peak}$, by an equal amount, and therefore so does the number of orbits close to the minimum periapsis.

Lower Bound on $a_\star$: However, once the Kozai period, or $t_{\rm merger}$, approaches the inspiral time of the secondary SMBH (likely $\lesssim 10^9$ yr) this EMRI formation mechanism clearly breaks. This puts a polite lower bound on $\bar{a}$:

$$\bar{a} > \left(\frac{4}{3q}\sqrt{\frac{a_\bullet^3}{GM}}\frac{1}{10^9\,{\rm yr}}\right)^{2/3}\,. \tag{A.30}$$

For $q = 0.1$, $a_\bullet = 0.01\,{\rm pc}$, and $M = 10^6 M_\odot$, the values relevant to the simulations found in this article, then

$$a_\star > 2\times 10^{-4}\,{\rm pc}\,. \tag{A.31}$$

Bibliography

[1] G. Nelemans, L. R. Yungelson, and S. F. P. Zwart, MNRAS **349**, 181 (2004), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2004MNRAS.349..181N&link_type=ABSTRACT.

[2] A. Stroeer and A. Vecchio, Classical and Quantum Gravity **23**, 809 (2006), http://arXiv.org/abs/astro-ph/0605227, URL http://adsabs.harvard.edu/abs/2006CQGra..23S.809S.

[3] C. Cutler and É. E. Flanagan, Phys. Rev. D **49**, 2658 (1994), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=1994PhRvD..49.2658C&link_type=ABSTRACT.

[4] P. Amaro-Seoane, J. R. Gair, M. Freitag, M. C. Miller, I. Mandel, C. J. Cutler, and S. Babak, arXiv **astro-ph** (2007), URL http://arxiv.org/abs/astro-ph/0703495v2.

[5] E. O'Connor and C. D. Ott, ApJ **730**, 70 (2011), http://arXiv.org/abs/1010.5550, URL http://adsabs.harvard.edu/abs/2011ApJ...730...70O.

[6] W. Zhang, S. E. Woosley, and A. Heger, ApJ **679**, 639 (2008), http://arXiv.org/abs/astro-ph/0701083, URL http://adsabs.harvard.edu/abs/2008ApJ...679..639Z.

[7] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, 1973), URL http://adsabs.harvard.edu/abs/1973grav.book.....M.

[8] S. Gillessen, F. Eisenhauer, S. Trippe, T. Alexander, R. Genzel, F. Martins, and T. Ott, ApJ **692**, 1075 (2009), URL http://iopscience.iop.org/0004-637X/692/2/1075.

[9] S. Gillessen, F. Eisenhauer, T. K. Fritz, H. Bartko, K. Dodds-Eden, O. Pfuhl, T. Ott, and R. Genzel, ApJL **707**, L114 (2009), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2009ApJ...707L.114G&link_type=ABSTRACT.

[10] A. M. Ghez, S. Salim, N. N. Weinberg, J. R. Lu, T. Do, J. K. Dunn, K. Matthews, M. R. Morris, S. Yelda, E. E. Becklin, et al., ApJ **689**, 1044 (2008), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2008ApJ...689.1044G&link_type=ABSTRACT.

[11] A. M. Ghez, S. Salim, S. D. Hornstein, A. Tanner, J. R. Lu, M. Morris, E. E. Becklin, and G. Duchêne, ApJ **620**, 744 (2005), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2005ApJ...620..744G&link_type=ABSTRACT.

[12] E. E. Salpeter, ApJ **140**, 796 (1964), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=1964ApJ...140..796S&link_type=GIF.

[13] J. H. Krolik, *Active galactic nuclei : from the central black hole to the galactic environment* (Princeton University Press, 1999), URL http://adsabs.harvard.edu/abs/1999agnc.book.....K.

[14] L. Ferrarese and H. Ford, Space Sci. Rev. **116**, 523 (2005), http://arXiv.org/abs/astro-ph/0411247, URL http://adsabs.harvard.edu/abs/2005SSRv..116..523F.

[15] K. Gültekin, D. O. Richstone, K. Gebhardt, T. R. Lauer, S. Tremaine, M. C. Aller, R. Bender, A. Dressler, S. M. Faber, A. V. Filippenko, et al., ApJ **698**, 198 (2009), URL http://adsabs.harvard.edu/abs/2009ApJ...698..198G.

[16] M. C. Aller and D. Richstone, AJ **124**, 3035 (2002), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2002AJ....124.3035A&link_type=ABSTRACT.

[17] N. Jarosik, C. L. Bennett, J. Dunkley, B. Gold, M. R. Greason, M. Halpern, R. S. Hill, G. Hinshaw, A. Kogut, E. Komatsu, et al., ApJS **192**, 14 (2011), URL http://adsabs.harvard.edu/abs/2011ApJS..192...14J.

[18] C. Wegg and J. N. Bode, arXiv (2010), URL http://arxiv.org/abs/1011.5874.

[19] R. N. Lang and S. A. Hughes, ApJ **677**, 1184 (2008), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2008ApJ...677.1184L&link_type=ABSTRACT.

[20] J. E. Pringle, MNRAS **248**, 754 (1991).

[21] A. I. MacFadyen and M. Milosavljević, arXiv (2006), http://arXiv.org/abs/astro-ph/0607467, URL http://adsabs.harvard.edu/abs/2006astro.ph..7467M.

[22] Y. T. Liu and S. L. Shapiro, arXiv **astro-ph.HE** (2010), URL http://arxiv.org/abs/1011.0002v1.

[23] S. M. O'Neill, M. C. Miller, T. Bogdanovic, C. S. Reynolds, and J. Schnittman, arXiv (2009), http://arXiv.org/abs/0812.4874, URL http://adsabs.harvard.edu/abs/2008arXiv0812.4874O.

[24] E. S. Phinney, astro2010: The Astronomy and Astrophysics Decadal Survey **2010**, 235 (2009), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2009astro2010S.235P&link_type=ABSTRACT.

[25] J. S. Bloom, D. E. Holz, S. A. Hughes, K. Menou, A. Adams, S. F. Anderson, A. Becker, G. C. Bower, N. Brandt, B. Cobb, et al., arXiv **0902**, 1527 (2009), URL http://arxiv.org/abs/0902.1527.

[26] P. Madau, T. Abel, P. Bender, T. D. Matteo, Z. Haiman, S. Hughes, A. Loeb, E. Phinney, J. Primack, T. Prince, et al., arXiv **astro-ph.CO** (2009), URL http://arxiv.org/abs/0903.0097v1.

[27] P. J. Armitage and P. Natarajan, ApJL **567**, L9 (2002), http://arXiv.org/abs/astro-ph/0201318, URL http://adsabs.harvard.edu/abs/2002ApJ...567L...9A.

[28] A. Loeb, Phys. Rev. Lett. **99**, 041103 (2007), http://arXiv.org/abs/astro-ph/0703722.

[29] B. Kocsis and A. Loeb, arXiv **803** (2008), http://arXiv.org/abs/0803.0003, URL http://adsabs.harvard.edu/abs/2008arXiv0803.0003K.

[30] P. Chang, L. E. Strubbe, K. Menou, and E. Quataert, MNRAS **407**, 2007 (2010), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2010MNRAS.407.2007C&link_type=ABSTRACT.

[31] N. Stone and A. Loeb, arXiv **astro-ph.CO** (2010), URL http://arxiv.org/abs/1004.4833v2.

[32] T. Bode, R. Haas, T. Bogdanović, P. Laguna, and D. Shoemaker, ApJ **715**, 1117 (2010), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2010ApJ...715.1117B&link_type=ABSTRACT.

[33] N. Seto and T. Muto, Phys. Rev. D **81**, 103004 (2010), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2010PhRvD..81j3004S&link_type=ABSTRACT.

[34] C. Palenzuela, T. Garrett, L. Lehner, and S. L. Liebling, Phys. Rev. D **gr-qc** (2010), URL http://arxiv.org/abs/1007.1198v1.

[35] C. Palenzuela, L. Lehner, and S. Liebling, Science **329**, 927 (2010), URL http://www.sciencemag.org/cgi/content/full/329/5994/927.

[36] M. Milosavljević and E. S. Phinney, ApJL **622**, L93 (2005), http://arXiv.org/abs/astro-ph/0410343, URL http://adsabs.harvard.edu/abs/2005ApJ...622L..93M.

[37] Z. Lippai, Z. Frei, and Z. Haiman, ApJL **676**, L5 (2008), http://arXiv.org/abs/0801.0739, URL http://adsabs.harvard.edu/abs/2008ApJ...676L...5L.

[38] M. Megevand, M. Anderson, J. Frank, E. W. Hirschmann, L. Lehner, S. L. Liebling, P. M. Motl, and D. Neilsen, arXiv (2009), http://arXiv.org/abs/0905.3390, URL http://adsabs.harvard.edu/abs/2009arXiv0905.3390M.

[39] L. R. Corrales, Z. Haiman, and A. MacFadyen, arXiv **astro-ph.HE** (2009), URL http://arxiv.org/abs/0910.0014v1.

[40] J. D. Schnittman and J. H. Krolik, arXiv **802** (2008), http://arXiv.org/abs/0802.3556, URL http://adsabs.harvard.edu/abs/2008arXiv0802.3556S.

[41] E. M. Rossi, G. Lodato, P. J. Armitage, J. E. Pringle, and A. R. King, arXiv **astro-ph.HE** (2009), URL http://arxiv.org/abs/0910.0002v1.

[42] J. Kormendy and D. Richstone, ARA&A **33**, 581 (1995), URL http://adsabs.harvard.edu/abs/1995ARA%26A..33..581K.

[43] J. Magorrian, S. Tremaine, D. Richstone, R. Bender, G. Bower, A. Dressler, S. M. Faber, K. Gebhardt, R. Green, C. Grillmair, et al., AJ **115**, 2285 (1998).

[44] M. C. Begelman, R. D. Blandford, and M. J. Rees, Nature **287**, 307 (1980), URL http://adsabs.harvard.edu/abs/1980Natur.287..307B.

[45] D. Merritt and M. Milosavljević, Living Reviews in Relativity **8**, 8 (2005), http://arXiv.org/abs/astro-ph/0410364, URL http://adsabs.harvard.edu/abs/2005LRR.....8....8M.

[46] Q. Yu, MNRAS **331**, 935 (2002), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2002MNRAS.331..935Y&link_type=ABSTRACT.

[47] M. Colpi and M. Dotti, arXiv **astro-ph.CO** (2009), URL http://arxiv.org/abs/0906.4339v1.

[48] J. M. Bardeen and J. A. Petterson, ApJL **195**, L65 (1975), URL http://adsabs.harvard.edu/abs/1975ApJ...195L..65B.

[49] N. I. Shakura and R. A. Sunyaev, A&A **24**, 337 (1973), URL http://adsabs.harvard.edu/abs/1973A%26A....24..337S.

[50] J. H. Jeans, MNRAS **85**, 2 (1924), URL http://adsabs.harvard.edu/abs/1924MNRAS..85....2J.

[51] M. P. Savedoff and S. Vila, AJ **69**, 242 (1964), http://arXiv.org/abs/0708.3999, URL http://adsabs.harvard.edu/abs/1964AJ.....69..242S.

[52] P. C. Peters and J. Mathews, Phys. Rev. D **131**, 435 (1963), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=1963PhRv..131..435P&link_type=ABSTRACT.

[53] Y. Mino, M. Sasaki, M. Shibata, H. Tagoshi, and T. Tanaka, Progress of Theoretical Physics Supplement **128**, 1 (1997), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=1997PThPS.128....1M&link_type=ABSTRACT.

[54] M. A. Scheel, M. Boyle, T. Chu, L. E. Kidder, K. D. Matthews, and H. P. Pfeiffer, Phys. Rev. D **79**, 024003 (2009), http://arXiv.org/abs/0810.1767, URL http://adsabs.harvard.edu/abs/2009PhRvD..79b4003S.

[55] F. Pretorius, Phys. Rev. Lett. **95**, 121101 (2005), http://arXiv.org/abs/gr-qc/0507014, URL http://adsabs.harvard.edu/abs/2005PhRvL..95l1101P.

[56] F. Pretorius, Class. Quant. Grav. **23**, S529 (2006), http://arXiv.org/abs/gr-qc/0602115.

[57] J. G. Baker, J. Centrella, D.-I. Choi, M. Koppitz, and J. van Meter, Phys. Rev. D **73**, 104002 (2006), http://arXiv.org/abs/gr-qc/0602026, URL http://adsabs.harvard.edu/abs/2006PhRvD..73j4002B.

[58] M. Campanelli, C. O. Lousto, and Y. Zlochower, Phys. Rev. D **73**, 061501 (2006), http://arxiv.org/abs/gr-qc/0601091.

[59] E. Berti, V. Cardoso, J. A. Gonzalez, U. Sperhake, M. Hannam, S. Husa, and B. Brügmann, Phys. Rev. D **76**, 064034 (2007), http://arXiv.org/abs/gr-qc/0703053, URL http://adsabs.harvard.edu/abs/2007PhRvD..76f4034B.

[60] L. Blanchet, Phys. Rev. D **65**, 124009 (2002), http://arXiv.org/abs/gr-qc/0112056, URL http://adsabs.harvard.edu/abs/2002PhRvD..65l4009B.

[61] M. Campanelli, C. O. Lousto, and Y. Zlochower, Phys. Rev. D **74**, 041501 (2006), http://arXiv.org/abs/gr-qc/0604012, URL http://adsabs.harvard.edu/abs/2006PhRvD..74d1501C.

[62] H. P. Pfeiffer, S. A. Teukolsky, and G. B. Cook, Phys. Rev. D **62**, 104018 (2000), http://arXiv.org/abs/gr-qc/0006084.

[63] F. Herrmann, I. Hinder, D. Shoemaker, P. Laguna, and R. A. Matzner, ApJ **661**, 430 (2007), http://arXiv.org/abs/gr-qc/0701143, URL http://adsabs.harvard.edu/abs/2007ApJ...661..430H.

[64] D. Pollney, C. Reisswig, L. Rezzolla, B. Szilágyi, M. Ansorg, B. Deris, P. Diener, E. N. Dorband, M. Koppitz, A. Nagar, et al., Phys. Rev. D **D76**, 124002 (2007), http://arXiv.org/abs/0707.2559.

[65] S. A. Hughes and R. D. Blandford, ApJL **585**, L101 (2003), http://arXiv.org/abs/astro-ph/0208484, URL http://adsabs.harvard.edu/abs/2003ApJ...585L.101H.

[66] A. Sesana, F. Haardt, P. Madau, and M. Volonteri, ApJ **623**, 23 (2005), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2005ApJ...623...23S&link_type=ABSTRACT.

[67] A. Sesana, M. Volonteri, and F. Haardt, MNRAS **377**, 1711 (2007), http://arXiv.org/abs/astro-ph/0701556, URL http://adsabs.harvard.edu/abs/2007MNRAS.377.1711S.

[68] A. Sesana, M. Volonteri, and F. Haardt, Classical and Quantum Gravity **26**, 4033 (2009), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2009CQGra..26i4033S&link_type=ABSTRACT.

[69] T. Bogdanović, C. S. Reynolds, and M. C. Miller, ApJ **661**, L147 (2007), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2007ApJ...661L.147B&link_type=ABSTRACT.

[70] J. A. González, U. Sperhake, B. Brügmann, M. Hannam, and S. Husa, Phys. Rev. Lett. **98**, 091101 (2007), http://arXiv.org/abs/gr-qc/0610154, URL http://adsabs.harvard.edu/abs/2007PhRvL..98i1101G.

[71] F. Herrmann, I. Hinder, D. Shoemaker, and P. Laguna, Classical and Quantum Gravity **24**, 33 (2007), http://arXiv.org/abs/gr-qc/0601026, URL http://adsabs.harvard.edu/abs/2007CQGra..24...33H.

[72] J. G. Baker, W. D. Boggs, J. Centrella, B. J. Kelly, S. T. McWilliams, M. C. Miller, and J. R. van Meter, ApJ **668**, 1140 (2007), http://arXiv.org/abs/astro-ph/0702390, URL http://adsabs.harvard.edu/abs/2007ApJ...668.1140B.

[73] M. J. Fitchett and S. Detweiler, MNRAS **211**, 933 (1984), URL http://adsabs.harvard.edu/abs/1984MNRAS.211..933F.

[74] M. Favata, S. A. Hughes, and D. E. Holz, ApJL **607**, L5 (2004), http://arXiv.org/abs/astro-ph/0402056, URL http://adsabs.harvard.edu/abs/2004ApJ...607L...5F.

[75] D. E. Holz and S. A. Hughes, ApJ **629**, 15 (2005).

[76] C. M. Hirata, D. E. Holz, and C. Cutler, Phys. Rev. D **81**, 124046 (2010), http://arXiv.org/abs/1004.3988, URL http://adsabs.harvard.edu/abs/2010PhRvD..81l4046H.

[77] M. Kesden, U. Sperhake, and E. Berti, ApJ **715**, 1006 (2010), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2010ApJ...715.1006K&link_type=ABSTRACT.

[78] M. Dotti, M. Ruszkowski, L. Paredi, M. Colpi, M. Volonteri, and F. Haardt, MNRAS **396**, 1640 (2009), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2009MNRAS.396.1640D&link_type=ABSTRACT.

[79] D. Mihalas and B. W. Mihalas, *Foundations of radiation hydrodynamics* (Oxford University Press, 1984), URL http://adsabs.harvard.edu/abs/1984oup..book.....M.

[80] E. F. Toro, *Riemann Solvers and Numerical Methods for Fluid Dynamics* (Springer, 1999).

[81] L. D. Landau and E. M. Lifshitz, *Fluid mechanics* (Butterworth-Heinemann, 1959), URL http://adsabs.harvard.edu/abs/1959flme.book.....L.

[82] L. Barack and C. Cutler, Phys. Rev. D **69**, 82005 (2004), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2004PhRvD..69h2005B&link_type=ABSTRACT.

[83] F. D. Ryan, Phys. Rev. D **56**, 1845 (1997), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=1997PhRvD..56.1845R&link_type=ABSTRACT.

[84] S. A. Hughes, ARA&A **47**, 107 (2009), 0903.4877.

[85] N. Yunes, M. C. Miller, and J. Thornburg, Phys. Rev. D **83**, 44030 (2011), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2011PhRvD..83d4030Y&link_type=ABSTRACT.

[86] R. Narayan, ApJ **536**, 663 (2000), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2000ApJ...536..663N&link_type=ABSTRACT.

[87] N. Yunes, B. Kocsis, A. Loeb, and Z. Haiman, arXiv **astro-ph.CO** (2011), URL http://arxiv.org/abs/1103.4609v1.

[88] A. Sesana, A. Vecchio, M. Eracleous, and S. Sigurdsson, MNRAS **391**, 718 (2008), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2008MNRAS.391..718S&link_type=ABSTRACT.

[89] K. Menou, Z. Haiman, and B. Kocsis, arXiv **astro-ph** (2008), URL http://arxiv.org/abs/0803.3627v1.

[90] D. Hils and P. L. Bender, ApJL **445**, L7 (1995).

[91] Y. Levin, arXiv **astro-ph** (2003), URL http://arxiv.org/abs/astro-ph/0307084v1.

[92] Y. Levin and A. M. Beloborodov, ApJ **590**, L33 (2003), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2003ApJ...590L..33L&link_type=ABSTRACT.

[93] M. C. Miller, M. Freitag, D. P. Hamilton, and V. M. Lauburg, ApJ **631**, L117 (2005), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2005ApJ...631L.117M&link_type=ABSTRACT.

[94] M. Milosavljević and D. Merritt, ApJ **596**, 860 (2003), http://arXiv.org/abs/astro-ph/0212459, URL http://adsabs.harvard.edu/abs/2003ApJ...596..860M.

[95] S. Tremaine, D. O. Richstone, Y.-I. Byun, A. Dressler, S. M. Faber, C. Grillmair, J. Kormendy, and T. R. Lauer, AJ **107**, 634 (1994), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=1994AJ....107..634T&link_type=ABSTRACT.

[96] A. Sesana, F. Haardt, and P. Madau, ApJ **686**, 432 (2008), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2008ApJ...686..432S&link_type=ABSTRACT.

[97] M. Freitag, P. Amaro-Seoane, and V. Kalogera, ApJ **649**, 91 (2006), http://arXiv.org/abs/astro-ph/0603280, URL http://adsabs.harvard.edu/abs/2006ApJ...649...91F.

[98] J. N. Bahcall and R. A. Wolf, ApJ **209**, 214 (1976), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=1976ApJ...209..214B&link_type=ABSTRACT.

[99] P. C. Peters, Phys. Rev. D **136**, 1224 (1964), URL http://adsabs.harvard.edu/abs/1964PhRv..136.1224P.

[100] D. Merritt, J. D. Schnittman, and S. Komossa, ApJ **699**, 1690 (2009), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2009ApJ...699.1690M&link_type=ABSTRACT.

[101] P. Côté, J. P. Blakeslee, L. Ferrarese, A. Jordán, S. Mei, D. Merritt, M. Milosavljević, E. W. Peng, J. L. Tonry, and M. J. West, ApJS **153**, 223 (2004), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2004ApJS..153..223C&link_type=ABSTRACT.

[102] T. Alexander, Physics Reports **419**, 65 (2005), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2005PhR...419...65A&link_type=ABSTRACT.

[103] M. L. Lidov, Planet. Space Sci. **9**, 719 (1962), URL http://adsabs.harvard.edu/abs/1962P%26SS....9..719L.

[104] Y. Kozai, AJ **67**, 591 (1962), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=1962AJ.....67..591K&link_type=ABSTRACT.

[105] M. Holman, J. Touma, and S. Tremaine, Nature **386**, 254 (1997), URL http://adsabs.harvard.edu/abs/1997Natur.386..254H.

[106] O. Blaes, M. H. Lee, and A. Socrates, ApJ **578**, 775 (2002), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2002ApJ...578..775B&link_type=ABSTRACT.

[107] P. B. Ivanov, A. G. Polnarev, and P. Saha, MNRAS **358**, 1361 (2005), http://arXiv.org/abs/astro-ph/0410610, URL http://adsabs.harvard.edu/abs/2005MNRAS.358.1361I.

[108] T. A. Thompson, arXiv **astro-ph.HE** (2010), URL http://arxiv.org/abs/1011.4322v1.

[109] J. R. Gair, D. J. Kennefick, and S. L. Larson, Phys. Rev. D **72**, 84009 (2005), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2005PhRvD..72h4009G&link_type=ABSTRACT.

[110] L. Wen, ApJ **598**, 419 (2003), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2003ApJ...598..419W&link_type=ABSTRACT.

[111] M. Preto and S. Tremaine, AJ **118**, 2532 (1999), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=1999AJ....118.2532P&link_type=ABSTRACT.

[112] A. H. G. Peter, Proquest Dissertations And Theses 2008. Section 0181 p. 1 (2008), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2008PhDT.....1P&link_type=ABSTRACT.

[113] A. H. G. Peter, Phys. Rev. D **79**, 103531 (2009), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2009PhRvD..79j3531P&link_type=ABSTRACT.

[114] C. Wegg and J. N. Bode (2011), in prep.

[115] S. Weinberg, Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity (1972), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=1972gcpa.book.....W&link_type=ABSTRACT.

[116] W. Dehnen, MNRAS **265**, 250 (1993), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=1993MNRAS.265..250D&link_type=ABSTRACT.

[117] J. R. Gair, D. J. Kennefick, and S. L. Larson, ApJ **639**, 999 (2006), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2006ApJ...639..999G&link_type=ABSTRACT.

[118] X. Chen, F. K. Liu, and J. Magorrian, ApJ **676**, 54 (2008), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2008ApJ...676...54C&link_type=ABSTRACT.

[119] X. Chen, P. Madau, A. Sesana, and F. K. Liu, ApJL **697**, L149 (2009), URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2009ApJ...697L.149C&link_type=ABSTRACT.

[120] X. Chen, A. Sesana, P. Madau, and F. K. Liu, ApJ **729**, 13 (2011), URL http://adsabs.harvard.edu/abs/2011ApJ...729...13C.

[121] B. C. Koo and C. F. McKee, ApJ **354**, 513 (1990), URL http://adsabs.harvard.edu/abs/1990ApJ...354..513K.

[122] R. A. Chevalier, ApJ **207**, 872 (1976), URL http://adsabs.harvard.edu/abs/1976ApJ...207..872C.

[123] L. I. Sedov, Prikl. Mat. i. Makh **10**, 245 (1946).

[124] J. Von-Neumann, Los Alamos Science Laboratory Technical Series, Los Alamos **7** (1947).

[125] G. I. Taylor, Proc. Roy. Soc. London **10**, 175 (1950).

[126] E. Waxman and D. Shvarts, Physics of Fluids **5**, 1035 (1993), URL http://adsabs.harvard.edu/abs/1993PhFl....5.1035W.

[127] R. D. Blandford and C. F. McKee, Physics of Fluids **19**, 1130 (1976), URL http://adsabs.harvard.edu/abs/1976PhFl...19.1130B.

[128] P. Best and R. Sari, Physics of Fluids **12**, 3029 (2000), URL http://adsabs.harvard.edu/abs/2000PhFl...12.3029B.

[129] R. Sari, Physics of Fluids **18**, 027106 (2006), http://arXiv.org/abs/astro-ph/0505174, URL http://adsabs.harvard.edu/abs/2006PhFl...18b7106S.

[130] M. Pan and R. Sari, ApJ **643**, 416 (2006), http://arXiv.org/abs/astro-ph/0505176, URL http://adsabs.harvard.edu/abs/2006ApJ...643..416P.

[131] Y. B. Zel'Dovich and Y. P. Raizer, *Physics of shock waves and high-temperature hydrodynamic phenomena* (New York: Academic Press, W.D. Hayes and Ronald F. Probstein Eds., 1967), URL http://adsabs.harvard.edu/abs/1967pswh.book.....Z.

[132] D. Ryu and E. T. Vishniac, ApJ **313**, 820 (1987), URL http://adsabs.harvard.edu/abs/1987ApJ...313..820R.

[133] J. Goodman, ApJ **358**, 214 (1990), URL http://adsabs.harvard.edu/abs/1990ApJ...358..214G.

[134] R. A. Chevalier, ApJ **359**, 463 (1990), URL http://adsabs.harvard.edu/abs/1990ApJ...359..463C.

[135] R. Sari, E. Waxman, and D. Shvarts, ApJS **127**, 475 (2000), URL http://adsabs.harvard.edu/abs/2000ApJS..127..475S.

[136] D. Kushnir, E. Waxman, and D. Shvarts, ApJ **634**, 407 (2005), http://arXiv.org/abs/astro-ph/0502125, URL http://adsabs.harvard.edu/abs/2005ApJ...634..407K.

[137] V. Simonsen and J. Meyer-Ter-Vehn, Physics of Fluids **9**, 1462 (1997), URL http://adsabs.harvard.edu/abs/1997PhFl....9.1462S.

[138] L. D. Landau and E. M. Lifshitz, *Fluid mechanics* (Course of Theoretical Physics, Oxford, Pergamon Press, 1987), URL http://adsabs.harvard.edu/abs/1959flme.book.....L.

[139] J. P. Cox, *Theory of stellar pulsation* (Research supported by the National Science Foundation. Princeton, NJ, Princeton University Press, 1980), URL http://adsabs.harvard.edu/abs/1980tsp..book.....C.

[140] Y. Oren and R. Sari, Physics of Fluids **21**, 056101 (2009), http://arXiv.org/abs/0907.1731, URL http://adsabs.harvard.edu/abs/2009PhFl...21e6101O.