Mergers of Supermassive Black Hole Binaries in Gas-rich Environments: Models of Event Rates and Electromagnetic Signatures

Takamitsu Tanaka

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2011

ABSTRACT

Mergers of Supermassive Black Hole Binaries in Gas-rich Environments: Models of Event Rates and Electromagnetic Signatures

Takamitsu Tanaka

Supermassive black holes permeate the observable Universe, residing in the nuclei of all or nearly all nearby massive galaxies and powering luminous quasars as far as ten billion light years away. These monstrous objects must grow through a combination of gas accretion and mergers of less massive black holes. The direct detection of the mergers by future gravitational-wave detectors will be a momentous scientific achievement, providing tests of general relativity and revealing the cosmic evolution of supermassive black holes. An additional — and arguably equally rewarding — challenge is the concomitant observation of merging supermassive black holes with both gravitational and electromagnetic waves. Such synergistic, "multi-messenger" studies can probe the expansion history of the Universe and shed light on the details of accretion astrophysics.

This thesis examines the mergers of supermassive black hole binaries and the observable signatures of these events. First, we consider the formation scenarios for the earliest supermassive black holes. This investigation is motivated by the *Sloan Digital Sky*

Survey observation of a quasar that appears to be powered by a supermassive black hole with a mass of billions of solar masses, already in place one billion years after the Big Bang. Second, we develop semianalytic, time-dependent models for the thermal emission from circumbinary gas disks around merging black holes. Our calculations corroborate the qualitative conclusion of a previous study that for black hole mergers detectable by a space-based gravitational-wave observatory, a gas disk near the merger remnant may exhibit a dramatic brightening of soft X-rays on timescales of several years. Our results suggest that this "afterglow" may become detectable more quickly after the merger than previously estimated. Third, we investigate whether these afterglow episodes could be observed serendipitously by forthcoming wide-field, high-cadence electromagnetic surveys. Fourth, we introduce a new subset of time-dependent solutions for the standard equation describing thin, viscous Keplerian disks. Finally, we apply these solutions to model the electromagnetic emission of accretion disks around supermassive black hole binaries that may be detectable with precision pulsar timing.

This page intentionally left blank.

Contents

List of Figures

This page intentionally left blank.

List of Tables

This page intentionally left blank.

ACKNOWLEDGMENTS

The path I took to becoming an astrophysicist was roundabout and somewhat unconventional. It was also a journey of personal growth and maturation as much as it was one of scientific education and professional apprenticeship. For these reasons I have become indebted to many people along the way: teachers, students, friends, family, loved ones, and on occasion, total strangers. I express my heartfelt gratitude here.

I thank my advisors, Kristen Menou and Zoltán Haiman, for their continuous mentorship, guidance, knowledge, encouragement, and seemingly inexhaustible patience. They always prioritized my professional and intellectual development, and granted me the time and space to learn and discover on my own. I could not have asked for a better person to have had as a teacher, let alone two. I must thank Kristen doubly, because I would most likely have remained outside of academic research if not for his taking a chance on me and supporting my return to astronomy.

I am also indebted to Yervant Terzian and Bhuvnesh Jain for encouraging my return to academia. In what was a period of great anxiety and uncertainty, your kindness meant a great deal to me.

I thank the faculty at the Department of Astronomy for maintaining a nurturing environment for its students; for your dedication to our professional development as independent researchers; and for being attentive to our well-being, and receptive to our suggestions and grievances. I thank Greg Bryan for being a generous source of kind advice and feedback; David Helfand, Jacqueline van Gorkom, Ed Spiegel and Kathryn Johnston

for sharing their knowledge, as well as their direction and counsel; Mordecai-Mark Mac Low for constructive criticism and professional advice; Arlin Crotts for stimulating conversations; and Joe Patterson and Jim Applegate for indulging my second passion, baseball, over many a morning coffee. I am thankful to Greg Bryan (again) and Jules Halpern, as well as Jeremy Goodman of Princeton University, for sharing their valuable time and expertise to sit on my dissertation committee. I am grateful to postdoctoral fellows, as well as associate and senior research scientists — among them Jeno Sokoloski, Josh Peek, Paul Bryans, Holger Kreckel, Elizabeth Tasker, Matthew Turk, Kelle Cruz and Daniel Wolf Savin — for being approachable resources of friendly tips and professional knowhow.

I thank my fellow graduate students for their camaraderie. Many thanks to Emily, Kathryn, Jana, Kyle, Cameron, Erika, Duane, Dan, Josh, Yuan, Lia, Neil, Destry, Christine, Jeff, Joo, Jia, Ivan, Jackie, Andreas, Maureen, Victoria, Jemma, Jeff, Dave, Colin, Munier, Brandon, Mari, Sarah, Ian, Maria, Sasha, Gabe and others for their friendship and advice. I am especially appreciative of my officemates through the years — Elizabeth, Franz, Roban, Stephanie, Josh, Ximena, Neil, Jennifer, Lauren, and Gray — for putting up with the sight of my messy desk, and with the growling, sighing and muttered curses directed at my computer.

To Millie and Ayoune, who run the department so smoothly and look after us helpless graduate students so well: a very special Thank You. Thanks also to Trudy, Donna, Krystal and Julie, for making sure that the students are able to receive reimbursements on time. Thank you, Francisco, for making sure that all our academic expenses are paid for. Thank you, Nelson, for making sure we have a clean and well-furnished working environment.

To my friends and acquaintances who lifted my spirits with their company and messages: I am very lucky to have you in my life, and I thank you all. I am especially grateful to Ross, Ted and Rachel for their lasting friendship and numerous shared adventures in New York City. I thank my bandmates, Ross, Tim, Glenn, Rochelle, Ricardo and Dave, for sharing a much-needed creative outlet.

I thank the Boston Red Sox for ending their 86-year championship drought within my lifetime. I sleep a little better at night not having to wonder whether I'll live to see them win a World Series.

I thank my parents for allowing me the freedom and space to choose my own path in life, and for continually prioritizing and supporting my education over the course of two decades. I thank my siblings for their love, and my grandparents for inspiration and strength. I thank my grandmother Rayko who, seeing my childhood fascination with astronomy, bought me my first telescope and thereby became the first to expose my strong predilection toward theoretical work. I thank my second set of parents, David and Sue, for their love, generosity and support. Thank you, Tyler and Megan, for being such fun siblings. I am grateful to Pazu for his companionship and for being a reliable source of endless entertainment.

I thank my wife, Melissa, for walking this journey with me. I may have given up long ago if not for your faith, support, understanding, patience and love. I thank you for putting up with my overnight stays at the office, travels away from home, and the countless dinner conversations that failed to materialize because I was too distracted thinking about research. I appreciate your always making sure that I am eating and sleeping properly,

even if I complain about the attention at times. As much as I love astrophysics, I am incredibly fortunate that my reprieves from it are even more wonderful.

I am grateful to every student, public-lecture attendee, friend and acquaintance that listened attentively to my ramblings about physics and astronomy. Your enthusiasm and thirst for knowledge make what I do feel truly worthwhile. I thank the taxpayers and benefactors around the world for funding research in astronomy and astrophysics, and hope that they will continue to do so.

Chapter 1

Introduction

"So, can you tell me... what happens when two black holes collide?" — My father, Autumn 2002

I begin my thesis with an overview of astrophysical supermassive black holes (hereafter SMBHs; 1.1), reviewing their properties as predicted by the general theory of relativity and as they are observed in nature. I then diskuss SMBH binaries (1.2), and outline the theoretical expectations for their formation and evolution (1.2.1). I review current observational evidence for close SMBH pairs and binary candidates (1.2.2), as well as some circumstantial evidence for their existence and influence. I provide a condensed theoretical summary of SMBH binaries as powerful sources of gravitational waves (GWs), and diskuss why these waves are expected to be observable at very large distances (1.3). I review the major types of current and planned GW detection experiments, their capabilities and astrophysical targets (1.3.1). I then describe the scientific motivations behind multi-messenger studies of compact and merging SMBH binaries — that is, the concomitant

observations of these systems in both electromagnetic (EM) and gravitational waves (1.4). I summarize the various mechanisms that have been suggested to elicit EM signatures from merging SMBHs (1.4.1). Many of these mechanisms require the existence of a gaseous accretion flow around the merging binary. Accretion physics is also a cornerstone for the theory of active galaxies and quasars. A brief review of standard accretion disk theory and its predictions is presented (1.5). Finally, I will summarize the remaining chapters of the thesis (1.6).

1.1 Supermassive Black Holes

SMBHs, as their name might suggest, are astronomical bodies in a mass class of their own. They are loosely defined as having masses of hundreds of thousands of solar masses ($M_\odot$) or more, with the most massive ones exceeding a billion solar masses. By comparison, "stellar-mass" black holes that form from the deaths of massive stars are predicted and observed to have masses of $\sim 10 M_\odot$ (e.g., Narayan 2005). Soon after the diskovery of radio quasars in the 1960s (Schmidt 1963), it was realized that the gravitational power of gas accreting onto SMBHs — and not nuclear burning — was the most plausible mechanism behind these extremely luminous phenomena billions of light-years away (Zel'Dovich 1964; Salpeter 1964; Lynden-Bell 1969, 1978). It has become widely accepted that SMBH gravitational potentials are also the engines of less powerful active galactic nuclei (AGN; Rees 1984) and the extended radio jets often found emanating from AGN (Begelman et al. 1984). Quasar activity is observed to have been much more prevalent when the Universe was younger, peaking at a cosmological redshift of $z \approx 2$ (Hopkins et al. 2007b).

Observations of luminous quasars at cosmological redshifts as high as $z \sim 6$ point to the existence of SMBHs with masses $M \sim 10^9 M_\odot$ as early as ~ 1 Gyr after the Big Bang (Fan et al. 2001; Fan 2006). Probes of stellar and gas dynamics in the nuclei of nearby galaxies indicate that virtually all of them — including our own — harbor a SMBH, but usually in a quiescent, non-AGN state (e.g., Kormendy & Richstone 1995; Magorrian et al. 1998; Richstone et al. 1998; Kormendy & Gebhardt 2001; Ferrarese & Ford 2005, and references therein). The consensus interpretation is that the nearby SMBHs are the dead relics of the engines of past quasars (Soltan 1982; Yu & Tremaine 2002).

Currently, it remains uncertain whether these supermassive objects are truly the black holes described by the general theory of relativity; indeed, one of the broader scientific motivations behind the theoretical work presented in this thesis is to test whether they are. For the purposes of explaining AGN power and the quiescent supermassive dark objects in the local Universe, however, the only essential requirements are that these objects have a very large mass and a very compact size.

According to general relativity, a single SMBH is characterized by two intrinsic quantities: its mass M and rotation parameter $0 \leq a < 1$ (the "No Hair" theorem[1]). The latter is defined as $a \equiv Jc/(GM^2)$, where J is the spin angular momentum of the hole, c is the speed of light and G is the gravitational constant. The mass and rotation parameter uniquely determine the gravitational field around the hole and the relevant spatial scales. We may associate the size of a black hole with either of two radii: the event horizon $R_{\rm H}$; or the location of the innermost stable circular orbit (ISCO), or $R_{\rm ISCO}$, inside which circular orbits are dynamically unstable to small perturbations. The event horizon

[1]We implicitly assume that astrophysical black holes have no electric charge.

is located at $R_{\rm H} = 2GM/c^2$ for non-rotating (i.e., $a = 0$) black holes, and has a lower bound of $R_{\rm H} > GM/c^2$ corresponding to maximum rotation ($a < 1$). The ISCO radius is equal to $6GM/c^2$ for a non-rotating black hole; around rotating holes, $R_{\rm ISCO}$ can be as small as GM/c^2 for orbits aligned with the spin axis of a maximally rotating hole, and as large as $9GM/c^2$ for those in counter-alignment (Bardeen 1970). It is usually assumed that an accretion flow onto a black hole cannot maintain a stable structure inside $R_{\rm ISCO}$, is truncated there. Even if could extend inside the ISCO radius (due to, e.g., magnetohydrodynamic forces or horizontal advection), it would terminate at the event horizon. These two radii differ by a factor of 3 or less for orbits in corotation with the black hole; both have a characteristic size $R \gtrsim GM/c^2$. Expressing this length scale for SMBHs in familiar astrophysical units,

$$\frac{GM}{c^2} = 2.1 \left(\frac{M}{10^6 M_\odot}\right) R_\odot = 9.9 \left(\frac{M}{10^9 M_\odot}\right) \text{ AU}, \tag{1.1}$$

where $R_\odot$ is the radius of the Sun, we see that it is typically comparable to the radii of massive stars.

Observational evidence strongly suggests that SMBH candidates indeed have sizes in the $\gtrsim GM/c^2$ range. Infrared observations of the nuclear star cluster in our galaxy have revealed that the innermost stars orbit a single dark object with a mass of $4 \times 10^6 M_\odot$ (Ghez et al. 2005). The stellar orbits alone constrain the size of the central dark body to within 10 AU, or $\sim 200GM/c^2$. Radio observations of the same region suggest that the intrinsic size of the radio source is at most $\sim 10GM/c^2$ (Shen et al. 2005), although an alternative interpretation is that the radio source is not located at the center of the accretion flow but rather represents a hot spot in an outer region (Doeleman et al. 2008). Relativistic

models of the broadening of the Fe K α lines from AGN and stellar-mass black hole candidates strongly indicate that gas accretion flows have kinetic properties consistent with an accretion flow extending down to the GM/c^2 scale (Fabian et al. 2000; Reynolds & Nowak 2003; Miller 2007). By associating the truncation radius of the accretion flow with the ISCO radius, it is possible to constrain the spin of the central black hole (McClintock et al. 2011). Variability studies of broadened spectral lines in AGN also point to these lines being produced by hot ionized gas orbiting supermassive objects within distances of $\sim 10^3 GM/c^2$ (Blandford & McKee 1982; Peterson 1993; Wandel et al. 1999; Peterson et al. 2004).

It is precisely this compact size that makes SMBHs such spectacular power sources. The maximum specific gravitational energy that can be liberated by the infall of matter down to the $R_{\rm ISCO}$, written as a fraction ϵ of its specific rest-mass energy c^2, is $1/9 < \epsilon < 1$. An elegant method to infer the value of ϵ in quasars is to compare the integrated luminosity of quasars across all redshifts with the mass of the nearby SMBH population. This suggests the cumulative average value of ϵ in quasars to be $\langle\epsilon\rangle \sim 0.1$ (Yu & Tremaine 2002; Marconi et al. 2004). However, this assumes that most of the mass growth of the SMBH population is observable as luminous quasars; the radiative efficiency could be higher if a significant portion of the growth occurs via less luminous or obscured modes of accretion. By comparison, the fraction of the rest mass that can be extracted from hydrogen burning is $\epsilon \approx 0.007$. Gravitational infall onto a compact object appears to be the only plausible theoretical explanation for quasar power (see, e.g, Lynden-Bell 1978, for diskussions and exclusions of several alternative mechanisms).

It is common to express AGN luminosities in terms of the Eddington luminosity $L_{\rm Edd}$ of the central SMBH. For an object of mass M surrounded by a spherically symmetric, homogeneous medium in hydrostatic equilibirum, $L_{\rm Edd}$ is defined as the limiting luminosity above which the outward radiative forces experienced by a particle in the medium exceeds the inward gravitational pull exerted on it. For a medium consisting of ionized hydrogen and whose opacity is dominated by electron scattering (Shapiro & Teukolsky 1986),

$$L_{\rm Edd}(M) \approx 1.3 \times 10^{38} \left(\frac{M}{M_\odot}\right) \text{ erg s}^{-1} \approx 3.3 \times 10^{10} \left(\frac{M}{10^6 M_\odot}\right) L_\odot, \tag{1.2}$$

where $L_\odot \approx 3.8 \times 10^{33}$ erg s^{-1} is the solar luminosity. Quasars are observed to have bolometric luminosities on the order of $L \sim 0.1-1 L_{\rm Edd}$ (e.g., Kollmeier et al. 2006; Alexander et al. 2008; Shankar et al. 2009b; Hickox et al. 2009, and references therein).

Given a value for the radiative efficiency ϵ, quasar power is related to the mass accretion rate $\dot{M}$ of the fuel through $L = \epsilon \dot{M} c^2$. Adopting $\epsilon \sim 0.1$, the value of the accretion rate corresponding to the Eddington limit is

$$\dot{M}_{\rm Edd} \equiv \frac{L_{\rm Edd}}{\epsilon c^2} \sim 0.022 \left(\frac{M}{10^6 M_\odot}\right) \left(\frac{\epsilon}{0.1}\right)^{-1} M_\odot \text{ yr}^{-1}. \tag{1.3}$$

Eddington-limited accretion satisfies $\dot{M} \propto M$, and thus the mass growth of the central object is exponential with an e-folding timescale

$$t_{\rm Sal} = \frac{M}{\dot{M}} = 4.5 \times 10^7 \left(\frac{\epsilon}{0.1}\right) \text{ yr} \tag{1.4}$$

known as the Salpeter time. This is comparable to the typical quasar lifetime derived

from observations (e.g., by dividing the time-integrated accretion by the space density of SMBHs Yu & Tremaine 2002; Martini 2004), $t_{\rm Q} \sim 10^{6-8}$ yr, suggesting that during a typical quasar phase a SMBH can increase its mass by a significant fraction or even by a factor much larger than unity. The quasar lifetime, however, likely depends on many factors (e.g., Hopkins & Hernquist 2009), including the mass of the SMBH and the amount of fuel available. Indeed, quasar lifetimes appears to depend on cosmological redshift, with SMBHs at higher redshifts spending larger fractions of their time as quasars (Shen et al. 2007; Shankar et al. 2010).

There is now little doubt that SMBHs play important roles in shaping their own surroundings. The ultraviolet (UV) and X-ray radiation from the earliest quasars likely contributed to the reionization of the Universe (Barkana & Loeb 2001). On smaller scales, empirical correlations have been diskovered between the masses of nuclear SMBHs and the properties of their host galaxies, suggesting that SMBHs regulate the properties of their stellar environments, or vice versa. The most striking of these correlations is the so-called $M - \sigma$ relation between the SMBH mass M and the stellar velocity dispersion σ of the bulge of the host galaxy (Figure 1.1; Ferrarese & Merritt 2000; Gebhardt et al. 2000; Tremaine et al. 2002):

$$\log_{10}\left(\frac{M}{M_\odot}\right) \sim 8 + 4\log_{10}\left(\frac{\sigma}{200\ {\rm km\ s^{-1}}}\right). \tag{1.5}$$

A large fraction of AGN host galaxies also show signs of having recently undergone intense episodes of star formation (Kauffmann et al. 2003), suggesting a causal link between the two processes.

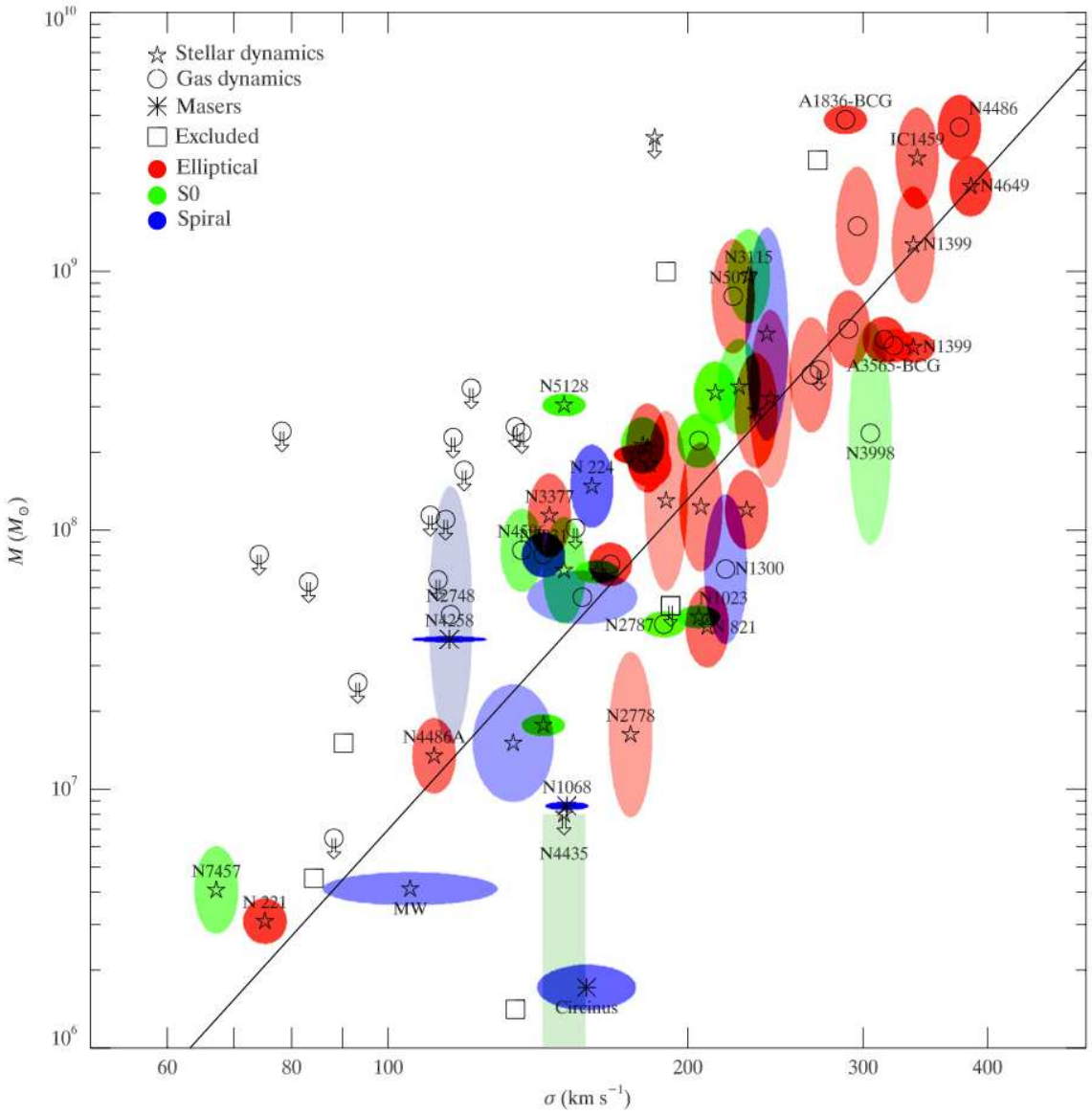

Figure 1.1 From Gültekin et al. (2009), the empirical correlation between the mass of the nuclear SMBH M and the stellar velocity dispersion σ of the bulge of the host galaxy. The slope of the relationship is roughly $M \propto \sigma^{4.2\pm0.4}$.

Despite the abundant evidence for the ubiquity and influence of SMBHs throughout cosmic time, the details of their coevolutionary relationship with their host galaxies remain unclear. A leading hypothesis is that both quasar activity and star formation are triggered by the merger of the host galaxy with another galaxy (Sanders et al. 1988; Hernquist 1989; Carlberg 1990; Barnes & Hernquist 1991; Hernquist & Mihos 1995; Mihos & Hernquist 1996; Kauffmann & Haehnelt 2000; Hopkins et al. 2007a, 2008); and that the energy feedback from the AGN (and perhaps supernovae) heats the gas to quench accretion and star formation in a self-regulating fashion (Haehnelt et al. 1998), while also imprinting the $M - \sigma$ relation (Wyithe & Loeb 2003b; Granato et al. 2004; Murray et al. 2005; Di Matteo et al. 2005; Robertson et al. 2006). Mergers offer a convenient mechanism for funneling the considerable amounts of gas necessary to fuel quasars (Hernquist 1989; Hernquist & Mihos 1995). They also naturally explain the downturn in AGN activity at low redshifts, via reduced merger rates, the depletion of available cold gas through consumption by SMBHs and conversion into stars, and the increase in the timescales necessary to bring gaseous fuel to SMBHs (Kauffmann & Haehnelt 2000). Quasar host galaxies are frequently found to have nearby galaxy companions (e.g., Bahcall et al. 1997), supporting the notion that gravitational interactions can serve as the trigger. On the other hand, surveys of the morphologies of AGN hosts do not reveal tell-tale signs of recent merger activity, suggesting that perhaps not all AGN activity is the result of a major galaxy merger (e.g., Cisternas et al. 2011; Schawinski et al. 2011).

1.2 Supermassive Black Hole Binaries

A merger of two galaxies has another significant consequence: if each galaxy contains a nuclear SMBH, the merger must result in the formation of one galaxy containing two SMBHs (Begelman et al. 1980). Since SMBHs appear to have resided in galactic nuclei for at least $\sim 90\%$ of the age of the Universe, in the standard paradigm of hierarchical structure formation the formation of such multiple-SMBH galaxies would appear to be an inevitable and plausibly frequent occurence. Even if SMBHs were present only in a small fraction of the first galaxies, galaxy mergers would increase their occupation fraction (Menou et al. 2001), so that by the quasar epoch ($z \sim 2$) a plurality of massive galaxies would host a minimum of one SMBH. Yet, the vast majority of galaxies appear to harbor one — and only one — SMBH. As I diskuss below, dual SMBHs may shape the observable characteristics of their host galaxies, in ways that solitary SMBHs cannot. For example, independently of whether its host has recently experienced a merger, a SMBH pair may act to enhance AGN activity (e.g., Gaskell 1985; Liu et al. 2011a) and trigger starbursts (Taniguchi & Wada 1996). To fully understand the relationship between SMBHs and galaxies, it is necessary to examine the fates of multiple SMBHs in a common galaxy.

1.2.1 Orbital evolution of compact SMBH binaries

SMBH pairs evolve by getting rid of their orbital angular momentum and energy. At least some pairs — depending on their masses and environment — are expected to form gravitationally bound binaries (Begelman et al. 1980; Milosavljević & Merritt 2001; Yu 2002; Merritt & Milosavljević 2005). Following Begelman et al. (1980) and Yu (2002), we

can divide the formation and evolution of a SMBH binary system into three stages:

1. The dynamical friction stage, in which the SMBHs migrate toward the center of the shared host potential, and ultimately form a bound binary.

2. The environment-driven binary stage, in which the binary's orbit is hardened through three-body interactions with stars, dissipative gaseous processes, and/or encounters with a third SMBH.

3. The GW-driven binary stage, in which the binary loses orbital energy and angular momentum through the emission of GWs, and ultimately coalesces.

Below, I review each of these stages and the relevant distances and timescales.

Because of their extreme masses, SMBHs in a stellar medium will experience a significantly higher acceleration due to dynamical friction compared to a star in the same background (Chandrasekhar 1943; Binney & Tremaine 1987). In a stellar core with a radius $R_{\rm core}$ and N stars with velocity dispersion $\sigma_{\rm core}$, the migration timescale of a SMBH with mass M is given by (Yu 2002)

$$t_{\rm DF} \sim \frac{2\times10^7}{\log N}\left(\frac{M}{10^7 M_\odot}\right)^{-1}\left(\frac{\sigma_{\rm core}}{100\ {\rm km\ s^{-1}}}\right)\left(\frac{R_{\rm core}}{100\ {\rm pc}}\right)^2\ {\rm yr}. \tag{1.6}$$

Note that equation 1.6 suggests that pairs in which one member has a mass much lower mass than the other may not efficiently form a binary. For example, if $\sigma_{\rm core}$ and $R_{\rm core}$ correlate with the mass of the more massive hole, say $M_1 \sim 10^8 M_\odot$, then an intermediate-mass black hole with a mass of $10^3 M_\odot$ may not migrate to the center within a Hubble time. Hierarchical formation models predict that such pairings can occur frequently if

intermediate-mass black holes are formed en route to the growth and assembly of SMBHs (Chapter 2; cf. Volonteri et al. 2003a; Lippai et al. 2009; Holley-Bockelmann et al. 2010).

Even after a bound binary is formed, dynamical friction will continue to dissipate its orbital energy. However, this mechanism becomes increasingly inefficient when the orbital velocity of the SMBHs reaches the circular velocity of the background stars, $\sim \sqrt{2}\sigma_{\rm core}$. The semimajor axis at which a binary of mass M reaches the end of the dynamical friction stage can thus be estimated as

$$a \sim \frac{GM}{\sigma_{\rm core}^2} = 4.3 \left(\frac{M}{10^7 M_\odot}\right)\left(\frac{\sigma_{\rm core}}{100\ {\rm km\ s^{-1}}}\right)^{-2}\ {\rm pc}. \tag{1.7}$$

Three-body interactions with the stellar background distribution have long been diskussed as a plausible mechanism to continue the orbital hardening of the SMBH binary past the dynamical-friction stage (Begelman et al. 1980; Roos 1981; Quinlan 1996; Quinlan & Hernquist 1997; Zier & Biermann 2001; Milosavljević & Merritt 2001). The binary ejects stars in its vicinity (the binary's "loss-cone") via gravitational slingshot, expending a fraction of its orbital energy with each encounter (Heggie 1975). However, analytic and N-body calculations using spherical stellar distributions reveal that the slingshot ejections become rapidly rarified as the stars inside the loss-cone are depleted, and are insufficient to merge a binary within a Hubble time. This theoretical barrier for binary evolution has been termed the "final parsec problem" (Makino 1997; Milosavljević & Merritt 2001, 2003; Makino & Funato 2004; Figure 1.2). The barrier is less severe for less massive SMBHs in low-dispersion stellar environments, and for eccentric binaries (Sesana et al. 2007a; Sesana 2010).

Recent N-body simulations using realistic models of triaxial galaxies have shown that binaries in these galaxies are able to overcome the final parsec problem (Merritt & Poon 2004; Berczik et al. 2006). The shape of these galaxies is due to centrophilic stellar orbits, which can continuously refill the SMBH binary's loss-cone, increase the scattering rate by several orders of magnitude, and prevent the orbital decay from stalling. A gas-rich environment may also accelerate the formation and orbital evolution of the binary, and help to surmount the final parsec barrier (Gould & Rix 2000; Armitage & Natarajan 2002; Escala et al. 2004, 2005; Kazantzidis et al. 2005; Mayer et al. 2007; Cuadra et al. 2009; Callegari et al. 2009; Hayasaki 2009; see, however, Lodato et al. 2009).

Galaxy mergers can result in the formation of a SMBH triple (Valtonen 1996), if a SMBH binary is unable to merge (see next section) before its host galaxy undergoes a subsequent merger. A third SMBH is capable of exciting high orbital eccentricities of the inner SMBH binary and facilitate its merger (e.g., Blaes et al. 2002; Wen 2003; Iwasawa et al. 2006). Orbital eccentricities induced by such interactions may be observable by future GW detectors (Amaro-Seoane et al. 2010; see 1.3.1 below).

At close separations, general relativity predicts that the binary's orbit will decay further via the emission of orbital energy as GWs. GW-driven orbital decay has been tested and confirmed to exquisite precision in the binary pulsar PSR 1913+16 (Taylor et al. 1979; Taylor & Weisberg 1989; Weisberg et al. 2010). The timescale of the orbital decay is predicted by a relatively simple formula (Peters 1964),

$$t_{\rm GW} \equiv -\frac{a}{da/dt} = \frac{5}{16}\frac{c^5}{G^3M^3}a^4\eta^{-1}\left[\frac{(1-e^2)^{7/2}}{1+73/24\,e^2+37/96\,e^4}\right]$$

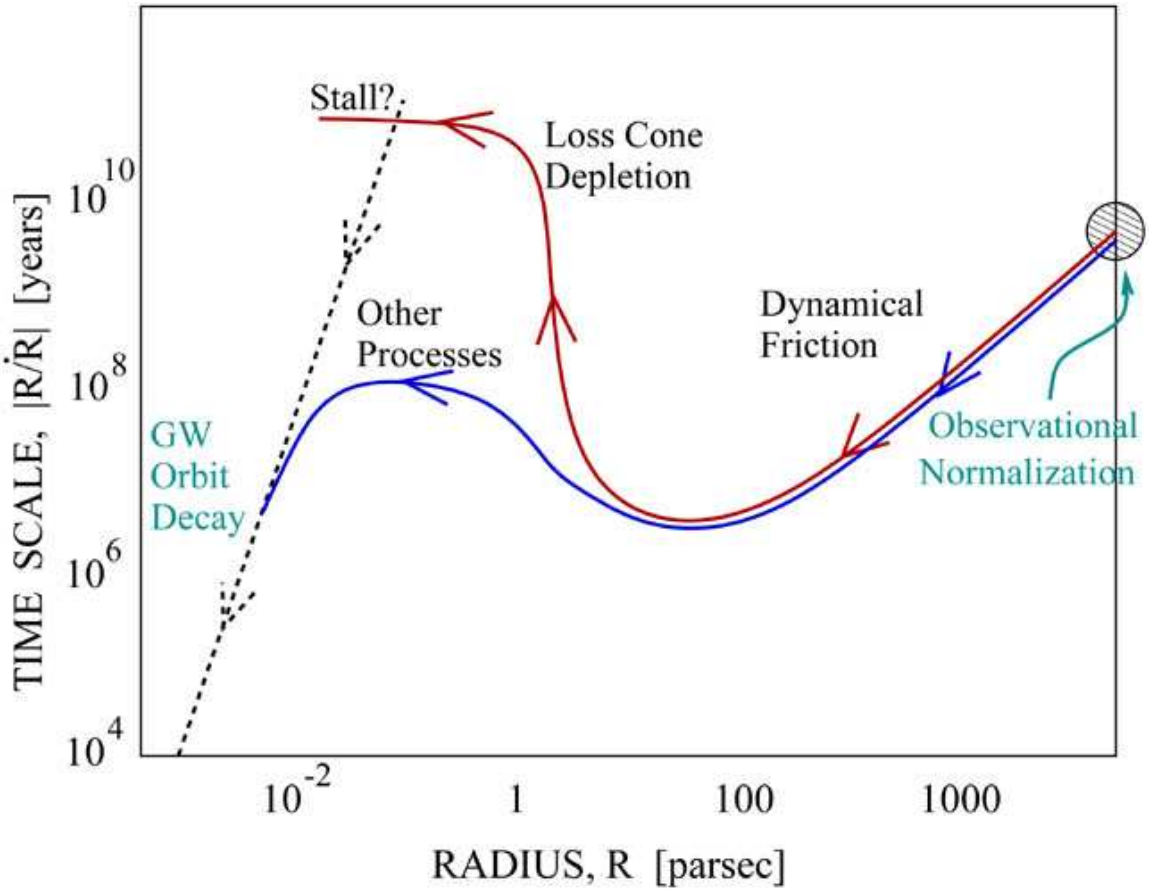

Figure 1.2 A schematic diagram of the evolution of a SMBH binary (Backer et al. 2004; see also Begelman et al. 1980, Yu 2002). The horizontal axis shows the orbital separation of a $\sim 10^{7-8} M_{\odot}$ SMBH binary whose members are similar in mass. The vertical axis shows the characteristic timescale spent at a given orbital separation. For massive binaries and galaxies, orbital decay via stellar processes is inefficient, possibly causing the binary to stall and never reach the separations at which the emission of gravitational radiation can effect a merger. Triaxial galaxy potentials and/or gas-rich environments may help to bridge this gap (i.e., serve as the "Other Processes" indicated in the figure; see references in text).

$$= \quad 9.2 \times 10^6 \left(\frac{M}{10^7 M_\odot} \right)^{-3} \left(\frac{a}{10^{-3} \text{ pc}} \right)^4 (4\eta)^{-1} [...] \text{ yr}, \tag{1.8}$$

where e is the orbital eccentricity and $\eta \equiv M_1 M_2 / M^2 \leq 1/4$ is the symmetric mass ratio of the binary. For circular orbits, the time to merger is given by $t_{\rm GW}/4$. Note that the orbital decay is faster for eccentric orbits, due to increased GW emission near periapsis. In many models of binary interaction with gas (e.g., Artymowicz et al. 1991; Armitage & Natarajan 2002; Cuadra et al. 2009), and in some models of interaction with stars (Sesana et al. 2004; Sesana 2010), the transport of the binary's orbital angular momentum via tidal torquing is more efficient than the extraction of its orbital energy (e.g., as spiral density waves). In such models, the binary's eccentricity increases as its orbital separation decays, a behavior that would in principle accelerate the onset as well as the progression of GW-driven orbital decay (e.g., Hayasaki 2009). In contrast, GW emission causes the orbital eccentricity to decay more rapidly than the semimajor axis, so that $|e/(de/dt)| < t_{\rm GW}$ always for nonzero e (Peters 1964). Thus, unless interactions prior to the GW-driven stage can excite eccentricities close to unity, SMBH binaries are expected to be nearly circular by the time they enter the final stages of the merger. Note that the time to merger is a very steep function of a — $t_{\rm GW} \propto a^4$ — and so as long as interactions with its environment can bring the binary separation to within a separation of $a < 0.01$ pc (i.e., $a \lesssim 10^{3-4} GM/c^2$), coalescence via GWs is rapid and almost unavoidable.

Recent studies have focused on possible observable effects of the coalescence on the SMBH binary remnant. The most notable of these is the gravitational recoil effect, in which the coalesced SMBH receives a "kick" during the merger due to asymmetric ejection of

linear momentum via GWs. This mechanism has long been known in Newtonian and post-Newtonian treatments of non-spinning black holes (Peres 1962; Bekenstein 1973; Fitchett 1983), where the maximum expected recoil velocities were estimated to be on the order of $\gtrsim 100$ km s^{-1}. It received renewed interest after post-Newtonian calculation including spin effects (Kidder 1995; Favata et al. 2004) suggested that recoil velocities could exceed 1000 km s^{-1}, and when such kick velocities were confirmed by calculations in full numerical general relativity (Baker et al. 2006b; Campanelli et al. 2007a,b; González et al. 2007). During inspiral and merger, the binary can also lose up to several percent of its mass as GWs. In addition, the coalesced product can end up with a significantly different spin magnitude and direction than either of the progenitor black holes (e.g., Tichy & Marronetti 2008, and references therein).

1.2.2 Observational Evidence for SMBH Binaries

Current observational evidence for galaxies harboring multiple SMBHs is scarce. This is unsurprising, for many reasons. For instance, the most straightforward method for identifying SMBH pairs is to observe them as dual AGN. Spatial separations below the kpc scale in the quasar epoch cannot be imaged with current instruments (e.g., the spatial resolution of the *Hubble Space Telescope* is ~ 1 kpc at $z \sim 0.5$), and so unambiguous identifications of close AGN pairs are severely distance-limited. Furthermore, depending on the orientation of the surrounding matter (e.g., if their molecular torii are misaligned), one member of an AGN pair may be obscured. Or, if one SMBH is much less massive than the other, emission from the lower-mass member can be buried in that of the more powerful

AGN. Orbital separations of SMBH binaries are expected to decrease at an increasing rate as they evolve (1.2.1 above), and so compact binaries are much rarer than those with large orbital separations. It has also been suggested (see Chapter 3 and references therein) that compact SMBH binaries can significantly suppress the luminosities of their accretion flows. Below, we give a condensed summary of the current observational evidence for SMBH pairs and binaries; for a more detailed account, see the review by Komossa (2006), and references therein.

Imaged pairs. The most compact SMBH binary to be observed is the dual AGN in the elliptical galaxy 0402+379 (Rodriguez et al. 2006), interpreted to be powered by two SMBHs with a combined mass of $M \sim 10^8 M_\odot$ and a separation of ~ 7 pc $\sim 10^6 GM/c^2$. This is by far the closest separation at which an AGN pair has been directly observed. There are several more identified pairs separated on ~ 1 kpc scales (Figure 1.3; Komossa et al. 2003; Bianchi et al. 2008; Comerford et al. 2009b; Green et al. 2010; Liu et al. 2010). Recent surveys have suggested that a small fraction ($0.1\% - 1\%$) of all nearby ($z \lesssim 1$) AGN may harbor SMBH at such separations (Comerford et al. 2009a; Liu et al. 2011c).

SMBH triples. There are currently three observed triple SMBH systems. The AGN triples SDSS J1027+1749 ($z \approx 0.07$; Liu et al. 2011b) and NGC3341 ($z \approx 0.03$; Barth et al. 2008) have separations of several kpc between the nuclei. The triple quasar candidate QQ 1429 - 008 ($z \approx 2$; Djorgovski et al. 2007) has projected separations of ~ 40 kpc.

Periodic emission. The object OJ 287 (Sillanpaa et al. 1988; Pursimo et al. 2000) is one of the most extensively studied quasars. It is known for its extremely regular optical outbursts, which have a period of approximately 12 years and have been observed as far

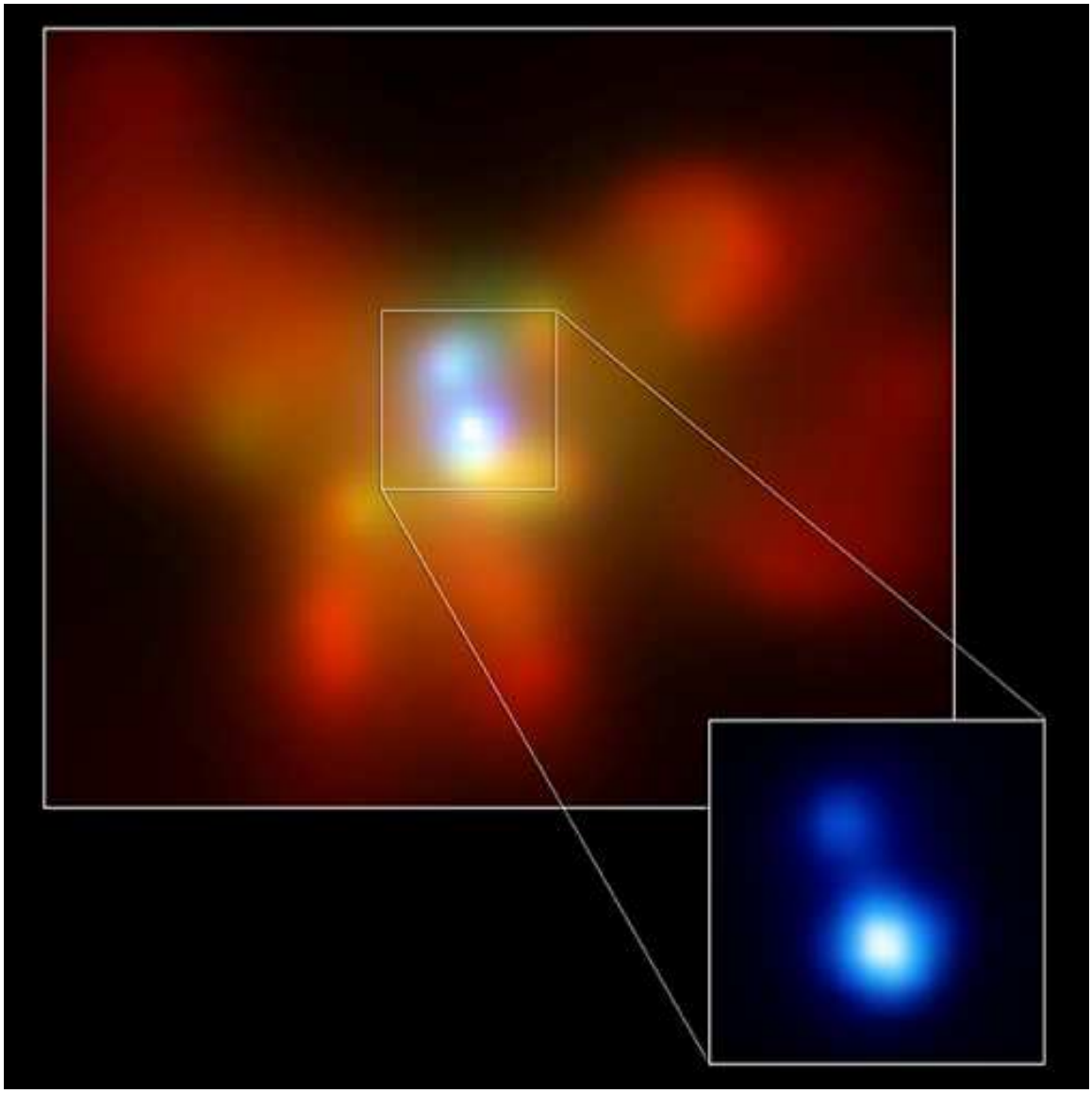

Figure 1.3 X-ray images of the nearby galaxy NGC 6240 (z = 0.024). The two clearly distinguishable nuclei, with a projected spatial separation of 730 pc, indicate the presence of a close SMBH pair. Image credit: NASA/CXC/MPE/Komossa et al. (2003)

back as 1890. Several models for the periodic outbursts have been proposed, among them the tidal excitation of a compact accretion disk around the more massive SMBH by the close encounter of the secondary (Sillanpaa et al. 1988); and the secondary crossing into an accretion disk that is misaligned with respect to the binary orbit (Lehto & Valtonen 1996). In both models, the 12-year period is interpreted as the orbital period of the binary, which has an inferred total mass of $M \sim 10^{10} M_{\odot}$(Lehto & Valtonen 1996). Careful timing of future outbursts could constrain the binary's orbital parameters and the mass and spin of the primary SMBH, since general-relativistic effects — such as spin-coupled orbital precession and orbital decay — should shift the time between the outbursts on $\sim 10 - 100$ day timescales (Valtonen et al. 2008, 2010). On the other hand, the accretion disk may also undergo dynamical changes resulting in similar shifts (Valtonen et al. 2006). There are several other candidates exhibiting evidence for periodic variability (e.g., Rieger & Mannheim 2000; Raiteri et al. 2001; Lobanov & Roland 2005; see also Komossa 2006 for additional references) though none as pronounced as in OJ 287. Interestingly, OJ 287 and most of the AGN candidates exhibiting periodicity are blazars, which are believed to be quasars whose jets are beamed almost directly along the observational lines-of-sight.

Double-peaked broad lines. Numerous AGN exhibit broad emission lines (interpreted to be generated in the inner accretion disk) that are double-peaked, with the peaks separated by velocities of $\sim 10^3$ km s^{-1}. One possible interpretation for these peaks is Doppler shifting of distinct broad-line regions, each around one member of a SMBH binary whose orbit lies close to the line of sight (e.g., Gaskell 1983; Peterson et al. 1987; Gaskell 1996; Boroson & Lauer 2009). This hypothesis can be tested, for some double-peaked

broad-line-region AGN, by monitoring the line separation over timescales comparable to the inferred orbital period (Halpern & Filippenko 1988). So far, the binary SMBH origin for these double peaks has been ruled out for dozens of systems (e.g., Eracleous et al. 1997; Halpern & Eracleous 2000; Eracleous & Halpern 2003; Chornock et al. 2009). In the vast majority of AGN with double-peaked broad lines, gas physics, such as reflection and reprocessing of the lines by an accretion disk, appears to be the most plausible explanation.

Shifted broad lines. It is also common for AGN to have broad lines that are redshifted or blueshifted by $\sim 100 - 1000$ km s^{-1} from the narrow lines (which are interpreted to originate farther out in the accretion disk). Such shifts have been attributed in some systems to the orbital motion of an accreting SMBH binary. Recently, several authors have proposed that the shifted broad lines are generated by gas bound to SMBHs that are escaping their galactic centers with velocities of $\sim 10^3$ km s^{-1} due to the gravitational recoil effect (Komossa et al. 2008; Shields et al. 2009b; Civano et al. 2010). The physical origin of these shifts is not well understood; for example, the quasar SDSS J092712.65+294344.0 has been interpreted as a recoiling SMBH (Komossa et al. 2008), a SMBH binary (Bogdanović et al. 2009; Dotti et al. 2009), an interacting pair of galaxies in a cluster (Heckman et al. 2009), and a superposition of two AGN (Shields et al. 2009a). Bogdanović et al. (2007) proposed that the dearth of recoiling SMBH candidates is due to the spin alignment of binary SMBHs through accretion from a common accretion disk, a configuration that is predicted to produce smaller recoil velocities. Recoil velocities decrease rapidly with mass ratio (see Figure 2.2), so an alternative explanation is that SMBH binaries with nearly-equal mass binaries are rare.

Cores of elliptical galaxies. Elliptical galaxies exhibit a bimodality in the central stellar distribution: one group shows dense stellar cusps with a power-law radial density profile, while another exhibits steep cutoffs in the light profile, with a nuclear stellar "core" (Ferrarese et al. 1994; Lauer et al. 1995). The predicted ejection of stars via three-body interactions with a SMBH binary is consistent with the stellar mass deficits found in cores (comparable to the central SMBH mass; Ebisuzaki et al. 1991; Makino & Ebisuzaki 1996; Quinlan 1996; Quinlan & Hernquist 1997; Milosavljević & Merritt 2001; Merritt 2002; Ravindranath et al. 2002; Milosavljević et al. 2002; Volonteri et al. 2003b). The oscillations of a recently merged, recoiling SMBH can also create similar stellar deficits (Merritt et al. 2004; Boylan-Kolchin et al. 2004). Hoffman & Loeb (2007) suggested that core-scouring by triple SMBH systems can help explain very large cores, such as the one in M87. Although the connection between stellar cores and SMBH binaries (or binary remnants or triples) is circumstantial, the observed data are wholly consistent with these hypotheses (Kormendy & Bender 2009; Hopkins & Hernquist 2010, and references therein).

Morphology of radio jets. It has also been suggested that SMBH binaries and mergers may be responsible for radio jets with peculiar shapes. The jet in the galaxy 1928+738 shows regularly spaced wiggles, which Roos et al. (1993) speculated may be caused by the precession of the jet caused by the orbital motion of a binary SMBH. Merritt & Ekers (2002) interpreted jets with X-shaped morphologies to be caused by a sudden change in the SMBH spin axis due to a merger event. Gopal-Krishna et al. (2003) also advocated a merged SMBH origin for X-shaped radio sources, and argued that binaries en route to coalescence are responsible for Z-symmetric lobes. Liu et al. (2003b) have suggested

that double-double radio lobes owe their shape to a compact binary stage of evolution, during which the binary torques suppress accretion and jet formation (see Chapter 3). However, other studies have pointed out that such morphologies could result from the outflows interacting with an anisotropic or asymmetric gaseous environment (e.g., Leahy & Williams 1984; Kraft et al. 2005; Saripalli & Subrahmanyan 2009).

Displaced AGN. Magain et al. (2005) reported the diskovery of an enigmatic quasar that was located near a companion galaxy but had little or no stellar host of its own. Hoffman & Loeb (2006) suggested that the quasar could have been ejected from the companion galaxy by triple-SMBH interactions (see, however, Merritt et al. 2006). Recently, Batcheldor et al. (2010) reported that the bright nuclear point-source of M87 is displaced from the light-center of the galaxy by ~ 7 pc; a possible explanation for the offset is that this central SMBH suffered a recoil event in a relatively recent merger ($\lesssim 1$ Gyr ago) .

1.3 Gravitational Waves from Merging SMBH Binaries

Inspiraling black holes and neutron stars have long been theorized to be extremely luminous sources of GWs (see, e.g., Misner et al. 1973; Shapiro & Teukolsky 1986; Blanchet 2006, for reviews). GW emission in compact binaries has been indirectly observed through the orbital decay in binary neutron stars (see the review by Weisberg et al. 2010). The first direct detections of GWs are anticipated, perhaps during this decade, following the next round of upgrades to operational ground-based detectors (e.g., Schutz 1999).

For the purposes of this thesis, we are interested only in the superficial consequences of GW emission from SMBH binaries; that is to say, the types of systems that can be

detected, the physical properties that can be determined, and the accuracy with which we can measure them. Below, I provide a cursory overview to suit our purposes (for a more detailed treatment, see Flanagan & Hughes 2005).

Most of the detection techniques for detecting GWs utilize laser interferometers with extremely long arms. In the simplest terms, the goal of these observatories is to continuously time the light signals traveling along the arms and thus measure as a function of time the fractional change $\delta L/L$ of the lengths of the arms. The change in the arm length is related to the perturbation of the spatial component of the metric h, or "strain", as $\delta L/L \sim h/2$. If we assume a linearized deviation from a flat metric, i.e. that the true metric $g_{\mu\nu}$ is the sum of the flat metric $\eta_{\mu\nu}$ and a small strain metric $h_{\mu\nu}$, then the Einstein equation reduces to a simple wave equation (Misner et al. 1973)

$$G_{\mu\nu} = 8\pi T_{\mu\nu} = \frac{1}{2}\Box\bar{h}_{\mu\nu}. \tag{1.9}$$

Above, $\Box \equiv \nabla^2 - \partial^2/\partial t^2$ is the usual wave operator, $G_{\mu\nu}$ is the Einstein tensor, $T_{\mu\nu}$ is the matter-energy source tensor and $\bar{h} = h_{\mu\nu} - (1/2)\eta_{\mu\nu}h$.

For a point source, the solution to the wave equation is easily obtained by use of the Green's function, (e.g., Flanagan & Hughes 2005). Skipping the mathematical details, the end result is that the spatial components of the strain $\bar{h}_{ij}$ are given in terms of the moment of inertia tensor I_{ij} of the source as

$$\bar{h}_{ij} = \frac{2G}{c^4}\frac{1}{r}\frac{d^2}{dt^2}I_{ij}(t - r/c), \tag{1.10}$$

where r is the distance to the source and I_{ij} is evaluated at the retarded time $t - r/c$. For a binary system with orbital separation a, the nonzero components of I_{ij} have a magnitude $I \sim \mu a^2$, where $\mu = M_1 M_2/M = \eta M$ is the reduced mass of the binary. The characteristic scale of its second time derivative is $d^2 I/dt^2 \sim \mu a^2 \Omega^2 = \eta G^{2/3} M^{5/3} \Omega^{2/3}$. Neglecting a few geometrical factors, the characteristic magnitude of the strain due to GWs from a rotating binary is

$$h \sim \frac{G^{5/3}}{c^4} \eta M^{5/3} \frac{\Omega^{2/3}}{r}. \tag{1.11}$$

For binaries that are decaying gradually, in the so-called adiabatic stage, $h(t)$ oscillates at a characteristic frequency $f_{\rm GW}$ that is twice the orbital frequency. If the source is located at cosmological distances, then the dependence on redshift z is introduced as

$$h \sim \frac{G^{5/3}}{c^4} \frac{\mathcal{M}_z^{5/3}}{D_L} f_{\rm GW}^{2/3}, \tag{1.12}$$

where we have substituted the redshifted "chirp mass" $\mathcal{M}_z \equiv \eta^{3/5} M(1+z)$. Note that the strain amplitude decays only as D_L^{-1}. This, coupled with the fact that GWs hardly interact at all with matter, is the reason why interferometry experiments are expected to detect GWs from extremely distant sources. Substituting representative values for the strain sensitivity and frequency window of the proposed *LISA* detector (see 1.3.1 below), we may estimate the maximum distance at which it can detect a compact, equal-mass SMBH binary:

$$D_L \lesssim 200 \left[\frac{M(1+z)}{10^6 M_\odot}\right]^{5/3} \left(\frac{h}{10^{-23}}\right)^{-1} \left[\frac{f_{\rm GW}}{0.01\ {\rm Hz}}\right]^{2/3} {\rm Gpc}, \tag{1.13}$$

which corresponds to $z \sim 15$ in a standard ΛCDM cosmology.

The frequency $f_{\rm GW}$ evolves with the binary, as the orbit of the latter decays with GW emission. For example, to leading Newtonian order, circular, non-spinning binaries in adiabatic inspiral satisfy (Peters & Mathews 1963; it also can be derived directly from equation 1.8)

$$\frac{df_{\rm GW}}{dt} = \frac{96\pi^{8/3}}{5}\mathcal{M}_z^{5/3} f_{\rm GW}^{11/3}. \tag{1.14}$$

In this simplified picture we have two equations (1.12, 1.14) with two unknowns: $\mathcal{M}_z$ and D_L. In principle, GW observations of an inspiraling binary can give a direct measurement of the luminosity distance to the source. In practice, the problem is much more complicated, as the strain is a three-dimensional quantity that depends on the orientation of the source, and affected by the spins and orbital eccentricity of the binary. The Newtonian and post-Newtonian approximations rapidly break down as the binary approaches coalescence, and numerical treatments are required to compute the waveform for the merger (see Figure 1.4).

The problem of calculating fully relativistic solutions to the waveform had been a longstanding problem (Ehlers et al. 1976; Thorne 1980) that was only recently solved (Pretorius 2005; Campanelli et al. 2006; Baker et al. 2006a). It is now possible, given a specific set of binary properties and orbital parameters, to calculate the exact waveform $h(t)$, albeit at considerable computational expense. Doing the reverse is a complex problem that is very much an active area of research (e.g., Cornish & Crowder 2005; Cornish & Porter 2007; Arun et al. 2009). It is believed that by observing the full wave form down to coalescence and the subsequent "ringdown", it will be possible to determine the masses, spins, distance, and the approximate sky location of the source (Echeverria 1989; Cutler &

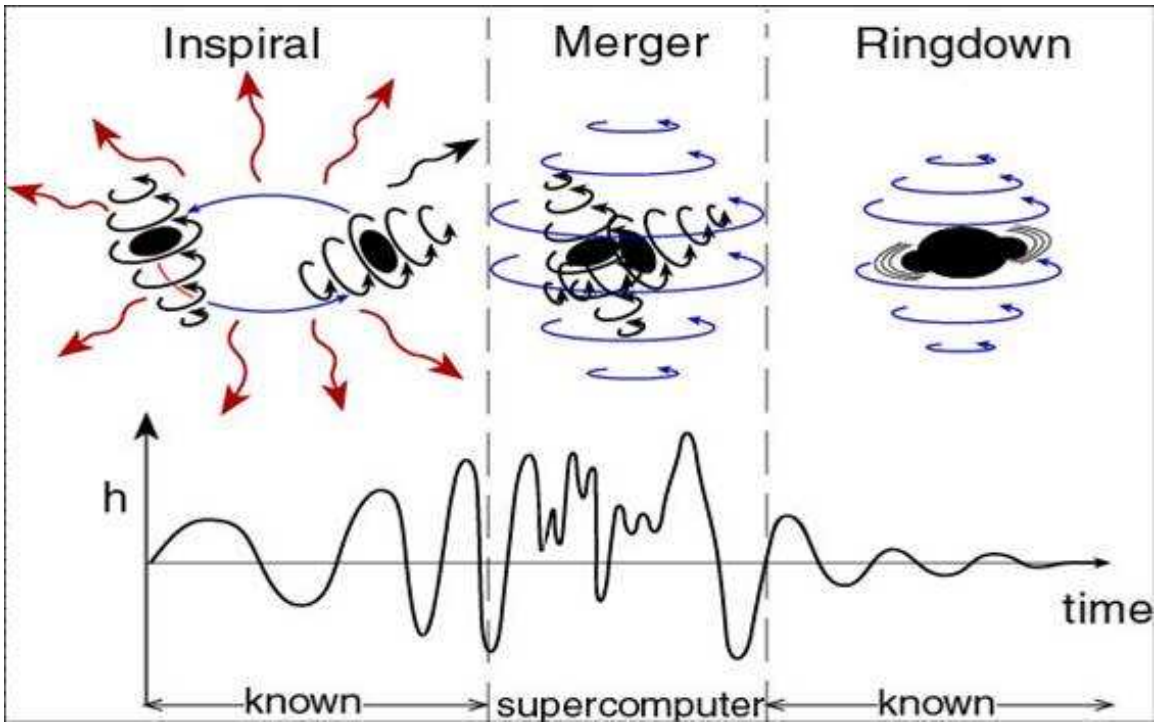

Figure 1.4 A cartoon of the three stages of black-hole coalescence. The GW waveform for the inspiral stage, in which the binary evolves adiabatically due to gradual emission of GWs, can be calculated with post-Newtonian methods. The waveform during the ringdown stage, which follows the merger and may be thought of as a "settling" stage for the newly formed hole, can also be calculated. Until 2005 (see references in text), the waveform for the merger itself could not be computed in a fully relativistic treatment. (Note that the wiggles of the merger waveform are exaggerated to emphasize the lack of understanding at the time this image was created.) Image credit: Kip Thorne.

Flanagan 1994; Flanagan & Hughes 1998; Vecchio 2004; Berti et al. 2005; Lang & Hughes 2006; Arun et al. 2009).

1.3.1 Proposed GW-detection Experiments

I review the currently operational and planned GW detection efforts below. For detecting SMBH binaries, the most relevant experiments are *LISA* (or a similar space-borne

interferometer) and pulsar timing arrays.

Space-based detector *LISA*. NASA and ESA had planned the joint GW-detection mission *LISA* (*Laser Interferometer Space Antennae*). The GW detector design consists of three interferometer modules placed in drag-free satellites in an equilateral-triangular orbit around the Sun. The extremely long (5 million km) "arm" lengths of the interferometer would allow for extremely sensitive measurements of the GW strain ($h \gtrsim 10^{-23}$). The detector design is expected to allow for high-fidelity detections of the GWs in the frequency range $\sim 10^{-4} - 0.1$ Hz. The science capabilities of the mission include:

- Detection of SMBH coalescences, in the mass range $\sim 10^{4-7}(1+z)^{-1}M_{\odot}$, out to redshifts of $z \sim 20$ (Figure 1.5; Baker et al. 2007). *LISA* observations can be used to determine the masses, luminosity distance and spins of the source SMBHs to percent-level precision (Echeverria 1989; Cutler & Flanagan 1994; Flanagan & Hughes 1998; Vecchio 2004; Berti et al. 2005; Lang & Hughes 2006; Arun et al. 2009), as well as the orbital eccentricity to $\sim 10^{-3}$ (Damour et al. 2004; Königsdörffer & Gopakumar 2006, and references therein). The data will reveal the cosmological evolution of the SMBH population, and determine how much mergers — as opposed to accretion — contribute to their growth (Sesana et al. 2005, 2007b). The measurements of spins of SMBHs in a wide range of redshifts can also determine how SMBH spins have evolved with cosmic time, and constrain accretion mechanisms (Berti & Volonteri 2008). The dominant source of systematic error is expected to be weak-lensing distortions of the signal (Holz & Hughes 2005; Kocsis et al. 2006). The weak-lensing error is expected to be non-Gaussian, with strong kurtosis, which may mitigate its

effects on *LISA* data (Hirata et al. 2010; Shang & Haiman 2011) *LISA* is expected to be able to localize the source to $\sim 1 - 10\,\mathrm{deg}^2$ on the sky (Cutler 1998; Kocsis et al. 2006; Lang & Hughes 2008). If the weak-lensing errors can be reduced, localization of the source on arcminute scales may be possible by incorporating effects due to spin precession (Lang & Hughes 2006) or higher-order harmonics (McWilliams et al. 2010) in the signal analysis.

- Detection of extreme-mass ratio inspiral (EMRI) events — i.e., the inspiral of $M \sim 1 - 10 M_\odot$ compact objects into a SMBH — out to $z \sim 1$, at rates of dozens to hundreds per year (Gair et al. 2004; Levin 2007; Amaro-Seoane et al. 2007, and references therein). Waveforms from EMRIs can constrain the masses of the SMBH and the inspiraling compact object, the SMBH spin to $\sim 10^{-4}$, and the location on the sky to $\sim 3\,\mathrm{deg}$ (Barack & Cutler 2004).

- Tests of alternative theories of gravity, in which damping of GWs and the graviton mass and propagation speed differ from general relativity (Will 1998; Will & Yunes 2004; Berti et al. 2005, e.g.,). *LISA*, along with the other detectors described here, may detect GWs produced by cosmic strings, or place observational constraints on string theories (Maggiore 2000; Damour & Vilenkin 2005; Siemens et al. 2007).

See Chapter 7 for recent developments and future uncertainty of the *LISA* mission.

Pulsar timing arrays (PTAs). Pulsars are extremely precise cosmic clocks, often having arrival time uncertainties of only $\sim 100\mu$s. Sazhin (1978) proposed that prolonged observations of pulsars could detect extremely long-period GWs ($P_{\mathrm{GW}} \sim 1 - 10$ yr; or $f_{\mathrm{GW}} \sim$ nHz). Detweiler (1979) suggested that pulsar timing can detect or place upper limits

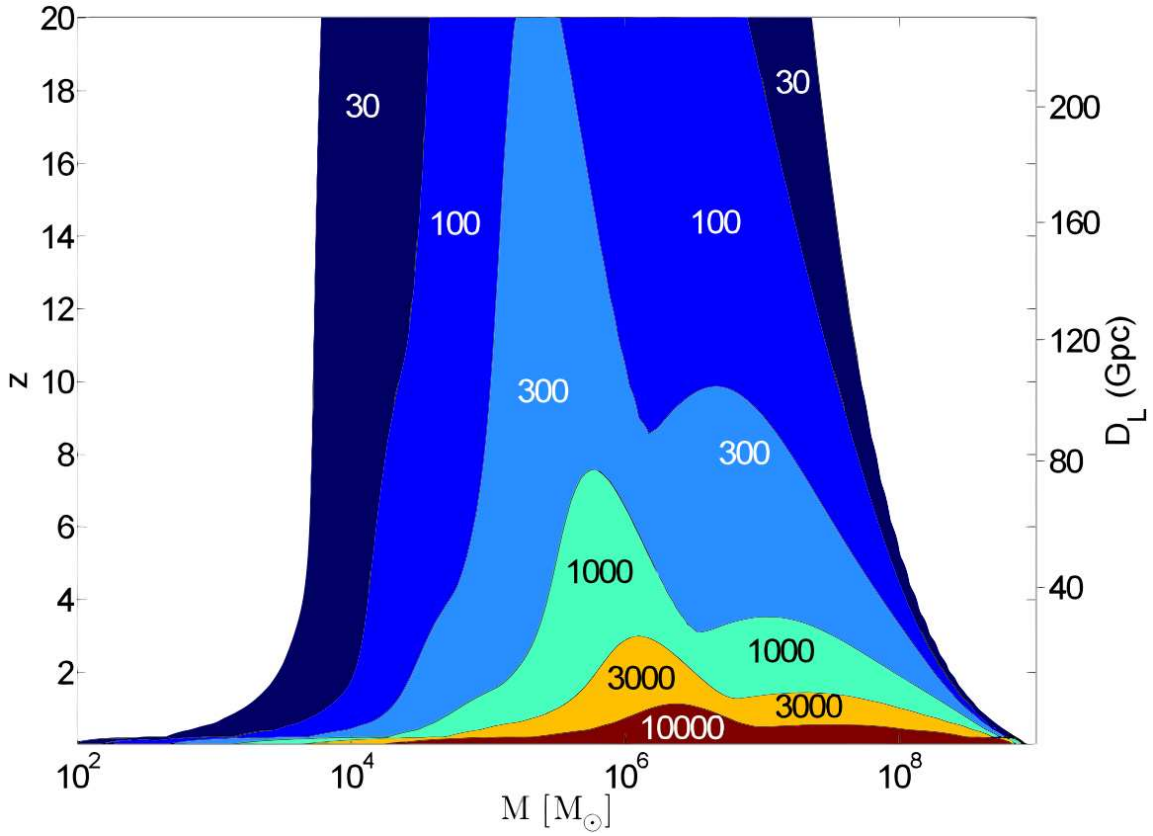

Figure 1.5 From Baker et al. (2007), model contours of the signal-to-noise ratio of the proposed *LISA* detector for equal-mass, non-spinning SMBH binaries, as a function of mass M and redshift z. Unequal masses will decrease the signal-to-noise of the merger and ring-down stages, roughly in proportion to the symmetric mass ratio $\eta = M_1 M_2 / M^2$. The signal-to-noise reduction in the adiabatic inspiral scales as $\sim \eta^{1/2}$. Spins can act to either increase or decrease the signal amplitude.

on the stochastic background of low-frequency GWs. Correlations of timing residuals from an array of multiple pulsars can dramatically enhance the fidelity of the detection (Foster & Backer 1990). Currently, an international effort is underway to perform long-term observations of the most accurately timed pulsars for this purpose (Manchester 2008; Jenet et al. 2009; Hobbs et al. 2010; for recent limits on the GW background, see Jenet et al. 2006). The most definitive source of such a background is the population of massive ($M > 10^7 M_\odot$) extragalactic SMBH binaries at orbital separations of $10^{2-3} GM/c^2$ (e.g., Wyithe

& Loeb 2003a; Sesana et al. 2004, 2008). Recently, Sesana et al. (2009) suggested that the most massive and/or nearby binaries may stick out above the stochastic background, and be individually resolved.

Ground-based detectors. There are currently several ground-based GW interferometers that are operational, including *LIGO*[2], *VIRGO*[3], and *GEO 600*[4]. These detectors are designed to detect high-frequency ($10 - 10^4$ Hz) GWs emitted by coalescing binaries of stellar-mass black holes and neutron stars in the Galaxy and in nearby galaxies. Future detectors of similar design are planned or under consideration (e.g., *AIGO*, *INDIGO*, *LCGT*). Having multiple detectors located far apart is important for determining the sky location of the source (since the strain is three-dimensional). Although no detections have been made, *LIGO* has been able to place upper limits on the emission of GWs from ~ 30 galactic pulsars (Abbott et al. 2005). It is anticipated that the planned upgrades of the *LIGO* detector to Advanced *LIGO* in 2014 could result in multiple and frequent direct detections.

Space-based detector *DECIGO***.** The proposed Japanese mission *DECIGO* (*DECi-hertz Interferometer Gravitational wave Observatory*; Kawamura et al. 2006), like *LISA*, is designed as three interferometer modules in orbit around the Sun. The arm length between the modules is planned to be $\sim 10^3$ km, with a target frequency range $0.1 - 10$ Hz specifically chosen to bridge the "blind spot" between *LISA* and ground-based detectors. (It is uncertain whether this mission will be redesigned if significant changes are made to the *LISA* mission goals.) The main astrophysical sources for *DECIGO* are expected to

[2]*Laser Interferometer Gravitational Wave Observatory*, `http://www.ligo.caltech.edu`
[3]`http://www.ego-gw.it`
[4]`http://www.geo600.org`

be the coalescence of intermediate-mass black holes ($M \sim 10^{3-4} M_{\odot}$) and compact neutron star binaries out to $z \sim 1$. The mission is tentatively planned for launch during the next decade, following two exploratory missions to test design and feasibility.

1.4 Multi-messenger Astronomy with Compact SMBH Binaries

Recently, intense attention has focused on the possibility of detecting a merging SMBH binary synergistically through both its GW and EM signatures (Cutler 1998; Holz & Hughes 2005; Dotti et al. 2006; Kocsis et al. 2006, 2007, 2008; Stubbs 2008; Phinney 2009; Bloom et al. 2009). The complementary information extracted from each type of astronomical messenger — from GWs, the masses and spins of the black holes, and the luminosity distance to the source; from EM signatures, the source redshift — can address questions that are difficult to answer with either messenger alone, and thus dramatically enhance the scientific yield of GW detectors:

- **The use of SMBH mergers as cosmological "standard sirens."** EM measurements of source redshifts, together with measurements of luminosity distances by *LISA*, can precisely determine the history of cosmic expansion (Schutz 1986; Holz & Hughes 2005) and constrain models of dark energy. Such a study would be independent from the use of type Ia supernovae as "standard candles," collect data at higher redshifts (out to $z \gtrsim 10$ for *LISA*), have significantly less systematic error, and possibly probe the existence of extra dimensions (Deffayet & Menou 2007).

- **Detailed probes of AGN physics.** The determination of SMBH masses, spins and orientations through GWs will allow tests of fundamental AGN hypotheses. For example, multi-messenger observations can more directly measure the Eddington ratios of AGN and calibrate mass-measurement techniques such as reverberation mapping. They can also test whether the power of AGN jets scales with the SMBH spin, and whether the jets are directed along the spin axis (Blandford & Znajek 1977; Narayan & Quataert 2005). If the EM emission exhibits an observable response to the orbital motion of a compact or merging SMBH binary or a recoiling merger remnant, this could shed light on the physical properties of the accreting gas.

- **Corroborate tentative GW detections.** EM detections can confirm putative GW sources with low signal-to-noise, such as distant lower-mass sources and binaries resolved by pulsar timing (Sesana et al. 2009; Kocsis & Sesana 2011).

Whereas GW signatures of merging SMBHs boast precise theoretical predictions that await future empirical tests, the state of science for the EM signatures is virtually the opposite. Current telescopes are already capable of *detecting* most of the EM signatures that have been proposed in the literature, at least at redshifts $z \lesssim 3$, where GW detections of SMBH binaries by *LISA* and PTAs are expected to be most common. The difficulty lies in correctly *identifying* an EM counterpart with a theoretical model.

The nature of the problem is twofold. First, from an observational standpoint, AGN exhibit intrinsic variability at virtually all wavelengths and timescales (Mushotzky et al. 1993; Ulrich et al. 1997). Thus, in order to be distinguished from ordinary variability, an EM signature must have a precise timing prediction, exhibit extraordinary characteristics,

or both.

Second, an EM counterpart is inherently a "hairy" problem. Consider the computation of the GW waveform from a compact black hole binary. That problem is complex, but cleanly defined: the evolution of the spacetime is computed in a vacuum, and depends only on the masses, spins, positions and velocities of the binary. GW predictions also enjoy the property that the messenger does not interact with matter; transferring the predictions of the waveform at the source to what is observed by the detector is relatively straightforward, with the chief systematic contaminant being the weak lensing distribution. In contrast, the intrinsic properties of the EM counterpart may depend on such factors as the stellar and gaseous distributions, inhomogeneities in the same, magnetohydrodynamics, radiative transport, etc. The *observed* properties could differ significantly from what is produced at the source, due to obscuration and reprocessing. Even if a theoretical model were able to account for all of the relevant physical processes at the appropriate levels of detail — this is a mountainous prerequisite, given that the detailed astrophysics of AGN emission is poorly understood even in the case of a solitary SMBH engine — uncertainty regarding the correct initial conditions and environment could severely undermine its predictive power.

The difficulties listed above may create the impression that understanding EM signatures of SMBH binaries is a nightmarishly intractable problem. However, all of these difficulties are common to many astrophysical processes, and especially acute for AGN. Exchanging a solitary SMBH engine with a binary SMBH does not add any fundamental theoretical uncertainty. In fact, a SMBH binary can add clarity to the problem because it

introduces additional non-free, "hairless" parameters that determine characteristic physical scales of the system. Just as the mass and spin set relevant scales in power ($L_{\rm Edd}$) and length ($R_{\rm ISCO}$), the orbital separation and eccentricity of the binary establish relevant timescales: the orbital period and the orbital decay timescale (equation 1.8).

1.4.1 Overview of Proposed Electromagnetic Signatures of Merger

Over the last decade, numerous models of EM signatures of merging SMBH binaries have been proposed. I briefly review some of these signatures and their underlying physical mechanisms[5] (cf. Schnittman 2011).

Tidal excitation of fossil gas. At sufficiently close separations, the tidal torques of a binary can inject orbital angular momentum into the surrounding material at a "tidal radius" $\sim 2a$ (Goldreich & Tremaine 1980; see Chapter 3). A geometrically thin circumbinary gas disk will be prevented from accreting inward of the tidal radius. As the binary accretes the gas inside this radius, it is expected to form and maintain a central cavity around its orbit (Artymowicz et al. 1991). Chang et al. (2010) showed that even if the amount of gas remaining in the cavity is small, it can experience strong tidal forcing as the binary "squeezes" the intrabinary gas during the final stages of inspiral. This may lead to a powerful luminous precursor several days before the merger.

Tidally induced periodic precursor. Several studies have pointed to the possibility that binary systems can elicit periodic emission signatures in their accretion flows. For example, the binary torques may modulate the accretion rate onto one of its members

[5]I exclude the X-ray afterglow mechanism of Milosavljević & Phinney (2005) from this review, as it is described at length in Chapters 3 and 4.

at some harmonic of the orbital frequency; this periodic behavior is strongly enhanced for eccentric binaries (Artymowicz & Lubow 1996; Hayasaki et al. 2007; Cuadra et al. 2009). A binary can also induce quasi-periodic spiral density waves in an accretion disk, even if it is circular (MacFadyen & Milosavljević 2008). Hayasaki & Okazaki (2009) suggested that the tidal excitation of the inner edge of a circumbinary disk could generate nonaxisymmetric, precessing waves, which may elicit periodic light curves in the disk. Such periodic behavior could constitute a pre-merger EM counterpart to a SMBH binary en route to merger (Haiman et al. 2009b). A population of such close, wet binaries could be detected statistically with future wide-field, high-cadence EM surveys (Haiman et al. 2009b).

Disk heating due to mass loss and recoil. As described above, during the final stages of the merger, GW emission can extract a few percent of the binary's mass and impart a recoil velocity as high as $\sim 10^3$ km s^{-1}. These changes occur on timescales that are much shorter than the response timescale of the gas. Any gas orbiting the binary will suddenly be in "wrong" orbits with "excess" eccentricity. The orbits cross to form shocks in the gas, whose energy is subsequently radiated as an EM afterglow. This mechanism has recently been investigated by several groups (Lippai et al. 2008; Schnittman & Krolik 2008; Shields & Bonning 2008; O'Neill et al. 2009; Megevand et al. 2009; Anderson et al. 2010; Corrales et al. 2010; Rossi et al. 2010). The rough consensus appears to be as follows:

- The onset of the afterglow and its lifetime are sensitive to the SMBH mass, with low-mass ($M \sim 10^6 M_\odot$) systems exhibiting prompt brightening on timescales of days to months. For high-mass ($M \sim 10^8 M_\odot$) binaries, the afterglow develops much more

slowly and may last for thousands of years.

- As can be inferred from simple analytic calculations (see Schnittman & Krolik 2008; Rossi et al. 2010), the afterglow is highly sensitive to the kick direction, with kicks along the plane of the disk resulting in the largest shocks and luminosities.

Emission from the binary's interactions with EM fields. Palenzuela et al. (2009, 2010b) noted that strong interactions of the binary with nearby EM fields could enhance or induce EM emission. Their preliminary calculations indicate that such interactions could be strongly modulated by the binary's orbital motion and lead to quasiperiodic EM signatures. Palenzuela et al. (2010a) showed that in a mechanism analogous to the Blandford-Znajek mechanism (Blandford & Znajek 1977) — in which magnetic fields propel jets that are powered by the spin energy of a solitary spinning black hole — a binary black hole can generate jets by extracting its orbital energy.

Tidal disruption events. Recently, Chen et al. (2009, 2011) noted that core scouring by a binary SMBH can dramatically enhance the rate of tidal disruption events — stars being ripped apart in the tidal field close to the SMBH (Rees 1988; Ayal et al. 2000; Zhao et al. 2002; Wang & Merritt 2004; Bogdanović et al. 2004). The rate of events could increase by as much as four orders of magnitude compared to the rate triggered by solitary SMBH, to $\gtrsim 0.1$ yr^{-1}, and a significant fraction ($\sim 10\%$) of all such events may occur in galaxies harboring SMBH binaries. Stone & Loeb (2011) pointed out that a recoiling SMBH could promptly cause similarly large increase in the event rate, compared to one that is stationary. Observations of tidal disruption events have only recently begun to be made (Gezari et al. 2006); ongoing and future transient surveys such as the *Palomar Transient Factory* and the

Large Synoptic Survey Telescope are likely to detect tidal disruption events (Gezari et al. 2009), making them exciting future probes of the population of dark SMBHs.

Coupling of GWs with viscous gas. Kocsis & Loeb (2008) suggested that the weak coupling of GWs with matter (Hawking 1966; Esposito 1971) could lead to small amounts of the GW-emitted energy being absorbed by surrounding gas, and viscously dissipated in the form of a prompt, observable afterglow.

1.5 Accretion Theory Overview

Above, I stressed that the compact size of the central accretion region, down to scales of $R_{\rm ISCO} \gtrsim GM/c^2$, is the key to explaining quasar power (1.1), and alluded several times to the presence of a disk-like accretion flow around the binary. However, I have neglected to address the important issue of how gas is able to descend to those scales; that is, the problem of angular momentum. Below, I present a short synopsis of astrophysical accretion theory, aimed at providing background for the rest of the thesis. For more comprehensive reviews, see Pringle (1981), Frank et al. (2002) and Blaes (2004).

1.5.1 Bondi-Hoyle-Lyttleton Accretion

Bondi (1952) considered the accretion of an ideal fluid onto a central object, which accretes matter inside the radius $\approx GM/c_{\rm s}^2$, where $c_{\rm s}$ is the characteristic sound speed in the fluid. In the standard solution to this problem, the fluid inside the above radius is supersonic and falls to the center along radial orbits. Hoyle & Lyttleton (1939) and Bondi & Hoyle

(1944) studied the accretion due to the motion of an object moving at a velocity v through a medium of uniform density ρ. In this problem, the object captures the mass inside a critical impact parameter $\approx 2GM/v^2$. The combination of these two scenarios is known as Bondi-Hoyle-Lyttleton (BHL) accretion (see Edgar 2004, for a concise review). The central object captures the fluid that crosses within a capture radius

$$R_{\rm BHL} \sim \frac{GM}{v^2 + c_{\rm s}^2}, \tag{1.15}$$

at an accretion rate of

$$\dot{M}_{\rm BHL} \sim 4\pi \frac{G^2 M^2 \rho}{(c_{\rm s}^2 + v^2)^{3/2}}. \tag{1.16}$$

We invoke BHL theory in Chapter 2 for the purposes of modeling the accretion rate onto recoiling black holes.

The BHL treatment gives a simple analytic estimate for the radius inside which gas becomes strongly bound to a massive object. However, because it does not consider at all the angular momentum of the captured gas, BHL theory cannot provide a realistic picture for what happens inside the capture radius. In virtually all realistic astrophysical scenarios $R_{\rm BHL} \gg GM/c^2$, and thus BHL accretion is ill-suited for modeling the inner regions of accretion flows onto black holes.

1.5.2 Thin Accretion Disks

For gas to accrete down to GM/c^2 scales, it must transfer its orbital angular momentum elsewhere. Let us consider viscous fluids, which can transport angular momentum inter-

nally. Below, I present an abbreviated derivation of the so-called thin-disk equation; more complete derivations may be found in Pringle (1981) and Frank et al. (2002).

Consider a homogeneous blob of a viscous fluid, possessing negligible self-gravity and in orbit around a central mass. Shear viscosity converts the kinetic energy of shearing motion in the fluid to thermal energy. If the fluid is *radiatively efficient* — that is, if it is able to quickly radiate away the locally generated thermal energy — then the blob will evolve toward a state in which each fluid element has the minimal possible orbital energy with respect to its angular momentum, i.e., toward circular orbits. If the timescale on which viscous torques can evolve the orbit is much longer than the orbital period, then the effect of the torques on the orbits will be to adiabatically move circular orbits inward or outward. Any vertical and non-orbital azimuthal bulk motion will also be viscously damped. All of the above imply that the long-term configuration of this ideal blob should be an axisymmetric disk consisting of fluid elements in Keplerian circular orbits. For the moment, let us ignore for simplicity the possibility of winds or radiative feedback.

What is the radial evolution of the disk due to internal viscous stresses? Everywhere in the flow, three conditions hold: mass is conserved, angular momentum is conserved, and angular momentum transport is mediated by shear stress. For the first two conditions, one can simply write down the general conservation equations. Some time can be saved by taking advantage of the azimuthal symmetry of the problem and the irrelevance of vertical structure: integrations along the ϕ and z coordinates are trivial, and the only velocity component that survives the divergence operator is the as-yet undetermined radial flow

speed v_R. Writing the mass distribution in terms of the surface density $\Sigma \equiv \int \rho \, dz$,

$$2\pi R \frac{\partial \Sigma}{\partial t} + \frac{\partial}{\partial R}(2\pi \Sigma v_R R) = 0, \tag{1.17}$$

$$\frac{\partial}{\partial t}\left(\Sigma \Omega R^2\right) + \frac{1}{R}\frac{\partial}{\partial R}\left(\Sigma \Omega v_R R^3\right) = \frac{1}{2\pi R}\frac{\partial \mathcal{T}}{\partial R}. \tag{1.18}$$

In equation 1.17, I have left the factor 2π to underscore the fact that the quantity in parentheses is the radial mass flux $\dot{M}$. Above, $\Omega(R)$ is the Keplerian angular orbital frequency, and $\partial \mathcal{T}/\partial R$ is the radial torque density. The latter can be calculated from the viscous shear stress $S = -\rho \nu R \, \partial(\Omega)/\partial R$:

$$\mathcal{T} = R \int R \, d\phi \int S \, dz = 3\pi \nu \Sigma \Omega R^2, \tag{1.19}$$

where ν is the kinematic viscosity.

After some manipulation to substitute for v_R, one obtains the thin-disk equation:

$$\frac{\partial}{\partial t}\Sigma(R,t) = \frac{3}{R}\frac{\partial}{\partial R}\left[R^{1/2}\frac{\partial}{\partial R}\left(\nu \Sigma R^{1/2}\right)\right]. \tag{1.20}$$

Time-dependent solutions of this equation are used extensively in much of the thesis. Chapter 5 is devoted to deriving new time-dependent solutions to equation 1.20.

The basic behavior of the thin disk is as follows (Pringle 1981). The viscous stress acts to dissipate differential rotation in the flow, doing so at the expense of orbital energy of the disk. Viscosity thus acts to drives mass inward to the center, while transporting angular momentum outward. Of course, a Keplerian flow is always differentially rotating, so a

radiatively efficient Keplerian disk must viscously spread indefinitely. The thin, viscous accretion disk provides a simple mathematical formalism that can consistently transport gas to some central massive object. The local inward mass flux is given by the expression

$$\dot{M}(R) = R^{1/2}\frac{\partial}{\partial R}\left(6\pi\nu\Sigma R^{1/2}\right). \tag{1.21}$$

Let us diskuss the steady-state solution ($\partial\Sigma/\partial t \to 0$) and some of its basic properties. For the quantity inside square brackets in equation 1.20 to vanish nontrivially, the quantity $\nu\Sigma$ must have a radial dependence of the form $\nu\Sigma = A + BR^{-1/2}$, where A and B are constants of integration. In a steady-state disk, there will be no local accumulation or depletion of mass, and so the mass flow $\dot{M}$ must be a constant with radius. The constant A sets the magnitude of the mass flux, whereas the term $BR^{-1/2}$ contributes a constant radial viscous torque while contributing zero radial transport. The general tendency of the disk is to transport mass inward and angular momentum outward; thus, unless there is inner source of angular momentum or a boundary condition prohibiting inward mass flux, solutions with $B > 0$ are physically untenable. In other words, if the gas is free to accrete inward, we expect a steady-state solution with $\dot{M} = 3\pi\nu\Sigma$ = constant.

What happens at the ISCO radius? The accretion disk must truncate at $R_{\rm ISCO}$, so there can be no viscous torques exerted at this radius. We also assume that the angular velocity near $R_{\rm ISCO}$ does not strongly deviate from the Keplerian value. Integrating the angular momentum equation from $R_{\rm ISCO}$ to some arbitrary radius R, and applying the steady-state

conditions $\partial\Sigma/\partial t = 0$ and $\dot{M} \equiv 2\pi R v_R \Sigma$ = constant, we obtain

$$\dot{M}R^2\Omega = \mathcal{T}(R) - \mathcal{T}(R_{\rm ISCO}) + \text{constant} = 3\pi\nu\Sigma\Omega R^2 + \text{constant}. \tag{1.22}$$

Since the disk truncates at $R_{\rm ISCO}$, the constant quantity above must equal $\dot{M}\Omega(R_{\rm ISCO})R_{\rm ISCO}^2 \propto R_{\rm ISCO}^{1/2}$. Therefore, the disk must satisfy

$$3\pi\nu\Sigma = \dot{M}\left(1 - \sqrt{\frac{R_{\rm ISCO}}{R}}\right). \tag{1.23}$$

The rate at which orbital energy is dissipated in the disk per unit area is

$$F = \frac{1}{2}\nu\Sigma\left(R\frac{\partial\Omega}{\partial R}\right)^2 = \frac{9}{4}\nu\Sigma\Omega^2 = \frac{3}{4\pi}\dot{M}\Omega^2\left(1 - \sqrt{\frac{R_{\rm ISCO}}{R}}\right). \tag{1.24}$$

Integrating across the entire disk, and multiplying by two to account for the fact that the disk dissipates energy from two faces,

$$\begin{aligned} L &= \int_{R_{\rm ISCO}}^{\infty} \frac{3}{4\pi}\dot{M}\Omega^2\left(1 - \sqrt{\frac{R_{\rm ISCO}}{R}}\right) 2\pi R\, dR \\ &= \frac{1}{2}\frac{GM\dot{M}}{R_{\rm ISCO}}. \end{aligned} \tag{1.25}$$

One can compute the spectral energy distribution of the disk by assigning a Planck function to the bolometric flux output in equation 1.24, then integrating across the area of the disk as a function of frequency (see Chapter 3). To do this it is also necessary to

calibrate the bolometric blackbody emission via the midplane temperature T:

$$2\sigma_{\rm SB}T^4 = \frac{9}{4}\nu\Sigma\Omega^2, \tag{1.26}$$

where $\sigma_{\rm SB}$ is the Stefan-Boltzmann constant. The factor of 2 on the left-hand side accounts for the fact that the disk emits from two faces. If the disk is optically thick, it will be necessary to compute the thermal emission at the photon escape height and characterize the optical depth τ, e.g. with a one-zone model for the disk atmosphere (Blaes 2004). The flux intensity is highest near the center of the disk, and so the gas closest to the black hole is expected to produce the brightest and hardest emission.

There is no vertical transport, so the disk should be in hydrostatic equilibrium along this direction. If the self-gravity of the disk is negligible, then the local gravitational field Φ is dominated by the central object, so that at a radius R and height z the vertical gravitational field per unit mass is given by

$$g = \nabla\Phi\cdot\hat{z} = \frac{\partial}{\partial z}\frac{GM}{\sqrt{R^2+z^2}} = -z\frac{GM}{(R^2+z^2)^{3/2}} \approx -\Omega^2 z, \tag{1.27}$$

where $\hat{z}$ is the unit vector in the vertical direction and the approximation is valid for $z \ll R$. Hydrodynamic equilibrium requires that

$$\frac{1}{\rho}\frac{\partial P}{\partial z} = -g \approx \Omega^2 z. \tag{1.28}$$

Integrating from the disk midplane to a scale height H where $P(z=H) \ll P(z=0)$, we see

that $P_\rho/\rho \equiv c_s^2 \sim H^2\Omega^2$, where c_s is the sound speed.

If the self-gravity of the disk is appreciable, the above approximation for H breaks down. Assuming that the mass of the disk is concentrated along the equator and that the surface density is a weak function of radius (Paczynski 1978),

$$g_{\rm disk} = \frac{\partial}{\partial z}\Phi_{\rm disk}(R,z) = -\frac{\partial}{\partial z}\int \frac{G\Sigma}{\sqrt{(R-R')^2+z^2}} 2\pi R' dR' \approx -2\pi G\Sigma. \tag{1.29}$$

Integrating the hydrostatic equilibrium equation and normalizing to our previous result, we obtain

$$c_s^2 \sim H^2\Omega^2 + 4\pi G\Sigma H \tag{1.30}$$

for a self-gravitating disk.

1.5.3 The Shakura-Sunyaev α Disk

The thin-disk model gives a practical framework in which to study viscous flows onto black holes. However, in the absence of a physical viscosity prescription, the surface density and temperature profiles of the disk cannot be computed.

Shakura & Sunyaev (1973) proposed a heuristic viscosity prescription, in which the shear stress S due to a kinematic shear viscosity ν scales with the pressure P multiplied by a dimensionless parameter α:

$$S \sim \rho\nu\Omega = \alpha P. \tag{1.31}$$

Given the mass of the central body M, the mass supply rate of the disk $\dot{M}$ and the free parameter α, equations 1.23, 1.26 and 1.31 can predict the radial profiles Σ, ν and T.

The disk scale height $H \sim c_s/\Omega$ and the volume density $\rho \sim \Sigma/H$ are also easily evaluated. Additional thermodynamical details can be included in the model (see, e.g., Chapter 3), as long as cooling in the disk is efficient and viscosity is the principal carrier of angular momentum. A remarkable outcome of the α-disk model is that many of the predicted characteristics of the disk (such as the midplane temperature) depend weakly on the free parameter α.

It is important to note that while the α-disk provides a viscosity prescription, it still does not provide a viscosity mechanism. That the angular momentum transport must be very effective points to turbulence as the likely source of viscosity. However, Keplerian flows satisfy the Rayleigh criterion for linear hydrodynamical stability, $\partial(R^2\Omega)/\partial R > 0$, indicating that they are laminar. Thus, if astrophysical accretion flows are turbulent, this turbulence is either due to a powerful nonlinear hydrodynamical instability, or not hydrodynamical in nature.

It has long been suggested that magnetic transport mechanisms could induce the necessary turbulence (Lynden-Bell 1969; Shakura & Sunyaev 1973). Balbus & Hawley (1991) demonstrated the existence of the magnetorotational instability (MRI) in a differentially rotating ionized disk satisfying $\partial\Omega/\partial R < 0$ threaded with weak poloidal magnetic fields. The instability generates rapidly proliferating turbulence that is able to transport angular momentum extremely efficiently. The efficacy and robustness of the MRI has been confirmed by numerical simulations (e.g., Proga & Begelman 2003; Sano et al. 2004, and references therein). It is uncertain whether MRI is actually the angular momentum transport mechanism in astrophysical accretion flows. However, that it requires only very

modest physical conditions and the efficiency with which it can transfer angular momentum in numerical simulations make it the most promising mechanism on the table.

1.5.4 Alternative Accretion Flow Models

A key requirement for self-consistency in thin-disk theory is that the disk be able to cool efficiently at its surface, i.e. that energy transport within the disk plane can be ignored. It was recognized early that the standard thin α-disk solution is thermally unstable in the inner regions of the disk where radiation pressure dominates gas pressure (Shakura & Sunyaev 1976; Pringle 1976; see, however, Hirose et al. 2009b). To see this, consider the viscous heating in a radiation-supported α-disk. In this regime, the disk scale height satisfies $H = aT^4/(3\Sigma\Omega^2)$, and the flux due to viscous heating is

$$Q_+ = \frac{9}{8}\nu\Sigma\Omega^2 = \frac{9}{8}\alpha H^2\Sigma\Omega^3 \propto T^8\Sigma^{-1}. \tag{1.32}$$

On the other hand, radiative cooling scales with temperature as $Q_- \propto P_{\rm rad}/\Sigma \propto T^{-4}/\Sigma$. If a disk is initially in thermal equilibrium ($Q_+ = Q_-$), then a slight increase in temperature will result in $Q_+ > Q_-$, and the disk will heat up further still. A slight decrease in temperature will likewise cause runaway cooling.

One way to solve this energy-balance problem is to parameterize the viscosity in a different way: thin disks in which the viscosity scales with the gas pressure (and is independent of radiation pressure) are thermally stable (Lightman & Eardley 1974; Piran 1978). Many of the thin-disk models (Milosavljević & Phinney 2005; Chang et al. 2010) for

merging SMBHs have used such a viscosity prescription, which is sometimes called the "β-disk" model.

Another approach is to develop steady-state solutions that allow for horizontal energy advection. Such solutions are able to achieve global energy balance by carrying the unradiated energy into the black hole event horizon (Narayan & Yi 1994). There are two main classes of advective solutions for accretion flows, both of which have been applied with considerable success in explaining the observed features of accretion-powered astrophysical phenomena.

The first class is the advection dominated accretion flow (ADAF) (Ichimaru 1977; Narayan & Yi 1994, 1995b), which provides excellent spectral fits for many low-luminosity AGN (Narayan & Raymond 1999; Ptak et al. 2004), as well as quiescent states of (stellar-mass) X-ray binaries (Menou et al. 1999; Kong et al. 2002), whose emission characteristics are poorly described by standard thin disks. ADAFs are geometrically thick and optically thin. Free-free emission is unable to balance the viscous heating of the gas, and the excess energy is advected directly into the black hole event horizon. The model is characterized by hard (but dim) emission due to thermal Comptonization, and predicts a low accretion rate onto the central black hole compared to the mass supply rate at outer radii. Note that where gravitational explanation for quasar power requires the accretion of gas down to scales comparable to the predicted location of the event horizon, the ADAF model assumes the actual presence of a horizon. The empirical success of the model provides circumstantial evidence that the engines of these flows truly are black holes (Narayan 2005), and not merely ultracompact massive bodies.

The second class of solutions is the "slim disk" model (Abramowicz et al. 1988). Slim disks are starkly contrastive to ADAFs, in that they are optically thick, relatively geometrically thin, and transport mass at super-Eddington accretion rates. Although they are radiatively inefficient compared to the canonical thin disk, due to their high accretion rates they are capable of near- and super-Eddington luminosities. Slim disks have been used to explain the soft X-ray excess of some AGN (Mineshige et al. 2000), as well as quasars with very high Eddington ratios.

The canonical thin-disk model also breaks down if the disk is strongly self-gravitating, as is the case for most disk models around SMBHs exceeding $\sim 10^8 M_\odot$. The outer region of the disk is locally unstable to gravitational collapse and will fragment on dynamical timescales, plausibly forming stars instead of accreting to the center (Shlosman & Begelman 1987, 1989; Heller & Shlosman 1994). Several studies have suggested that local turbulence due to the gravitational instability could mediate horizontal transport and help stabilize the disk (e.g., Paczynski 1978; Lin & Pringle 1987; Gammie 2001). Feedback in the disk, such as heating by the stars formed from gravitational fragmentation and their supernovae, may be able to preserve a marginally stable outer disk (Collin & Zahn 1999; Sirko & Goodman 2003; Thompson et al. 2005). However, there appears to be no plausible physical mechanism that could make a quasar disk robustly gravitationally stable beyond $\gtrsim 10^3 GM/c^2$ (Goodman 2003).

1.5.5 Jets from Accretion Flows

Bipolar outflows, including collimated jets, are observed to be a regular structure of accretion flows (e.g., Ouyed et al. 1997; Konigl & Pudritz 2000). The most prominent examples are the bipolar extended radio structures pointing away from AGN (e.g., Rees 1978; Blandford & Konigl 1979); however, jets are also seen in galactic X-ray binaries and protostars (Lada 1985; Bachiller 1996), as well as white dwarfs (Galloway & Sokoloski 2004).

The propulsion of ionized plasma along the magnetic field lines of the central object has long been a leading hypothesis for the mechanism behind these ubiquitous astrophysical phenomena (e.g., Blandford & Payne 1982; Konigl 1982; Uchida & Shibata 1985); however, the details remain murky. It has also been argued that a hot, radiatively inefficient ADAF-like flow has a propensity to expel outflows because the sum of its kinetic and gravitational energies and its enthalpy is positive (Narayan & Yi 1995a; Blandford & Begelman 1999). Observationally, many astrophysical systems that are well-described by ADAFs also are known to frequently have associated winds or relativistic radio jets (Quataert & Narayan 1999; Yuan et al. 2002; Fabbiano et al. 2003). Like ADAFs, slim disks have also been suggested to generate intense outflows (e.g., Takeuchi et al. 2009, and references therein). Outflows may also be aided by thermodynamical processes, such as thermal conduction (Tanaka & Menou 2006; Johnson & Quataert 2007), within the accretion flow. Historically, the ergosphere of spinning black holes has often been attributed as the likely jet-propelling mechanism (Blandford & Znajek 1977; Rees et al. 1982). However, that accretion-powered jets are observed in protostars and white dwarfs would appear

to rule out a purely general-relativistic origin of astrophysical jets.[6] As diskussed above (1.2.2) and later in Chapter 3, jets powered by SMBHs may provide key observational clues for SMBH binaries and their mergers.

1.6 Thesis Summary

I begin my thesis by considering the origins of the earliest known SMBHs. In Chapter 2, I investigate several different scenarios for the origins of the $M \sim 10^9 M_\odot$ SMBHs powering $z \sim 6$ quasars (Tanaka & Haiman 2009). Much of the remainder of the thesis is devoted to modeling the evolution of thin accretion disks around merging SMBH binaries. The goal of these studies is to provide simple models for possible EM accretion signatures of these powerful GW sources. In Chapter 3, I present a semianalytic time-dependent model for the dramatic brightening and spectral hardening of a thin accretion disk following the coalescence of the central SMBH binary (Tanaka & Menou 2010). The focus in this chapter is on $M \sim 10^6 M_\odot$ binaries whose mergers may be detected by a space-based GW detector. In Chapter 4, I ask whether these "afterglow" events could be detected by future wide-angle and high-cadence EM surveys, without the aid of GW observations (Tanaka et al. 2010). As a theoretical aside, I derive in Chapter 5 new exact, time-dependent solutions for the thin-disk equation (equation 1.20) that are applicable to disks around compact binary systems (Tanaka 2011). In Chapter 6, I apply these solutions to model the EM appearance of $M \sim 10^9 M_\odot$ binaries that may be diskovered via pulsar timing (Tanaka et al. 2011, in prep.). The potential for multi-messenger observations of such systems are considered.

[6] It is still possible that the Blandford-Znajek mechanism could act to enhance jet power.

In Chapter 7, I conclude with thoughts on the direction of this field and future prospects.

1.6.1 Chapter 2: Assembly of the First SMBHs

The SMBHs massive enough ($\gtrsim$ few$\times 10^9 M_\odot$) to power the bright redshift $z \approx 6$ quasars observed in the *Sloan Digital Sky Survey (SDSS)* are thought to have assembled by mergers and/or gas accretion from less massive "seed" black holes. If the seeds are the $\sim 10^2 M_\odot$ remnants from the first generation of stars, they must be in place well before redshift $z = 6$, and must avoid being ejected from their parent proto–galaxies by the large (several$\times 10^2$ km s^{-1}) kicks they suffer from gravitational–radiation induced recoil during mergers with other black holes. We simulate the SMBH mass function at redshift $z > 6$ using dark matter halo merger trees, coupled with a prescription for the halo occupation fraction, accretion histories, and radial recoil trajectories of the growing black holes. Our purpose is (i) to map out plausible scenarios for successful assembly of the $z \approx 6$ quasar SMBHs by exploring a wide region of parameter space, and (ii) to predict the rate of low–frequency GW events detectable by the proposed *LISA* detector for each such scenario. Our main findings are as follows: (1) $\sim 100 M_\odot$ seed black holes can grow into the *SDSS* quasar SMBHs without super–Eddington accretion, but only if they form in minihalos at $z \gtrsim 30$ and subsequently accrete $\gtrsim 60\%$ of the time; (2) the scenarios with optimistic assumptions required to explain the *SDSS* quasar SMBHs overproduce the mass density in lower–mass (few$\times 10^5 M_\odot \lesssim M_{\rm bh} \lesssim$ few$\times 10^7 M_\odot$) black holes by a factor of $10^2 - 10^3$, unless seeds stop forming, or accrete at a severely diminished rates or duty cycles (e.g. due to feedback), at $z \lesssim 20 - 30$. We also present several successful assembly models and

their *LISA* detection rates, including a "maximal" model that gives the highest rate (~ 30 yr^{-1} at $z = 6$) without overproducing the total SMBH density.

1.6.2 Chapter 3: Afterglows of SMBH Mergers

The detector *LISA* will detect GW signals from coalescing pairs of SMBHs in the total mass range $(10^5 - 10^7)/(1 + z)\ M_\odot$ out to cosmological distances. Identifying and monitoring the EM counterparts of these events would enable cosmological studies and offer new probes of gas physics around well-characterized SMBHs. Milosavljević & Phinney (2005) proposed that a circumbinary disk around a binary of mass $\sim 10^6 M_\odot$ will emit an accretion-powered X-ray afterglow approximately one decade after the GW event. We revisit this scenario by using Green's function solutions to calculate the temporal viscous evolution and the corresponding EM signature of the circumbinary disk. Our calculations suggest that an EM counterpart may become observable as a rapidly brightening source soon after the merger, i.e. several years earlier than previously thought. The afterglow can reach super-Eddington luminosities without violating the local Eddington flux limit. It is emitted in the soft X-ray by the innermost circumbinary disk, but it may be partially reprocessed at optical and infrared frequencies. We also find that the spreading disk becomes increasingly geometrically thick close to the central object as it evolves, indicating that the innermost flow could become advective and radiatively inefficient, and generate a powerful outflow. We conclude that the mergers of SMBHs detected by *LISA* offer unique opportunities for monitoring on humanly tractable timescales the viscous evolution of accretion flows and the emergence of outflows around SMBHs with precisely known

masses, spins and orientations.

1.6.3 Chapter 4: Afterglows as Birthing Quasars

The coalescence of a SMBH binary is thought to be accompanied by an EM afterglow, produced by the viscous infall of the surrounding circumbinary gas disk after the merger. It has been proposed that once the merger has been detected in GWs by *LISA*, follow-up EM observations can search for this afterglow and thus help identify the EM counterpart of the *LISA* source. Here we study whether the afterglows may be sufficiently bright and numerous to be detectable in EM surveys alone. The viscous afterglow is characterized by an initially rapid increase in both the bolometric luminosity and in the spectral hardness of the source. For binaries with a total mass of $10^5 - 10^8 M_\odot$, this phase can last for years to decades, and if quasar activity is triggered by the same major galaxy mergers that produce SMBHBs, then it could be interpreted as the birth of a quasar. Using an idealized model for the post-merger viscous spreading of the circumbinary disk and the resulting light curve, and using the observed luminosity function of quasars as a proxy for the SMBHB merger rate, we delineate the survey requirements for identifying such birthing quasars. If circumbinary disks have a high disk surface density and viscosity, an all-sky soft X-ray survey with a sensitivity of $F_X \lesssim 3 \times 10^{-14}$ erg s^{-1} cm^{-2} which maps the full sky at least once per several months, could identify a few dozen birthing quasars with a brightening rate $d \ln F_X / dt > 10\%$ yr^{-1} maintained for at least several years. If $> 1\%$ of the X-ray emission is reprocessed into optical frequencies, several dozen birthing quasars could also be identified in optical transient surveys, such as the *Large Synoptic Survey Telescope*.

Distinguishing a birthing quasar from other variable sources may be facilitated by the monotonic hardening of its spectrum, but will likely remain challenging. This reinforces the notion that observational strategies based on joint EM-plus-GW measurements offer the best prospects for the successful identification of the EM signatures of SMBHB mergers.

1.6.4 Chapter 5: Time-dependent Solutions of Accretion Flows

We diskuss Green's-function solutions of the equation for a geometrically thin, axisymmetric Keplerian accretion disk with a viscosity prescription $\nu \propto R^n$. The mathematical problem was solved by Lynden-Bell & Pringle (1974) for the special cases of zero viscous torque and zero mass flow at the disk center. While it has been widely established that the observational appearance of astrophysical disks depend on the physical size of the central object(s), exact time-dependent solutions with boundary conditions imposed at finite radius have not been published for a general value of the power-law index n. We derive exact Green's-function solutions that satisfy either a zero-torque or a zero-flux condition at a nonzero inner boundary $R_{\rm in} > 0$, for an arbitrary initial surface density profile. The new solutions are applicable to the evolution of the innermost regions of thin accretion disks.

1.6.5 Chapter 6: Electromagnetic Counterparts of Pulsar Timing Array Sources

PTAs are expected to detect GWs from individual low–redshift ($z \lesssim 1.5$) compact SMBH binaries with orbital periods of $\sim 0.1 - 10$ yr. Identifying the EM counterparts of these

sources would provide confirmation of putative direct detections of GWs, present a rare opportunity to study the environments of compact SMBH binaries, and could enable the use of these sources as standard sirens for cosmology. Here we consider the feasibility of such an EM identification. We show that because the host galaxies of resolved PTA sources are expected to be exceptionally massive and rare, it should be possible to find unique hosts of resolved sources out to $z \approx 0.2$. At higher redshifts, the PTA error boxes are larger, and may contain as many as ~ 100 massive-galaxy interlopers. The number of candidates, however, remains tractable for follow-up searches in upcoming wide-field EM surveys. We develop a toy model to characterize the dynamics and the thermal emission from a geometrically thin gaseous disk accreting onto a PTA-source SMBH binary. Our model predicts that at optical and infrared frequencies, the source should appear similar to a typical luminous AGN. However, owing to the evacuation of the accretion flow by the binary's tidal torques, the source might have an unusually low soft X-ray luminosity and weak UV and broad optical emission lines, as compared to an AGN powered by a single SMBH with the same total mass. For sources at $z \sim 1$, the decrement in the rest-frame UV should be observable as an extremely red optical color. These properties would make the PTA sources stand out among optically luminous AGN, and could allow their unique identification.

Chapter 2

The Assembly of Supermassive Black Holes at High Redshifts[1]

2.1 Introduction

The discovery of bright quasars at redshifts $z \gtrsim 6$ in the Sloan Digital Sky Survey (SDSS) implies that black holes (BHs) as massive as several $\times 10^9 M_\odot$ were already assembled when the age of the universe was less than ≈ 1 Gyr (see the recent review by Fan 2006). These objects are among the oldest detected discrete sources of radiation in the Universe. The likelihood that all of these quasars are significantly magnified by gravitational lensing, without producing detectable multiple images (Richards et al. 2004), is exceedingly small (Keeton et al. 2005), and if their luminosities are powered by accretion at or below the

[1]This chapter is a reformatted version of an article by the same name by T. Tanaka and Z.Haiman that can be found in The Astrophysical Journal, Volume 696, Issue 2, pp. 1798-1822. The abstract for this paper is reproduced in Section 1.6.1.

Eddington rate, the central objects must be $\sim 10^9 M_\odot$ supermassive black holes (SMBHs). In particular, the quasar SDSS J1148+5251 (Fan et al. 2001, 2003) is likely to be powered by a SMBH with a mass of $\approx 10^{9.5} M_\odot$ (Willott et al. 2003).

The mechanism by which such massive BHs formed within 1 Gyr after the Big Bang remains poorly understood. Generically, these SMBHs are thought to have assembled by mergers with other BHs and/or by gas accretion [2] onto less massive BHs. If the first ("seed") BHs are the $\sim 10^2 M_\odot$ remnant BHs of the first generation of stars (e.g. Heger et al. 2003), they must be in place well before redshift $z = 6$. If accretion onto BHs is limited at the Eddington rate with radiative efficiency ϵ, defined as the fraction of the rest mass energy of matter falling onto the BH that is released as radiation, then $1 - \epsilon$ of the matter is accreted and the growth of the BH mass m is given by

$$\frac{d \ln m}{dt} = \frac{1-\epsilon}{\epsilon} \frac{4\pi G \mu m_{\rm p}}{\sigma_e c}, \tag{2.1}$$

where G is the gravitational constant, c is the speed of light, $\mu \approx 1.15$ is the mean atomic weight per electron for a primordial gas, and σ_e is the Thompson electron cross section. The e–folding time scale for mass growth is $t_{\rm Edd} \approx 4.4 \times 10^7$ yr for $\epsilon = 0.1$. In the concordance cosmological model (see below) the time elapsed between redshifts $z = 30$ (when the first seeds may form) and $z = 6.4$ (the redshift of the most distant quasar) is ≈ 0.77 Gyr, allowing for a mass growth by a factor of $\approx 10^{7.7}$. Therefore, individual $\sim 100 M_\odot$ seeds can grow into the SDSS quasar BHs through gas accretion alone, provided the accretion is uninterrupted at close to the Eddington rate and $\epsilon \lesssim 0.1$. A higher efficiency and/or

[2] In this paper, "accretion" onto BHs should be assumed to mean gas accretion, unless otherwise noted.

a lower time–averaged accretion rate will require many seed BHs to merge together; the number of required mergers increases exponentially for lower time-averaged accretion rates.

The discovery of the bright quasars at $z \gtrsim 6$ were followed by the first successful numerical calculations in full general relativity of the coalescence of a BH binary and the corresponding emission of gravitational waves (GWs; Pretorius 2005; Campanelli et al. 2006; Baker et al. 2006a). These calculations also confirmed a result previously known from post-Newtonian (Kidder 1995) and perturbation-theory treatments (Favata et al. 2004; Schnittman & Buonanno 2007): the coalesced product receives a large center-of-mass recoil imparted by the net linear momentum accumulated by the asymmetric gravitational wave emission (see Schnittman et al. 2008 for a recent detailed discussion of the physics of the recoil, and for further references). Typical velocities for this gravitational recoil (or "kick") are in excess of ~ 100 km s^{-1}. This is likely more than sufficient to eject the BHs residing in the low–mass protogalaxies in the early Universe, since the escape velocities from the DM halos of these galaxies are only a few km s^{-1}.

Several recent works have studied the role of gravitational kicks as an impediment to merger-driven modes of SMBH assembly. Simple semi–analytic models show that if every merger resulted in a kick large enough to remove the seed BHs from halos with velocity dispersions up to ≈ 50 km s^{-1}, then super-Eddington accretion would be required to build SMBHs of the required mass in the available time (Haiman 2004; Shapiro 2005). Monte–Carlo merger tree models that exclude kicks entirely (Bromley et al. 2004) or which include a distribution of kick velocities extending to low values (e.g. Yoo & Miralda-Escudé 2004,

Volonteri & Rees 2006) give a slightly more optimistic picture, showing that if seed BHs form in most minihalos in the early Universe, and especially if ejected seeds are rapidly replaced by new seeds (Yoo & Miralda-Escudé 2004), then SMBH assembly is just possible before $z \approx 6$ without exceeding the Eddington accretion rate. These works are encouraging steps toward demonstrating that there are plausible physical models that lead to the timely assembly of SMBHs massive enough to power the $z > 6$ SDSS quasars.

At present, we have no direct observational constraints on SMBH assembly at $z > 6$, and there is, in principle, a large range of "physically plausible" possibilities. The *Laser Interferometer Space Antenna* (*LISA*) is expected to be able to detect mergers of SMBHs in the mass range $\sim (10^4\text{–}10^7)\,M_\odot/(1+z)$ out to $z \sim 30$, and to extract binary spins and BH masses with high precision up to $z \sim 10$ (Vecchio 2004; Lang & Hughes 2006). It is also likely that by the time *LISA* is operational, there will be additional independent constraints on the demography of high-redshift SMBHs. It is therefore a useful exercise to calculate the expected *LISA* event rate from high–redshift SMBH mergers (e.g. Wyithe & Loeb 2003a; Sesana et al. 2004, 2005, 2007b) in a range of plausible models. Note that published estimates (Menou et al. 2001; Heger et al. 2003; Menou 2003; Haehnelt 2003; Wyithe & Loeb 2003a; Sesana et al. 2004; Islam et al. 2004; Koushiappas & Zentner 2006; Micic et al. 2007; Lippai et al. 2008; Arun et al. 2009) for the *LISA* event rate, even at lower redshifts, vary by orders of magnitude, from ~ 1 to as high as $\sim 10^4$ yr^{-1}; there is a large range even among models that explicitly fit the evolution of the quasar luminosity function (Lippai et al. 2008). A related open question is to what degree the *LISA* data stream can help pinpoint the actual SMBH assembly scenario. One aim of this paper is to

understand the model degeneracies that can lead to similar *LISA* data streams. Another is to explore as much as possible the full variety of *LISA* event rates arising from various "physically plausible" assembly models.

The physical factors that determine the growth of SMBHs at high redshift fall broadly into four categories: (1) the properties of the initial seed BHs, such as their redshift, mass, and abundance; (2) the time–averaged gas accretion rate of individual seeds; (3) the merger rates of BHs; and (4) effects governing the fate of gravitationally kicked BHs. The first category determines the number and mass of seed BHs available for assembly, and depends primarily on the behavior of gas in the host DM halos, and the mass and metallicity of the first stars. The second category measures the subsequent growth through accretion, and depends on the availability of fuel over the Hubble time, and its ability to shed angular momentum and accrete onto the BH. The third category is a combination of the halo merger rate, the rarity of seeds, and the timescale for the formation, orbital decay, and ultimate coalescence of a SMBH binary. Finally, the recoil velocity of the coalesced binary is determined by the mass ratio and spin vectors of the BHs, and its subsequent orbit – and whether it is retained or ejected before the next merger – will be determined by the overall depth of the gravitational potential of the DM host halo, and on the spatial distribution of gas and DM in the central region of the halo, which determine drag forces on the kicked BH.

Our approach to model the above effects closely follows those in earlier works (e.g., Yoo & Miralda-Escudé 2004; Bromley et al. 2004; Sesana et al. 2004; Volonteri & Rees 2006): we use Monte Carlo merger trees to track the hierarchical growth of DM halos,

and a simple semi-analytic model to follow the growth and dynamics of BHs. We expand over earlier works by adding an explicit calculation of the orbits of kicked BHs, and self–consistently include their corresponding time–dependent accretion rate. Additionally, we extend the merger tree to redshifts beyond $z > 40$, and we examine a large range of different models. For each set of model parameters, we apply our "tree-plus-orbits" algorithm to the entire halo population at $z \approx 6$ to construct a full population of SMBHs at this redshift, and calculate physical quantities of interest: the mass function, the SMBH-to-halo mass ratio, the fraction of DM halos hosting SMBHs, and the expected detection rate of SMBH mergers by *LISA*.

This algorithm assembles SMBHs through simple prescriptions of the aforementioned four categories of physical contributions to SMBH formation. We model the seed population by assuming that some fraction $f_{\rm seed}$ of DM halos reaching a threshold virial temperature $T_{\rm seed}$ forms a Pop III remnant BH. In–between mergers, the BHs are assumed to accrete gas at a rate of $f_{\rm duty}$ times the Eddington rate $\dot{m}_{\rm Edd} \equiv (1-\epsilon)/\epsilon \times L_{\rm Edd}/c^2$. Here $f_{\rm duty}$ should be interpreted as the mean gas accretion rate (averaged over time–scales comparable to the Hubble time) in units of the Eddington rate. Note that this prescription makes no distinction between episodic accretion near the Eddington rate during a fraction $f_{\rm duty}$ of the time (with no accretion in-between), and constant accretion at all times at a fraction $f_{\rm duty}$ of the Eddington rate. We assume that the mergers of BH binaries closely follow the mergers of their host halos (but allow for a delay in the latter due to dynamical friction). Finally, we simulate the orbits of recoiling BHs under different assumptions about the baryon density profile and binary spin orientation. We discuss the relative importance of

assembly model parameters on the final SMBH mass function and the *LISA* data stream, and ask whether *LISA* will be able to uniquely determine the underlying assembly model from data. We also examine several variants of the above scenario, in which (i) the seed BHs are massive, $\sim 10^5 M_\odot$, and formed from the super-Eddington accretion of a collapsed gaseous core; (ii) the DM halo is initially devoid of gas when the seed BHs is formed; (iii) seed BHs stop forming below some redshift; and (iv) models which maintain the so–called $M_{\rm bh} - \sigma$ relation between BH mass and (halo) velocity dispersion (Ferrarese & Merritt 2000; Gebhardt et al. 2000) at all redshifts.

This paper is organized as follows. In § 2, we detail our methodology by describing our assumptions, the assembly algorithm, including the prescriptions of the aforementioned physical effects, and the different assembly scenarios we consider. We present and discuss our main results in § 3. In § 4, we summarize our most important results, and comment on future prospects to understand SMBH assembly at high redshift. To keep our notation as simple as possible, throughout this paper the capital M will be used to denote halo masses, and m will refer to BH masses. In this paper, we adopt a ΛCDM cosmology, with the parameters inferred by Komatsu et al. (2009) using the five-year data from the *Wilkinson Microwave Anisotropy Probe* (*WMAP5*): $\Omega_{\rm CDM} = 0.233$, $\Omega_b = 0.046$, $\Omega_\Lambda = 0.721$, $h = 0.70$, and $\sigma_8 = 0.82$. We use $n_{\rm s} = 1$ for the scalar index.

2.2 Assumptions and Model Description

Our strategy is as follows: (1) We use Monte Carlo merger trees to construct the hierarchical merger history of DM halos with masses $M > 10^8 M_\odot$ at redshift $z = 6$, i.e. those that can

host SMBHs of mass $m \gtrsim 10^5 M_\odot$; (2) We insert seed BHs of mass $m_{\rm seed}$ into the tree in some fraction $f_{\rm seed}$ of halos that reach a threshold temperature $T_{\rm seed}$; (3) We follow the subsequent BH assembly history by allowing the BHs to grow by gas accretion in–between mergers, and by calculating the orbit and accretion history of each recoiling BH in its host halo. We assume that BHs add their masses linearly upon merging, and ignore mass losses due to gravitational radiation, as these losses never accumulate to significant levels, even through repeated mergers (Menou & Haiman 2004). We prescribe spin distributions of the BHs and gas distributions within their host halos. We repeat this procedure for different mass bins of the halo mass function, until we have a statistically robust sample to represent the global SMBH mass function at redshift $z = 6$ to an accuracy of a factor of two. We also record the BH binary mergers whose masses lie within the mass range $\sim (10^4$–$10^7)\,{\rm M}_\odot/(1+z)$, corresponding roughly to *LISA*'s sensitivity range.

2.2.1 The Merger Tree

We construct DM merger trees based on the algorithm by Volonteri et al. (2003a), which allows only binary mergers. Similar numerical algorithms (e.g., Somerville & Kolatt 1999; Cole et al. 2000) give somewhat different results, as have been discussed recently by Zhang et al. (2008). We will not reproduce the Volonteri et al. (2003a) recipe here in full; instead we present a brief review. We take the extended Press-Schechter (EPS) mass function (Press & Schechter 1974; Lacey & Cole 1993),

$$\frac{dN}{dM} = \frac{1}{\sqrt{2\pi}} \frac{M_0}{M} \frac{d\sigma_M^2}{dM} \frac{\delta_c(z) - \delta_c(z_0)}{\left(\sigma_M^2 - \sigma_{M_0}^2\right)^{3/2}} \exp\left\{-\frac{1}{2}\frac{[\delta_c(z) - \delta_c(z_0)]^2}{\sigma_M^2 - \sigma_{M_0}^2}\right\}, \tag{2.2}$$

which gives for a parent halo of mass M_0 at redshift z_0 the number of progenitor halos in the range $M \pm dM/2$ at a higher redshift z. Here σ_M is the amplitude of the linear matter fluctuations at redshift $z = 0$, smoothed by a top hat window function whose scale is such that the enclosed mass at the mean density is M (computed using the fitting formula for the transfer function provided in Eisenstein & Hu 1999), and δ_c is the redshift–dependent critical overdensity for collapse. We take the limit as $\Delta z \equiv z - z_0 \to 0$:

$$\frac{dN}{dM} = \frac{1}{\sqrt{2\pi}} \frac{M_0}{M} \frac{d\sigma_M^2}{dM} \left(\sigma_M^2 - \sigma_{M_0}^2\right)^{-3/2} \frac{d\delta_c}{dz} \Delta z. \tag{2.3}$$

The two advantages of taking this limit are that (i) by linearizing the expression, the $z-$ and M–dependences separate, allowing a tabulation as functions of the parent & progenitor halo mass, and (ii) separating Δz allows for a simple algorithm for adaptive timesteps to make sure that fragmentations produce binaries at most (no triplets).

For a parent halo of mass M_p after a small step Δz, the mean number of "minor" fragments in the mass range $M_{\rm lo} < m < M_p/2$ is given by

$$N_p = \int_{M_{\rm lo}}^{M_p/2} \frac{dN}{dM} dM \propto \Delta z. \tag{2.4}$$

We choose Δz adaptively such that $N_p \ll 1$, which ensures that multiple fragmentations are unlikely in a given single time step. We place a lower limit of 10^{-3} (in redshift units) for the timestep to keep computation times manageable. The integral in Equation (2.4) diverges as $M_{\rm lo} \to 0$, making it computationally prohibitive to compute the merger history of arbitrarily small halos. To avoid this problem, all progenitors below a fixed

mass resolution $M_{\rm res}$ are considered jointly as accreted mass and not tracked individually. Any progenitor with $M < M_{\rm res}(z)$ is thus discarded from the tree and its prior history is disregarded. For our calculations, we choose $M_{\rm res}(z)$ to be the mass corresponding to a virial temperature of 1200K, corresponding to $M_{\rm res} \sim 4.7 \times 10^5 M_\odot$ at $z = 40$ and $3.4 \times 10^6 M_\odot$ at $z = 10$. Theoretical studies have concluded that Pop III stars can start forming at lower virial temperatures,[3] but numerical considerations have forced us to adopt a somewhat larger threshold. We do not impose an explicit upper redshift limit, and we run the tree until our halos at $z = 6$ are entirely broken up into progenitors with $M < M_{\rm res}$. As a check on our Monte–Carlo merger tree algorithm, in Figure 2.1, we present the progenitor mass functions of a $10^{12} M_\odot$, $z_0 = 6$ parent halo at redshifts of $z = 8$, 13, 21 and 34, together with the Poisson errors of the merger tree output and the predictions from the EPS conditional mass function (eq. 2). Our merger tree results are consistent with the EPS conditional mass function up to redshift $z \approx 40$, with agreement within a factor of two for most mass bins and redshift values. In particular, the numerical mass function agrees well with the EPS prediction for the low-mass progenitors, even at very high redshit, but the higher-mass progenitors are under-predicted by a factor of up to two. We note that Cole et al. (2000) used a numerical algorithm similar to the one we adopted, to construct a halo merger-tree. As discussed in Zhang et al. (2008), that algorithm results in a similar inaccuracy.

[3]Haiman et al. (1996), Tegmark et al. (1997) and Machacek et al. (2001) suggest a threshold virial temperature of $\sim$ 400K for collapse. In their recent high-resolution numerical simulations, O'Shea & Norman (2007) find star formation in halos of masses $(1.5 - 7) \times 10^5 M_\odot$ between $19 < z < 33$, with no significant redshift dependence on the mass scatter. These values correspond to virial temperatures of 260 – 1300K.

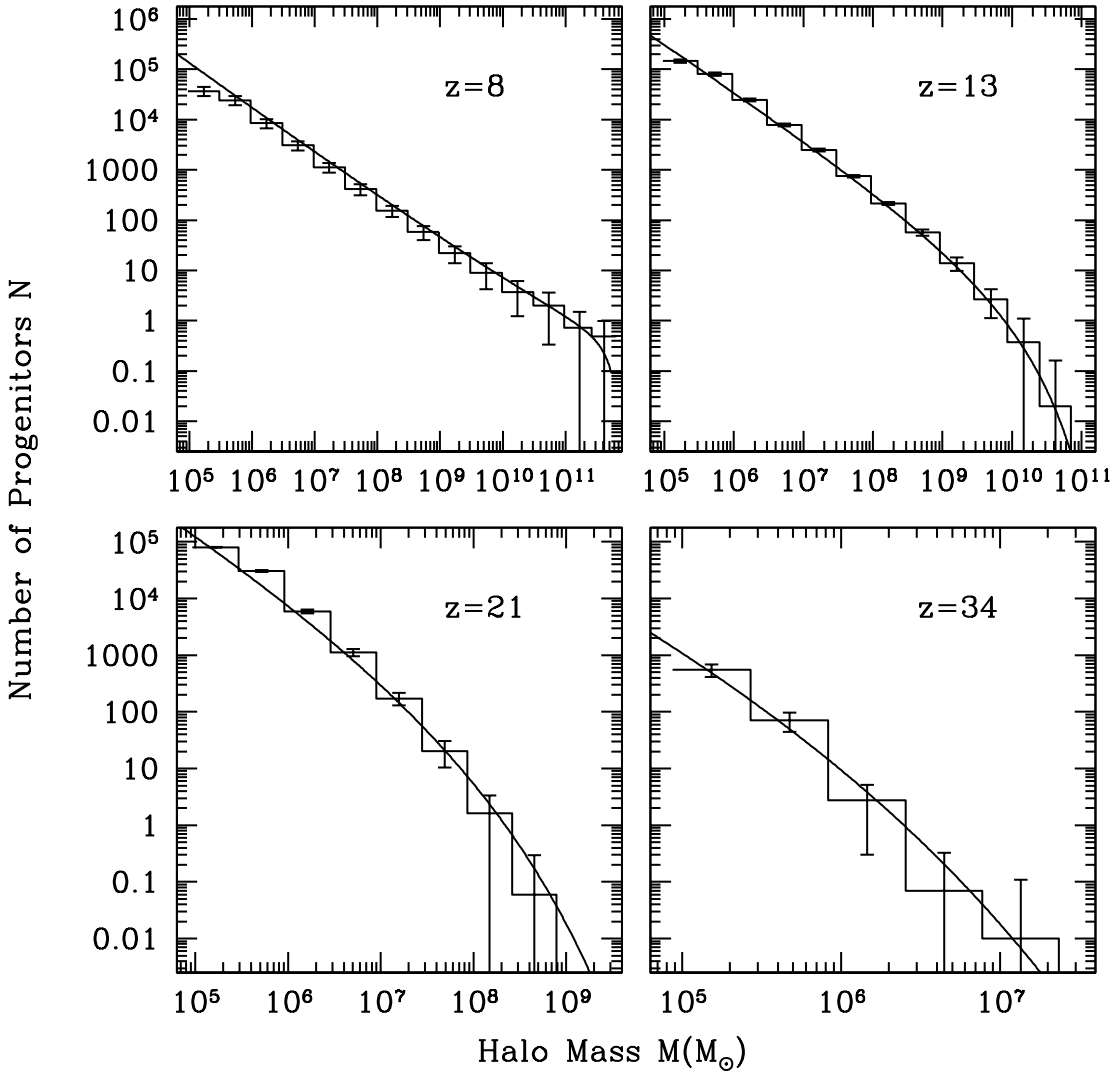

Figure 2.1 The Monte–Carlo–generated mass function of progenitors for a $10^{12}M_{\odot}$, $z_0 = 6$ parent halo at $z = 8$, 13, 21 and 34. The histogram is the mean number of 100 realizations, and the error bars demarcate the Poisson errors. The solid curve is the EPS prediction.

2.2.2 The Initial Black Hole Population

The conditions under which the first black holes form are highly uncertain, though numerical simulations (Abel et al. 2002; O'Shea & Norman 2007) do provide useful indications. We parametrize our ignorance in terms of a seeding fraction, such that a fraction $f_{\rm seed}$ of all halos reaching the critical virial temperature $T_{\rm seed}$ form a seed BH. There are physical mechanisms that make a low seeding fraction plausible: the first stars may form only in rare, baryon-rich overdense regions with unusually low angular momentum, and seed remnants may receive ejecting kicks from collapse asymmetry mechanisms similar to those responsible for high-velocity pulsars. Furthermore, radiative and other feedback processes may prohibit H_2–formation, cooling, and star–formation in the majority of low–mass minihalos at high redshift (e.g. Haiman et al. 1997; Mesinger et al. 2006). Since the *LISA* event rate, especially at the earliest epochs, will depend primarily on the abundance of BHs present, it is highly sensitive to the seeding function.

We choose two fiducial seeding models, the first with $T_{\rm seed} = 1200$K (the minimum value required for metal-free molecular line cooling and star formation) for a Pop-III remnant seed BH with $m_{\rm seed} = 100 M_\odot$. The second model has $T_{\rm seed} = 1.5 \times 10^4$K and $m_{\rm seed} = 10^5 M_\odot$, inspired by the "direct collapse" models of more massive BHs from the central gas in halos with a deep enough potential to allow atomic cooling (Oh & Haiman 2002; Bromm & Loeb 2003; Volonteri & Rees 2005; Begelman et al. 2006; Spaans & Silk 2006; Lodato & Natarajan 2006). If Eddington accretion is the main mode of growth, then we do not expect the choice of seed mass for each type of model to qualitatively affect our results, other than the obvious linear scaling of the overall BH mass function with the

initial seed mass. Only the binary mass ratios affect recoil magnitudes, and the subsequent orbital dynamics depends minimally on the BH mass.

2.2.3 Baryonic and Dark Matter Halo Profiles

The DM profile for the earliest halos is found to be similar to the NFW (Navarro et al. 1997) profile of lower–redshift, more massive DM halos (Abel et al. 2000; Bromm et al. 2002; Yoshida et al. 2003). However, the composition and spatial distribution of the baryons, at the time when the seed BH appears in these halos, is poorly understood, and is unconstrained by observations. This is unfortunate, since these quantities play a pivotal role in determining the orbital dynamics and growth rate of a recoiling BH.

A steep profile with a cusp will retain BHs more effectively, owing both to a deeper gravitational potential well and a larger dynamical drag force at the halo center. The baryon distribution will also determine the accretion history of the central BH by determining the accretion rate as the BH rests near the halo's potential center, or as it oscillates in a damped orbit through the halo following a recoil displacement event.

In addition, whether the baryons are gaseous or stellar has nontrivial consequences, owing to the difference in the dynamical friction force between the two cases. A collisional medium provides a greater dynamical friction force than a stellar or DM medium with the same density profile (Ostriker 1999; Escala et al. 2004). Because of the difference in the drag force, an environment dominated by gas, and not by stars (or dark matter), has several possible consequences on BHs: (1) binaries coalesce more rapidly; (2) a BH recoiling in gas has a higher likelihood of being retained in its parent halo; and (3) any

"vagrant" BH that is displaced from the baryon-rich center of the gravitational potential of its host halo takes less time to return there, reducing episodes of suppressed accretion. In three–dimensional simulations of star–formation in metal–free minihalos suggest that star–formation is inefficient, with either a single star, or at most a few stars, forming at the center of the halo (Abel et al. 2000; Bromm et al. 2002; Yoshida et al. 2003, 2008). Since in the context of this paper we are concerned with the pre-reionization Universe, we work with the assumption that stars are rare before $z \gtrsim 6$ and that the baryons in our halos are mostly gaseous.

We model each galaxy as a spherically symmetric mass distribution with two components: a DM halo with an NFW profile, and a superimposed baryonic component. Previous studies on this subject (see e.g. Volonteri et al. 2003a; Madau & Quataert 2004) have often assumed a non-collisional singular isothermal sphere (SIS) profile for the mass distribution. This is justified if the gas does not cool significantly below the virial temperature of the DM halo, and if it has little angular momentum (so that it is supported thermally, rather than by rotation). In most halos whose virial temperature is above 10^4K, this assumption is less justified, and a disk may form at the core of the DM halo (Oh & Haiman 2002). The direct collapse scenario in Begelman et al. (2006) and also Volonteri & Rees (2006) adopt such a "fat disk" configuration. However, the central densities of such disks are within the range of our adopted spherical profiles. For simplicity, here we only consider three different prescriptions for spherical gas distribution. Our fiducial gas profile is a cuspy, $\rho \propto r^{-2.2}$ power law, where we have taken the power–law index of 2.2 as suggested by numerical simulations of the first star-forming minihalos (Machacek

et al. 2001). This profile is established in halos that are able to cool their gas via H_2, and describes the gas distribution at the time of the first star–formation.

It is possible, however, that the typical seed BHs are surrounded by a very different gas distribution, at the time of their formation. First, the progenitor Pop-III stars of the first seed BHs are here assumed to form in DM halos of mass $\sim 10^{5-6} M_{\odot}$. The UV radiation from the star will photo–heat, and easily blow out most of the gas from low–mass minihalos, even before the star collapses to leave behind a seed BH (e.g. Whalen et al. 2004). In this case, the remnant BH will find itself in a DM halo devoid of gas, and can only start accreting once a merger with another, gas-rich halo has taken place, or until the parent halo has accreted enough mass to replenish its gas (e.g. Alvarez et al. 2009). We therefore make the simple assumption that *no* gas is present, until the minihalo containing the newly–formed seed BH merges with another halo, or grows sufficiently – assumed here to be a factor of 10 – in mass. However, we will examine the consequences of this assumption below, by performing runs without such a blow–out phase.

Second, as mentioned above, feedback processes may prohibit H_2–formation and cooling in the majority of the low–mass minihalos (e.g. Haiman et al. 1997; Mesinger et al. 2006). The gas in such minihalos remains nearly adiabatic, and can not contract to high densities. To allow for this possibility, we will study a variant for the effective gas profile. Specifically, we adopt the gas distribution in these halos by the truncated isothermal sphere (TIS) profile proposed in Shapiro et al. (1999), which has an r^{-2} profile at large radii, but has a flat core at the center owing to the central gas pressure. The density profile is normalized (here, and also in our fiducial gas profile above) such that the total

baryon-to-DM mass ratio inside the virial radius r_{200} equals the cosmological value. Both the DM and the baryonic components are assumed to extend out to $10r_{\rm vir}$, at which point the density falls to the background value. This is consistent with infall–collapse models of Barkana (2004).

2.2.4 Mergers of Dark Matter Halos and Black Holes

We next have to make important assumptions about the treatment of mergers between dark matter halos and their resident BHs.

First, we consider the merger between two DM halos, with the more massive halo referred to as the "host" and the less massive as the "satellite" halo. The Press-Schechter formalism and our merger tree consider as "merged" two halos that become closely gravitationally bound to each other. However, if the mass ratio of a halo pair is large, then in reality the smaller halo can end up as a satellite halo, and its central BH will never merge with that of the more massive halo. We therefore require in our models that for BHs in such halo pairs to coalesce, the halo merger timescale must be shorter than the Hubble time. We take the standard parametrization of the merger time:

$$\tau_{\rm merge} \approx x\frac{M_1}{M_2}\tau_{\rm dyn} \approx 0.1x\frac{M_1}{M_2}t_{\rm Hub},$$

where $M_1/M_2 > 1$ is the ratio of the halo masses, $x \lesssim 1$ is some dimensionless factor that encodes the orbit geometry (e.g. circularity) and dynamical friction, and $\tau_{\rm dyn}$ and $t_{\rm Hub}$ are the dynamical and Hubble times, respectively. It has been suggested (Boylan-Kolchin

et al. 2006) that radial infall along filaments may be preferred in the mergers of elliptical galaxies. Boylan-Kolchin et al. (2008) give a fitting formula for the merger time based on numerical simulations. Their Equation (5) reduces approximately to $\tau_{\rm merger}/\tau_{\rm dyn} \approx 0.45$ for a moderately non-circular orbit with circularity (defined as the ratio of the orbit's specific angular momentum to the angular momentum of the circular orbit with the same energy) of 0.5. We take a moderate value of $x \approx 0.5$. That is, if $M_1/M_2 < 20$, then the BH in the smaller halo is considered to be "stuck" out in the orbiting satellite halo and never merges with the central BH of the more massive halo. This choice also ensures that the vast majority of BH binaries in our simulation do not have extreme mass ratios, as the BH masses co-evolve with the host halos.

We next make assumptions regarding the timescales involving BH dynamics in their host halos, as follows: (1) if the two merging halos each contain a central BH, the two holes are assumed to form a binary efficiently, i.e. we assume there is no delay, in addition to the time taken by the DM halos to complete their merger; (2) the binary is then assumed to coalesce in a timely fashion, prior to the interaction with a third hole; and (3) the binary coalescence is assumed to take place at the center of the potential of the newly merged halo. The first assumption has been addressed by Mayer et al. (2007), who report that the increased drag force of gas in wet mergers allows the timely formation of supermassive BH binaries. The second assumption is valid for binaries in extremely gas-rich environments (see Milosavljević & Merritt 2003 for a review), or in triaxial galaxies (Berczik et al. 2006). As for the third assumption: given that the timescales of orbital damping are comparable to the intra-merger timescale for BH velocities of $\gtrsim \sigma_{\rm SIS}$, unperturbed BHs free-falling

during the halo merger process are likely to end up near the center of the potential within the merging timescales of their hosts. We do not include triple-BH interactions in our analysis.

The assumption that BH binaries merge efficiently following the mergers of their host halos may be unjustified in our models in which initially, the DM halo is devoid of gas, since gas is generally believed to be necessary for prompt coalescence. However, this inconsistency will not affect our conclusions, for the following reasons. First, we find that a BH merger in a gasless environment is a rare event, as it occurs only if both parent halos are low-mass halos that had formed seed BHs relatively recently (i.e. neither halos have yet grown in mass by a factor of 10). Second, members of such binaries will have equal (or, in actuality, similar) masses, since they have not been able to add to their seed mass by accretion. If the BHs can merge efficiently without gas, the coalesced product will likely be ejected, owing to a shallow halo potential and a relatively large kick of an equal-mass merger. If the binary does not coalesce efficiently, it will coalesce once the parent halo merges with a gas-rich, BH-free halo, or once the parent halo accretes enough gas to facilitate the merger. Such belated mergers will also presumably take place with a mass ratio of close to unity and will likely result in ejection, regardless of whether significant gas accretion takes place before coalescence. Now, consider the case of a stalled binary encountering a third BH before gas enrichment of the halo. If the third BH is much more massive, it will not be ejected by gravitational interaction or recoil. There will be a massive BH in the center of the host halo, and whether the two smaller seed BHs were ejected or swallowed by the larger BH is of little consequence to our analysis, especially

given the rarity of double-seed binaries. Thus, inefficient binary coalescence is of concern only when a double-seed binary encounters a third BH of comparable mass before the host halo is gas-enriched. Such triple-seed systems are likely to be extremely rare indeed, and unlikely to affect the overall mass function at $z = 6$. We anticipate that the main effects of a gas-depleted host halo will be increasing the number of similar-mass mergers following the initial epoch of seed formation, and slightly reducing the time available for gas accretion.

2.2.5 Gravitational Recoil

2.2.5.1 Probability Distribution of Kick Velocities

To calculate the recoil velocities of coalesced BHs, we employ the formulae provided in Baker et al. (2008), which are used to fit their numerical results,

$$v^2_{\rm recoil} = v^2_m + v^2_\perp + v^2_\parallel + 2v_\perp \left(v_m \cos\xi + v_\parallel \sin\xi\right), \tag{2.5}$$

$$v_m = A\eta^2 \sqrt{1-4\eta}\,(1 + B\eta), \tag{2.6}$$

$$v_{a\parallel} = H\tfrac{\eta^2}{1+q}\left(a_1 \cos\theta_1 - q a_2 \cos\theta_2\right), \tag{2.7}$$

$$v_{a\perp} = -K\tfrac{\eta^3}{1+q}\left[a_1 \sin\theta_1 \cos\left(\phi_1 - \Phi_1\right) - q a_2 \sin\theta_2 \cos\left(\phi_2 - \Phi_2\right)\right], \tag{2.8}$$

where $q \equiv m_2/m_1 \leq 1$ is the binary mass ratio, $\eta \equiv q/(1+q)^2$ is the symmetric mass ratio, and $\theta_{1,2}$ are the angles between the BH spin vectors $\vec{a}_{1,2} = \vec{S}_{1,2}/m_{1,2}$ and the binary orbital angular momentum vector. The angles $\phi_{1,2}$ denote the projection of the spin vectors onto

the orbital plane, measured with respect to a fixed reference angle. As seen from the equations themselves, v_m is the kick component that depends only on the symmetric mass ratio; $v_{a\parallel}$ and $v_{a^\perp}$ depend on the mass ratio and the projection of the binary spins parallel and perpendicular, respectively, to the orbital angular momentum. [4] $\Phi_1(q) = \Phi_2(1/q)$ are constants for a given value of q that encode the dependence of the kick and orbital precession on the initial spin configuration. We use the mean values given in Baker et al. (2008) for the fitting parameters: $A = 1.35 \times 10^4$ km s^{-1}, $B = -1.48$, $H = 7540$ km s^{-1}, $K = 2.4 \times 10^5$ km s^{-1}, and $\xi = 215°$. We assume spherical symmetry in our host DM halos, so we are concerned only with the recoil magnitudes and not with the kick orientations.

Following Schnittman & Buonanno (2007), for every recoil event, we assign to both members of the binary spin magnitudes in the range $0.0 \leq a_{1,2} \leq 0.9$, randomly selected from a uniform distribution. We consider two scenarios for the spin orientation: a case where the orientation is completely random, with $0 \leq \theta_{1,2} \leq \pi$ and $0 \leq \phi_{1,2} \leq 2\pi$ chosen randomly from a uniform distribution; and one where the spins are completely aligned with the orbital angular momentum vector.[5] The latter scenario is motivated by Bogdanović et al. (2007), who argued that external torques (such as those provided by a circumbinary accretion flow) may help align the binary spins with the orbital angular momentum prior to coalescence, making kicks of $\gtrsim 200$ km s^{-1} physically unfavored. While the argument was originally used to explain the lack of quasars recoiling along

[4]Note that this notation differs slightly from Baker et al. (2008) – we have simplified their notation to be more transparent in our spherically symmetric geometry.

[5]Both cases have the computational advantage that one does not require the values for $\Phi_{1,2}(q)$. For a totally random spin orientation and a given value of q, choosing $\phi_{1,2}$ randomly is equivalent to choosing $\phi_{1,2} - \Phi_{1,2}$ randomly. When the spins are aligned with the orbital angular momentum, $\Phi_{1,2}$ terms are irrelevant because they are always multiplied by $\sin\theta_{1,2} = 0$.

the line of sight at lower redshifts (Bonning et al. 2007; although see Komossa et al. 2008 for a candidate recoil detection), the same spin–orbit alignment will also impact the early assembly history of SMBHs, by allowing less massive halos to retain recoiling BHs.

Berti & Volonteri (2008) have provided a merger–tree model to follow the spin evolution of BHs through gas accretion and binary merger events. In this work, we opt not to track the spins of individual BHs due to the uncertainties involved. For instance, if circumbinary disks can act to align the spins of each binary member, this would present a scenario significantly different from the scenario presented by Berti & Volonteri. As we will show in later sections, the spin prescription does not appear to play a primary role in determining the mass function of $z = 6$ BHs.

We show the recoil velocity distribution for both orientation scenarios as a function of the mass ratio q in Figure 2.2. The figure shows the mean, 1-σ, and maximum values for the recoil velocity magnitude from 10^6 random realizations at a given value of $0.01 \leq q \leq 1$. For $q > 0.1$, if spins are randomly oriented then kicks for similar-mass mergers are in the $100 - 1000$ km s^{-1} range, with a handful of kicks above 1000 km s^{-1} and a maximum possible kick of ≈ 3000 km s^{-1}; for spins aligned with the orbital angular momentum, kicks are typically below 200 km s^{-1}, with the maximum allowed kick no more than 300 km s^{-1}. For $q \gtrsim 0.1$ and for random spin orientations, the maximum kick $v_{\rm max}(q)$ is achieved close to where $v_{\parallel} \gg v_{\perp}$ is maximized; this occurs when $a_{1,2} = 0.9$, $\cos(\phi_{1,2} - \Phi_{1,2}) = 1$, and $\sin\theta_1 \approx -\sin\theta_2 \approx 1$. At these spin parameter values, $v_{\rm max}(q)$ is well approximated by $\sqrt{v_m^2 + v_{\parallel}^2}$, and is a monotonically increasing function of q. If the spins are aligned with the orbital angular momentum vector, then $v_{\parallel} = 0$ and the maximum kick

occurs when $a_1 = 0.0, a_2 = 0.9$. Also, in this aligned case, the spin-independent component v_m is the dominant term for $q \lesssim 0.6$ and peaks at $q \approx 0.345$, resulting in the maximum and mean values for the recoil speed peaking between these q values.

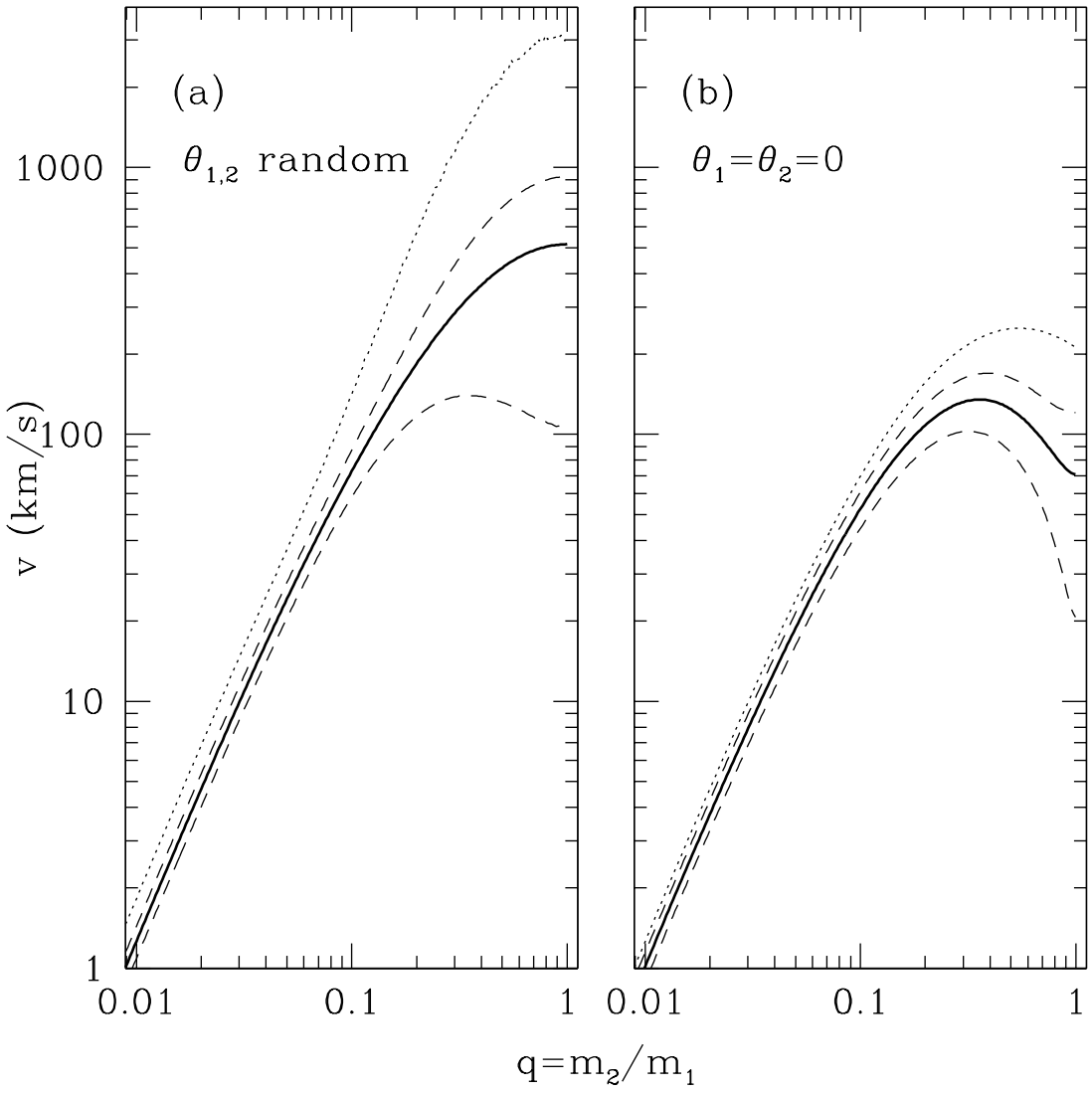

Figure 2.2 The kick velocity distribution as a function of the BH binary mass ratio, after 10^6 realizations of Equations (2.5-8) at each value of q. The left panel shows kicks for random spin orientations, while the right panel shows kicks when both BH spins are aligned with the orbital angular momentum. In each case, the spin magnitudes are chosen from a uniform random distribution in the range $0.0 \leq a_{1,2} \leq 0.9$. The solid curves show the mean, the dashed curves show the 1-σ range, and the dotted curve gives the maximum value generated in the 10^6 realizations.

2.2.5.2 Trajectories of Kicked Black Holes

Given the mass distribution of the host halo and a recoil speed generated from the method detailed in the previous subsection, we numerically integrate the equation for the radial motion of a BH with mass m,

$$\frac{dv}{dt} = -\frac{GM(r)}{r^2} + a_{\rm DF} - v\frac{\dot{m}}{m}, \tag{2.9}$$

where $r(t)$ is the radial displacement of the BH from the center of the host halo and $v(t)$ is the BH's radial velocity. The first term on the right–hand side is the acceleration due to Newtonian gravity with $M(r)$ the total (dark matter + baryon) mass enclosed inside spherical radius r; the second is the drag deceleration due to dynamical friction; and the third is the deceleration due to mass accretion. A similar calculation of the kicked BH's trajectory has been performed by Madau & Quataert (2004) – the main difference from our prescription is that they assumed the halo to have a collisionless SIS profile, and adopted parameters describing galactic bulges in the local Universe, whereas we use the hybrid DM+gas profile described above, and adopt parameters relevant to low–mass halos at high redshifts.

For a non-collisional medium (in our case, for dark matter), the dynamical friction is described by the standard Chandrasekhar formula (see e.g. Binney & Tremaine 1987),

$$a_{\rm DF}^{\rm star}(r, v) = -4\pi G^2 m \rho(r)\, \frac{1}{v}\, \ln\Lambda \left[{\rm erf}(X) - \frac{2}{\sqrt{\pi}} X \exp(-X^2) \right], \tag{2.10}$$

where $\ln\Lambda$ is the Coulomb logarithm and $X = v/(\sqrt{2}\sigma_{\rm DM})$, with $\sigma_{\rm DM}$ the velocity disper-

sion of the DM halo. In a collisional medium, the density wave in the wake of an object traveling at near or above the sound speed is enhanced via resonance, an effect that has no counterpart in collisionless media. This results in an enhancement of the dynamical friction force, for which Ostriker (1999) has derived an analytic formula. However, the Ostriker prescription is known to overpredict the drag force at slightly supersonic velocities when compared with numerical simulations. While Escala et al. (2004) provides a fitting formula that better fits the numerical results at low speeds, their formula suffers from the opposite problem, and over-predicts the drag for highly supersonic motion. We therefore adopt a hybrid prescription, adopting the Escala et al. (2004) formula for motion with $\mathcal{M} < \mathcal{M}_{eq}$ and the Ostriker formula for $\mathcal{M} > \mathcal{M}_{eq}$, where $\mathcal{M}_{eq}$ is the Mach number ($= |v|/c_s$) where the two prescriptions predict equal drag. The entire prescription is described by

$$a_{\rm DF}^{\rm gas}(r, v) = -4\pi G^2 m \rho(r) \frac{1}{v} \times f(\mathcal{M}), \qquad \text{where :} \tag{2.11}$$

$$f(\mathcal{M}) = \begin{cases} 0.5 \ln\Lambda \left[\mathrm{erf}\left(\frac{\mathcal{M}}{\sqrt{2}}\right) - \sqrt{\frac{2}{\pi}} \mathcal{M} \exp(-\mathcal{M}^2/2) \right] & \text{if} \quad 0 \le \mathcal{M} \le 0.8; \\ 1.5 \ln\Lambda \left[\mathrm{erf}\left(\frac{\mathcal{M}}{\sqrt{2}}\right) - \sqrt{\frac{2}{\pi}} \mathcal{M} \exp(-\mathcal{M}^2/2) \right] & \text{if } 0.8 < \mathcal{M} \le \mathcal{M}_{eq}; \\ \frac{1}{2} \ln\left(\frac{\mathcal{M}+1}{\mathcal{M}-1}\right) - \ln\Lambda & \text{if} \qquad \mathcal{M} > \mathcal{M}_{eq}. \end{cases} \tag{2.12}$$

The Coulomb logarithm $\ln\Lambda$ is not a precisely known parameter, but is generally agreed to be $\gtrsim 1$ for both the stellar and the gaseous cases. We adopt the value $\ln\Lambda = 3.1$ used in Escala et al. (2004), which yields $\mathcal{M}_{eq} \approx 1.5$.

The drag force depends on the local gas sound speed. Instead of attempting to compute a temperature profile explicitly, we make the approximation that the gas sound speed is constant and given by the isothermal sound speed of the halo virial temperature. We be-

lieve this to be justified from numerical simulation results that show the gas temperature to vary by at most a factor of ≈ 3 within the virial radius despite a steeper-than-isothermal ($\propto r^{-2.2}$) density profile (see, e.g., Machacek et al. 2001). While local variations in the sound speed may have significant effects when $v \sim c_s$, the recoil events of interest here are for the most part highly supersonic. Recoil events with $v \sim c_s$ will result in quick damping of the BH orbit and for the purposes of this paper will in all likelihood be indistinguishable from the zero-recoil calculation in terms of their accretion history.

The virial temperature is given by (e.g., Barkana & Loeb 2001)

$$T_{\rm vir} \approx 370(1+z)\left(\frac{\mu}{0.6}\right)\left(\frac{M}{10^7 M_\odot}\right)^{2/3}\left(\frac{\Omega_0 h^2}{0.14}\right)^{1/3} {\rm K}, \tag{2.13}$$

and the isothermal sound speed is

$$c_s \approx 1.8\ (1+z)^{1/2}\left(\frac{M}{10^7 M_\odot}\right)^{1/3}\left(\frac{\Omega_0 h^2}{0.14}\right)^{1/6}\ {\rm km\ s^{-1}} \tag{2.14}$$

We also employ a simplified prescription for the velocity dispersion of non-collisional matter, given by the SIS expression evaluated at the virial radius: $\sigma_{\rm SIS} = \sqrt{GM/2r_{200}}$. The simplified expression agrees with the exact velocity dispersion for the NFW profile within $\sim 20\%$ inside the virial radius.

Because matter that is bound to the BH does not contribute to dynamical friction, we follow Madau & Quataert (2004) and truncate the density profiles at the BH radius of influence, $r_{\rm BH} \approx Gm/\sigma^2_{\rm SIS}$. The density is furthermore assumed to be constant inside this radius. Although the BH will drag with it the surrounding gravitationally bound matter,

effectively increasing its mass, the additional mass is small owing to the large initial recoil velocity (e.g., Lippai et al. 2008), and we have ignored this mass–enhancement here.

Figure 2.3 illustrates the effect of the halo matter distribution on the BH orbit. For each of the three orbits shown in the Figure, we adopt $M = 10^8 M_\odot$, $m = 10^5 M_\odot$, $z = 20$ and $v_{\rm kick} = 100$ km s^{-1}. The black curve shows, for reference, the radial orbit for a pure NFW profile. The red curve corresponds to the case which includes an NFW DM component and a power-law gas profile with $\rho \propto r^{-2.2}$. The blue curve is for a halo with an NFW DM component and a TIS gas profile.

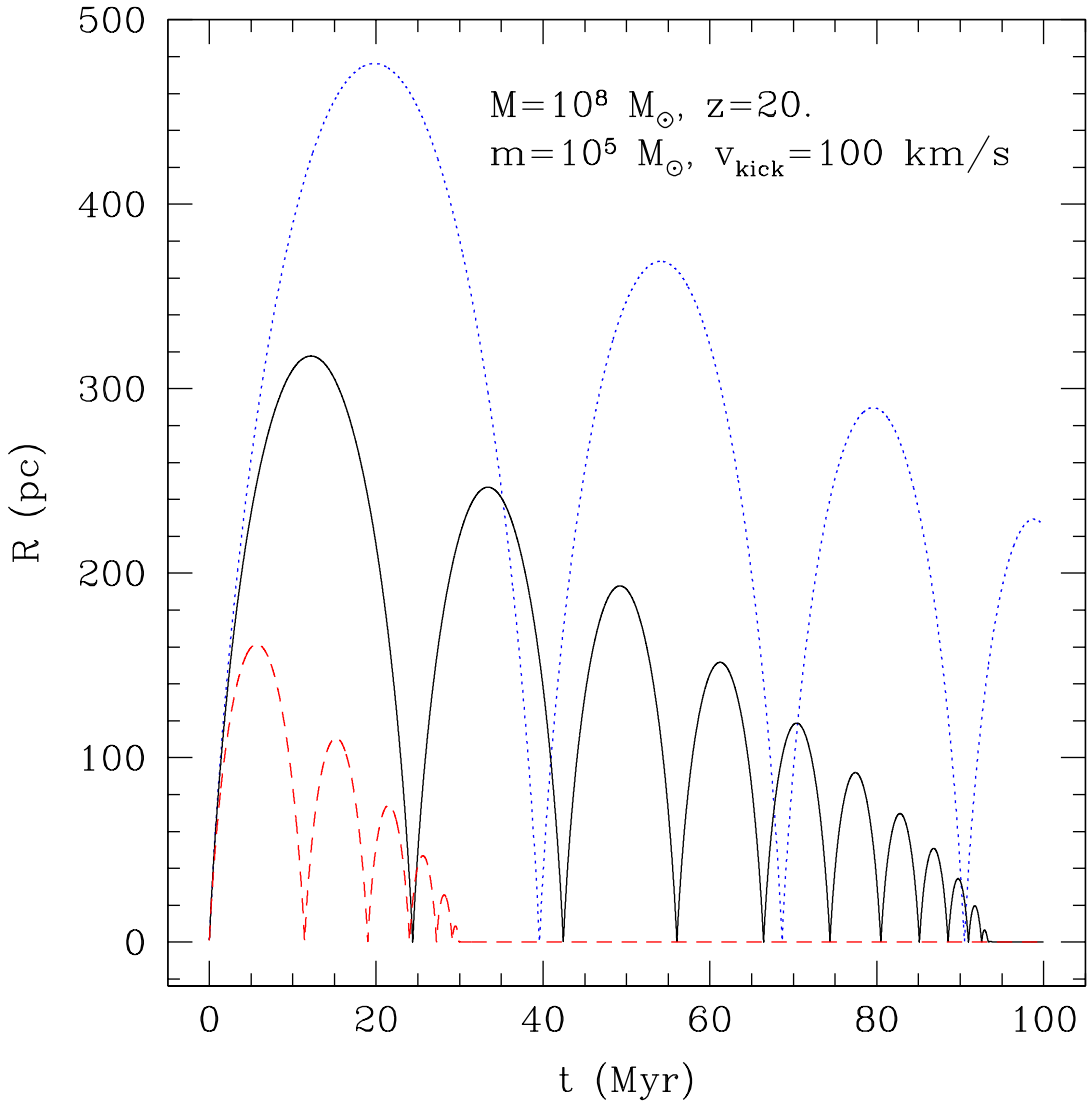

Figure 2.3 Examples of the radial motion of a recoiling black hole, for three different mass profiles for the host halo. The black curve shows the motion in a pure NFW halo; the halo of the blue (dotted) curve assumes that the host galaxy has a DM halo with an NFW profile and a corey gas component; the red (dashed) curve assumes a DM halo and a cuspy $r^{-2.2}$ power law gas profile. In all cases, the halo mass is $10^8 M_\odot$, the BH mass is $10^5 M_\odot$, the redshift is $z = 20$ and the kick velocity is 100 km s^{-1}.

In our calculations, we are mainly interested in whether the kicked BH is ejected and lost (i.e., can not contribute to the final SMBH mass at $z = 6$), or is retained (i.e., eventually returns to the nucleus, and can be incorporated into the $z = 6$ SMBH). We place the following retention condition for recoiling BHs: the BH must return to within 1/10th of the virial radius of the newly merged host halo within 1/10th of the Hubble time. The fate of a (SM)BH placed in an orbit extending to the outskirts of its host halo is uncertain: even if it is not lost through tidal interactions with an incoming merging halo, it is not likely to form a binary that hardens efficiently. We therefore impose the above conservative cutoff, in order to avoid tracking these vagrant BHs. Our retention threshold velocity, $v_{\rm ret}$, above which recoiling BHs do not return within the prescribed time limit and are considered lost, is a numerically calculable function of M, m, z and halo composition. In order to minimize computation time, we tabulate $v_{\rm ret}$ in the range $5 < z < 40$, $10^5 M_\odot < M < 10^{15} M_\odot$ and $10^{-6} < m/M < 1$ and approximate the result with a fitting formula that accurately reproduces the exact numerical results within 5% in the tabulated range. In principle, we should compute the retention velocity as a function of the time to the next merger experienced by the halo. However, we find the time dependence to be weak. The distribution of the time intervals between halo mergers in a given merger tree has a sharp peak at $\approx 10^{-1} t_{\rm Hub}$, with far fewer mergers occurring at $\sim 10^{-2} t_{\rm Hub}$ and $\sim t_{\rm Hub}$. At these tails of the distribution, $v_{\rm ret}$ varies by $\lesssim 10\%$ from the value for $10^{-1} t_{\rm Hub}$. We find that $v_{\rm ret} \sim 5 - 8 \times \sigma_{\rm SIS}$. This is comparable to the escape velocity for a *non-dissipative* pure SIS profile that is truncated at the BH radius of influence, $v_{\rm esc} \approx 5\sigma_{\rm SIS}$ if $m = 10^{-3} M$ (Yoo & Miralda-Escudé 2004).

The weak dependence on $v_{\rm ret}$ on the return time limit is a counterintuitive result, but it can be understood as follows. There is a minimum kick speed that is required to displace the hole beyond $0.1r_{200}$, which represents $\lim_{t\to 0} v_{\rm ret}$; and there is a maximum kick, $\lim_{t\to\infty} v_{\rm ret}$, beyond which the BH remains completely unbound from the halo, even in the presence of drag. $v_{\rm ret}(t)$, then is a function of time that is always between these two extreme values. However, owing to the high central density of our gas–dominated halos, the difference between these two limits is small, $\sim 10\%$. Since this difference is smaller than other model uncertainties, such as those stemming from discrepancies from the actual density profile of high-redshift DM halos (e.g. clumping or triaxiality) and the merger tree prescription, we simply use the retention velocity computed for the approximate median time limit, $0.1t_{\rm Hub}$.

Finally, for simplicity, we treat the halo as a static mass distribution during each recoil event. That is, we ignore the cosmological evolution of the DM potential, and we assume that the recoiling BH does not affect the medium by clumping or heating it. Note, however that the latter feedback may play a nontrivial role in real systems, since the kinetic energy of a recoiling hole can be comparable to the gravitational binding energy of the entire host halo and can be expected to cause significant disruption of the surrounding matter.

2.2.6 The Black Hole Accretion Rate

We turn now to our prescription for the accretion rate of the BHs in our model. Of particular interest is the possibility that the gravitational recoil effect will significantly limit the ability of kicked BHs to accrete gas, by displacing them into low-density regions

for prolonged periods, and/or by limiting through high relative velocities the amount of gas that can be gravitationally captured. One can imagine a scenario in which a SMBH whose progenitors have survived numerous kicks but have spent long episodes in underdense regions may have a final mass much less than that predicted by simple Eddington growth. We therefore follow the accretion rate self-consistently, as the recoiling holes proceed along their radial orbits. Specifically, in our models a BH embedded in gas is assumed to undergo standard Bondi-Hoyle-Lyttleton (BHL) accretion,

$$\dot{m}(r,v) = \frac{4\pi G^2 \rho_b(r)}{(c_s^2 + v^2)^{3/2}} m^2. \tag{2.15}$$

The accretion rate is capped at the Eddington rate,

$$\dot{m} = \frac{1-\epsilon}{\epsilon} \frac{m}{t_{\rm Edd}}, \tag{2.16}$$

where $t_{\rm Edd} = 44$Myr and ϵ is the radiative efficiency, for which we assume $\epsilon = 0.10$.

If the gas density is too low, or if the sound speed or its velocity with respect to the gas disk too high, a BH may not be able to accrete at the Eddington rate even if it is close to the center of a halo. Because the BHL rate is proportional to m^2, an underfed BH with initial mass m_0 will eventually reach the Eddington accretion rate ($\propto m$) at a threshold mass

$$m_{\rm Edd} = \frac{1-\epsilon}{\epsilon} \frac{c_s^3}{4\pi G^2 \rho_b \, t_{\rm Edd}} \approx 3500 \left(\frac{c_s}{4 \text{ km s}^{-1}}\right)^3 \left(\frac{\rho_b}{M_\odot \text{pc}^{-3}}\right)^{-1} M_\odot, \tag{2.17}$$

where 4 km s^{-1} is the isothermal sound speed for an ionized hydrogen gas at 1200K. The typical central density of a 1200K halo with a TIS gas profile is $\sim 5 \times 10^{-3}(1+z)^{7/3} M_\odot \text{pc}^{-3}$.

Sub-Eddington accretion rates are not an issue for BHs in a power-law gas profile, as the steep profile provides a sufficient central density for immediate Eddington accretion.

The time it takes for a BH with $m_0 < m_{\rm Edd}$ to reach this threshold mass, assuming that the local density remains constant, is

$$t_{\rm crit} = \left(\frac{m_{\rm Edd}}{m} - 1\right)\frac{\epsilon}{1-\epsilon}t_{\rm Edd} \approx 2.7\left(\frac{m_0}{100M_\odot}\right)^{-1}\left(\frac{c_s}{4\ {\rm km\ s^{-1}}}\right)^3\left(\frac{\rho_b}{M_\odot {\rm pc}^{-3}}\right)^{-1}{\rm Gyr}. \quad (2.18)$$

$t_{\rm crit} \sim$ a few 100 Myr for $100M_\odot$ halos embedded in 1200K TIS halos at $z \gtrsim 20$, and $t_{\rm crit} \gtrsim$ Gyr for $z \lesssim 14$. This means that for TIS gas profiles, seed holes will spend a significant fraction or all of the available time prior to $z \approx 6$ accreting below the Eddington rate.

The difference between BHL and Eddington accretion rates as they relate to BH growth is also discussed in Volonteri & Rees (2006). However, in that paper the context is for the direct formation of $m > 10^4 M_\odot$ intermediate-mass BHs through super-Eddington BHL accretion. We here adopt the opposite extreme assumption, i.e. that the BH radiates efficiently at all times, and its accretion rate obeys the Eddington limit. The BHL rate can then initially be sub-Eddington in TIS halos, owing to the low BH mass and low gas density (the baryon density required to fuel BHL accretion at the Eddington rate was also discussed by Turner 1991).

To illustrate the impact of extended sub–Eddington growth phases, we perform a simple analytic calculation. Suppose that a BH is formed with mass $m_{\rm seed}$ at redshift z in a halo with virial temperature 1200K, and that the local gas density is held constant at the value when the BH was formed. Note that for this exercise, we assume that gas density

is constant even as the halo around the BH is growing by merging with other halos – in other words, we assume these mergers deliver gas to the nucleus containing the original BH, roughly maintaining a constant density at the Bondi radius around the BH. Figure 2.4 shows the maximum possible mass that can be attained by such a BH growing in isolation through gas accretion alone, assuming that the accretion rate is determined solely by Equations (2.15) and (2.16) and that accretion is not supply–limited. If the host halo has a steep power-law cusp, the accretion rate is Eddington throughout regardless of when the seed BH is formed. However, if the central fuel density is low, then it is possible for the local BHL accretion rate to be significantly sub-Eddington initially. In such a scenario, the earliest forming seeds are the only ones able to reach the Eddington rate; the late-forming seeds are unable to reach the Eddington rate before $z = 6$. In this scenario the late-forming seeds, which are easily identifiable by the drastically shallower growth slope in the figure, cannot grow rapidly enough to contribute to the SMBH population. Note that assuming a constant gas density in this calculation gives an optimistic accretion rate for the TIS case, as in those profiles the central gas density generally decreases with Hubble expansion and significantly reduce the BHL rate.

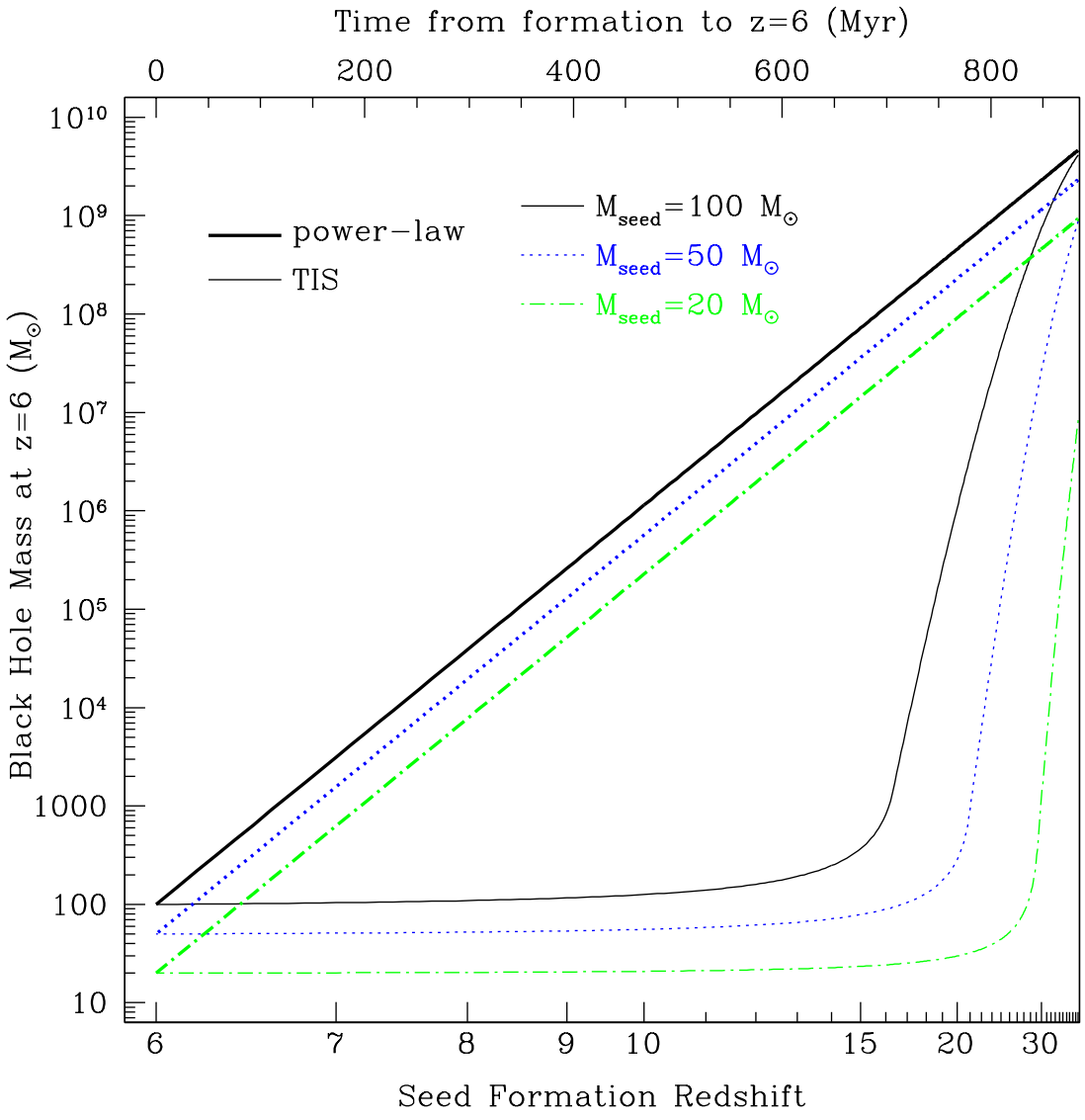

Figure 2.4 The maximum possible accreted mass by redshift $z = 6$ for a seed BH born with a seed mass $m_{\rm seed}$ in a halo with virial temperature $T_{\rm vir} = 1200$K at redshift z. If the BH is always surrounded by a steep gas profile, with a power-law cusp, the growth remains Eddington throughout (which appears as a straight line in this log-log plot). If the gas profile has a flat core (as in the TIS profile), the central density is initially insufficient to feed the BH at the Eddington rate, resulting in much slower growth. The cuspy and corey gas distributions are demarcated by thick and thin lines, respectively, and different seed BH masses are shown in different colors (line styles).

It is computationally expensive to numerically integrate the individual orbits and accretion histories of every recoiling BH in our simulations. We therefore tabulate the accretion growth in Eddington units during the first $0.1t_{\rm Hub}$ of the orbit, and describe the results in a fitting formula in the same manner as we have done for the retention velocity. Given a specific prescription for the baryon distribution, we tabulate across four relevant variables in the following ranges: $10^5 M_\odot < M < 10^{15} M_\odot$, $10^{-6} < m/M < 1$, $5 < z < 40$, and $0 < v_{\rm kick}/v_{\rm ret} < 1$.

In the absence of a kick, and if the accretion rate were always at the Eddington limit, the SMBH mass in a given halo at $z \approx 6$ is easily approximated by

$$m_{\rm SMBH} \approx N_{\rm seed}\ \exp\left(f_{\rm duty}\frac{\epsilon}{1-\epsilon}\frac{t_{\rm seed,6}}{t_{\rm Edd}}\right) m_{\rm seed}, \tag{2.19}$$

where $m_{\rm seed}$ is the seed BH mass, $N_{\rm seed}$ is the number of seeds in the merger history of the halo, $t_{\rm seed,6}$ is the available time between the typical seed formation time and $z \approx 6$ and $f_{\rm duty}$ is the time–averaged duty cycle for accretion. Equation (2.19) represents the ideal, maximally efficient scenario for SMBH assembly, and we can use it to effectively measure the cumulative impact of underfed accretion, recoil-induced ejection, and other factors that limit the assembly efficiency.

The measurements of clustering of quasars in the SDSS suggest that the duty cycle of active (luminous) accretion increases steeply with redshift at $3 \lesssim z \lesssim 6$, with the most active quasar BHs at $z \approx 6$ showing $0.6 \lesssim f_{\rm duty} \lesssim 0.9$ (Shen et al. 2007; Shankar et al. 2010). We therefore adopt duty cycles of ≥ 0.6. Although it is likely that SMBHs regulate their own growth through feedback mechanisms, we do not address such scenarios a priori. To

keep our models as simple as possible, we will only impose the loosest upper limit on the SMBH accretion rate: they must not accrete more mass than there the total mass of baryons in the host halo. However, we will discuss an alternative ad–hoc model below, which is able to reproduce the well-known relation between SMBHs and the velocity dispersions of the bulges of host galaxies (the m-σ relation).

2.2.7 Putting Together the $z = 6$ SMBH Mass Function

Explicitly constructing the SMBH mass function at $z = 6$ over a wide mass range is computationally intractable. The host halo mass inferred from the observed quasar space density from the $z \sim 6$ quasars is several $\times 10^{12} M_{\odot}$ (e.g., Fan 2006). For every halo with this mass, there are $\sim 10^7$ halos with $10^8 - 10^9 M_{\odot}$. Blindly calculating the SMBH mass for every halo with $M > 10^8 M_{\odot}$ using our trees-plus-orbits algorithm would be prohibitively expensive computationally.

We therefore carry out a piecewise calculation of the SMBH mass function that is computationally tractable. The procedure is as follows: (1) We group the halo population into logarithmic mass bins of size $x < \log_{10} M < x + \Delta x$; (2) For each bin, we select $\gtrsim 10^2 - 10^4$ individual Monte-Carlo-generated halos with masses randomly generated from the Press-Schechter distribution at $z = 6$; (3) We simulate the BH population for each such halo using our trees-plus-orbits algorithm, and assume the resulting sample is representative of all $z = 6$ halos in the mass bin; (4) For each bin we multiply the sample by the appropriate weight to construct the entire Press-Schechter halo mass function, $\int_x^{x+\Delta x} dN/d\ln M d\ln M$; and finally (5) Sum the contributions from each bin. The result is

a Press-Schechter distribution of halos with $M > 10^8 M_\odot$ at $z = 6$, with a statistical representation of the corresponding SMBH mass function. The bins used and their relevant properties, including the number of Monte-Carlo halos that were cloned to populate the full mass function, are listed in Table 2.1. This numerical shortcut is not used for the most massive halos. 40 halos are expected above $10^{12.85} M_\odot$, and these are simulated individually.

Table 2.1 Masses and quantities of simulated dark matter halos.

$M_{\rm lo} < M < M_{\rm hi}$	$\log_{10} N_{\rm bin}$	$\log_{10}\langle M_{\rm bin}\rangle$	$N_{\rm sim}$	$W_{\rm bin}$
$8.0 < \log_{10} M < 8.5$	12.33	8.22	50000	4.28×10^7
$8.5 < \log_{10} M < 9.0$	11.78	8.72	27000	2.23×10^7
$9.0 < \log_{10} M < 9.5$	11.20	9.22	15000	1.06×10^7
$9.5 < \log_{10} M < 10.0$	10.57	9.71	9000	4.13×10^6
$10.0 < \log_{10} M < 10.5$	9.87	10.2	5000	1.48×10^6
$10.5 < \log_{10} M < 11.0$	9.08	10.7	2700	4.45×10^5
$11.0 < \log_{10} M < 11.5$	8.14	11.2	1500	9.20×10^4
$11.5 < \log_{10} M < 12.0$	6.98	11.7	900	1.06×10^4
$12.00 < \log_{10} M < 12.5$	5.50	12.1	500	632
$12.50 < \log_{10} M < 12.85$	3.48	12.6	303	10.
$12.85 < \log_{10} M < \infty$	1.60	12.9	40	1.0

For each mass bin, BH assembly was simulated by creating a merger tree for $N_{\rm sim}$ Monte-Carlo halos, and the results were multiplied by the weighting factors $W_{\rm bin}$ to represent the Press-Schecter mass function for DM halos at $z = 6$. $N_{\rm bin} = \int_{M_{\rm lo}}^{M_{\rm hi}} dN/dM\, dM$ is the expected Press-Schechter number of halos in each bin, and $\langle M_{\rm bin}\rangle$ is the number-weighted mean halo mass in each bin.

This method allows a fast calculation of the BH mass function, with the caveat that there must be enough BHs in each bin to keep statistical uncertainties to a minimum. Unfortunately, this is not always preventable for models with extremely low $f_{\rm seed}$, and the reader will notice statistical noise in the results of such models. In some cases, we increase the halo sample by a factor of 10 in an attempt to reduce the errors.

In all, our simulations represent the Press-Schechter population of DM halos in a comoving volume of $\approx 280\mathrm{Gpc}^3$, roughly equal to the comoving volume that was probed by the SDSS between $z \approx 5.7-6.4$. We have simulated a total of $\approx 1.17\times10^5$ DM halos (see the column $N_{\rm sim}$ in Table 2.1), and through the procedure described above extrapolated the results to represent the $\approx 3\times10^{12}$ Press-Schechter halos (with $M > 10^8 M_\odot$) expected to be present in the $280\mathrm{Gpc}^3$ volume.

2.3 Results and Discussion

2.3.1 Building the $> 10^9\mathrm{M}_\odot$ SMBHs

As stated in the Introduction, our primary goals are (i) to understand the possible ways in which the $> 10^9\mathrm{M}_\odot$ SMBHs may have been assembled at redshift $z > 6$, and (ii) whether the *LISA* event rates are sufficiently different in competing models so that one can disentangle the physical assembly scenario from the *LISA* data stream. As a first step toward these goals, we would like to survey all feasible combinations of the physical assembly parameters, understand the impact of each model parameter, and look for corresponding give–away features in the predicted *LISA* stream.

Although a broad simulation survey is beyond the scope of this paper, we have undertaken a cursory tour of the basic parameter space. We begin by considering two basic seed models: $100M_\odot$ seeds forming as remnants of Pop III stars in minihalos when they first reach the virial temperature $T_{\rm vir} \geq T_{\rm seed} = 1200$K and $10^5 M_\odot$ seeds forming as a result of direct collapse of gas in halos when they first reach $T_{\rm vir} \geq 1.5\times10^4$K. For each

case, we consider halos with an NFW DM component and a gaseous component with either a cuspy $\rho_{\rm gas} \propto r^{-2.2}$ power-law or a corey TIS profile. The TIS models consistently failed to produce SMBHs by $z = 6$, with typical maximum BH masses of $\lesssim 10^3 M_\odot$. In those models the central gas densities are too low to allow for prolonged episodes of accretion near the Eddington rate, as noted in Section 2.5 above.

For each type of seed, we vary three sets of parameters: (i) the seeding fraction $10^{-3} \leq f_{\rm seed} \leq 1$, i.e. the probability that a halo reaching the critical temperature will form a seed BH; (ii) the time–averaged accretion rate in Eddington units, characterized by the duty cycle $f_{\rm duty}$, for which we use $f_{\rm duty} = 0.6$, 0.8 and 1.0 (note that $f_{\rm duty}$ and the radiative efficiency ϵ are degenerate in our prescription; see below); and (iii) whether at the time of their merger, the BH spins are randomly oriented or aligned with the orbital angular momentum of the binary. The spin magnitudes are chosen from a uniform random distribution $0 < a_{1,2} < 0.9$ (Schnittman & Buonanno 2007). We do not track the evolution of BH spins in our models. While more rapidly spinning BHs are capable of accreting at higher efficiency, we neglect this effect and assume a global efficiency coefficient in our models. Of the four main ingredients of the SMBH assembly introduced in § 1 above, we here therefore vary three: the seed rarity, the accretion rate, and the recoil dynamics. For the fourth, the merger rate, we have simply assumed that BHs merge when their parent halos do, as extreme–mass BH binaries are rare in our simulation, given the threshold we have imposed on the mass ratio for halo mergers. We will call the above our fiducial set of parameters.

For each realization, we compute the mass function of BHs at $z = 6$ with $m > 10^5 M_\odot$;

the rate of events in the *LISA* observational mass range, $\sim 10^{4-7}/(1+z)M_\odot$; the fraction of DM halos hosting a central massive BH; and m/M, the ratio between the mass of the SMBH and its host DM halo, which serves as a proxy for the m-σ relation.

We begin with Figure 2.5, showing the mass function for the $m_{\rm seed} = 100M_\odot$ Pop III seed model. This and the companion figures are organized with different values for $f_{\rm seed}$ in different columns, the two spin prescriptions in different rows, and the different duty cycles in different line styles (and different colors, in the online version). This will be the default organization of our 8-panel figures unless noted otherwise.

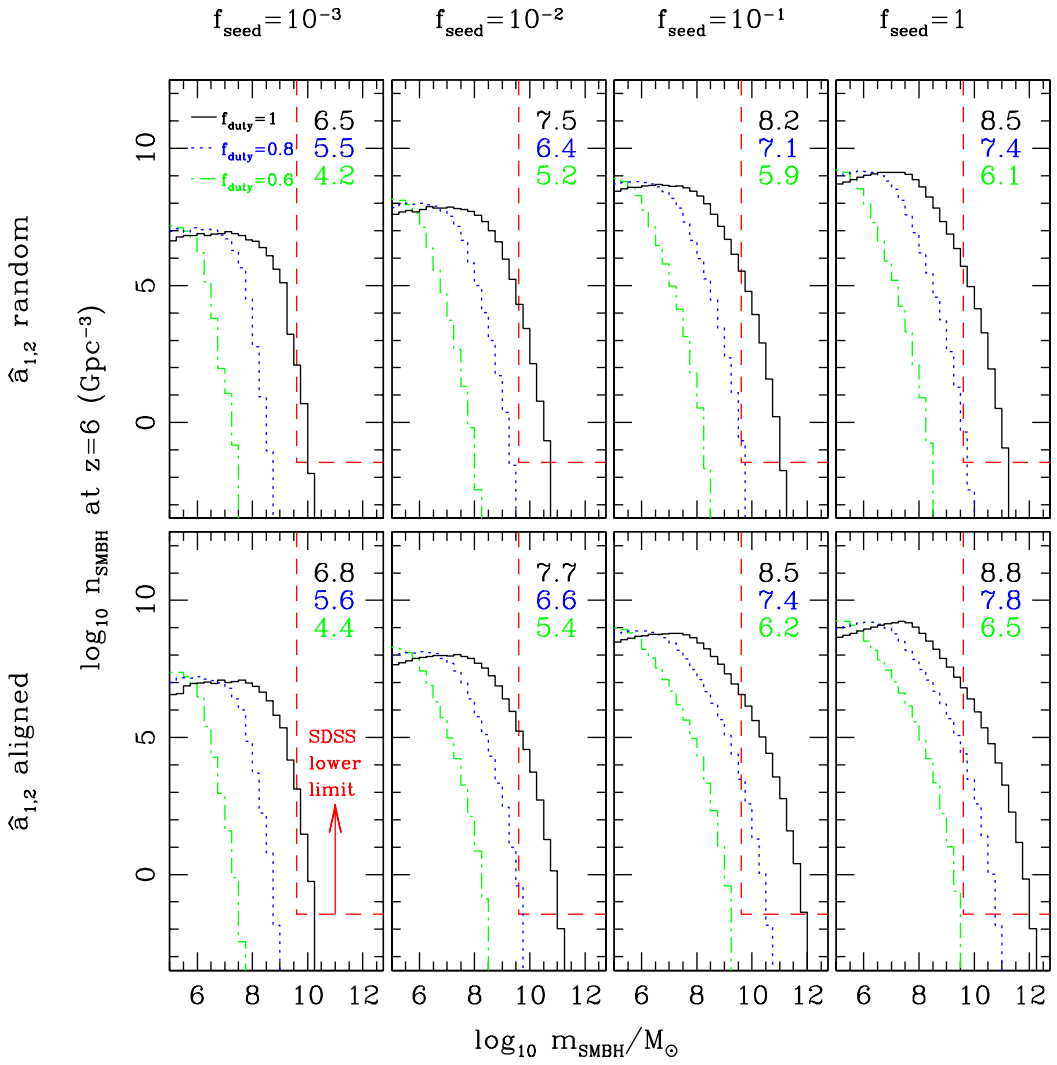

Figure 2.5 The comoving number densities of SMBHs in different mass bins at redshift $z = 6$. Colored figures are available in the online version. The 24 different models shown in the figure assume different parameter combinations as follows. The columns, from left to right, adopt $f_{\rm seed} = 10^{-3}$, 10^{-2}, 10^{-1}, 1. The top row is for simulations with random binary spin orientation, and the bottom row is for spins aligned with the binary's orbital angular momentum. Time–averaged accretion rates are distinguished by color: black (solid, $f_{\rm duty} = 1$), blue (dot, $f_{\rm duty} = 0.8$), and green (dash-dot, $f_{\rm duty} = 0.6$). The numbers in the upper-right corners represent $\log_{10}[\rho_\bullet/(M_\odot\ {\rm Mpc}^{-3})]$ for each model, in descending order of $f_{\rm duty}$. The red (dashed) line demarcates the rough indication for the minimum number of $z \approx 6$ SMBHs in the observable universe with $m \gtrsim 10^{9.6} M_\odot$ given the area surveyed by SDSS.

The numbers in the upper-right-hand corner of each panel represents the total SMBH mass density, $\log_{10}[\rho_\bullet/(M_\odot\ \mathrm{Mpc}^{-3})]$, for all BHs with $m > 10^5 M_\odot$. Because of statistical fluctuations, for multiple model realizations with identical parameters this value can vary by $\lesssim 10\%$ for $f_{\rm seed} \gtrsim 10^{-1}$, and as much by a factor of a few for $f_{\rm seed} \lesssim 10^{-3}$. For each of these simple models, this density is exceedingly high compared to the SMBH density of the local universe. For the present discussion, we will overlook this point as we address the effects of the various model factors; we will introduce the additional constraint from $\rho_\bullet$ in subsequent sections.

The most stringent observational requirement for the high-mass end of the $z \approx 6$ SMBH mass function comes from the SDSS observation of the $z \approx 6.4$ quasar J1054+1024, which has an inferred mass of $\sim 4\times10^9 M_\odot$. Since this object was detected in a region $\sim 10\%$ of the sky, we estimate that $\gtrsim 10$ similar objects may exist at $z \sim 6$. We have adopted this as a rough indication of the lower limit of the SMBH mass function at redshift $z = 6$, and represent this limit by the upper right quadrangle in each of the panels, delineated by the red dashed lines. Note that these lower limits are conservative, since they do not require the presence of any additional SMBHs with comparable mass that are "off". Since high values for the duty cycle – near unity – are suggested by quasar clustering measurements (as discussed above), and are also required for growing the most massive SMBHs (as we find below), this correction is only of order a factor of ~two. As seen in the figure, the high–mass end of the SMBH mass function is so steep that increasing the lower limit on the required SMBH space density by even $\sim$ 2 orders of magnitude would make little difference to our conclusions.

The first conclusion one can draw from this figure is that several combinations of parameters can produce SMBHs massive enough to power the quasar J1054+1024. We will return to this and other observational constraints in the next subsection, but let us for now focus on understanding the effects of our various parameters and their combinations. In general, the effect of varying each of the parameters is relatively straightforward to understand. Increasing the accretion rate, increasing the seed occupation fraction, and aligning the spins all tend to result in more massive BHs. Note that the accretion rate in Eddington units, $f_{\rm duty}$, is degenerate with the accretion efficiency, as $\dot{m} \propto f_{\rm duty} \times (1-\epsilon)/\epsilon$. For example, a model with $\epsilon = 0.15$ and $f_{\rm duty} = 1.0$ is equivalent to $\epsilon = 0.10$ and $f_{\rm duty} = 0.63$. We have used $\epsilon = 0.10$ throughout the results presented here. With this value for the efficiency, it is possible to build the most massive SDSS quasar SMBHs by $z \approx 6$, starting with $100 M_{\odot}$ seeds. If the efficiency is ≈ 0.15, however, it is only possible to build the SMBHs in question with the most optimistic assumptions: $f_{\rm duty} \approx 1$, $f_{\rm seed} \gtrsim 0.1$, and spin alignment would all be required.

We note two non–trivial observations to be made from Figure 2.5. First, if the seed fraction is low, the spin orientation has a minimal effect on the BH mass function. This is because if seeds are extremely rare, they are likely to grow in isolation for much of their existence along with their host halos. The first mergers are not likely to occur until the gravitational potentials of their host halos are deep enough to retain them from any recoil event. Second, the increase in the SMBH abundance from increasing the seeding fraction has a tendency to plateau. This is the reverse situation compared to the low $f_{\rm seed}$ limit: if seeds are very common, they are likely to experience multiple BH mergers very early

in the merger tree, when their masses are still comparable, and ejections become very common. As $f_{\rm seed}$ increases, assembly becomes increasingly inefficient at early times. In models where the seeding halo temperatures are lower, the efficiency saturates at lower values of $f_{\rm seed}$. Furthermore, in models with the shallower TIS gas profiles, we find that increasing the occupation fraction beyond a certain "sweet spot" value rapidly decreases the final SMBH masses. The reasons for this are that (i) the seed BHs in these models barely grow, making near–equal mass ($q \sim 1$) BH mergers, and therefore large kicks, much more common and (ii) the gas drag is reduced, making it easier to kick holes out of these halos. We conclude that arbitrarily increasing the number of seed holes contributing to the assembly process is not necessarily an efficient solution to the SMBH assembly problem. In fact, excessive seeding can lead to a different conflict with observations – overproducing the mass density in lower–mass SMBHs –, to which we will return in the next subsection.

Let us turn now to Figure 2.6, which shows the BH occupation fraction in $M > 10^8 M_\odot$ DM halos at $z = 6$, and the BH-to-halo mass ratio (which here serves as a proxy for the m-σ relation; see, e.g. Ferrarese 2002) for each of the models. Here, we have arbitrarily defined our occupation fraction to mean the fraction of DM halos that host a BH with a minimum mass of $10^4 M_\odot$, as we are interested in the population of nuclear BHs and not stellar remnants. Menou et al. (2001) have shown that, in the low-redshift Universe, the fraction of DM halos hosting a SMBH will approach unity, regardless of the initial occupation fraction at earlier times. We find that at $z \approx 6$, the occupation fraction in halos with $M \lesssim 10^{11} M_\odot$ can still be significantly below unity, depending primarily on $f_{\rm seed}$. In principle, a future survey that can determine the quasar luminosity function

(LF) at $z = 6$ to several magnitudes deeper than the SDSS could look for this drop in the occupation fraction, since it will produce a flattening of the LF at magnitudes below some threshold. The spin prescription has a noticeable effect on the $z = 6$ occupation fraction for $f_{\rm seed} \gtrsim 10^{-2}$. On the other hand, the duty cycle essentially only affects the mass of the BHs, and not their presence or absence, and so has a minimal effect on the occupation fraction, within the parameter range shown. Note that we do not expect to reproduce the m-σ relation in our simulations, as we employ simple and z-independent accretion and seeding prescriptions, and our models have no feedback to enforce the relation. Instead, the m/M relation should be taken as a sanity check that we are producing a physically viable BH population. Note that in some of the models shown in Figure 5, the BHs grow much larger than the m/M relation observed in nearby galaxies. In particular, as the figure reveals, SMBHs in the lower–mass ($\sim 10^8$ $M_\odot$) halos in the models with $f_{\rm duty} = 1$ tend to consume most of the gas in their parent halos, clearly an improbable result.

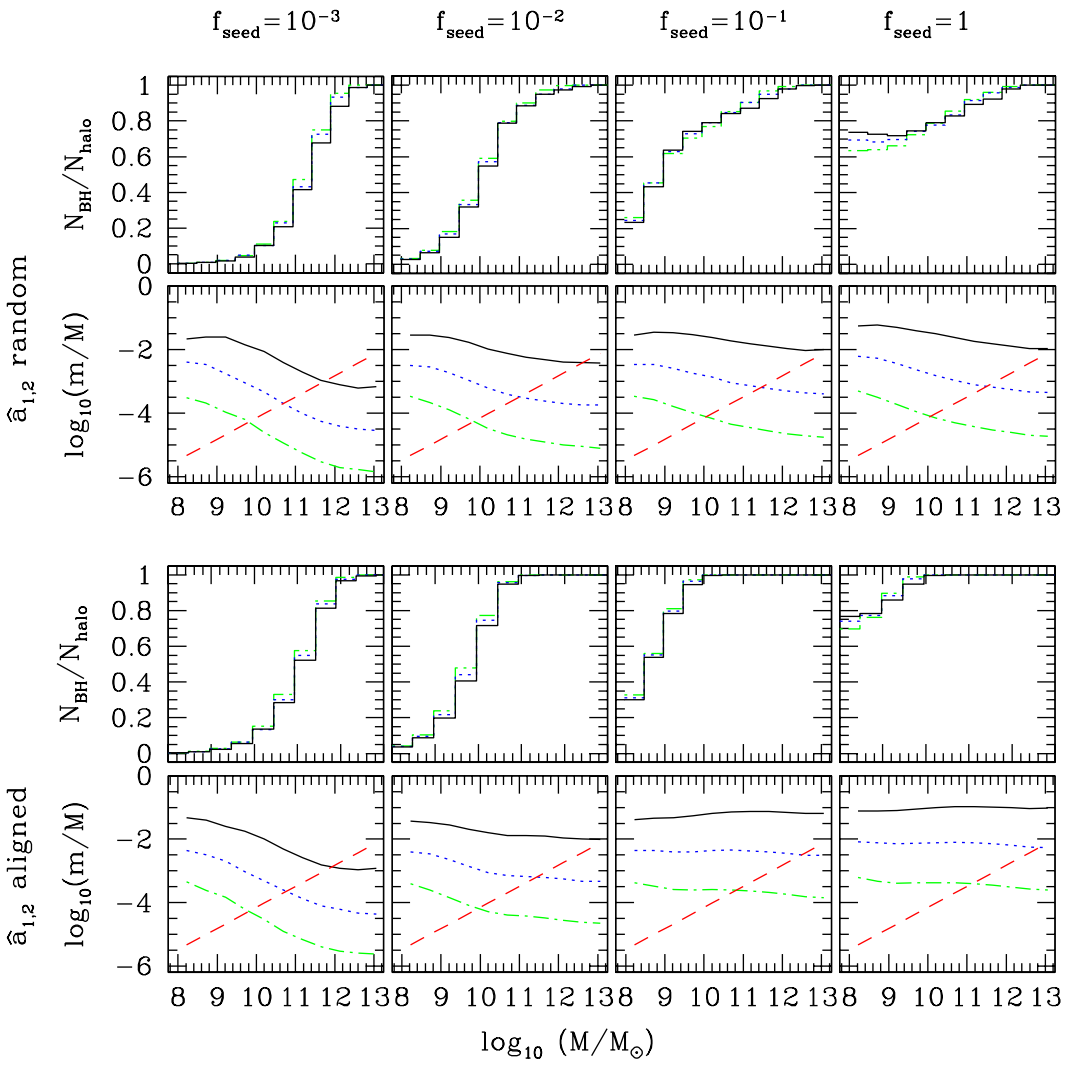

Figure 2.6 Properties of the SMBH population at $z = 6$ as a function of the halo mass M: the percentage of DM halos hosting a central BH (assumed at most to be one BH per halo; top rows in both the upper and lower panels) with $m \geq 10^{-5}M$, and the mean BH–to–halo mass ratio m/M for the halos that do host a BH (bottom rows). Color (line-style) and panel schemes are the same as in Figure 2.5. The red (dashed) line is the empirical m/M relation extrapolated to $z = 6$ (see Equation 2.20; Wyithe & Loeb 2003a; Ferrarese 2002). Our m/M relation has the opposite trend with respect to halo mass from the trend observed in the local universe. Note that in some cases, the central BH consumes most of the baryonic mass in the halo.

For an explicit comparison, alongside the m/M relation produced by our models, we have plotted the the value expected based on measurements of the m-σ relation in local galaxies. Specifically, we show the expression from Wyithe & Loeb (2003a),

$$m = 10^{12}\epsilon_0 \left(\frac{M}{10^{12}M_\odot}\right)^{\gamma/3} \left(\Omega_0 h^2\right)^{\gamma/6} (1+z)^{\gamma/2} \approx 1.16 \times 10^7 \left(\frac{M}{10^{12}M_\odot}\right)^{\gamma/3} (1+z)^{\gamma/2}. \quad (2.20)$$

We adopt their parameter choices of $\epsilon_0 = 10^{-5.4}$ and $\gamma = 5$. This expression satisfies the SDSS constraints at the high–mass end of the mass function. It also agrees well with the relation suggested by Ferrarese (2002) for SMBHs in the local universe, $m \sim 10^7 (M/10^{12}M_\odot)^{1.65}$. As the figure shows, our predicted m/M relation tends to have the opposite slope than the one inferred from the observed m-σ relation: in our results m/M decreases with mass or stays roughly constant as M increases, but this is the opposite of the empirical trend. This is due mainly to the difference in the growth rates of halos and holes: our simple prescription for steady, exponential accretion for the BHs can significantly exceed the growth rate of DM halo masses due to the accretion of unresolved low–mass halos in the EPS merger tree. As a result, in some cases, the host halo mass predicted for the $z = 6$ quasars is as low as $10^{11}M_\odot$, which is an order of magnitude lower than would be predicted from the extrapolation of the locally measured m/M relation, or from the inferred space density of the host halos (e.g., Haiman & Loeb 2001). However, any extrapolation of the $m - \sigma$ or m/M relation to high redshift, and the masses of halos that host the brightest $z > 6$ quasars, at present, have large uncertainties, and do not robustly preclude such low values. As will be discussed below, the overly rapid growth of SMBHs in this suite of models motivates modifications to the modeling, including a

model in which the extrapolated m/M relation holds by assumption.

Figure 2.7 shows estimates for the *LISA* event rate, calculated for all binary mergers satisfying $10^4 M_\odot < (m_1 + m_2)(1+z) < 10^7 M_\odot$ as (see, e.g., Menou et al. 2001),

$$\frac{d^2N}{dz\,dt} = 4\pi c\, d_{\rm com}^2(z) \frac{\Delta N}{\Delta z\,\Delta V}, \tag{2.21}$$

where ΔN is the number of SMBH merger events in the tree in a time step Δz and a simulated comoving volume ΔV, and $d_{\rm com}$ is the comoving distance. Although there is a mild dependence on the duty cycle / accretion rate and the kick prescription, it is evident that for our assembly models $f_{\rm seed}$ has the greatest effect in setting the rate of SMBH binary mergers detectable by *LISA*. Because the initial merger rates scale as $f_{\rm seed}^2$, the measured event rate is extremely sensitive to the BH number population. Since the merger activity peaks near $z \lesssim 10$, *LISA* should be able to measure the masses of most of these SMBH binaries to high prevision (Hughes 2002; Arun et al. 2009). Although the raw detection rate – without any information on BH spin or mass ratio – alone will be richly informative, the ability to determine the source masses with high fidelity should be very useful in further constraining the growth rate of the first SMBHs, and breaking degeneracies between the allowed parameter–combinations. The observed distribution of binary masses as a function of redshift will provide direct snapshots of the mass function and shed light on its evolution, independently from quasar luminosity surveys (e.g., Yu & Tremaine 2002).

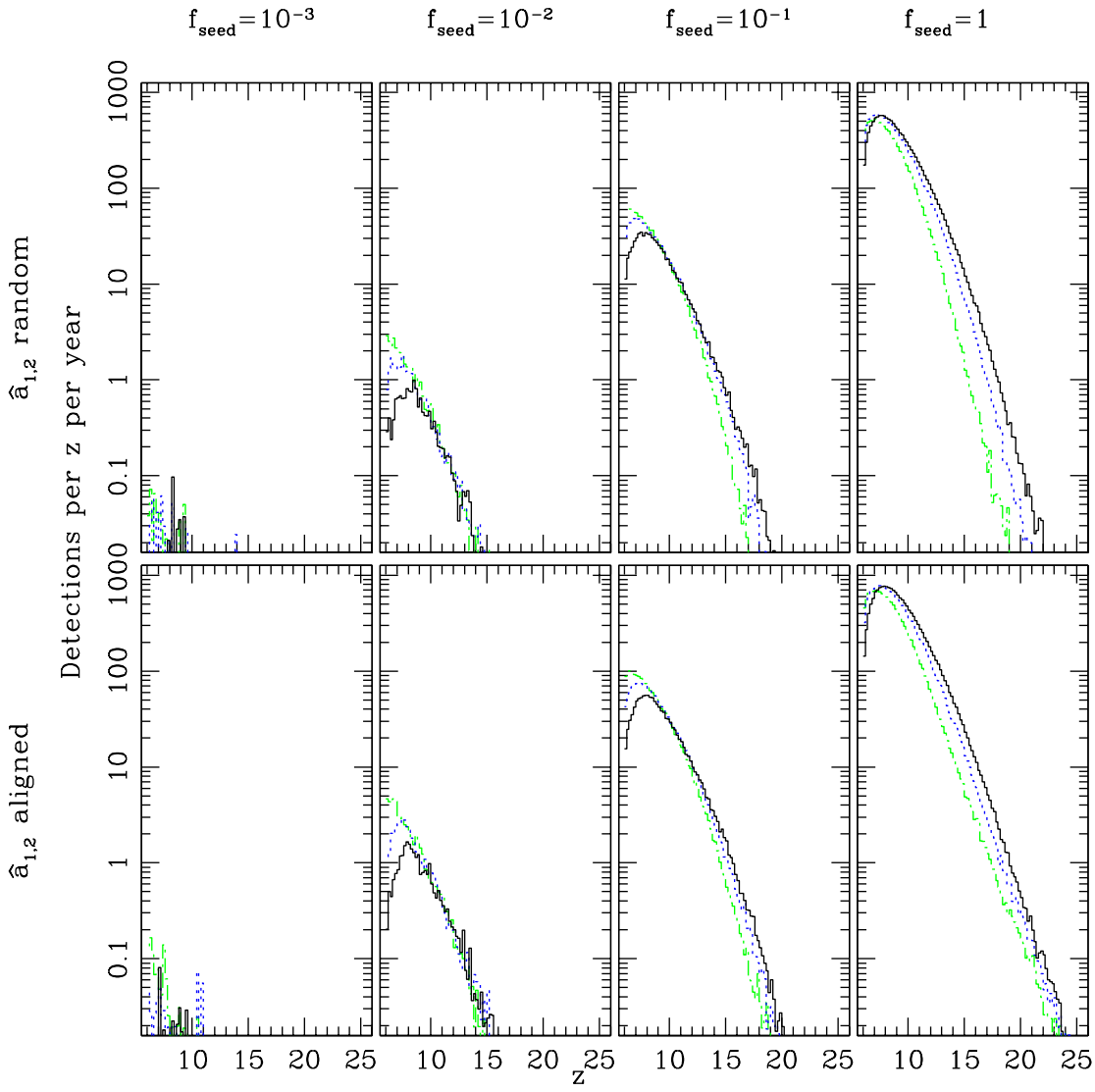

Figure 2.7 Expected number of *LISA* detections per redshift per year due to SMBH mergers with binary mass $10^4 M_\odot \leq (m_1+m_2)(1+z) \leq 10^7 M_\odot$. The color, line-type and panel schemes are the same as in Figures 2.5 and 2.6.

Another family of assembly models that has been frequently discussed in the literature is the so-called "direct collapse" model, wherein BHs with $m \sim 10^{4-6} M_\odot$ are formed rapidly from gas cooling via atomic H in halos with virial temperature $T \gtrsim 10^4$K (Oh & Haiman 2002; Bromm & Loeb 2003; Volonteri & Rees 2005; Begelman et al. 2006; Spaans & Silk 2006; Lodato & Natarajan 2006). We simulate such a family of models, for the same set of the parameters $f_{\rm seed}$ and $f_{\rm duty}$ and the same spin alignments. We choose $m_{\rm seed} = 10^5 M_\odot$ and $T_{\rm seed} = 1.5 \times 10^4$K. We show the mass functions, the hole-halo relations and the *LISA* rates (the same information as in Figures 2.5, 2.6 and 2.7) in Figures 2.8, 2.9 and 2.10. Although the main differences are all fairly intuitive, it is instructive to address how the direct-collapse models differ from the Pop-III seed models.

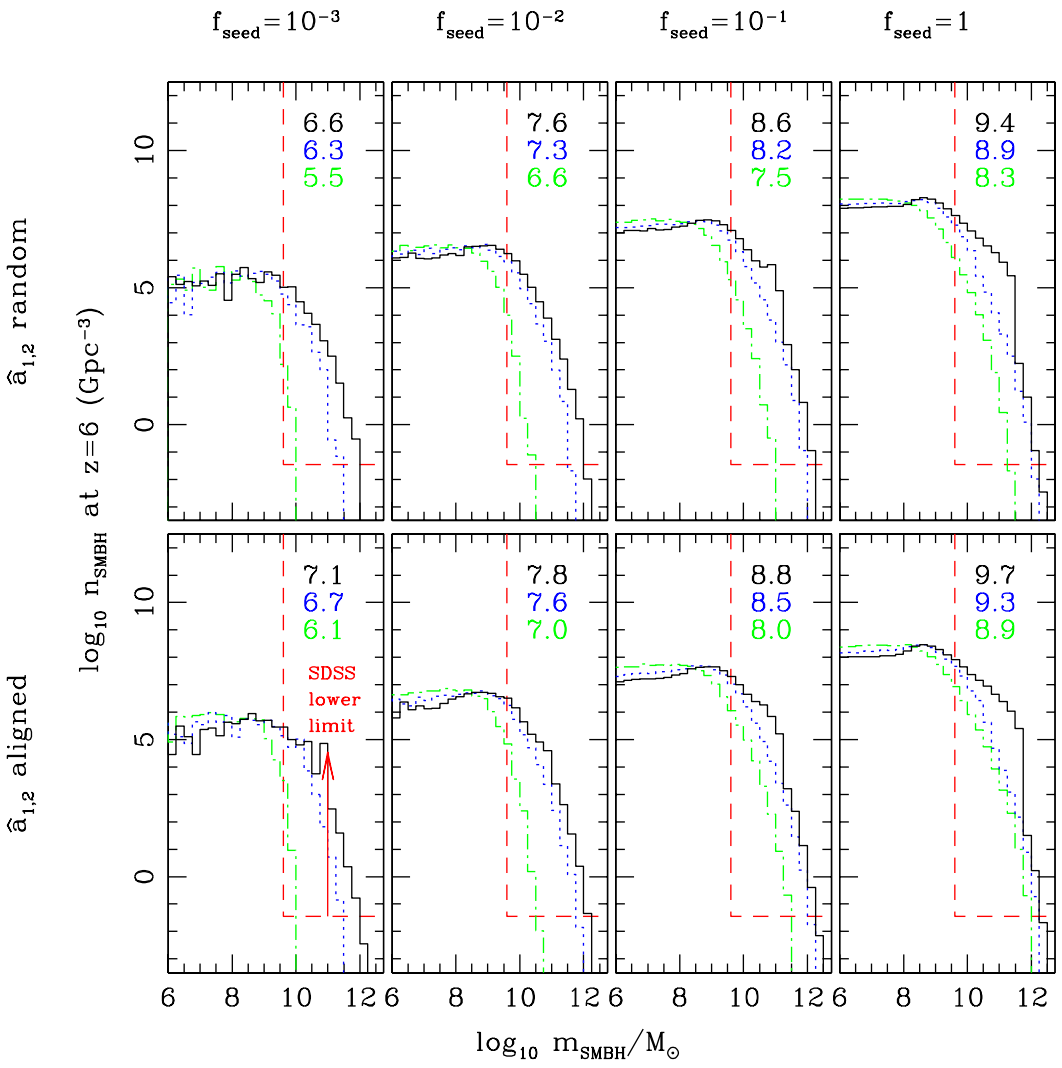

Figure 2.8 The $z = 6$ SMBH mass function in the direct collapse scenarios, with $m_{\rm seed} = 10^4 M_\odot$ and $T_{\rm seed} = 1.5\times10^4$K. Color and panel organization for accretion rate, seed fraction and spin alignment is the same as in Figure 2.5.

First, from Figure 2.8 we see that it is much easier to construct more massive SMBHs owing to the larger seed masses. In fact, the models with $f_{\rm seed} \gtrsim 0.1$ become unphysical, since the SMBHs in these models would exceed the baryon mass $(\Omega_b/\Omega_m)M$ of their parent halos. The second point is that the mass function is very differently distributed between the Pop-III and direct-collapse scenarios. The reader will note that for each set of parameters, the overall SMBH density does not differ significantly between the corresponding seed models. However, this is deceiving as the mass function is clearly different, with the direct-collapse seeds resulting in a more "top heavy" SMBH population. For the range of parameters surveyed, the Pop-III model has $\gtrsim 90\%$ of the SMBH mass in the $m < 10^7 M_\odot$ range if $D = 0.6$, compared to $\lesssim 30\%$ for the $D = 0.6$ direct-collapse seed models. For $D = 1.0$ and random spin alignment, $\sim 2 - 5\%$ of the SMBH density resides in the most massive $m > 10^9 M_\odot$ BHs if the seeds are Pop III; for the same parameters, $\sim 60 - 70\%$ of the mass is in the billion-plus solar mass BHs in the corresponding direct-collapse models. Note that the total BH mass density remains extremely high; again, we will address this point shortly.

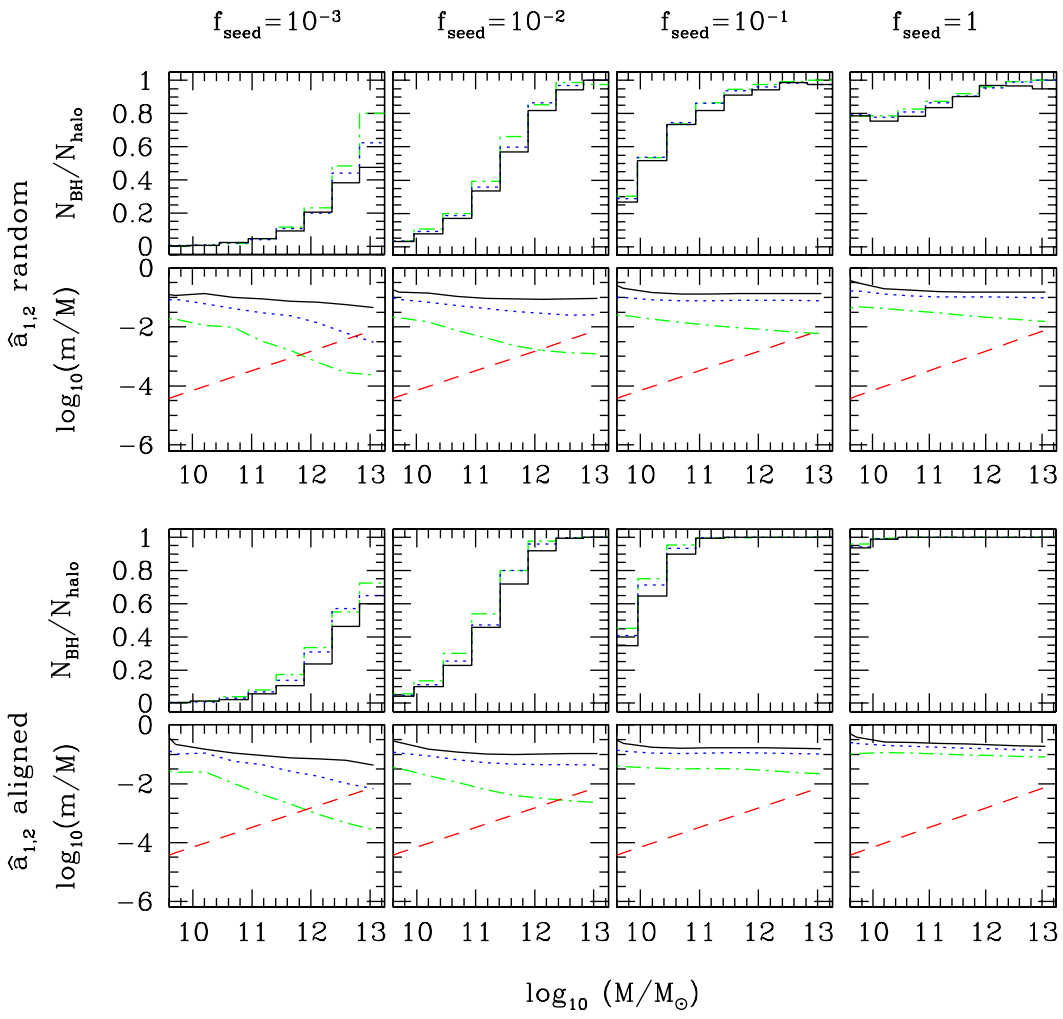

Figure 2.9 The SMBH occupation fraction and the m/M ratios in the direct collapse models. Refer to Figure 2.6 for color and panel organization. The dotted line is the extrapolated empirical m/M relation.

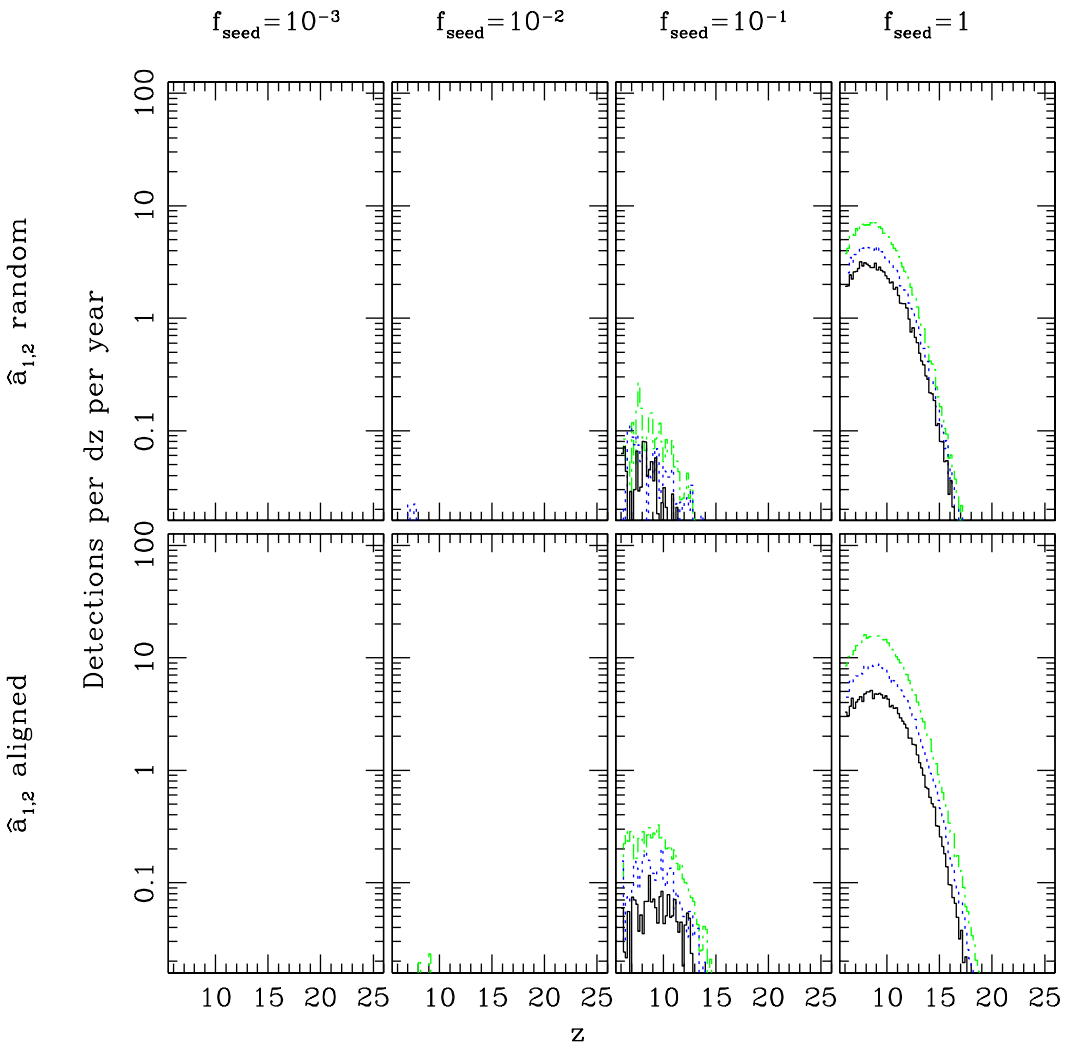

Figure 2.10 *LISA* event rates in the direct collapse scenarios. Refer to previous figures for color and panel organization. Note the significant reduction in the event rates relative to the pop–III seed models, owing to the smaller number of seed-forming halos in the tree.

Third, there is a slightly weaker dependence on the spin orientation of the BH binaries. This is most apparent by comparing the $z = 6$ occupation fractions in Figures 2.6 and 2.9, and is due to the deeper potentials of the host halos in the direct collapse case. Fourth, even though the m/M relation continues to have a slope opposite to the locally observed trend, there appears to be a break in the relation at $M \sim 10^9 M_\odot$, accompanied by a drop in the occupation fraction. This is due to the simple fact that halos below this mass threshold cannot have many $T > 1.5 \times 10^4$K progenitors and the model similarly prohibits intermediate-mass BHs. This cutoff contributes to the top-heaviness of the mass function for these models. Fifth, *LISA* rates are lower by one to two orders of magnitude than in the Pop-III seed models, because there are fewer 1.5×10^4K halos than 1200K halos (another factor is that the seed BHs are already born with a mass near the middle of *LISA*'s logarithmic mass range, so they spend only $\sim$ half the time in the *LISA* band, compared to the Pop-III seeds). It is worth noting, in particular, that it is possible to build the SDSS quasar BHs in ways that produce no detectable GW events for observation with *LISA* beyond $z \sim 6$ (in contrast, in the successful pop–III models, a minimum of a few events are predicted). In such scenarios, SMBHs are extremely rare until $z \sim 6$, at which point the SMBH occupation fraction will evolve toward unity fairly rapidly, as described by Menou et al. (2001).

Finally, in Figure 2.11 we present the mass function and the *LISA* detection rate in five variants of a fiducial $f_{\rm duty} = 0.65$, $f_{\rm seed} = 1$, aligned spin model with pop-III BH remnant seeds. We choose these values because they just barely are able to match the abundance of the most massive SDSS quasar BHs, and because they produce the most

BH mergers, and thus the effects of varying the other recoil-related parameters will be the most discernible. We show the fiducial model in bold. In the modified models, we (i) allow the gas density profile to be shallower, with a r^{-2} power law – this is to allow for the possibility that the gas surrounding the BH has not cooled and condensed to high density (dark blue curves); (ii) require the BH spins to be near-maximal at $a_{1,2} = 0.9$, instead of choosing them randomly from the range 0.0 to 0.9 – this is to allow for the possibility that all SMBHs at high z are rapidly spinning, e.g. due to coherent accretion (green curves); (iii) allow halos of all mass ratios to merge, where previously we had considered a halo with mass less than 1/20th that of its merger companion to become a satellite (yellow curves); (iv) assume that the Pop III seed progenitors are not able to blow out the gas in the host minihalo (red curves); and (v) ignore the effects of accretion suppression due to episodes of recoil-induced wandering, and instead assume that all BHs accrete at $f_{\rm duty}$ unless ejected (light blue curves).

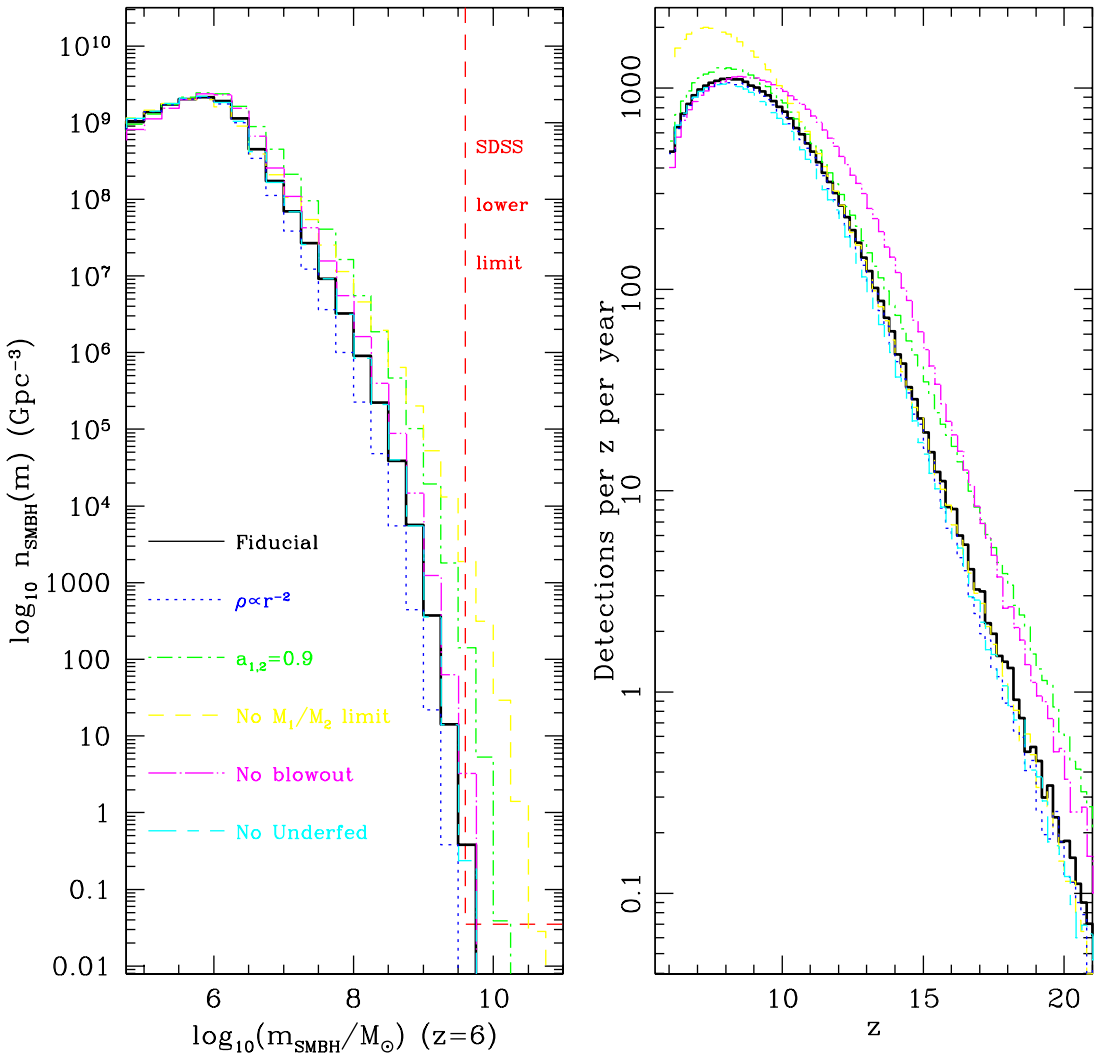

Figure 2.11 Properties of the SMBH population under several variants of our fiducial models. All models plotted have $f_{\rm duty} = 0.65$, $f_{\rm seed} = 1$, and aligned binary spins, and modify a single aspect of the basic fiducial model prescription, as labeled: the gas density is an isothermal power–law instead of the fiducial $\rho \propto r^{-2.2}$ (dark blue, dotted curve); spin magnitudes are near-maximal at $a_{1,2} = 0.9$ (green, dot-short-dash curve); halos are allowed to merge and form BH binaries irrespective of their mass ratio (yellow, dashed curve); Pop III stars do not blow out the gas from their host halos prior to leaving a seed BH (magenta, dot-long-dash curve); and recoiling BHs wandering in low-density regions continue to accrete efficiently (light blue, long-short-dash curve). We also ran simulations with a corey, TIS gas profile for the host halos. We were unable to produce SMBHs of even $\gtrsim 10^7 M_\odot$ in those models even when prescribing the most optimistic values for the other assembly parameters (therefore results from these models are not shown in the figure).

We also ran models with a TIS profile for the gas component. We found that these models always failed to produce any SMBHs above $10^6 M_\odot$ by redshift $z = 6$ due to the initial phase of sub–Eddington accretion of seed BHs, and their results are not shown in Figure 2.11. This implies that SMBHs must be continuously surrounded by dense cores of gas that was able to cool at the centers of DM halos – feeding holes with the low–density gas in DM halos whose gas was unable to cool does not allow for high enough accretion rates (Turner 1991 emphasized the same issue for the growth of $z \sim 4$ quasar BHs).

The results shown in Figure 2.11 give insight to the importance of the assumptions that went into our models. In particular, of all the parameters that we have varied for fixed $f_{\rm seed}$ and $f_{\rm duty}$, the most significant for the SMBH mass function, by far, are the spin orientation and the limit on the halo mass ratio for timely mergers. Both of these have a similar effect of increasing the number of BHs, especially at the massive end. The former result – that maximizing the spin increases the SMBH mass function – may seem surprising at first, since generally large spins imply larger kicks. However, this is not always the case, as can be understood from equations (5-8). Under the assumption that both spins are aligned with the orbital angular momentum, $v_\parallel = 0$, and $v_\perp \propto (a_1 - q a_2)$ is maximized for *unequal* spins (i.e. $a_1 = 1$ and $a_2 = 0$ for a typical $q \sim 1$); setting $a_1 = a_2 = 0.9$ therefore eliminates the largest kicks, and allows more BHs to be retained. Comparatively smaller differences are visible in the mass function for the other model variants. Figure 2.11 also shows that adding in the unequal–mass halo mergers and increasing the spins affect the *LISA* event rates differently: the former adds new events mostly at $z \lesssim 10$, where unequal–mass mergers are more common, whereas increasing the spin mostly adds new

events at $z \gtrsim 10$. Interestingly, ignoring the blow–out has little impact on the SMBH mass function at $z = 6$, but it does shift the *LISA* events to higher redshifts, especially at $z \gtrsim 10$. Figure 2.11 suggests that the *LISA* event rate can be useful in disentangling these three effects.

Perhaps the most surprising inconsequential variation is ignoring episodes of reduced accretion due to the BH wandering in low-density outskirts following recoil. The reason for this is simply that lengthy oscillating orbits are relatively rare if the central gas density is high; the BHs either return quickly, or are ejected (by assumption). Recall that our definition for a "retained" BH is that the recoiled hole must return to within 1/10th of the halo's virial radius within 1/10th of the Hubble time. For low kick speeds, the BH does not get very far, because the gas provides both a steep gravitational potential barrier and a strong dissipative sink. The orbit will thus be rapidly damped, with only very brief periods of underfeeding. If a BH is kicked hard enough to reach the outskirts of the halo, there is very little dissipation there, to tug it toward the center or to further damp its oscillation. Since the radial velocity at this point is low, it's likely to remain outside 1/10th of the virial radius for a significant period. The bottom line is that the range of initial kicks that would take the BH to the low–density outskirts to cause significant underfeeding, while still allowing it to return quickly enough to be "retained" is simply negligibly small.

Blecha & Loeb (2008) recently performed a more detailed analysis of the orbits of recoiling SMBHs that include a multi-component halo mass distribution and three–dimensional orbits. They report that recoil velocities of between 100 km s^{-1} and the escape velocity lead to significant suppression of the accretion rate, with SMBHs accreting

only $\sim 10\%$ of its initial mass over $10^6 - 10^9$ years of wandering through the halo. We note that (1) in our simulations we ignore the longest–wandering BHs through our prescribed retention threshold; (2) typical kick magnitudes for SMBHs are lower than 100 km s^{-1} in our simulations, a point we explain below; and (3) their prescription of the baryon distribution results in a lower central density than our models, as in that paper they are concerned with typical galaxies at low redshift, and not with minihalos and protogalaxies. We conclude that prolonged periods of wandering and underfeeding are unlikely to have a significant effect on the mass function of high–z SMBHs as a whole. Oscillations may play a more prominent role in the growth history of SMBHs if large kicks are more common (and the retained holes tend to be marginally retained), if the halo gravitational potential is more shallow, or if the effect of dynamical friction on BH orbits is less than what we have considered in this paper. Also, halo triaxiality (Blecha & Loeb 2008) and/or a clumpy mass distribution (Guedes et al. 2008) could increase the time that kicked holes take to return to the halo center (or they may never return). In principle, this may increase the impact of these oscillations. However, in practice, the inner, baryon-dominated regions of the galaxies are likely close to smooth and spherical. The BHs that are not ejected in our models typically do not leave these regions and so will not be subject to large changes in their orbits from these effects.

The results presented thus far seem to paint a relatively simple set of relevant parameters for SMBH assembly. There is the seeding function $f_{\rm seed}$, which governs the BH merger rate and therefore to a large extent the *LISA* event rate. The event rate also depends on the time–averaged accretion rate, parameterized here by the duty cycle $f_{\rm duty}$, and the

initial seed mass $M_{\rm seed}$, as these set the evolution of the observable mass spectrum. The seeding prescription and $f_{\rm duty}$ determine the mass function almost entirely if $f_{\rm seed} \ll 1$. If $f_{\rm seed} \gtrsim 0.1$, then the spins of the BH binary play a role in setting the recoil speeds and the subsequent evolution.

Once typical spin and mass parameters of merging SMBHs are determined by *LISA*, either by direct measurement, or perhaps by extrapolating from detections at lower redshift, combined with the event rate and what is known about the upper end of the mass function, this information is likely sufficient to give at least a strong indication on the typical SMBH growth rate and initial seed mass.

Our simulations above also confirm a result reported by Volonteri & Rees (2006), namely that SMBHs form primarily through repeated mergers of the most massive SMBHs merging with less-massive (SM)BHs. This is because the gravitational rocket speeds decrease rapidly as the mass ratio q decreases. The first few BHs that "outweigh" their neighbors – be it through being endowed with more mass at birth, accreting faster or being fortunate enough to survive the first mergers – will be less likely to be ejected from their host halos. This survival advantage becomes a runaway effect, as each subsequent merger will result in a lower value of q for the next merger.

2.3.2 Constraints on the SMBH Mass Function

Now that we have a first-glance grasp of the assembly parameters and their most basic observational characteristics, we can turn to identifying actual candidate models for the formation of $m \gtrsim 10^9 M_\odot$ SMBHs before $z \approx 6$.

The suite of models discussed above has demonstrated that there are several feasible ways to build the SMBHs. These models have so far focused only on the number density of $m \gtrsim 10^9 M_\odot$ SMBHs, and ignored indirect constraints that exist on the mass function at lower masses. In particular, the total mass density in SMBHs with masses in the range $10^6 M_\odot \lesssim m \lesssim 10^9 M_\odot$ in the local Universe has been estimated by several authors, who find $\sim 4 \times 10^5 M_\odot$ Mpc^{-3} (to within a factor of ~two; see, e.g., the recent paper by Shankar et al. 2009b and references therein). Furthermore, a comparison of the locally observed mass density with the mass density inferred from accretion by the evolving quasar population suggests that $\sim 90\%$ of the total local SMBH mass density is attributable to quasar accretion. This implies that the total SMBH mass density increased by a factor of ~ 10 between $z \sim 6$ and the local Universe (Yu & Tremaine 2002; Haiman et al. 2004; Shankar et al. 2009b), which then places an indirect constraint on the SMBH mass function, down to $m \sim 10^5 M_\odot$, at $z = 6$.

To be specific, we set the following upper limit on the expected total $z \approx 6$ SMBH mass density in $m \gtrsim 10^5 M_\odot$ SMBHs:

$$\rho_{\rm SMBH,5+}(z = 6) \sim 0.1 \times \rho_{\rm SMBH,6+}(z \approx 0) \sim 4 \times 10^4 M_\odot \ {\rm Mpc}^{-3}, \tag{2.22}$$

That is, we assume that the total mass density of SMBHs with mass $m > 10^5 M_\odot$ at $z = 6$ is at most 10 percent of the total inferred mass density of SMBHs with mass $m > 10^6 M_\odot$ in the local universe. The major caveat to making such an expectation is that we assume that the low–mass end of the BH mass function has grown by a factor of 10, and that high–redshift quasar luminosity surveys have sufficiently accounted for selection effects

of any SMBHs that may be hidden by inactivity or by being too dim.

The analysis that follows below is similar to that of Bromley et al. (2004), who considered the upper limit to the $z \approx 6$ SMBH mass density in weighing the plausibility of various $z = 6$ quasar BH assembly models. The main difference is that Bromley et al. (2004) did not consider the gravitational recoil effect. Adding in this recoil makes the problem significantly worse. This is because the recoil necessitates more optimistic assumptions in order to build up the $\sim 10^9 M_\odot$ SMBHs, which tends to increase, by a large factor, the predicted number of lower–mass $\sim 10^6 M_\odot$ SMBHs, which arise later, and whose growth is therefore less sensitive to the kicks. In other words, the kicks preferentially suppress the abundance of the most massive SMBHs which arise from the earliest seeds; as a consequence, we predict steeper SMBH mass functions than the models considered in Bromley et al. (2004). We performed a series of assembly calculations explicitly without recoil, and find that these indeed produce mass functions with a flatter slope and a higher normalization, owing to greater SMBH masses at the high–mass end, and a higher occupation fractions at all masses. If there are no kicks, less optimistic assumptions are required to produce the SDSS quasars, and the overall mass density is lower in no-kick scenarios that produce the minimum number of $\gtrsim 10^9 M_\odot$ SMBHs.

The basic result we find is that among the models presented thus far, *all of those that match the SDSS abundance of the most massive SMBHs overshoot the above value by two or more orders of magnitude.* One possible solution to this apparent problem is that there is no problem at all: we have simply set our constraint too severely. Perhaps not all $10^5 M_\odot$ SMBHs at $z \approx 6$ have evolved to become $m > 10^6 M_\odot$ BHs in galactic nuclei in the local

universe. Still, there is no obvious way to "hide" these low–mass SMBHs locally, and the over–prediction in the SMBH densities in our simulations are very large: we find that in our fiducial models, we typically need to reduce the population of SMBHs in the mass range $10^{5-7}M_{\odot}$ by a factor of $\gtrsim 100$, and those in the $10^{7-9}M_{\odot}$range by a factor of $\gtrsim 10$.

2.3.3 Successful Models I: BH Seeds Stop Forming Early

One logical solution, and one that has been suggested by Bromley et al. (2004), is to simply stop making seeds below some cutoff redshift. The earliest seeds contribute the majority of the total mass of the most massive SMBHs, and late-forming seeds tend to end up in lower-mass SMBHs. In principle, then, by introducing a cutoff redshift for seed formation, one can suppress the low-mass end of the mass function, while still allowing for the formation of the most massive SMBHs. Such a model would also be in line with our physical understanding of seed BHs. We know Pop III stars stopped forming relatively early on in the universe, with the halt in production being due to trace metal contamination (e.g., Omukai et al. 2008), radiative feedback from the UV and/or X-ray background during the early stages of reionization (e.g., Haiman et al. 2000), the higher turbulence of gas in the centers of later halos, or a combination of these factors.

This proposed solution amounts to keeping $f_{\rm duty}$ a constant while allowing $f_{\rm seed}$ to evolve with z; in the simplest case, as a step function dropping to $f_{\rm seed} = 0$ below some redshift. Figure 2.12 shows the fractional contribution $(dM/dz_{\rm seed})/M$ from seeds forming at different redshifts to the final mass at $z = 6$ in three different bins of the $z = 6$ SMBH mass function, for two different models that are marginally able to assemble the SMBHs

powering J1054+1024. The model in the left panel has $f_{\rm seed} = 10^{-3}$, while the model on the right has $f_{\rm seed} = 1$. Note that contributions to each mass bin are normalized to integrate to unity, but the two lower mass bins $\rho(10^{5-7} M_\odot)$ and $\rho(10^{7-9} M_\odot)$ make up the vast majority of the total mass density. The formation epochs of the seeds contributing the majority of the mass of the different SMBHs are very distinct in the model with lower $f_{\rm seed}$, but overlap significantly for $f_{\rm seed} = 1$. What accounts for this qualitative difference in the assembly epochs? There are two ways to build $10^9 M_\odot$ holes: accretion onto the earliest-forming holes that undergo few mergers and grow mostly in isolation, and the mashed-together product of many seeds that may have formed later. If the occupation fraction is high, one expects both populations to contribute, and therefore a wide distribution for $dM/dz_{\rm seed}$. If the occupation fraction is low, SMBHs with many progenitors contributing become rare, and we find that the most massive SMBHs can only be formed from isolated seeds in the earliest minihalos, through few mergers, and so the dM/dz curves in these models are sharply peaked.

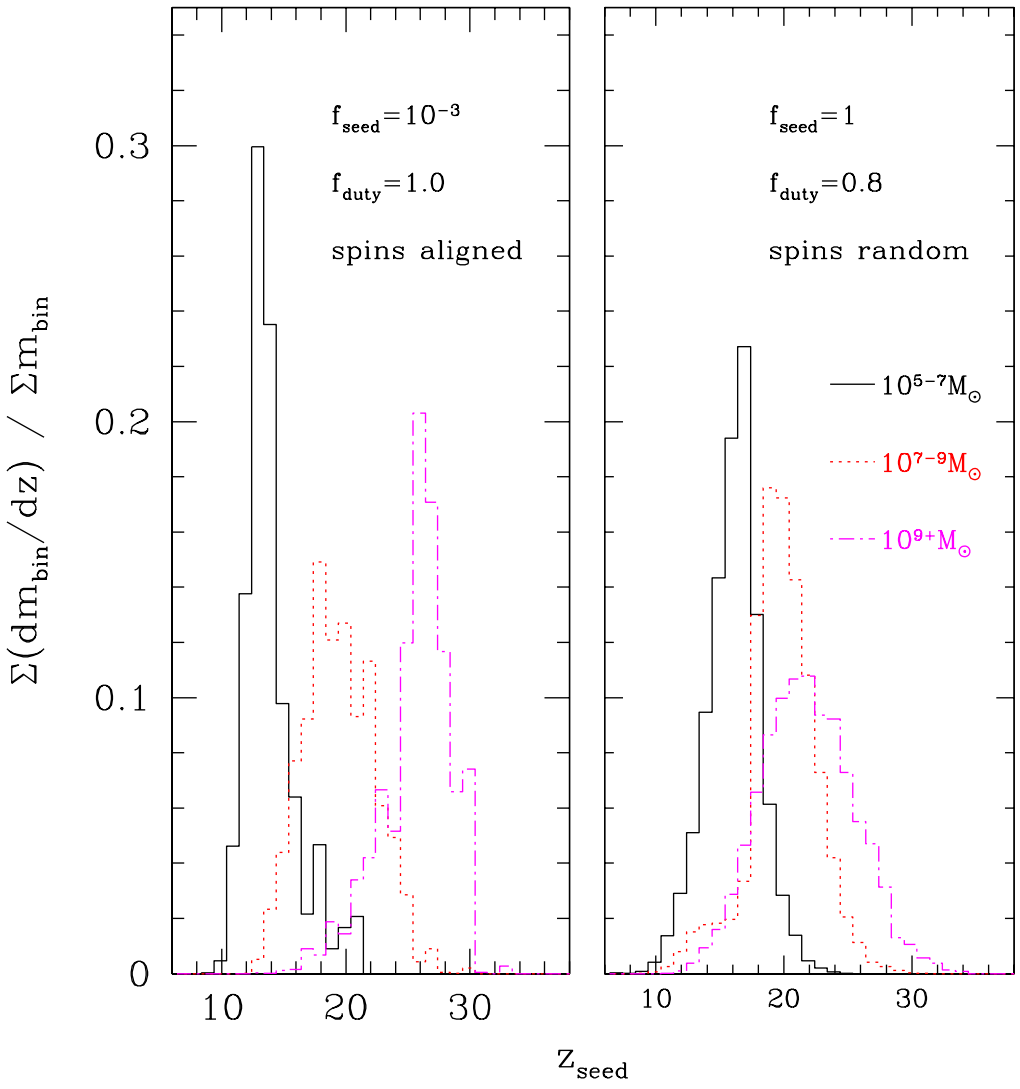

Figure 2.12 The fractional contribution to the mass of $z = 6$ SMBHs from $100M_\odot$ seed BHs that form at different redshifts $z_{\rm seed}$. The contributions are computed in three mass bins of $z = 6$ SMBHs: $10^5M_\odot \leq m < 10^7M_\odot$ (black, solid lines), $10^7M_\odot \leq m < 10^9M_\odot$ (red, dotted), and $m \geq 10^9M_\odot$ (magenta, dash-dot). The most massive $z = 6$ SMBHs arise mainly from the earliest 1200K progenitors of the most massive halos ($z \gtrsim 20$), whereas the seeds contributing most of the mass of lower–mass SMBHs formed later ($z \lesssim 20$).

Our task is to eliminate $\approx 99\%$ of the BH mass in the lower–mass bins, while leaving most of the $m \gtrsim 10^9 M_\odot$ holes unaffected. By simply examining the normalized dM/dz curves in Figure 2.12, one can see that simply cutting off seed formation at an arbitrary redshift for the $f_{\rm seed} = 1$ model is not a viable solution to the overproduction problem, as there is no way to eliminate the lower-mass SMBHs without eliminating a significant fraction of the $10^9 M_\odot$ holes. Cutting off the seed production can produce successful mass functions for models with $f_{\rm seed} \ll 1$, but only if the seed cutoff occurs at very high redshifts, typically $z \sim 30$. Essentially, the solution calls for a very brief and early period of seed BH formation, and very rare seed formation there after. An unfortunate generic consequence of this early cutoff redshift is that it quickly chokes off the *LISA* event rates.

We also simulated models where seed production continues beyond the cutoff redshift – with the same probability $f_{\rm seed}$ as in the minihalos at $z > z_{\rm cut}$ – in halos with $T_{\rm vir} > 10^5$K. Such halos could continue to form BHs if heating by the UV background is the primary mechanism for seed suppression, as they are able to shield their central gas from the UV radiation (Dijkstra et al. 2004). We find that models with such a partial cutoff overproduce the SMBH mass density if $f_{\rm seed} \gtrsim 10^{-4}$. This result implies that either the initial occupation fraction is very low (it is still possible to make the SDSS BHs with this low $f_{\rm seed}$; see Table 2 below), or else some other feedback beyond UV radiation, such as metal–enrichment, stops seed BHs from forming in all halos at $z < z_{\rm cut}$, even in the rare, more massive ones.

In short, the requirement in models in which the duty cycle is a constant, but seeds stop forming suddenly below some redshift, can be simply summarized as follows: the only way to build SMBH mass functions that satisfy observational constraints and indi-

cations at both the low-mass and the high-mass ends is from extremely rare seeds that form during a brief and very early epoch. We also note that these models are attractive because (i) there are physical reasons for the seeds to stop forming below some redshift, and (ii) there is independent empirical evidence, from constraints on the reionization history from *WMAP* measurements of the cosmic microwave background polarization anisotropies, that the ionizing luminosity in high–redshift minihalos was suppressed by a factor of $\gtrsim 10$ (Haiman & Bryan 2006).

We present in Table 2.2 the parameters for four such successful models. While it is not computationally feasible to search the entire parameter space for such models, we present two typical examples for both the Pop–III–remnant and direct–collapse seed BH scenarios. $f_{\rm seed}$ is low in each of these examples, as we have argued above that they must be. Although we have listed the spin prescriptions, they are relatively unimportant because seeds are rare (see Section 3.1 above). Given a particular value for $f_{\rm seed}$, the only free parameters are the accretion rate $f_{\rm duty}$ and the seed cutoff redshift $z_{\rm cut}$. For the Pop III models, we find in the range $f_{\rm seed} \geq 10^{-4}$ that seeds must typically stop forming at $z_{\rm cut} > 30$, with a lower cutoff $z_{\rm cut} \sim 20$ for the direct–collapse models. An important conclusion is that in each of these models, GW events are too rare ($< 10^{-3}/dz/yr$ for $z < 30$) for *LISA* detections beyond $z \sim 6$ to occur within the mission's lifetime.

The mass density of *ejected* BHs is exceedingly low in each of the successful scenarios, $< 10^{-3} M_\odot$ Mpc^{-3}. Because the total number of seeds is small, so are the number of ejected holes. We only give an upper limit here, because the ejected holes are too rare for us to give a robust value given the statistical limitations of our "halo cloning" method for

Table 2.2 Properties of four successful (A-D) and two failed (X and Y) models for SMBH growth

Model	$m_{\rm seed}$	$T_{\rm seed}$	$f_{\rm seed}$	$f_{\rm duty}$	spin	$z_{\rm cut}$	$\rho_{\rm SMBH,5+}(z=6)$
A	$200M_\odot$	1200K	10^{-4}	1	aligned	25	$3.4\times10^4 M_\odot$ Mpc^{-3}
B	$100M_\odot$	1200K	10^{-2}	0.95	aligned	28	$5.1\times10^4 M_\odot$ Mpc^{-3}
C	$10^5 M_\odot$	1.5×10^4K	10^{-4}	0.6	random	13	$6.2\times10^4 M_\odot$ Mpc^{-3}
D	$2\times10^5 M_\odot$	1.5×10^4K	10^{-2}	0.55	aligned	18	$7.0\times10^4 M_\odot$ Mpc^{-3}
X	$100M_\odot$	1200K	1	0.8	random	0	$2.9\times10^8 M_\odot$ Mpc^{-3}
Y	$10^5 M_\odot$	1.5×10^4K	10^{-3}	0.6	aligned	0	$1.1\times10^6 M_\odot$ Mpc^{-3}

The above shows parameters for four models that (1) have constant accretion rates of $f_{\rm duty}$ times the Eddington rate; (2) produce by $z = 6$ SMBHs massive enough to power the SDSS quasars; and (3) do not overproduce the overall SMBH population. In Models A and B the seed BHs are Pop III remnants, and in Models C and D the seeds are formed through direct–collapse in more massive halos. Models X and Y are unsuccessful models that barely produce the $m \gtrsim 10^9 M_\odot$ SMBHs by $z = 6$ but also far overproduce the lower-mass SMBHs.

populating the entire halo population at $z = 6$ (the mass density in ejected BHs can be large in models with large $f_{\rm seed}$; see below).

These models represent the simplest scenarios for SMBH formation, requiring a very brief period of seed formation and a prolonged period of accretion at rates comparable to the Eddington rate, and consequentially they represent the most pessimistic predictions for *LISA*'s observational prospects.

2.3.4 Successful Models II: Feedback Adjusted to Maintain m-σ Relation

While the suppression of BH seed formation is an attractive possibility that fits constraints on the $z = 6$ SMBH mass function, it is clearly not unique. One alternative solution to the over-production problem is to simply reduce the accretion rate of lower–mass BHs

at lower redshifts. This is again physically plausible: accretion could be choked off as a result of the baryonic gas being churned into stars, being heated and dispersed by reionization feedback, or through self-induced negative feedback where the BH's own accretion-powered radiation stops the gas supply. Rather than try to model such a time– (and probably mass–) dependent mass accretion scenario, we examined several model variants, in which BHs are allowed to accrete just enough mass to match the value inferred by the m-σ relation between BH mass and host halo velocity dispersion. That is, at each timestep $t \rightarrow t + \Delta t$, all BHs were assumed to grow in mass by $m \rightarrow m + \Delta m$ such that the $m - M$ relation is satisfied at the new host halo mass and redshift. However, whenever this requires super-Eddington growth, i.e. if $(m + \Delta m)/m > \exp(\Delta t/t_{\rm Edd})$, then Eddington growth is applied instead. The main additional assumption here is that the m-σ relation remains valid at all redshifts (which is at least consistent with a comparison between the evolution of quasars and early–type galaxies at $0 < z \lesssim 6$; Haiman et al. 2007; Shankar et al. 2009a. As above, we adopt Equation (2.20) as our extrapolated m/M relation.

The BHs in these models form as $100 M_{\odot}$ seeds, and, given their host halo mass and redshift, accrete to match this relation as closely as possible without exceeding the Eddington accretion rate. If the simulation completes with the mean BH accretion rate well below the Eddington rate at all redshifts, then it is consistent with satisfying the Eddington accretion rate and the $m - M$ relation inferred by Equation (2.20). As we shall find below, in our models the maintenance of the m-σ relation does not typically require that the Eddington accretion rate is saturated (see Figure 2.15 below) as long as the $f_{\rm seed} \gtrsim 10^{-2}$. If both the seed mass and the seeding fraction are low, it is increasingly difficult to satisfy

Equation (2.20) at higher redshifts while simultaneously satisfying the Eddington upper limit. We find that for $f_{\rm seed} \lesssim 10^{-2}$, accretion must saturate at the Eddington rate for much of $z \gtrsim 15$ until the extrapolated m-σ relation is satisfied, with the mass function falling below this relation at earlier stages of growth.

Recoil velocities are calculated with the spin magnitudes chosen uniformly between 0.0 and 0.9. As with our previous models, we run simulations where the spins are either randomly oriented or completely aligned with the angular momentum vector of the binary orbit. This class of models in effect represents the most optimistic *LISA* expectations, as it allows us to keep numerous seeds, while simply adjusting the accretion rate, as described above, to keep the mass function within bounds. Note that the recoil speeds are also minimized by our choice for the spin alignment.

We show the mass functions and occupation fractions for this model in Figure 2.13, for three different values of the cutoff redshift below which new seeds are not formed, $z_{\rm cut} = 0$, 12 and 18. Note that it is still possible to form the SDSS quasar-SMBHs via Eddington-limited accretion by $z \approx 6$ even if seeds form in just 0.1% of all 1200K halos and only before $z = 18$. We do not plot the m/M relation, as it is satisfied in the form of Equation (2.20) in all cases shown here, by construction. These models also satisfy, by construction, the upper limit for the SMBH mass density. In all of the models shown in Figure 2.13, $\rho_{\rm SMBH,5+}(z = 6) \gtrsim 1.3 \times 10^4 M_\odot$ Mpc^{-3} if $f_{\rm seed} \geq 10^{-2}$. In the $f_{\rm seed} = 10^{-3}$ models, we find $\rho_{\rm SMBH,5+}(z = 6) \sim 1.0 \times 10^4 M_\odot$ Mpc^{-3}. The difference in ρ_{SMBH} is due to varying occupation fractions at the low end of the halo mass function. The mass functions in Figure 2.13 have a much shallower slope overall than those in Figures 2.5 and 2.8. For

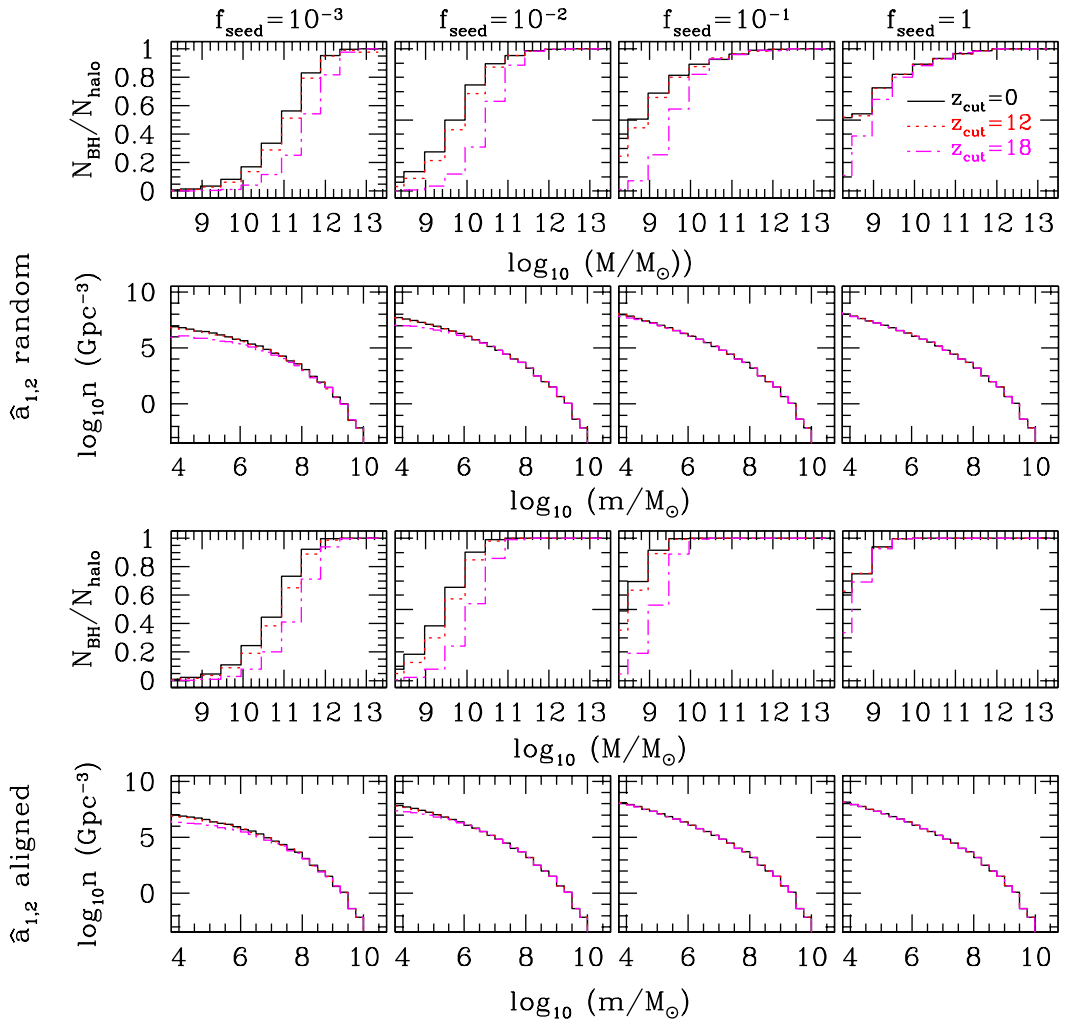

Figure 2.13 The SMBH mass functions and occupation fractions at $z = 6$ for various models satisfying the m-σ relation. These mass functions are much less steep in comparison to those shown in Figures 2.5 and 2.8. Occupation fractions are higher than those in Figures 2.6 and 2.9, as the BH binary mass ratios are generally lower compared to those models, resulting in less powerful recoil kicks.

the mass functions shown earlier, the steeper slopes were due to the ratio m/M increasing with decreasing host halo mass; the observed m/M relation has the opposite trend.

We show the *LISA* event rates in these alternative models in Figure 2.14. Most significantly, we note that in the $f_{\rm seed} = 1$ versions of these successful models, the *LISA* event rate can be as high as 30 $\rm yr^{-1}$. (Note that this number can be even higher if seeds can form in minihalos down to a virial temperature that is significantly lower than our assumed fiducial value of 1200K.) The rate is somewhat suppressed when compared to the earlier Pop III seed models (Figure 2.7) that exceeded realistic indications on the SMBH mass density, because the massive BHs in the *LISA* band are rarer due to the more modest growth rates. We draw the attention of the reader to the apparent independence of the detection rate on the seed fraction and seed formation cutoff in the cases where $f_{\rm seed} \gtrsim 10^{-1}$ and $z_{\rm cut} \lesssim 12$. Because BH ejections probabilities are lower in these models when compared to the constant-accretion scenarios of Figures 2.7 and 2.10, and because the m/M ratios are the same function of M in all models shown, the *LISA* rates saturate and converge once the occupation fraction in the most massive halos approach unity. Because the $T = 1200$K halos form in greatest abundance at $z \sim 20$, the $z_{\rm cut} = 12$ case hardly differs from the case with no seed cutoff.

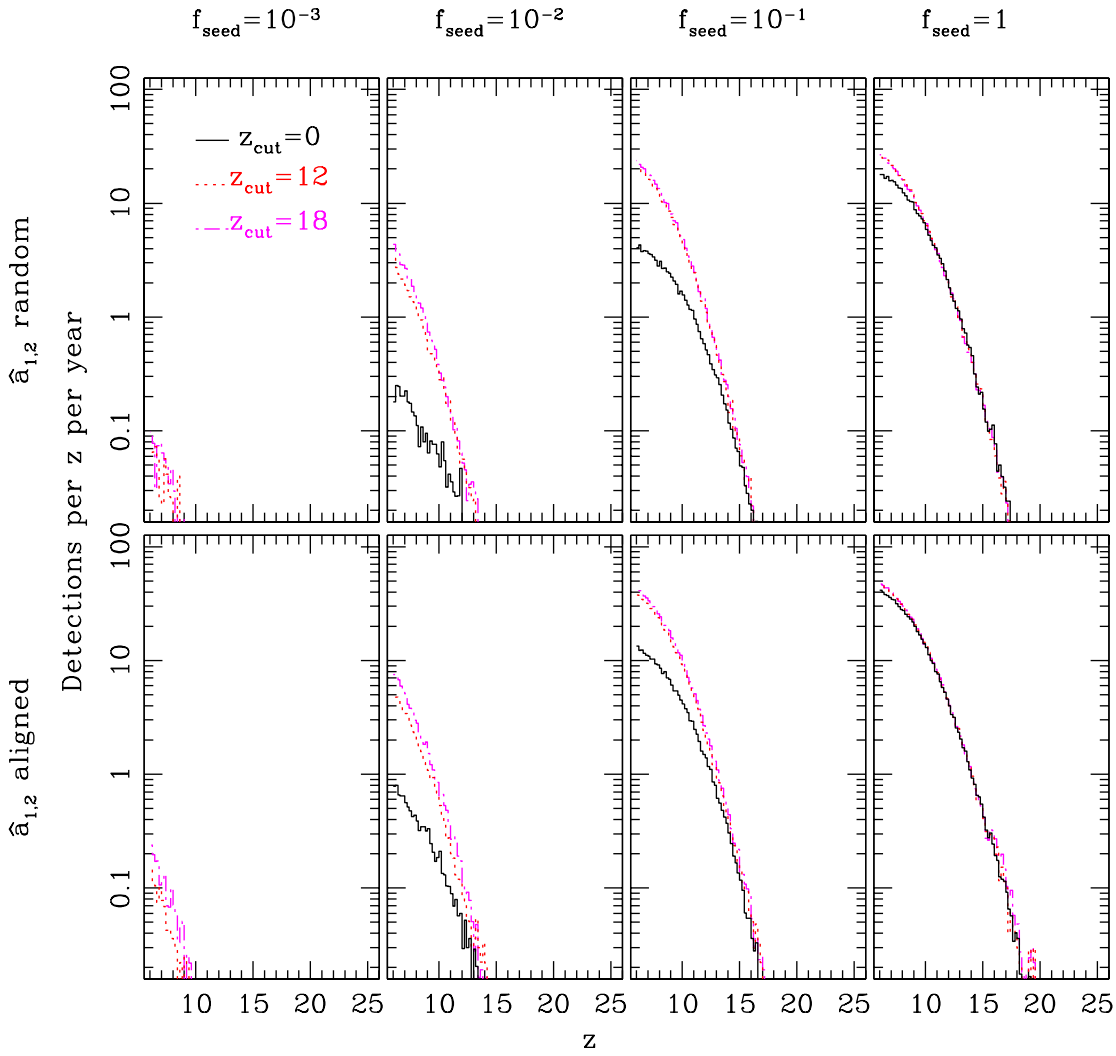

Figure 2.14 The *LISA* rates in models that satisfy the m-σ relation at all redshifts as closely as possible without exceeding the Eddington accretion rate. Note that the rate saturates at 30 yr^{-1} at $z \gtrsim 6$ for high seed fractions and low redshifts for seed cutoff.

A key characteristic of any SMBH assembly scenario is the balance between growth through BH mergers and growth through gas accretion. As discussed above, the two must strike a balance such that they are able to account for the most massive observed quasar-SMBHs at $z \sim 6$, while also not exceeding the total observed SMBH mass density. In Table 2.3, we illustrate the relative importance of mergers *vs.* growth in the models presented in this paper: the four successful constant-accretion models from Table 2.2; two of the unsuccessful constant-accretion models, also in Table 2.2, which overproduce the universal SMBH mass density; and four of the models that explicitly follow the extrapolated m-σ relation via Equation (2.20). The values shown in the table (in $\log_{10}$) are the sum of the initial masses of all the seed BHs that enter our merger trees; the total mass of galactic BHs at $z = 6$ [6]; and the total mass of BHs ejected before $z = 6$. We also calculate the ratio of the total (galactic and ejected) BH mass at $z = 6$ to the total initial seed BH mass, which gives a simple measure of the growth through gas accretion. We see immediately the contrast between the two types of successful models: the constant-accretion scenario relies on gas accretion for much of the growth, typically several orders of magnitude in the total BH mass; where the self-regulating models essentially describe the most heavily merger-driven scenarios possible, requiring accretion-driven growth of as little as a factor of a few.

These models also produce a significant population of ejected BHs. Even though ejection rates are lower on the whole than our constant-accretion models (compared to

[6]Note that the values in Table 2.3 include BHs of all masses equal to and above the seed mass, where we have considered only those with $m \geq 10^5 M_\odot$ in computing the universal "SMBH" mass density in previous sections. Also note that we do not generate trees for DM halos with $M(z = 6) < 10^8 M_\odot$, and throughout this paper we do not account for any BHs that may reside in such halos.

Table 2.3 The total seed, total ejected, and total retained BH masses

Model	$m_{\rm seed,\,tot}$	$m_{\rm GN}(z=6)$	$m_{\rm ejected,\,tot}$	$m_{\rm BH,\,tot}/m_{\rm seed,\,tot}$
A	9.3	16.0	6.2	6.7
B	10.1	16.1	7.7	6.0
C	12.7	16.2	6.4	3.4
D	13.2	16.3	6.2	3.1
X	16.0	20.0	19.1	4.0
Y	14.6	17.5	12.7	2.9
m-σ, $z_{\rm cut}=0$, $f_{\rm seed}=1$, aligned	15.7	15.7	16.0	0.48
m-σ, $z_{\rm cut}=0$, $f_{\rm seed}=1$, random	15.7	15.7	16.3	0.70
m-σ, $z_{\rm cut}=0$, $f_{\rm seed}=10^{-2}$, aligned	13.7	15.6	12.9	1.9
m-σ, $z_{\rm cut}=18$, $f_{\rm seed}=1$, aligned	14.4	15.7	14.7	1.34

The decomposition of the final SMBH mass into the contributions from the initial stellar seed BHs and subsequent gas accretion. Masses are in $\log_{10}$, and in units of $M_\odot$. The columns, starting from the second and from left to right, show: the total initial seed mass; the total SMBH mass retained in nuclei at $z=6$; the total ejected BH mass, and $\log_{10}$ of the ratio of the total (i.e. the sum of the ejected and nuclear) $z=6$ SMBH mass to the initial seed mass. The last ratio is a measure of the total growth due to gas accretion. The first four rows (A-D) show values for the four successful models. Also shown for comparison are values from two of the unsuccessful, constant-accretion models (X and Y) that overproduce the total BH mass function. The model parameters for Models A, B, C, D, X and Y can be found in Table 2.2. The models where the m-σ relation is enforced by hand can grow primarily through mergers, with gas accretion adding as little as a factor of a few to the total SMBH mass at $z=6$. If $f_{\rm seed}$ is sufficiently high, they also eject a total mass in low-mass BHs that is comparable to the retained nuclear population (most of the ejected holes have a mass near the seed mass).

the unsuccessful constant accretion Model X in Tables 2 and 3), two factors contribute to the ejected BH mass being comparable to the galactic BH mass at $z=6$. First, seed BHs are allowed to be very common, especially in contrast to the successful constant-accretion Models A through D; this results in a far greater number of total merger events, and a high total number of ejections despite the lower ejection probabilities. Second, the surviving BHs do not grow nearly as rapidly in these models as in the constant-accretion scenarios, so that the ratio between the total galactic BH mass and the total ejected mass can remain low (whereas in the constant accretion scenarios, retained holes can rapidly outgrow their

escaped counterparts.) In the $f_{\rm seed} = 1$ models, the ejected holes can outnumber and outweigh their retained galactic counterparts with mass densities of $\sim 3 \times 10^4 M_\odot$ Mpc^{-3} and number densities of ~ 100 Mpc^{-3}. The mass function of the ejected holes is peaked slightly above the seed mass (because holes are most likely to be ejected at the earliest stages of their evolution, when their host halos are the least massive). All of the ejected BHs are $\lesssim 10^4 M_\odot$ if spins are aligned, but in rare instances, SMBHs as massive as $\sim 10^8 M_\odot$ are ejected in our models with randomly oriented spins (the ejected SMBHs with masses above $m > 10^6 M_\odot$ have a very low number density, SMBHs of $\sim 4 \times 10^{-5}$ Mpc^{-3}, even in the model with the most ejections (random orientation, no cutoff redshift, $f_{\rm seed} = 1$).

These self-regulating accretion models work by adjusting the BH accretion rates according to the mass growth of their host halos. We plot the accretion rates in units of the Eddington rate for this new set of models in Figure 2.15. We do so for the $z_{\rm cut} = 12$ case with four combinations for $f_{\rm seed}$ and spin alignment, and for three different BH mass ranges: $10^3 M_\odot \leq m \leq 10^6 M_\odot$, $10^6 M_\odot \leq m \leq 10^8 M_\odot$, and $m \geq 10^8 M_\odot$. Note that the accretion rates must be slightly higher if BH binary spins are randomly oriented, in order to compensate for the higher ejection rates. Similarly, accretion rates are higher if seeds are less common, in order to compensate for the reduced merger-driven growth. For the models shown, the duty cycle (the mean accretion rate in Eddington units) for the most massive SMBHs converge to ~ 0.2 at $z \approx 6$, though it can be as high as $\gtrsim 0.5$ at $z \gtrsim 8$ if merger-driven growth is hindered by low occupation fraction or recoil-induced ejections.

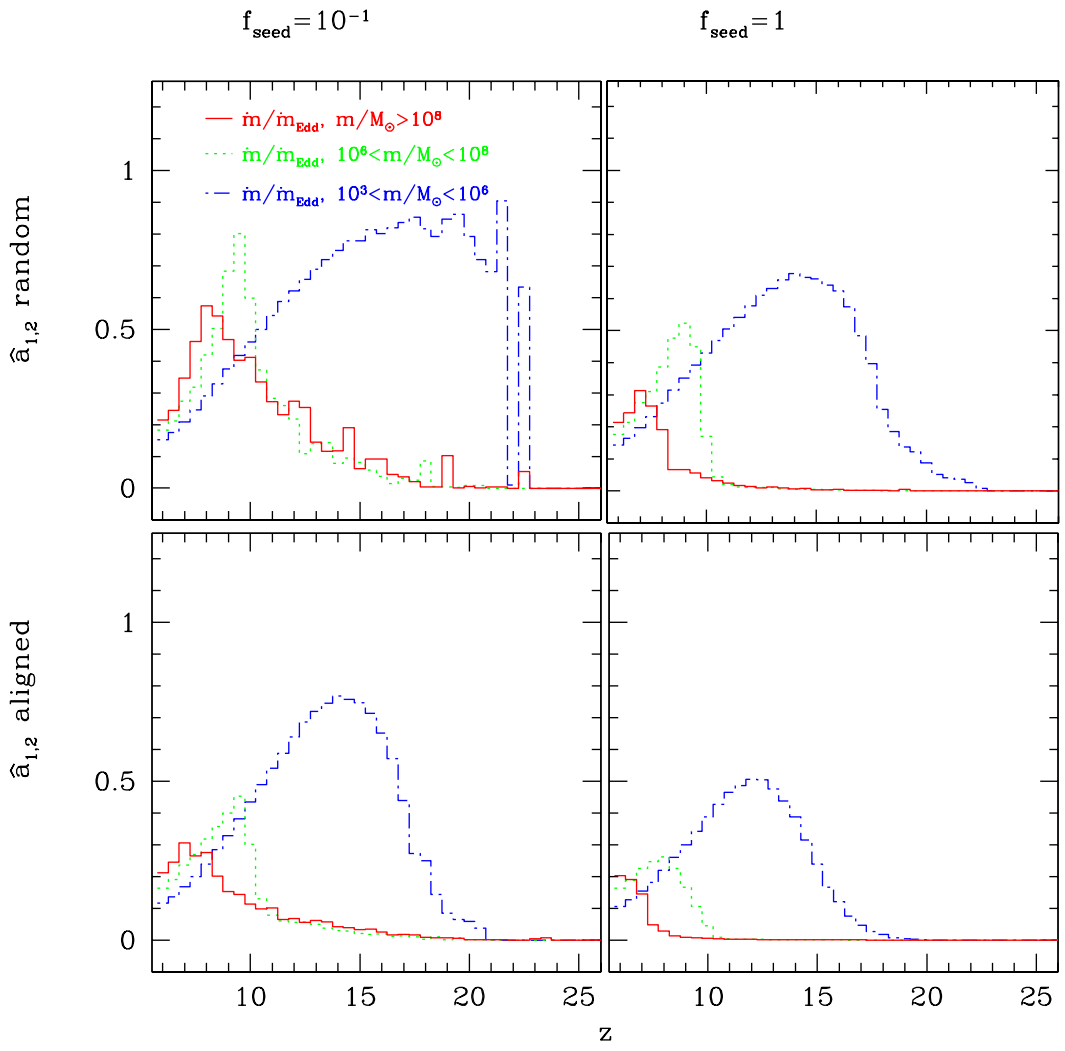

Figure 2.15 The accretion rate (thick lines) in Eddington limits in models that match the m-σ relation (Equation 2.20) at all redshifts, for different ranges of BH masses. We also plot the merger rates (thin lines) in units of mergers per halo for different halo mass bins corresponding to $M_1 + M_2 = 10^4 m$. Note that masses are defined instantaneously at each redshift interval, rather than tracking the histories of the $z = 6$ holes and halos.

We note that similar merger–tree models tracking SMBH growth have been published. For example, Koushiappas et al. (2004) presented a similar model where the SMBHs are assembled primarily through mergers of directly-collapsed halo cores. Bromley et al. (2004)also presented SMBH assembly model wherein gas accretion activity was triggered by major mergers of the BHs' host halos; in their model, a set fraction of the baryonic mass of the host was fed to the BH at each major merger. Their prescription (albeit without gravitational recoil) successfully produced the most massive SDSS SMBHs before redshift $z = 6$ without overproducing the mass density. In general, this type of assembly model is fairly easily tuned to broadly reproduce the $m - \sigma$ relation, as the parallel mass growth of BHs and their host halos is built in.

2.4 Conclusions

In this paper, we have attempted to map out plausible ways to assemble the $\gtrsim 10^9 M_\odot$ SMBHs that power the bright redshift $z \approx 6$ quasars observed in the SDSS, without overproducing the mass density in lower–mass ($\sim 10^{5-7} M_\odot$) BHs. We also computed the event rates expected for *LISA* in each of the successful models.

The physical effects governing SMBH assembly depend on the answers to four basic questions: (1) how common are the initial BH seeds; (2) how much mass in gas do they accrete, and therefore how much they contribute individually to the final SMBH's mass; (3) how often do they merge; and (4) what happens to SMBH binaries when they do merge? Currently, we do not have empirical constraints to offer definitive answers to any of these questions. However, we are capable of predicting the final outcome,

starting with a set of assumptions for the underlying physics. Our trees-plus-orbits algorithm simulates the formation history of SMBHs and the subsequent detection rate expectations for *LISA* by isolating and prescribing answers to the above four questions. It is a powerful simulation tool, as it can incorporate a detailed modeling of individual physical prescriptions without a significant increase in the computational load, as long as the prescriptions can be described by fitting formulae, tabulated in a lookup table of reasonable size or summarized in a statistical manner.

Using this tool, we have surveyed a wide range of candidate assembly models, and reported on common and distinguishing traits in the resulting SMBH mass functions and the corresponding *LISA* detection rates. In particular, we have shown that SMBHs can form in a manner consistent with other observational evidence either through the rapid growth of rare, massive seeds, or through ultra–early production of numerous Pop-III remnant seeds, provided these seeds stop forming below a redshift $z_{\rm cut} \sim 20 - 30$. We reach the pessimistic conclusion that these scenarios do not produce *any* detectable *LISA* events at $z > 6$ (at least not in a few year's operation). An alternative model, in which we assume that the extrapolation of the local $m - M$ relation holds at all redshifts (e.g. due to internal feedback), on the other hand, can produce up to ~ 30 *LISA* events per year, with a characteristic mass spectrum.

Our major findings can be summarized more specifically as follows:

- SMBHs must be continuously surrounded by dense gas that was able to cool at the centers of DM halos. Feeding holes with the low–density gas in DM halos whose gas was unable to cool does not allow for high enough accretion rates to explain the

SDSS quasar BHs.

- If embedded in dense gas nearly continuously, $\sim 100 M_{\odot}$ seed BHs can grow into the SDSS quasar BHs without super–Eddington accretion, but only if they form in minihalos at $z \gtrsim 30$ and subsequently accrete $\gtrsim 60\%$ of the time. However, these optimistic assumptions, required to explain the SDSS quasar BHs, overproduce the mass density in lower–mass (few$\times 10^5 M_{\odot} \lesssim M_{\rm bh} \lesssim$ few$\times 10^7 M_{\odot}$) BHs by a factor of $10^2 - 10^3$. We find that two conditions need to be satisfied to alleviate this overprediction: the initial occupation fraction of seed BHs has to be low ($f_{\rm occ} \lesssim 10^{-2}$), and new seeds must stop forming, or the seeds must accrete at severely diminished rates or duty cycles, at $z \lesssim 20 - 30$. We argued that models in which BH seeds stop forming at $z \sim 20$ are attractive because there are physical reasons for the seeds to stop forming below some redshift (such as metal pollution and/or radiative feedback that suppresses pop-III star formation), and because there is independent empirical evidence, from *WMAP* constraints on the reionization history, that star and/or BH formation in high–redshift minihalos was suppressed by a factor of $\gtrsim 10$ (Haiman & Bryan 2006).

- The simplest SMBH assembly scenarios, which have constant accretion rates, but in which BH seed formation stops abruptly at some redshift, and which meet constraints at both the high–mass and low–mass end of the $z = 6$ SMBH mass function, predict negligibly low *LISA* event rates. The reason for the low rates is as follows: in these models, the BHs that grow into the most massive, highest-redshift quasar-SMBHs accrete at the same (exponential) rate as all the other BHs, typically resulting

in a vast overproduction of massive ($m \sim 10^6 M_\odot$) holes. In order to offset this overproduction, seeds must be made very rare, and this diminishes the *LISA* rates. It is difficult to envision a scenario for high ($\gtrsim 10$ per year per unit redshift) detection rates unless a vast number of SMBHs in the $10^{5-7} M_\odot$ range lurk in the universe at all redshifts, which the current electromagnetic surveys have missed.

- A different class of successful models, in which the SMBH masses are self–regulated by internal feedback, can evade this constraint, and produce *LISA* rates as high as 30 yr^{-1}. The key difference in these models that predict higher *LISA* rates is that the SMBH growth is driven by a large number of seed BHs and far lower gas accretion rates than those required in the constant-accretion models. The majority of these events occur at $z \approx 6$ and in the low end ($10^3 - 10^4$ $\mathrm{M}_\odot$) of *LISA*'s mass range for detection. Also, for these models we find the ejected BH mass density can exceed that of the galactic BH population at $z = 6$. Most ejected holes are expected to have masses similar to the seed mass, but an ejected BH can be as massive as $\sim 10^8 M_\odot$ if large recoil velocities are allowed (e.g. if spins are not always aligned with the orbital angular momentum of the binary).

In addition to the above, our modeling reveals a number of interesting aspects of SMBH assembly. We find that in the successful models the initial seeds are rare, and the most massive SMBHs grow primarily from the few 'lucky' early seeds that avoided ejection due to kicks. The precise assumptions regarding the kick velocity distribution (such as the assumed spin orientations or the resulting oscillation of the BH) tend to have only a modest effect on the final results in these models. This is because, at least in our

simple prescription, BHs either return quickly to the gas-rich nucleus or are left wandering in the outer regions.

Our results suggest that *LISA* will be capable of narrowing the field of plausible SMBH assembly models from the raw event rate, even without detailed measurements of the binary spins or mass ratios. The spin and mass ratio measurements will further constrain the evolution of SMBH properties. While the component prescriptions explored in this paper are admittedly crude, exercises similar to the one performed in our study will be crucial in understanding the limits and possibilities offered by *LISA*, and ultimately to interpret the detected *LISA* events. The scarcity of empirical constraints on the various pieces of physics that determines the SMBH growth leaves us with a large range of "plausible" scenarios and free parameters for SMBH assembly.

Chapter 3

Time-Dependent Models for the Afterglows of Supermassive Black Hole Mergers[1]

3.1 Introduction

Supermassive black holes (SMBHs) are abundant at all observable redshifts, manifesting themselves most often as active galactic nuclei (AGN) and quasars (Kormendy & Richstone 1995; Richstone et al. 1998). A merger of a pair of galaxies each containing a SMBH will result in the formation of a SMBH binary (Begelman et al. 1980). Given the hierarchical structure formation history of the universe, and that most or all formed galaxies appear to

[1]This chapter is a reformatted version of an article titled "Time-Dependent Models for the Afterglows of Massive Black Hole Mergers" by T. Tanaka and K. Menou that can be found in The Astrophysical Journal, Volume 714, Issue 1, pp. 404-422. The abstract for this paper is reproduced in Section 1.6.2.

harbor a SMBH in their nuclei (e.g., Magorrian et al. 1998), there should be many SMBH binaries formed throughout cosmic time (Haehnelt 1994; Menou et al. 2001; Volonteri et al. 2003a; Wyithe & Loeb 2003a; Sesana et al. 2007b; Tanaka & Haiman 2009). Once compact, these binaries rapidly lose orbital energy via gravitational radiation (GW) and coalesce. The *Laser Interferometer Space Antenna* (*LISA*) will detect gravitational waves resulting from SMBH mergers of binaries with total mass $M = (10^5 - 10^7)/(1+z)M_\odot$ out to redshift $z \sim 20$ (e.g., Schutz 2009). *LISA* is expected to be able to constrain the individual redshifted mass $(1+z)M$, the mass ratio $q \leq 1$, spins of the SMBHs, and the luminosity distance of the source with precision, thereby providing an unprecedented test of general relativity and probing the assembly history and demography of SMBHs (e.g., Hughes 2002; Vecchio 2004; Lang & Hughes 2006). However, the space-based detector will be unable to measure the redshift (due to the fundamental degeneracy between source mass and redshift) or the precise angular location (due to the fact that sources are located using the modulation of the signal due to orbital motion and separation of the detector elements) of the source (Cutler 1998; Holz & Hughes 2005; Kocsis et al. 2006, 2007; Lang & Hughes 2008).

The coalescence of a SMBH binary is not by itself expected to emit an observable electromagnetic (EM) signal. If, however, SMBH mergers have *associated* and readily identifiable EM signatures, their detection would complement *LISA* by helping determine source redshifts and locations on the sky (Holz & Hughes 2005; Kocsis et al. 2006; Bloom et al. 2009; Phinney 2009). An obvious candidate to effect such emission is the abundant gas linked to galactic SMBHs. Rapid, sustained gas accretion onto SMBHs is required to explain the luminosities and number of observed AGN and quasars. Numerous studies

suggest that a gas-rich environment aids the formation of close SMBH binaries in recently merged galaxies (Escala et al. 2005; Dotti et al. 2007; Callegari et al. 2009; Colpi et al. 2009), and may facilitate the further shrinking of the orbit to where GW emission can enact the merger (Ivanov et al. 1999; Armitage & Natarajan 2002; MacFadyen & Milosavljević 2008; Cuadra et al. 2009; Lodato et al. 2009). Various types of interactions between SMBH binaries and their gas environments have been proposed as viable mechanisms for observable EM emission. Numerous studies have calculated the EM emission from the response of the gas disk to the mass loss and gravitational recoil (Peres 1962; Bekenstein 1973) effects which accompany the merger process (e.g., Bode & Phinney 2007; Schnittman & Krolik 2008; Lippai et al. 2008; Shields & Bonning 2008; O'Neill et al. 2009; Megevand et al. 2009; Corrales et al. 2010; Rossi et al. 2010; Anderson et al. 2010). Chang et al. (2010) have suggested a luminous EM signal from tidal and viscous excitation of fossil gas trapped inside the binary's orbit. Krolik (2010) pointed out that even small amounts of gas that is present in the immediate vicinity of a merging binary can power short afterglows on Eddington luminosity scales. Kocsis & Loeb (2008) have proposed an infrared afterglow from the dissipation of GW through a surrounding gas disk (see, however, Krolik 2010). Palenzuela et al. (2009, 2010b); Mösta et al. (2010) have emphasized the possibility of variable emission due to the perturbation and enhancement of the local EM fields. Haiman et al. (2009a) have raised the possibility of detecting the binary as a periodic variable source prior to the merger.

In this paper we revisit the accretion afterglow mechanism proposed by Milosavljević & Phinney (2005; hereafter MP05; see also Liu et al. 2003b; Liu 2004) for a geometrically

thin circumbinary disk. We summarize their model as follows:

1. The tidal torques from the binary open a gap in the circumbinary gas. The gas inside the gap accretes, while the gas outside is held at bay by tidal torques. What results is a circumbinary disk with the binary inside the central cavity. Because it is largely "missing" the inner, highest-temperature region, the disk cannot easily produce thermal X-rays via viscous dissipation (see, however, Chang et al. 2010). The system can remain in this configuration for an extended period of time (Ivanov et al. 1999; MacFadyen & Milosavljević 2008) as the timescale for the extraction of angular momentum from the binary's orbit is long.

2. Once the binary reaches a semimajor axis $a \sim 100GM/c^2$, it rapidly loses orbital angular momentum and energy via GW emission and the orbit closes faster than the surrounding gas can viscously follow (see also Armitage & Natarajan 2005; Haiman et al. 2009a; Chang et al. 2010). The binary coalesces, producing the GW signature detectable by *LISA*.

3. The gas (no longer held back by binary torques) reaches the center of the cavity by viscously spreading, and the corresponding accretion flow, deep in the potential well of the SMBH remnant, emits an X-ray afterglow.

By taking the difference between the time it takes the binary to merge and the time it takes for the *bulk* of the gas at the cavity edge to reach the central remnant, MP05 estimated that the X-ray afterglow would occur $t_{\rm EM} \sim 7(M/10^6 M_\odot)^{1.3}(1+z)$ yr after the GW signal (in the observer's rest frame).

In this study, we use the term "accretion afterglow" to denote the emission described above. We stress that this is a distinct mechanism from the various afterglow mechanisms that are powered by the response of the accretion disk to the mass loss and/or recoil of the central SMBH remnant (references above). In the latter scenarios, mass loss and/or recoil introduce additional eccentricities in the gas orbits, and the emission is powered by the circularization of the orbits and shock/wave dissipation. In the accretion afterglow scenario considered by MP05 and in this paper, the emission results from the deepest parts of the SMBH potential well becoming accessible to the circumbinary disk as the binary shrinks and merges. Before the merger the disk is deficient in the hard UV and X-ray frequency range because of the central cavity, while after the merger this high-frequency emission emerges as the cavity fills with gas. Although distinct, the various classes of emission mechanisms considered are not entirely unrelated. For example, mass loss introduces a roughly constant eccentricity everywhere in the disk, and the resulting luminosity of a circularization-powered afterglow (references above) would depend delicately on the innermost density profile of the circumbinary disk at the time of merger (see Corrales et al. 2010 for a comparison of afterglows for different density profiles). A high surface density for the innermost disk (where the available specific orbital energy is highest) at merger would enhance the mechanism emphasized by O'Neill et al. (2009) while it would suppress the pre- and post-merger emission contrast considered in this paper; a low surface density would have the reverse consequences for the two mechanisms.

The purpose of this work is to investigate the onset and time dependence of the

afterglow in greater detail. We use an idealized semi-analytic framework to model the time dependence of the electromagnetic signal. We present explicit, integral solutions for the time evolution of one-dimensional (viz. geometrically thin and azimuthally symmetric), viscous Keplerian disks. (See similar work by Shapiro 2010.) Our calculation method is valid insofar as the disk kinematic viscosity can be described as a simple function of radius, and as long as after decoupling the inner gas is minimally affected by the gravitational torques from the binary. This last point is worth emphasizing, as in general the decoupling of the binary from the gaseous influence does not guarantee that the gas is entirely free from the binary influence. Even after GW emission has become the dominant mechanism driving the orbital evolution of the binary, the binary's tidal torques can still influence any gas that is able to remain in the vicinity of the binary's orbit.

This paper is organized as follows. In §3.2 we review the disk properties as the GW-driven closing of the binary decouples from the viscosity-driven spreading of the circumbinary disk, and introduce a semi-analytic model for the subsequent evolution of the disk's surface density. Various derivations and intermediate results used in this section are provided in the Appendices 3.A & 3.B. We present in §3.3 the predictions of the model for the light curve and spectral evolution of the resulting electromagnetic emission. We also discuss in that section the possibility that the X-ray afterglow could be reprocessed by the outer disk, as well as possible effects of advection as the inner disk becomes geometrically thick. We conclude in §3.4.

3.2 Modeling the Binary-Disk System

Below, we first discuss the properties introduced by MP05 for the evacuated circumbinary gas disk at the time when the evolution of the binary orbit becomes dominated by GW emission and thus decoupled from the evolution of the surrounding disk, which evolves viscously (see also Haiman et al. 2009b). We then describe the semianalytic integral formalism for the viscous evolution of an accretion disk with an arbitrary initial surface density profile. Additional derivations and intermediate results for both the disk properties and the semianalytic model are detailed in Appendices A and B, respectively. Throughout this paper, R is the radial distance from the center of mass of the binary in the plane of its orbit; a is the binary's semimajor axis; M is the total mass of the binary; and $q \leq 1$ is the binary mass ratio. The quantities c, G, m_p, h, k, σ and $\sigma_{\rm T}$ are the speed of light, the gravitational constant, the mass of the proton, Planck's constant, the Boltzmann constant, the Stefan-Boltzmann constant and the Thomson cross section, respectively. Times, frequencies and rates are in the rest frame of the binary, unless noted otherwise.

3.2.1 The Circumbinary Disk at Decoupling

The "cold" gas in the nucleus is assumed to settle into a geometrically thin, rotationally supported disk. In this paper, we are concerned with the properties of the innermost disk and the binary in the very last stages of its evolution. We assume that the angular momentum of the disk is aligned with the orbital angular momentum of the binary (i.e., no warps; Ivanov et al. 1999; however we shall revisit the possibility of warping later), and that the binary has been circularized via gravitational wave emission (Peters 1964;

see Armitage & Natarajan 2005, MacFadyen & Milosavljević 2008 and Cuadra et al. 2009 for possible complications).

The disk is able to evolve locally on length scales $\sim R$ on the viscous timescale

$$t_{\rm visc}(R) = \frac{2}{3}\frac{R^2}{\nu}, \tag{3.1}$$

where $\nu(R) = \frac{2}{3}\alpha P_{\rm gas}/(\rho\Omega) = \frac{2}{3}\alpha kT/(\mu m_{\rm p}\Omega)$ and $P_{\rm gas}, \rho$, T and $\mu = 0.6$ are the pressure, density, temperature and mean molecular weight of the gas, respectively. The quantity Ω is the angular velocity, assumed to be approximately Keplerian: $\Omega \approx \sqrt{GM/R^3}$. The parameter α is the viscosity parameter in the disk model of (Shakura & Sunyaev 1973). We follow MP05 in adopting a viscosity prescription where ν scales only with the gas pressure, as opposed to the total (gas, radiation plus magnetic) pressure. This choice is motivated by calculations that suggest that disks where viscosity scales with the total pressure may be thermally unstable (Shakura & Sunyaev 1976; Pringle 1976; see, however, Hirose et al. 2009b). [2]

At a distance of $R \sim 2a$, a gap is opened in the circumbinary disk through tidal interactions (see, e.g., Artymowicz et al. 1991, MacFadyen & Milosavljević 2008 and Haiman et al. 2009b). In this region the input of angular momentum from the tidal torques exerted by the binary counteracts the viscous spreading of the disk. For simplicity, we neglect the small amount of fossil gas that may be present inside the binary's orbit (Chang et al. 2010). The gap, at first annular, becomes a roughly circular central cavity after the

[2]However, radiation pressure dominated disks may still be viscously unstable (Lightman & Eardley 1974; Piran 1978; Hirose et al. 2009a). Because the afterglow mechanism discussed in this paper takes place in the radiation-dominated part of the disk, the findings reported here are contingent on the answer to this open theoretical question.

inner gas is consumed. The boundary between the cavity and the disk is characterized by a steep surface density gradient, a pile-up of gas caused by tidal *decretion* in the vicinity of the binary's orbit and viscous accretion outside the decreting region. Whereas a steady-state accretion disk has a mass accretion rate $\dot{M}(R) = 3\pi\nu\Sigma$ that is constant with radius, a decretion disk with a steady inner torque instead satisfies constant viscous torque, i.e. $3\pi R^2\nu\Sigma\Omega$ = constant (e.g., Rudak & Paczynski 1981; Pringle 1991). It is a priori unclear how close to accretion or decretion a certain system is at decoupling, as this will depend on the system parameters and its past accretion history (e.g., Ivanov et al. 1999; Chang et al. 2010). We thus consider various possibilities for the circumbinary disk density profile at decoupling. Some of the innermost gas may be still able to accrete across the binary's orbit in narrow streams, but the accretion rate is expected to be only a small fraction of that of a comparable accretion disk around a single black hole (e.g., Lubow et al. 1999; Günther et al. 2004; Hayasaki et al. 2007, 2008; MacFadyen & Milosavljević 2008).

The timescale for the binary separation to shrink via GW emission is

$$t_{\rm GW}(a) \equiv \frac{a}{da/dt} = \frac{5}{16}\frac{c^5}{G^3M^3}\frac{a^4}{\zeta} \approx 4.90\left(\frac{M}{10^6 M_\odot}\right)\zeta^{-1}\left(\frac{a}{100GM/c^2}\right)^4 \ {\rm yr}, \tag{3.2}$$

where $\zeta \equiv 4q/(1+q)^2$ is the symmetric mass ratio scaled to unity for equal-mass binaries. Equation 3.2 is valid for binaries with non-extreme mass ratios and circular orbits. Because $t_{\rm GW} \gg t_{\rm visc}$ initially in the circumbinary disk, the gas is able to respond promptly to the gradual and relatively slow shrinking of the binary, probably maintaining a roughly constant geometric ratio between the disk inner edge and the orbital semimajor axis, i.e. $\lambda \equiv R/(2a) \sim 1$.

Because the disk viscosity ν is a weak function of radius outside the inner edge (see Appendix 3.A), roughly speaking the viscous time there scales as $t_{\rm visc}(R) \propto R^2$. As the binary closes, $t_{\rm GW}(a) \propto a^4$ will inevitably become shorter than the disk viscous time at the inner edge, and the binary will begin to close faster than the bulk of the gas at the inner edge can follow. The subsequent evolution of the binary is driven by GW emission and thus causally decoupled from that of the disk. In the absence of binary torques, the gas at the cavity edge would fill the cavity in a time $\sim \beta t_{\rm visc}$, where the approximate boost factor $\beta \sim 0.1$ was introduced by MP05 to account for the limit of a very steep surface density gradient at the edge (Lynden-Bell & Pringle 1974). Decoupling occurs, then, when $t_{\rm GW}(a) \sim \beta t_{\rm visc}(2\lambda a)$. This condition, along with the assumption that the viscously dissipated energy is locally emitted as thermal radiation, specifies the various properties of the inner edge of the disk described by MP05. We closely reproduce their results in Table 3.1, and detail the various intermediate calculations in Appendix 3.A. Note that while we recover the parameter dependencies exactly, the magnitudes of our disk properties differ somewhat from those of MP05. This is due to differences in how we have calculated the thermal structure of the disk, the most significant being that the flux expression in MP05 is a factor of two lower than ours (compare their equation [2] and preceding formula for F_ν with our equation [3.13] below). Throughout the rest of this paper, we use the subscript "0" to denote the value of a quantity at decoupling. When applied to a disk quantity with radial dependence, the subscript shall also denote the value at the inner edge of the disk. For example, a_0 is the semimajor axis of the binary orbit at decoupling and Ω_0 is the angular velocity at the inner edge of the disk at decoupling.

Table 3.1. Properties of the circumbinary disk at decoupling

Variable	Factor	α_{-1}	S	λ	M_6	β_{-1}, ζ	$\theta_{0.2}$
$a_0/(GM/c^2)$	126	-0.34	-0.24	0.70	0.08	0.42	-0.08
$t_{\rm EM}$ (yr)	9.2	-1.36	-0.98	2.80	1.32	1.7, 0.7	-0.34
Σ_0(g cm^{-2})	6.2×10^5	-0.68	0.51	-0.60	0.16	-0.15	-0.17
T_0 (10^6 K)	1.3	0.19	0.86	-1.95	-0.28	-0.49	0.30
H_0/R_0	0.17	0.76	2.43	-3.80	-0.12	-0.95	1.19
$P_{\rm rad,0}/P_{\rm gas,0}$	430	1.67	4.25	-7.35	-0.04	-1.84	2.17
$Q_{\rm adv,0}/Q_{\rm rad,0}$	2.0×10^{-2}	1.52	4.86	-7.60	-0.24	-1.90	2.38

At decoupling, the variable in column 1 equals the factor in column 2 multiplied by the column head parameters raised to the powers indicated in columns 3–8. All quantities except for a_0 are evaluated at the inner edge of the circumbinary disk. We derive power-law dependencies identical to those found by MP05. Note, however, that we derive a somewhat lower temperature than MP05, and estimate that the disk is significantly less geometrically thick than they did. See Appendix 3.A for detailed calculations.

The surface density at the inner edge of the circumbinary disk at decoupling cannot be determined through scaling arguments alone. However, we expect it to be greater than the value expected for a steady thin accretion disk. Just outside the edge, the gas may resemble a decretion disk, with $3\pi\nu\Sigma \propto R^{-1/2}$ approximately, because of mass accumulation. Sufficiently far from the binary and its tidal torques, we expect the gas to behave more like a steady thin accretion disk, with a local mass supply rate $\dot{M} = 3\pi\nu\Sigma$ roughly constant and determined by external conditions (e.g., Ivanov et al. 1999; Chang et al. 2010). It follows that if $3\pi\nu\Sigma$ is monotonic outside the edge then it must be greater than the outer mass supply rate of the disk. Therefore, we parametrize the surface density at the inner edge via the arbitrary relation $\Sigma_0 = S\dot{M}_{\rm Edd}/(3\pi\nu_0)$, where S is a dimensionless

parameter and $\dot{M}_{\rm Edd} = 4\pi GMm_{\rm p}/(\eta c\sigma_{\rm T})$ is the Eddington accretion rate onto a SMBH of mass M for a radiative efficiency $\eta = 0.1$. Effectively, S is the product of the gas supply rate onto the disk in Eddington units and the enhancement of the local surface density at the disk edge due to the binary torques and mass accumulation.[3] In general, S can exceed unity even if the disk mass supply rate and the local accretion rate are both sub-Eddington.

3.2.2 A Simple Model for the Viscous Evolution

Following decoupling, the binary outruns the surrounding gas to the center and merges in a time $t_{\rm GW,0}/4$. MP05 estimated that an accretion-powered X-ray afterglow would take place when the central cavity is filled, viz. after a time $t_{\rm EM} = \beta t_{\rm visc,0} - t_{\rm GW,0}/4 \sim (3/4)t_{\rm GW,0}$ after the GW event. Below, we revisit this estimate by calculating the time dependence of the signal with a simple model for the disk evolution. We proceed by assuming that following decoupling the binary's orbit closes faster than the gas can follow, viz. viscosity is the only source of torque in the disk and the gas orbits remain circular and Keplerian. Numerical simulations (e.g., Armitage & Natarajan 2005; Hayasaki et al. 2007; MacFadyen & Milosavljević 2008; Cuadra et al. 2009) of SMBH circumbinary disks indicate that both the eccentricity and the deviation of the angular velocity from the Keplerian value are small (of order a few percent) at the radius where the binary tides truncate the disk. Likewise, the mass reduction of the binary at merger due to GW emission will only introduce orbital eccentricities (roughly equal to the fractional mass loss) of $\gtrsim 10^{-2}$ (Tichy & Marronetti

[3]For practical purposes S is equivalent to the parameter "$\dot{m}$" used in MP05. We avoid using the latter notation to prevent confusion with the actual local accretion rate, which is discussed below in §3.2.2. Also note that Σ_0 and S do not relate linearly (see Table 3.1). The relationship is more complex because the disk gas viscosity ν_0 and the decoupling orbital radius a_0 also depend on S through the midplane temperature.

2008; O'Neill et al. 2009; Reisswig et al. 2009). At radii and binary masses of interest here, orbital speeds are too great for the gas to be significantly affected by gravitational recoil of the SMBH remnant: $\Omega \sim \sqrt{Rc^2/GM}c \gg v_{\rm recoil} \sim 300$ km s^{-1}. After the binary has merged, the central potential is that of a rotating black hole and circular orbits near the center will have somewhat super-Keplerian angular velocities. Approximate "pseudo-Newtonian" expressions of the Kerr potential (e.g., Artemova et al. 1996) suggest that for a merger remnant with less than maximal spin, the deviation from the Keplerian value is of order ten percent at the radius of innermost stable circular orbit and falls off as roughly $\propto R^{-1}$. All of these modest deviations and perturbations are neglected in our models.

The viscous evolution of a geometrically thin, cylindrically symmetric Keplerian disk can be described by the standard equation (e.g. see Pringle 1981 and Frank et al. 2002)

$$2\pi R \frac{\partial}{\partial t}\Sigma(R,t) = \frac{\partial}{\partial R}\left[2R^{1/2}\frac{\partial}{\partial R}\left(3\pi\nu\Sigma R^{1/2}\right)\right], \tag{3.3}$$

where Σ is the surface density of the disk and the innermost derivative on the right-hand side describes the angular momentum gradient. The left-hand side of 3.3 describes the rate of change of the radial mass distribution, while the right-hand side gives the radial gradient of the local mass flow

$$\dot{M} = 3\pi\nu\Sigma\,(1+2m)\,, \tag{3.4}$$

where $m(r) \equiv \partial \ln(\nu\Sigma)/\partial \ln R$. In the above equation, the dimensionless factor in parentheses equals unity for steady-state accretion solutions ($m \approx 0$) while it vanishes in steadily

decreting solutions ($m \approx -1/2$).

If the kinematic viscosity is a function of radius only, equation 3.3 is a linear diffusion equation. In the special case $\nu \propto R^n$, the solution for subsequent viscous evolution (see, e.g. Lynden-Bell & Pringle 1974; a derivation is provided in our Appendix 3.B) is given by

$$\Sigma(r,t) = \int_0^\infty G(r,r',t)\Sigma(r',t=0)dr', \tag{3.5}$$

where

$$\begin{aligned} G(r,r',t) &= \frac{2-n}{\tau} r^{-1/4-n} r'^{5/4} I_{1/(4-2n)}\left(\frac{2r'^{1-n/2}r^{1-n/2}}{\tau}\right) \\ &\quad \times \exp\left(-\frac{r'^{2-n}+r^{2-n}}{\tau}\right)\left\{1 - \sqrt{\frac{r_*}{r}}\exp\left[-\frac{(r-r_*)^{2-n}}{\mathcal{R}(\tau-\tau_{\rm merge})}\right]\right\} \end{aligned} \tag{3.6}$$

is the Green function, $\Sigma(r, t = 0)$ is the initial surface density profile (at decoupling), I_m is the modified Bessel function of the first kind, $\mathcal{R}$ is the ramp function, and $r \equiv R/R_0$, $r' \equiv R'/R_0$ and $\tau \equiv 8(1-n/2)^2 t/t_{\rm visc,0}$ are dimensionless variables. At decoupling, $\tau = 0$; the binary merges when $\tau = \tau_{\rm merge} = 2\beta(1-n/2)^2$; roughly speaking, gas concentrated at $r = 1$ diffuses to the center in a time $\tau \sim 1$ (hence the value $\beta \sim 0.1$ adopted by MP05).

The second term in the curled brackets of equation 3.6 accounts for the fact that, after merger, the SMBH remnant imposes a zero-torque boundary condition at some finite radius $R_* = R_0 r_*$, which we associate with the remnant's innermost radius for marginally stable circular orbits [4] ($R_{\rm ISCO}$). This radius depends on the SMBH spin as $a_{\rm spin} = (1/3)\sqrt{R_{\rm ISCO}c^2/GM}(4 - \sqrt{3R_{\rm ISCO}c^2/GM - 2})$ (Bardeen 1970), where $-1 \le a_{\rm spin} \le 1$

[4]See, however, Krolik & Hawley (2002), who showed that there are several ways to define an effective inner radius, some of which are quite different from $R_{\rm ISCO}$.

is the standard dimensionless spin parameter of the SMBH. For test particle orbits, one finds $R_* = 6GM/c^2$ if $a_{\rm spin} = 0$, and $R_* = GM/c^2$ if $a_{\rm spin} = 1$ and the spin of the hole is perfectly aligned with the test particle's orbital angular momentum. Immediately after the merger, the SMBH remnant is expected to have moderately high but non-maximal spin for a wide range of physically plausible scenarios (e.g. Hughes & Blandford 2003; Baker et al. 2004; Berti & Volonteri 2008; see however the "chaotic" accretion scenario of King & Pringle 2006, which predicts lower SMBH spins). We thus adopt the intermediate value $R_* = 3GM/c^2$, which is accurate to within $\sim 20\%$ in the range of spin $0.65 \lesssim a_{\rm spin} \lesssim 0.9$. Our solutions given by equation 3.6 depend only on the initial surface density profile (which may be arbitrarily complex and need not be differentiable), as well as on the values of $n \leq 2$ and r_*.

Figure 3.1 shows time-dependent solutions calculated following equation 3.5 for two different initial surface density profiles at decoupling (upper left panel). In solid lines, we evolve for demonstrative purposes the plainest possible model: a surface density profile obeying the decretion power-law $\Sigma \propto r^{-1/2-n}$, truncated by a step function at the cavity edge. We also evolve a second, more plausible profile (dotted lines):

$$\frac{\Sigma(r,0)}{\Sigma_0} = \frac{1}{r^{n+1/2} + \exp[-11.9\lambda(r-1)]}. \tag{3.7}$$

This analytic profile is inspired by the hydrodynamic simulation results of MacFadyen & Milosavljević (2008), who found an azimuthally averaged inner surface density profile with $\Sigma(R \lesssim 2a) \propto \exp(-5.95R/a)$ for a thin disk around an equal-mass binary. Equation 3.7 reproduces this exponential behavior for $r \lesssim 1$ and the decretion power-law behavior

$\Sigma \approx \Sigma_0 r^{-1/2-n}$ for $r \gtrsim 1$. Note that this profile does not peak at the inner edge $r = 1$, nor is its peak value equal to the parametrized quantity Σ_0. In the context of equation 3.7, the value of Σ_0 sets the scale for the power-law region of the surface density profile just outside the inner edge.

In Figure 3.1, the physical length scale and the time with respect to the binary merger is calculated using the relations in Table 3.1 and the fiducial parameter values $M_6 = q = \alpha_{-1} = \beta_{-1}\lambda = s = \theta_{0.2} = 1$. We take as $n = 0.4$ as our fiducial viscosity power-law. For our disk solutions, this value leads to agreement between the late-time disk profile $\nu\Sigma \propto 1 - \sqrt{R_*/R}$ and the physical viscosity prescription $\nu \propto P_{\rm gas}/(\rho\Omega)$ of the $\alpha_{\rm gas}$ disk. (We refer the reader to Appendix 3.A for a brief discussion on the parameter n.) In Figure 3.2, we also evolve a second case with $n = 11/170 \approx 0.065$, which is consistent with $\nu \propto P_{\rm gas}/(\rho\Omega)$ at early times just outside the inner edge for a disk that has the decretion profile $m \approx -1/2$ there. In both figures, we graph the profiles at decoupling (3.1 years before the merger), 1 year before merger, at merger, and 9.2 years after merger. The last snapshot corresponds to a time $\beta t_{\rm visc,0}$ after decoupling, our re-evaluation of MP05's nominal estimate for the onset of the X-ray afterglow.

These two figures suggest that the qualitative evolution of the disk does not depend very sensitively on n. The viscous evolution timescale governing the Green's function scales as $(1-n/2)^2$, and thus the temporal dependence on n is small as long as the quantity $1-n/2$ is of order unity. In both figures, the qualitative evolution is such that at early times and large radii ($t < t_{\rm visc}(R)$), the disk maintains the initial mass distribution. At late times and small radii ($t \gtrsim t_{\rm visc}(R)$), the disk approaches an analytic quasi-steady density profile.

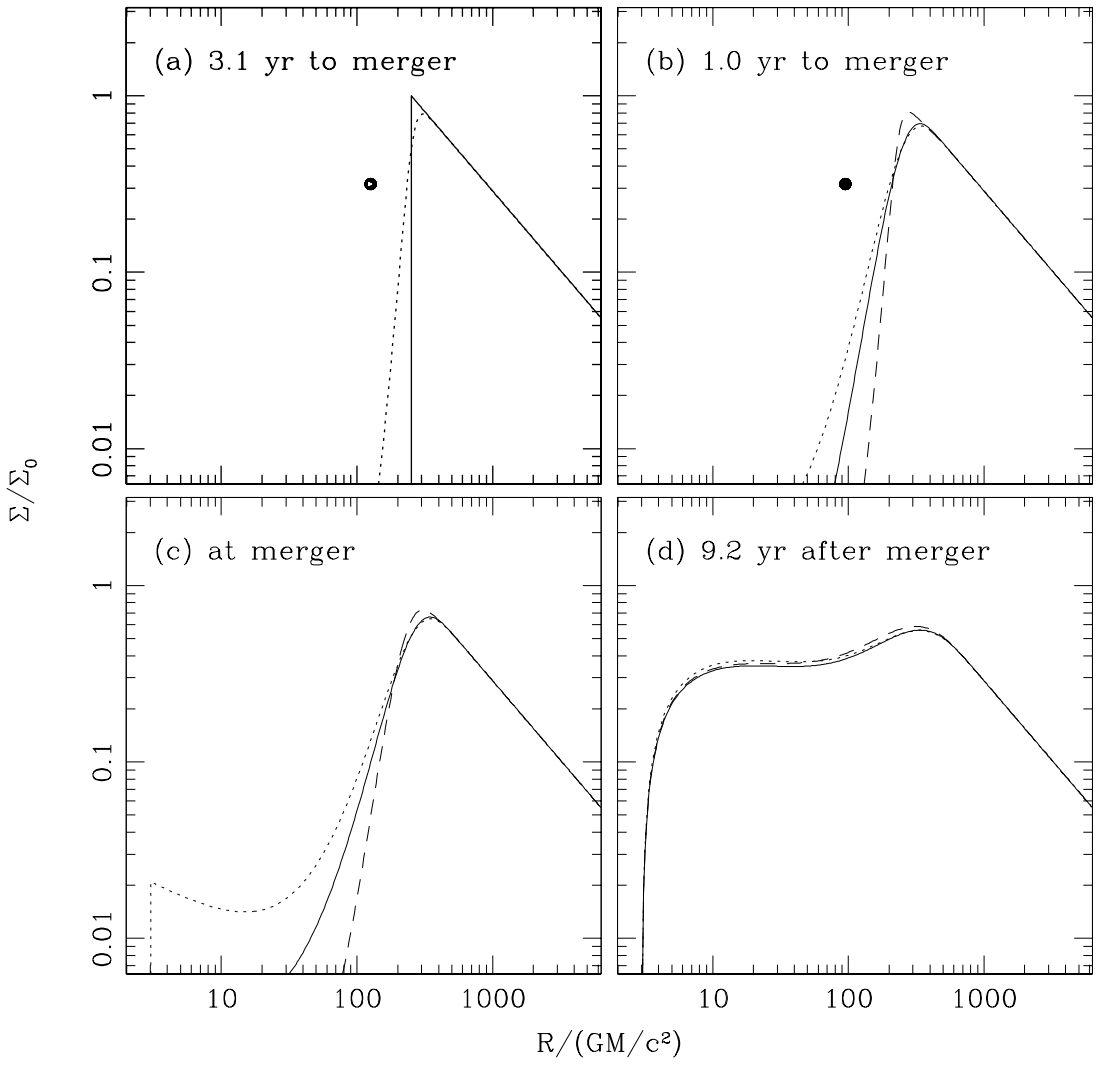

Figure 3.1 Snapshots of the time-dependent solutions to equation 3.5 for the surface density Σ/Σ_0 for fiducial disk parameters. We evolve the solution by applying 3.6 with $n = 0.4$, which is roughly consistent with the physical prescription $\nu \propto T/\Omega$ at late times and small radii, when the disk approaches a quasi-steady accretion track satisfying $\Sigma\nu \propto (1-\sqrt{R_*/R})$. Times in the figure are scaled to the merger of an equal-mass, $10^6 M_\odot$ binary. Panel (a) shows the initial surface density profiles at decoupling. Panels (b), (c) and (d) show the evolved profiles at 1 year before the merger, at merger, and 9.2 years after the merger, respectively. We also evolve a third profile (dashed lines) that qualitatively describes a scenario where the binary continues to open a gap even after the nominal decoupling condition (see text for details). In panels (a) and (b), the semimajor axis of the binary is shown schematically with black circles (the orbit shrinks by $\approx 25\%$ between the panels). After the merger, we impose a zero-torque condition at $R_* = 3GM/c^2$; surface density profiles at merger (Panel c) are truncated at this radius and subsequently evolved with the new boundary condition. All three profiles have the same qualitative time-dependent behavior, suggesting that neither the initial surface density profile at decoupling nor the precise time when the binary ceases to influence the disk edge are critical for our main conclusions on the afterglow signatures.

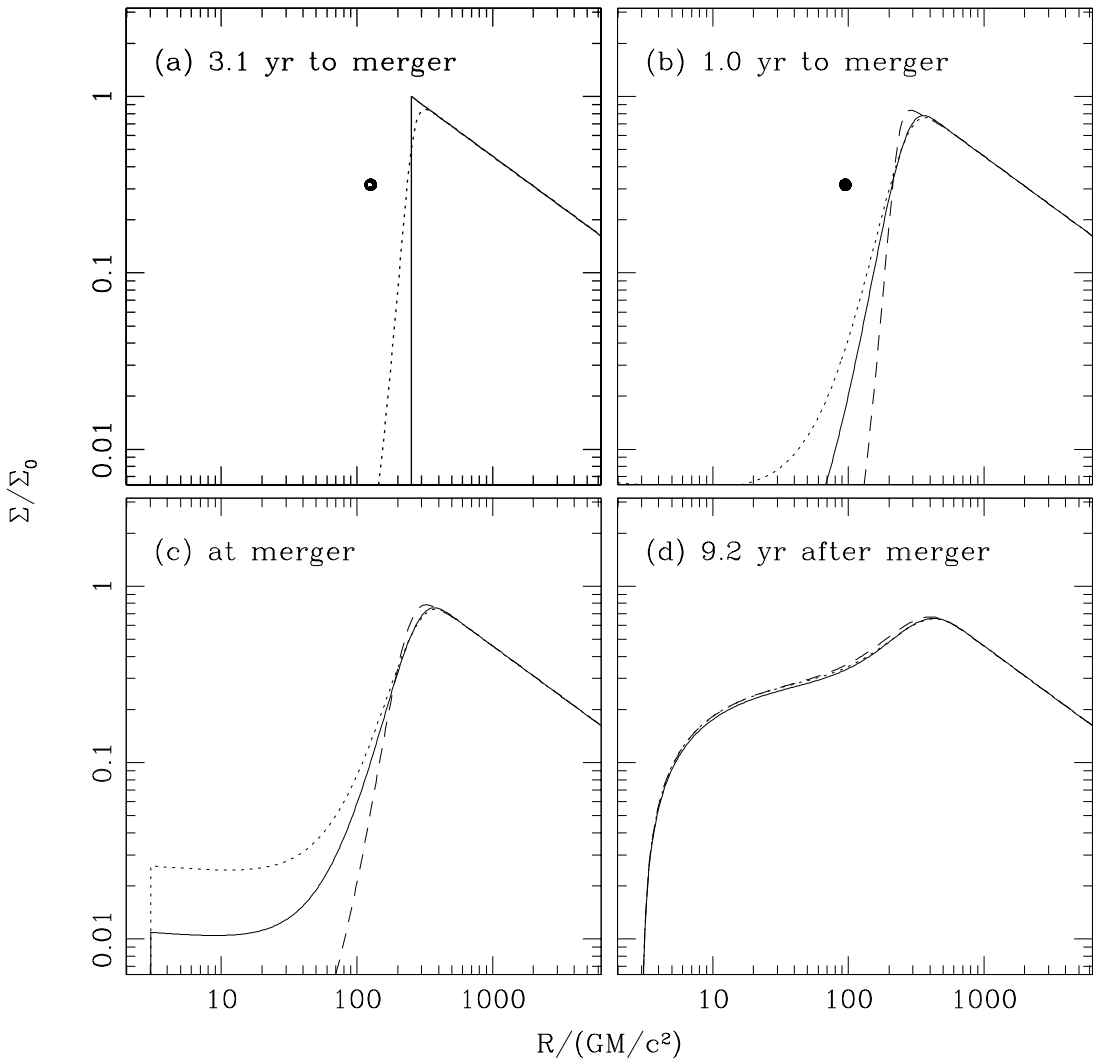

Figure 3.2 Same as Figure 3.1, but with $n = 11/170 \approx 0.065$, which is the value consistent with a physical prescription for viscosity, $\nu \propto T/\Omega$, just outside the inner edge of the disk. We find that our qualitative findings are insensitive to the value of n, as long as the quantity $2 - n$ is not much smaller than unity.

Before the merger, we do not specify an inner boundary condition and so the quasi-steady surface density profile satisfies $\Sigma \propto \nu^{-1} \propto R^{-n}$. After the merger, however, we assume that the SMBH remnant imposes a zero-torque boundary condition at finite radius R_*, which leads to a quasi-steady surface density profile satisfying $\Sigma \propto (1 - \sqrt{R_*/R})R^{-n}$. The two initial profiles evolve to become largely indistinguishable even before the binary merges. For a variety of initial profiles, we have confirmed that the solutions converge on timescales shorter than the time to merger if the profiles share the same value of λ, which is closely related to the location of the maximum of the surface density gradient, and have identical profiles just outside the inner edge.

We note that the physical circumbinary disk is not likely maintain a $m \approx -1/2$ decretion profile at radii far outside of the inner edge. Where the gas is causally detached from the binary torques and the mass accumulation of the inner edge, it is expected to have the $m \approx 0$ profile of a steady accretion disk. The simulations of MacFadyen & Milosavljević (2008) suggest that this transition to a steady accreting solution does not occur until the radius is at least several times the size of the central cavity. Because the viscous time scales as $t_{\rm visc}(R) \propto R^{2-n}$, we do not expect the gas profile at decoupling at these outer radii to contribute to the EM afterglow on timescales of observational interest.

Despite decoupling, some gas is able to follow the binary closely. This raises the possibility that some gas could continue to partially influence the subsequent orbital evolution of the binary. Such a dynamical effect will be proportional to the gaseous mass present in the vicinity of the binary. Since the surface density profile at decoupling cannot viscously evolve significantly at radii beyond several times R_0, we choose to conservatively

evaluate the maximum dynamical influence of the circumbinary gas based on the mass inside $10R_0$. All of the disk profiles in Figures 3.1 and 3.2 have a total mass inside $R < 10R_0$ of

$$M_{\rm disk}(R < 10R_0) \sim 10^{-5} M \times \alpha_{-1}^{-1.36} S^{0.02} \lambda^{2.8} M_6^{1.32} (\beta\zeta)^{0.7} \theta^{-0.34}, \tag{3.8}$$

where the parameter dependencies above are calculated from Table 3.1 and Appendix 3.A using the fact that $M_{\rm disk} \propto \Sigma_0 \lambda^2 a_0^2$. The low, conservative value of the disk to binary mass ratio in equation 3.8 suggests that the disk is unlikely to have a dynamical effect on the binary merger, and to contaminate the GW signal at levels detectable by *LISA*. (Note, however, the scenario of Armitage & Natarajan 2005, in which the disk can imprint *LISA*-observable perturbations on the binary's orbit prior to the decoupling time.) However, such gas may still contribute an observable EM signature because the available specific gravitational energy is large (Chang et al. 2010; Krolik 2010).

Even if the gas cannot influence the binary, it is possible for the binary to continue to influence the evolution of the innermost gaseous disk, even after decoupling. Our models assume that the evolution of the gas and that of the binary can effectively be treated independently, i.e. that the region where the potential is non-Keplerian shrinks faster than the gas can follow. Although this is a useful first approximation, it is not strictly guaranteed by a decoupling criterion based on GW and viscous timescales, as defined earlier. Indeed, our solutions allow for gas as dense as $\gtrsim 10^{-2}\Sigma_0$ to flow to $R \lesssim a(t)$ prior to the merger (see, e.g., Figure 3.1b). To understand why this happens, one can compare the radial gas velocity $V_R = -\dot{M}/(2\pi R\Sigma) = 3(1+2m)\nu/(2R)$ with the shrinking rate of the hypothetical disk edge, $d(2\lambda a)/dt$. The decoupling condition $a_0/(da/dt)_0 = \beta t_{\rm visc,0}$

directly implies $d(2\lambda a)/dt = \beta^{-1}R_0/t_{\rm visc,0} = 3\beta^{-1}\nu_0/(2R_0)$. We thus find that the nominal radial velocity (ignoring binary torques) at the inner edge at decoupling is given by the following expression:

$$V_{R,0} = 0.1\beta_{-1}\,(1 + 2m_0)\,\frac{d(2\lambda a)}{dt}. \tag{3.9}$$

Above, m_0 is the index $\partial \ln(\nu\Sigma)/\partial \ln R$ evaluated at the disk edge, which in principle may be much greater than unity. For sufficiently steep surface density profiles, the innermost gas can continue to flow into the binary's vicinity[5] even after decoupling as defined earlier. Thus, the binary can continue to open a gap in the disk for some time until the binary is truly able to outrun the surrounding gas. This happens when the binary's orbital radius has shrunk below the decoupling value a_0 by a factor $[0.1\beta_{-1}(1 + 2m)]^{1/(n+2)}$. The profiles given by the analytic form in equation 3.7 have $m_0 \gtrsim 10$ for each of the viscosity prescriptions shown in Figures 3.1 and 3.2. We thus estimate the binary can outrun the gas in these profiles when $a \lesssim 0.8a_0\beta_{-1}^{-1/(n+2)}$, at a time approximately $1M_6^{1.3}$ yr before the merger. This suggests that the time when the binary torques cease to influence the circumbinary gas may be as late as $\sim 0.05t_{\rm visc,0}$ before the merger.

To evaluate the possible effect of the binary's continued influence on the gas after formal decoupling, we consider a scenario where the inner gas has maintained an exponential inner density profile similar to the one in equation 3.7, scaled to the new semimajor axis of the binary, until one year before the merger as estimated above. This revised configuration is shown as dashed lines in Figures 3.1 and 3.2. It is evolved viscously in the same way

[5]In fact, because our viscosity prescription is not causal, V_R can be supersonic for large positive values of m_0. Flow speeds are, however, generally subsonic for realistic Σ profiles, as long as $\partial \ln \Sigma/\partial \ln R \lesssim 10^3$ at decoupling. Note that the step-function profile shown in Figure 3.1(a) rapidly evolves to a profile with subsonic radial speeds throughout.

as the other two scenarios in panels (c) and (d) of each figure. While the first two classes of profiles represent the assumption that the binary ceased to tidally interact with the circumbinary gas at decoupling ($t_{\rm GW} = \beta t_{\rm visc}$), the third class of profiles is evolved assuming the tidal interaction continues for a while longer after decoupling, until the binary is able to truly outrun the disk edge as estimated above. We find that the three types of profiles are unlikely to be observationally distinguishable from each other (compare long-dashed and dotted lines in Panels b, c, d in Figures 3.1 and 3.2); there are small differences until merger time, but they are quickly wiped clean by the boundary condition associated with the single SMBH remnant. Figures 3.1 and 3.2 illustrate that neither the density gradient at the inner edge at decoupling nor the subsequent gas-binary interaction inside this radius are likely to have an appreciable effect on the disk evolution on timescales of interest. At the level of accuracy of our idealized models, we do not expect either of these factors to affect too strongly the observable properties of the viscously spreading disk, to which problem we now turn.

3.3 Observable Features of the Time-Dependent Afterglow

3.3.1 Bolometric Light Curve

The power per unit area viscously dissipated from each face of a thin Keplerian disk is equal to $F \sim (9/8)\nu\Sigma\Omega^2$. Substituting our expressions for $\nu \propto r^n$ and $\Omega \propto r^{-3/2}$, and

ignoring advective loss, the power radiated by the disk is

$$\begin{aligned} L_{\rm visc}(t) &\approx \frac{9}{4}\nu_0\Omega_0^2\int \Sigma(r,t)\,2\pi r^{n-2}\,dr \\ &\approx 0.059\,L_{\rm Edd}\times\alpha_{-1}^{0.34}S^{1.2}\lambda^{-1.7}M_6^{-0.08}(\beta_{-1}\zeta)^{-0.42}\theta^{0.08}\int\frac{\Sigma(r,t)}{\Sigma_0}r^{n-2}dr, \end{aligned} \quad (3.10)$$

where $L_{\rm Edd} \equiv 4\pi cGm_{\rm p}M/\sigma_{\rm T}$ is the Eddington luminosity for an object with a mass equal to that of the binary. The dissipated power depends weakly on α and θ, scales roughly linearly with the binary mass, and is most sensitive to the gas distribution parameters S and λ. At decoupling, the second integral in equation 3.10 evaluates to a value $\approx 1/(1-m)$, and thus the initial luminosity of the disk at decoupling is expected to be $\sim 0.04L_{\rm Edd}$ for fiducial parameters and $m \approx -1/2$. In integrating equation 3.10 before the binary merger, we do not calculate emission from radii inside $R < 2\lambda a(t)$. In this region, the potential is highly time dependent and non-axisymmetric, and the assumption that the gas is in nearly circular Keplerian orbits about the binary's center of mass breaks down. Hydrodynamical simulations (e.g., MacFadyen & Milosavljević 2008; Hayasaki et al. 2008; Cuadra et al. 2009) suggest that the gas at these radii will form relatively dim, quasi-periodic accretion flows around the individual SMBHs. We do not address these circumsecondary and circumprimary accretion flows in this paper. The potential role of circumprimary gas inside of the binary's orbit is discussed by Chang et al. (2010).

In Figure 3.3 we present several bolometric rest-frame light curves for the evolving disk, for three different values of the viscosity power-law index n, using equation 3.7 for the surface density profile at decoupling. We show light curves for two different surface density profiles immediately outside the disk edge at decoupling: a decretion-like

disk ($m = -1/2$; thick lines), and an accretion-like disk ($m = 0$; thin lines). The vertical scale is the brightening $L_{\rm visc}(t)$, in units of the disk luminosity at decoupling L_0, while the horizontal scale is the time relative to the merger. In each case, the dissipated luminosity increases steadily by over an order of magnitude within ~ 20 yr of the merger. There is significant evolution before and after our re-evaluation of MP05's estimate for the onset for the afterglow, $t_{\rm EM} = (3/4)t_{\rm GW,0} \sim 9M_6^{1.3}$ yr after the merger.

The sudden brightening that accompanies the merger in Figure 3.3 is due to the fact that we only integrate the circumbinary disk emission outward of the radius $R = 2\lambda a(t)$, as described above: since the binary will merge within ~ 48 hours once it reaches a separation of $\sim 50GM/c^2$, the integrated area rises sharply just before the merger. This is a somewhat artificial ingredient of our model. We note, however, that at merger the depth of the potential well near the center of mass does increase drastically, as does the area where the gas is free to spread viscously. This sudden change in the central potential well could be associated with significant emission, for instance if even trace amounts of gas remain present between the components of the merging binary (Chang et al. 2010).

The disk emission can exceed $L_{\rm Edd}(M)$, the Eddington luminosity limit associated with the binary mass M, without violating local Eddington flux limits. This result may appear contradictory, but it can be understood as follows. The Eddington flux limit associated with the vertical gravitational field of the disk (ignoring self-gravity) is

$$F_{{\rm Edd},z} = \left\{ \frac{H}{R} \left[1 + \left(\frac{H}{R} \right) \right]^{-3/2} \right\} \frac{c}{\kappa_{\rm es}} \frac{GM}{R^2}, \tag{3.11}$$

where H is the scale height of the disk and $\kappa_{\rm es} \approx 0.40$ cm^{2}g^{-1} is the electron-scattering

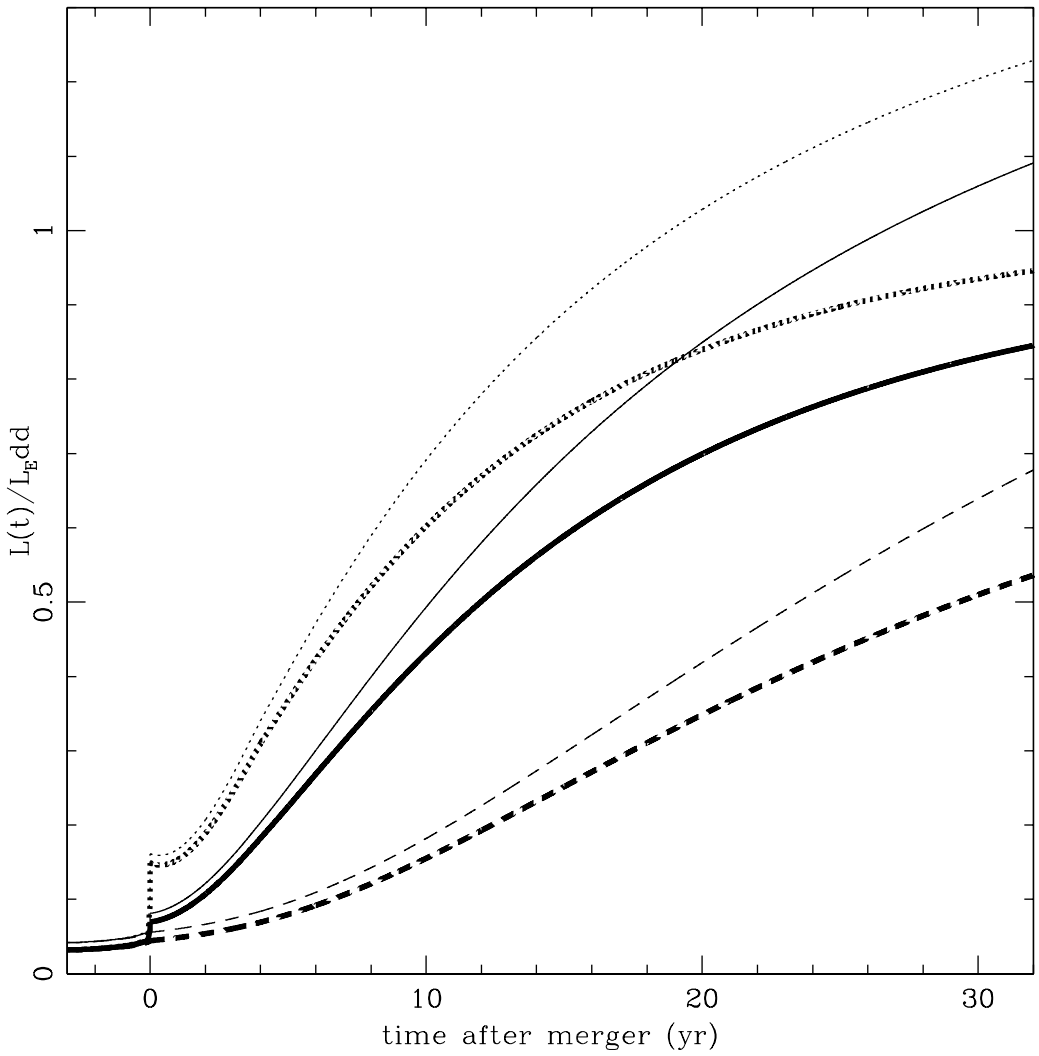

Figure 3.3 The viscously dissipated bolometric disk luminosity for fiducial parameters, in units of the Eddington luminosity. The time is relative to merger. The thick lines depict the light curve for density profiles that have a decretion power-law $m = \partial \ln(\nu\Sigma)/\partial \ln R = -1/2$ outside the inner edge. The thin lines are for profiles with $m = 0$ outside the inner edge. The light curves are somewhat sensitive to the value of $n \equiv \partial \ln \nu / \partial \ln R$, but have the same qualitative behavior. The light curves shown here correspond to disks with viscosity index $n = 0.4$ (solid lines), $n = 11/170$ (dashed lines) and $n = 1$ (dotted lines). Our model predicts significant evolution before and after $t_{\rm EM} \sim 9 M_6^{1.3}$ yr, our re-evaluation of the nominal time proposed by MP05 for the onset of the viscosity-powered afterglow. The sharp increase in luminosity at merger time is a somewhat artificial effect that arises from the fact that we account only for the flux dissipated outside the radius $2\lambda a(t)$ (see text), which shrinks rapidly in the days before merger. Note that the luminosity scales with the mass-accumulation parameter S as $L \propto S^{1.2}$, and so could be significantly enhanced if the inner disk is significantly more massive than a comparable steady-state Shakura-Sunyaev disk.

opacity. The maximum luminosity of a steady disk is then $L_z \sim 4\pi \int F_{\mathrm{Edd},z}\, R\, dR$, where we have introduced an additional factor of two to account for the fact that the disk emits from two faces. The leading geometric factor in curly brackets in equation 3.11 is a monotonically increasing function of H/R. It approaches the value H/R in the limit $H/R \ll 1$, reaches a maximum of $2/(3\sqrt{3}) \approx 0.38$ at $H/R = 1/\sqrt{2}$, and goes to zero as H/R goes to infinity. The luminosity L emitted by an annulus with inner and outer radii R_{in} and R_{out} is capped by the inequality

$$L < \int_{R_{\mathrm{in}}}^{R_{\mathrm{out}}} 4\pi \frac{2}{3\sqrt{3}} \frac{cGM}{\kappa_{\mathrm{es}}} R^{-1}\, dR = \frac{2}{3\sqrt{3}} \ln\left(\frac{R_{\mathrm{out}}}{R_{\mathrm{in}}}\right) L_{\mathrm{Edd}}. \tag{3.12}$$

Thus, the bolometric luminosity can reach several Eddington luminosities for reasonable disk sizes.

We find that the evolution of the light curve is largely determined by the viscosity profile in the disk and has a relatively weak dependence on the surface density profile beyond the disk edge. Figure 3.3 suggests that the accretion luminosity of the viscously spreading inner disk may be a viable candidate for observational follow-up almost immediately after the main GW event. If the electromagnetic accretion signature associated with the spreading inner disk can be detected in the "rise" phase, the light curve may provide clues about the nature of viscosity in the accretion disk. We find that our disk solutions can develop bolometric luminosities in excess of the Eddington luminosity of the merged SMBH remnant. These luminosities can be generated even if the locally dissipated flux is sub-Eddington (§3.3.4 below), and are enhanced strongly if the value of the mass-accumulation parameter S greater than unity.

3.3.2 Spectral Evolution

We now turn to the spectral evolution of the viscously spreading disk. Below, we present an abbreviated derivation of the thermal structure of the inner circumbinary disk and discuss its spectral features. We refer the reader to MP05 and our Appendix 3.A for more detailed derivations.

The spectrum differs from that of an exact blackbody because the photons of different frequencies are thermalized at different depths above the disk. This thermalizing region is referred to the "thermalization photosphere" or the "effective photosphere." The temperature at the bottom of the photosphere is equal to the effective temperature as seen by an observer far above the photosphere, viz. the effective optical depth above this height is of order unity. We call this temperature the photospheric temperature, $T_{\rm p}$.

There are two sources of opacity in the photosphere: electron scattering, which has the frequency-independent opacity $\kappa_{\rm es}$; and absorption, which is dominated by the bound-free process. We follow MP05 and prescribe a Kramer's functional form for the absorption opacity $\kappa_{{\rm abs},\nu} \propto \rho T^{-7/2} f_\nu(\xi)$, where $f_\nu(\xi) \equiv \xi^{-3}(1 - e^{-\xi})$ and $\xi \equiv h\nu/kT$. We scale $\kappa_{{\rm abs},\nu}$ so that its Rosseland mean recovers the standard Kramer's bound-free opacity for solar metallicity, $1.6 \times 10^{24} (\rho \ {\rm cm^3 g^{-1}}) (T/{\rm K})^{-7/2} \ {\rm cm^2 g^{-1}}$. For a wide range of system parameters, we find that scattering is the dominant source of opacity in the photosphere at the inner edge at decoupling. In this limit, $\kappa_{{\rm abs},\nu} \ll \kappa_{\rm es}$, we find $\kappa_{{\rm abs},\nu} \approx \kappa_{{\rm abs},*} f_\nu^{1/2}$, with $\kappa_{{\rm abs},*} \approx 4.7 \times 10^{20} \ {\rm cm^2 g^{-1}} \times (\Omega \ {\rm s}) (T_{\rm p}/{\rm K})^{-15/4}$ a frequency-independent quantity. (See Appendix 3.A for the intermediate calculations for the absorption opacity, including the general case allowing for $\kappa_{{\rm abs},\nu} \sim \kappa_{\rm es}$.)

The so-called "graybody" flux emitted by each face of the disk is given by the expression (e.g., Rybicki & Lightman 1986; Blaes 2004)

$$F_\nu \sim \pi \frac{2\epsilon_\nu^{1/2}}{1+\epsilon_\nu^{1/2}} B_\nu, \tag{3.13}$$

where $B_\nu(T_\nu)$ is the Planck function and $\epsilon_\nu \equiv \kappa_{\mathrm{abs},\nu}/(\kappa_{\mathrm{abs},\nu} + \kappa_{\mathrm{es}}) \leq 1$ is the ratio of the absorption to the total opacity. Both quantities are evaluated at T_p, the temperature at the bottom of the photosphere. Since $\kappa_{\mathrm{abs},\nu}$ is a monotonically decreasing function of ν, the photosphere emits increasingly less efficiently at higher frequencies, relative to a blackbody with the same temperature. In other words, the photosphere has a higher temperature relative to a blackbody with the same radiant flux, and thus emits at higher frequencies. The expression given in 3.13 recovers the blackbody flux $F_\nu = \pi B_\nu$ in the limit where absorption is the dominant source of opacity, i.e. when $\epsilon_\nu \to 1$. This limit is relevant for our disk solutions, in which ϵ_ν can span the full range between 0 and 1 in radii and frequencies of interest. Note that the flux expression used by MP05 is lower than ours by a factor of two.

If again advective losses are ignored, integrating the flux of equation 3.13 over frequency must give the flux viscously dissipated by half of the disk, $(9/8)\nu\Sigma\Omega^2$. We obtain

$$\frac{F_\mathrm{visc}}{2} = \int_0^\infty F_\nu \, d\nu = \Xi(T_\mathrm{p}, \Omega)\, \sigma T_\mathrm{p}^4 = \frac{9}{8}\nu\Sigma\Omega^2, \tag{3.14}$$

where

$$\Xi(T_\mathrm{p}, \Omega) \equiv \frac{15}{\pi^4} \int_0^\infty \frac{2\epsilon_\nu^{1/2}}{1+\epsilon_\nu^{1/2}(\xi)} \frac{e^{-\xi} d\xi}{f_\nu(\xi)} \approx \frac{4}{5}\epsilon_*, \tag{3.15}$$

is the deviation of the bolometric flux from blackbody and $\epsilon_*(T_\mathrm{p}, \Omega) \equiv \kappa_{\mathrm{abs},*}/(\kappa_{\mathrm{abs},*}+\kappa_\mathrm{es}) < 1$. The above approximation for Ξ is accurate to within 10% in the range $0 \leq \epsilon_* \lesssim 0.9$. (See Appendix 3.A for a more accurate fit to the integral.)

To calculate the spectrum, we solve equation 3.14 numerically for T_p, substitute the result into equation 3.13 and integrate:

$$L_\nu(R,t) \approx 2\pi^2 R_0^2 \int \frac{2\epsilon_\nu^{1/2}(\Omega, T_\mathrm{p})}{1+\epsilon_\nu^{1/2}(\Omega, T_\mathrm{p})} B_\nu(T_\mathrm{p})\, r\, dr. \tag{3.16}$$

The temperatures at the midplane and the bottom of the photosphere are related through

$$T^4 = \frac{3\tau}{4} T_\mathrm{p}^4, \tag{3.17}$$

if the region between the two heights is optically thick and can be treated as a one-zone gray atmosphere. Here $\tau = \theta\kappa_\mathrm{es}\Sigma$ is the scattering-dominated optical thickness between the two heights, and $\theta \leq 1$ is a porosity correction factor. We adopt MP05's interpretation of the simulation results of Turner (2004) and use $\theta = 0.2$ as our fiducial value, and define $\theta_{0.2} \equiv \theta/0.2$.

Neglecting the advected flux, and in the limit where the circumbinary disk is dominated by radiation pressure and electron scattering, we may estimate the frequency for the peak monochromatic flux at decoupling, and at late times after the merger. At decoupling, most of the emission comes from the inner edge, and the spectrum in equation 3.16 peaks

at the Wien frequency $\xi = h\nu/kT_{\rm p,0} \approx 2.8$, from which we estimate

$$h\nu_{\rm peak} \approx 15\ {\rm eV} \times \alpha_{-1}^{0.36} S^{0.73} \lambda^{2.1} M_6^{-0.32} (\beta_{-1}\zeta)^{-0.45} \theta_{0.2}^{0.09}. \tag{3.18}$$

Long after the merger, the inner disk approaches a quasi-steady accretion track satisfying $3\pi\nu\Sigma = \dot{M}_{\rm out}(1 - \sqrt{R_*/R})$, where $\dot{M}_{\rm out}$ is a radially constant mass supply rate of the inner disk. The bolometric emission is brightest where the quantity $2\pi R\nu\Sigma\Omega^2$ is maximal in the disk, which corresponds to $R = 25/16R_*$. Applying this to equation 3.14, we calculate the photospheric temperature and find the corresponding peak frequency for the monochromatic flux at very late times to be

$$h\nu_{\rm peak} \approx 0.71\ {\rm keV} \times M_6^{-8/34} \left(\frac{\dot{M}}{\dot{M}_{\rm Edd}} \frac{0.1}{\eta}\right)^{8/17} \left(\frac{R_* c^2}{3GM}\right)^{-18/17}. \tag{3.19}$$

Above, the dependence on S is replaced by a dependence on the mass supply rate, as at late times the disk loses memory of earlier accumulation near the inner edge. For our choice of $R_* = 3GM/c^2$, the radiative accretion efficiency is $\eta = 1 - \sqrt{1 - 2GM/(R_* c^2)} \approx 0.12$ (Novikov & Thorne 1973). The circumbinary disk is expected to reach this quasi-steady track after at least $t_{\rm visc,0} \approx 120 M_6^{1.3}$ yr has elapsed since merger.

Figure 3.4 shows the spectrum of our fiducial model disk around an equal-mass, $10^6 M_\odot$ binary as calculated at different times with respect to the merger, from decoupling time ($\approx$ 3 yr before the merger) up to 120 yr after the merger. We show the exact results using equation 3.16 in solid lines[6] , and the blackbody spectrum in dotted lines. In the

[6]The function νF_ν peaks at $\xi = 3.9$, so replace the leading factor 15 eV in equation 3.18 above with 21 eV to derive the locations of the peaks in the plots of Figure 3.4.

figure, we only account for emission out to a radius of $1000GM/c^2$, as we do not expect significant contribution of high-frequency photons from regions far outside R_0. We find that the spectrum of the circumbinary disk does not evolve significantly from that at decoupling until several hours before the merger, because the deepest parts of the central potential are not accessible to the circumbinary gas until that time. Our initial and final spectra agree qualitatively with those given in MP05. As with the bolometric light curve, however, significant evolution is ongoing well before and after the nominal estimate for the onset of the afterglow. Much of the evolution, however, is complete $\sim 10M_6^{1.3}$ yr after the merger.

3.3.3 Possible Reprocessing of the X-ray Signature

The frequencies calculated above suggest that an evolving afterglow at low redshift or low binary mass may be observable by existing and future X-ray and UV telescopes shortly after the GW event. MP05 discussed the possibility that the afterglow could be reprocessed to IR frequencies if the merged binary is enshrouded by gas and dust. Here we note that the circumbinary disk itself is a plausible candidate to reprocess the UV/X-ray afterglow to lower frequencies.

Ignoring general relativistic effects, the self-irradiating flux of the disk with viscously dissipated flux $F_{\rm visc}$ can be written (see, e.g., Blaes 2004)

$$F_{\rm irr}(R) = \frac{1-A}{4\pi R^2} \int 2\pi R' \sin\vartheta F_{\rm visc}(R')\, dR', \tag{3.20}$$

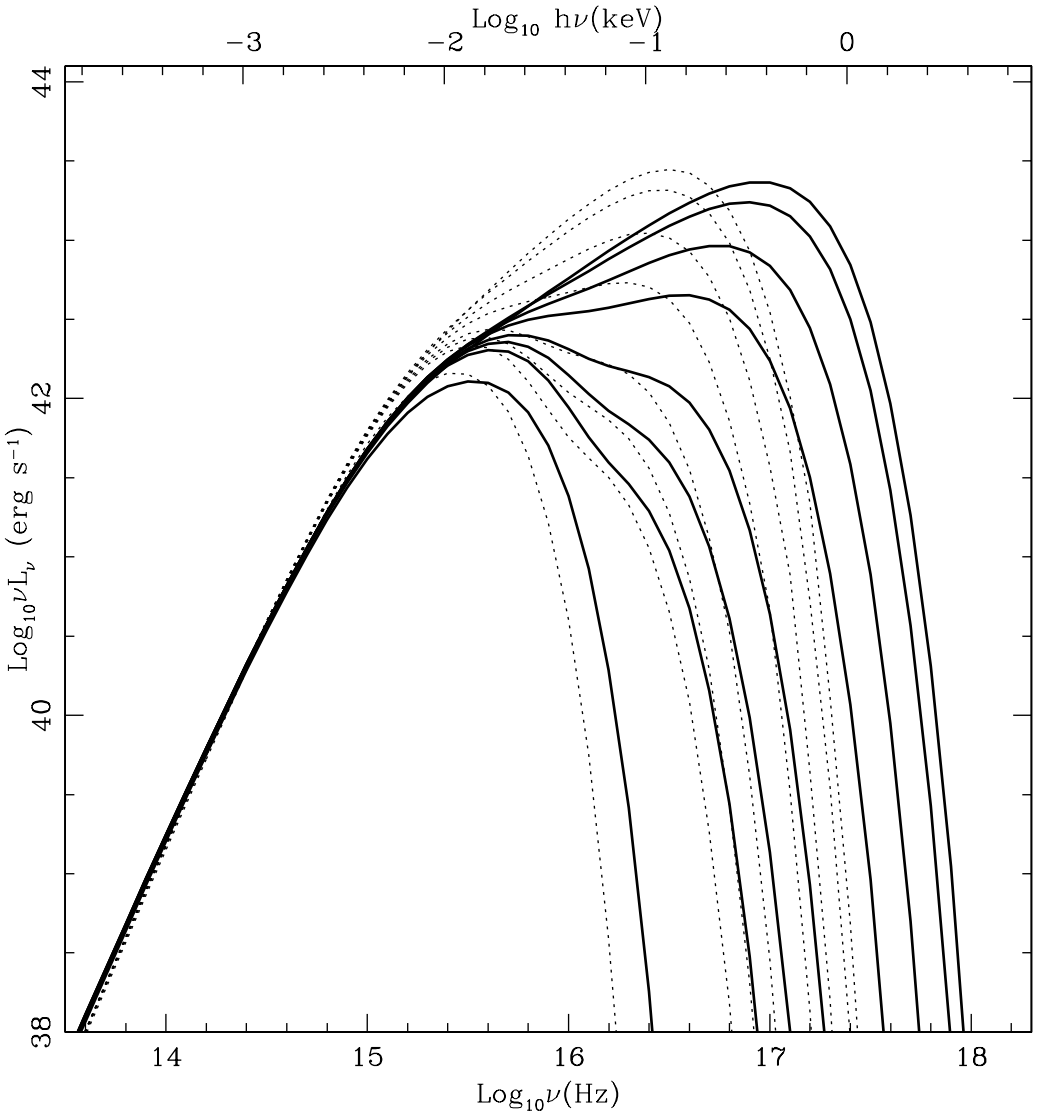

Figure 3.4 The thermal disk spectra emitted at (from left to right): decoupling (3.1 years before the merger), 1 month after, 1 year after, 2 years after, 5 years after, 9.2 years after, 20 years after, and 120 years after the merger. The thick, solid lines show the spectra calculated using the "graybody" formulation in equation 3.16, while the thin, dotted lines show the spectra calculated assuming the disk instead emits as a perfect blackbody. For any given snapshot, the two sets of lines show different spectral distributions of the same bolometric luminosity. For the gray-body case, the 9.2-year snapshot corresponds to our re-evaluation of the nominal estimate by MP05 for the onset of the X-ray afterglow, and 120 years is the viscous time at the disk edge at decoupling.

where A is the albedo of the disk and ϑ is the angle between the disk surface and the irradiating light ray. In our disk solutions, most of the dissipated disk luminosity $L_{\rm visc}$ is produced at small radii due to the fact that the viscously dissipated flux is proportional to Ω^2. Thus, if R is much greater than the size of the central emitting region then it is a reasonable approximation to replace the integral in the above equation with $L_{\rm visc} \sin\vartheta$. If one further assumes that A is small and that the irradiating flux is large compared to the local value of $F_{\rm visc}$, the luminosity reprocessed by the outer disk will be $L_{\rm irr} \lesssim L_{\rm visc} \sin\vartheta$. The light-travel time is short ($\sim 10^{-2} M_6$ yr across distances $\sim 10^5 GM/c^2$) compared to the timescale for viscous evolution of the inner disk, and thus we expect the reprocessing processes below to be effectively instantaneous.

One way that the factor $\sin\vartheta$ can be large is if there is a warp in the disk that offsets the planes of the outer and the innermost regions. The circumbinary disk is expected to be aligned with the binary orbit in the binary's vicinity, where tidal torques dominate viscous torques. Farther out in the disk, the disk may have a quasi-steady warp with respect to the binary (Ivanov et al. 1999). Because the warp would dissipate on timescales longer than the afterglow itself, it could act to promptly reprocess the afterglow into longer wavelengths.

We also find that the evolution of the innermost accretion flow itself could effect a disk geometry that is conducive to reprocessing. If there is no warp and the emitting region is in the equatorial plane, then (e.g. Blaes 2004)

$$\sin\vartheta \sim \frac{H}{R}\left(\frac{d\ln H}{d\ln R} - 1\right). \tag{3.21}$$

As we show below (Figures 3.5 and 3.7), just outside the regions of brightest emission ($R \lesssim R_0$) and around the time of merger our disk solutions are often moderately geometrically thick ($H/R \gtrsim 0.1$) and steeply flared geometrically (large positive $d \ln H/d \ln R$). That is, the circumbinary disk itself may serve as a shroud of gas that can capture and reprocess the high-frequency accretion signature.

Either of the two effects considered above could cause the luminous inner disk around the binary remnant to irradiate the outer disk at a sufficiently large angle ϑ to effect a reprocessed IR/optical signature of substantial luminosity.

3.3.4 Possible Effects of Advection and Super-Eddington Winds

Above, we largely confirmed the findings of MP05, with the additional suggestion that accretion afterglows of SMBH mergers may be observable somewhat earlier than estimated in that study. However, we also find that at late times the spreading accretion flow becomes geometrically thick near the SMBH remnant, with the scale-height-to-radius ratio H/R formally exceeding unity. Here, $H = \Sigma/\rho = c_s/\Omega$, where c_s is the sound speed $\sqrt{\gamma P/\rho}$ and $\gamma \approx 4/3$ is the adiabatic index for a radiation-pressure-dominated gas. The radial profile of ρ is calculated from the disk temperature profile, which is obtained through equations 3.14 and 3.17 and the total pressure in the disk $P_{\rm gas}(\rho, T) + P_{\rm rad}(T)$.

For some binary and disk parameters, the inner edge of the disk may already be geometrically thick at decoupling (MP05; see Table 3.1). In such a scenario, the circumbinary gas may be free to accrete off-binary plane, and the inner cavity may not be as evacuated as we have assumed here, and the difference between the pre- and post-merger disk spectra

may be less easily discernible. In addition, after the merger, it is also possible for the accreting gas to become radiatively inefficient, as in the advection-dominated accretion flow (ADAF; Narayan & Yi 1994; Abramowicz et al. 1988) model, which would make it more difficult to observe[7].

As argued in MP05, if horizontal advection is competitive with viscous dissipation then the physical disk would be thinner than suggested by our simple estimate of the ratio H/R, as in the "slim disk" model of Abramowicz et al. (1988). In this paper, we do not attempt to develop a self-consistent time-dependent model that incorporates vertical disk structure and radial advection. Rather, we proceed below as if the disk evolution can be approximated by the thin-disk evolution described by equation 3.6 while keeping the above caveats in mind.

The absolute value of the advected flux can be estimated as (e.g., Abramowicz et al. 1988)

$$Q_{\rm adv} = \frac{\dot{M}}{\pi R^2}\frac{P_{\rm rad}}{\rho}\left|\frac{d\ln\rho}{d\ln R} - 3\frac{d\ln T}{d\ln R}\right|. \tag{3.22}$$

Substituting equation 3.4 for the local accretion rate and equation 3.43 for the radiated flux, we crudely estimate the advected to radiated flux ratio as:

$$\frac{Q_{\rm adv}}{Q_{\rm rad}} \approx \frac{4}{3}\frac{P_{\rm rad,}H}{\Sigma R^2\Omega^2}\left[1 + \frac{d\ln(\nu\Sigma)}{d\ln R}\right]\left|\frac{d\ln\rho}{d\ln R} - 3\frac{d\ln T}{d\ln R}\right| \tag{3.23}$$

The simple calculations of the evolving disk spectrum performed in §3.3.2 are not highly dependent on the disk temperature (our Figure 3.4 agrees very well with the

[7]Here, we mean a single-temperature, collisional, radiatively inefficient flow (as opposed to the collisionless, optically thin, two-temperature models in the literature, e.g., Shapiro et al. 1976; Rees et al. 1982; Narayan & Yi 1995b).

spectra shown in MP05, despite the fact that we have calculated a somewhat lower value of T_0). However, the structure of the disk at the inner edge at decoupling and the subsequent evolution of the accretion flow is likely to depend much more sensitively on the temperature profile. In the radiation-dominated limit, H is proportional to $P_{\rm rad}\Sigma^{-1}\Omega^{-2}$ so from equation 3.22 we see that the advected flux is roughly proportional to $\nu T^8\Sigma^{-1}\Omega^{-2}R^{-2}$. This implies that near the inner edge, $Q_{\rm adv} \propto T^{62/7}$ approximately (see Appendix 3.A for how disk quantities such as Ω scale with the temperature near the edge). Thus, we expect the possible transition of the disk from a thin disk to an advective/slim disk or a geometrically thick ADAF − and the corresponding spectral evolution − to be a delicate function of the thermal structure of the disk at decoupling. For example, MP05 estimated a midplane temperature at the inner edge that was ~ 1.3 times higher than we did, with most of the discrepancy due to the aforementioned factor of two in the flux formalization. Although this does not result in significant differences between our values and theirs for the decoupling radius or the relevant timescales, it does lead those authors to estimate that the advective flux is more significant at the disk edge at decoupling than we do. We estimate that immediately outside the disk edge,

$$\frac{Q_{\rm adv}}{Q_{\rm rad}} \gtrsim 10^{-2}\alpha_{-1}^{1.5}S^{4.9}\lambda^{-7.6}M_6^{-0.24}\left(\beta_{-1}\zeta\right)^{-1.9}\theta_{0.2}^{2.4}, \tag{3.24}$$

where MP05 estimated $Q_{\rm adv}/Q_{\rm rad} \sim 0.44$. The flux ratio is higher at the edge due to the steep temperature and density gradients there. The equation above suggests that it is highly sensitive to all of the disk and binary parameters save for the binary mass.

In Figure 3.5 we show the radial profiles of several additional disk quantities for

the same parameters, profile and times as shown in Figure 3.1. Along with the surface density, we also plot: the simple estimate from equation 3.23 for the advected-to-radiated flux ratio, $Q_{\rm adv}/Q_{\rm rad}$; the estimated scale-height-to-radius ratio H/R; and $F/F_{\rm Edd}$, the ratio of the locally dissipated flux to $F_{\rm Edd} \approx H\Omega^2 c/\kappa_{\rm es}$, the local Eddington flux limit associated with the vertical component of gravity. For the last ratio, we find

$$\frac{F}{F_{\rm Edd}} \approx 0.06 \frac{\Sigma(R,t)}{\Sigma_0} \sqrt{1+(H/R)^{-2}}\, r^{n-1} \alpha_{-1}^{0.34} S^{1.2} \lambda^{-1.70} M_6^{-0.08} \left(\beta_{-1}\zeta\right)^{-0.42} \theta_{0.2}^{0.08}. \tag{3.25}$$

Figure 3.5 confirms that for fiducial parameters, the circumbinary accretion flow remains locally sub-Eddington. For fiducial parameter values, we find that a relatively high surface density $\Sigma/\Sigma_0 \gtrsim 1$ (or large S) is required to reach a super-Eddington flux near $R = R_*$, and the disk is thus not necessarily likely to develop super-Eddington winds at levels which would significantly affect the light curves or spectra presented in Figures 3.3 and 3.4. At late times, however, the nominal advective flux exceeds the locally dissipated flux as the disk becomes geometrically thick, which may lead the disk to become radiatively inefficient.

This hypothetical transition of the innermost accretion flow from a thin disk to a geometrically thick flow could act to suppress the X-ray afterglow. We estimate very simply the effect of advection on the emitted spectrum as follows. We assume that the emitted flux is suppressed by a factor $f \equiv \max\{0, 1 - Q_{\rm adv}/Q_{\rm rad}\} \leq 1$, and that the advected energy is not re-emitted. We also assume that advection acts to suppress emission evenly across all photon frequencies. The resulting advection-limited spectrum is shown in Figure 3.6 alongside a spectrum calculated without accounting for advection (i.e. the

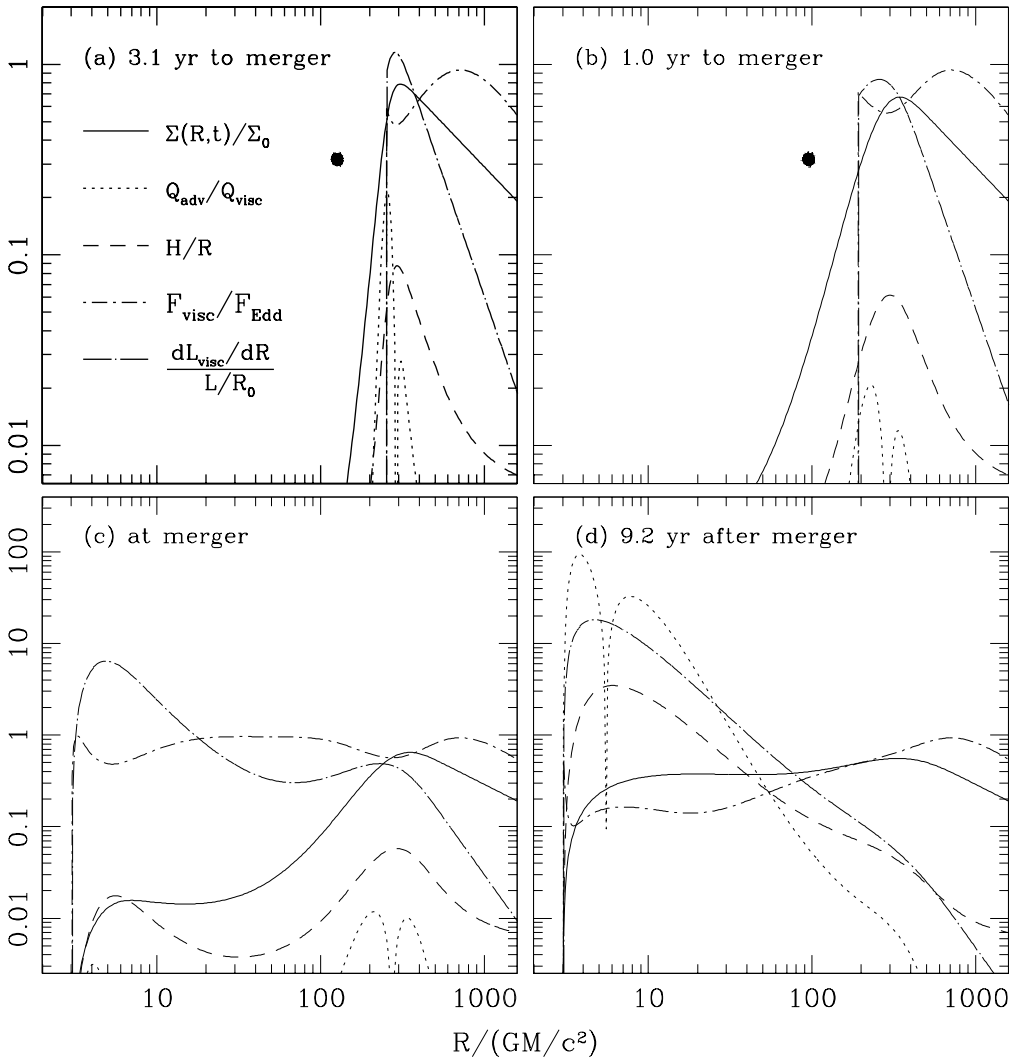

Figure 3.5 For our fiducial disk model: the surface density, scale-height-to-radius ratio H/R; estimated ratio of the advected to radiated flux $Q_{\rm adv}/Q_{\rm rad}$; the ratio of the local flux to the Eddington limit $F_{\rm visc}/F_{\rm Edd}$; and the flux contribution $dL_{\rm visc}/dR = 2\pi RF_{\rm visc}$ normalized to L/R_0. The disk is geometrically thin and advection is largely insignificant until the merger. After the merger, however, the inner disk becomes formally geometrically thick and advection-dominated (Panel d). Note that in panels (c) and (d) the vertical scale is different from the other two panels. The steep dips in $Q_{\rm adv}/Q_{\rm rad}$ are due to the advected flux changing sign (see equation [3.23]).

same disk spectra as those shown in Figure 3.4). from decoupling to 20 years after the merger. We compare the two classes of spectra at decoupling, and 1 month, 2 yr, 5 yr, 9.2 yr and 20 yr after the merger, with thicker lines in the figure denoting later times. If the disk were to remain radiatively efficient at all radii and times, the emission at high frequencies would steadily increase until the quasi-steady thin-disk track is reached. If instead the innermost disk becomes advective, this suppresses the high-energy emission at late times. Our results indicate that our fiducial disk model may become increasingly less bright as the inner disk becomes increasingly geometrically thick. However, the calculations shown in Figure 3.6 suggest that a brightening afterglow may still be visible in the soft X-ray for a short time $\sim 10^{-2} t_{\rm visc,0}$ after the merger. This suggests that the circumbinary gas may have two observable signatures of interest: an initial brightening phase with a soft X-ray afterglow, and a subsequent dimming phase as it transitions into an ADAF.

To further investigate the possible consequences of advection on the circumbinary disk evolution, we turn to a different disk model whose inner edge is already geometrically thick at decoupling. We take $S = 5$ and $\zeta = 1/3$ (binary mass ratio $q \approx 1/10$), keeping all other parameters the same as in the fiducial model. The new choices for S and ζ are physically plausible ones. If the mass supply rate of the outer disk is comparable to the Eddington limit of the central binary, we can expect $S > 1$ at decoupling as a result of mass accumulation. We may also expect $q \sim 1/10$ mergers to be more common[8] than near-equal-mass mergers because merging galaxies (dark matter halos) have unequal masses and because the mass contrast between the SMBHs is expected to be higher than that of the host halos (i.e. the observationally inferred SMBH-to-host-halo mass ratio relation

[8]Provided that they can overcome the "last-parsec problem."

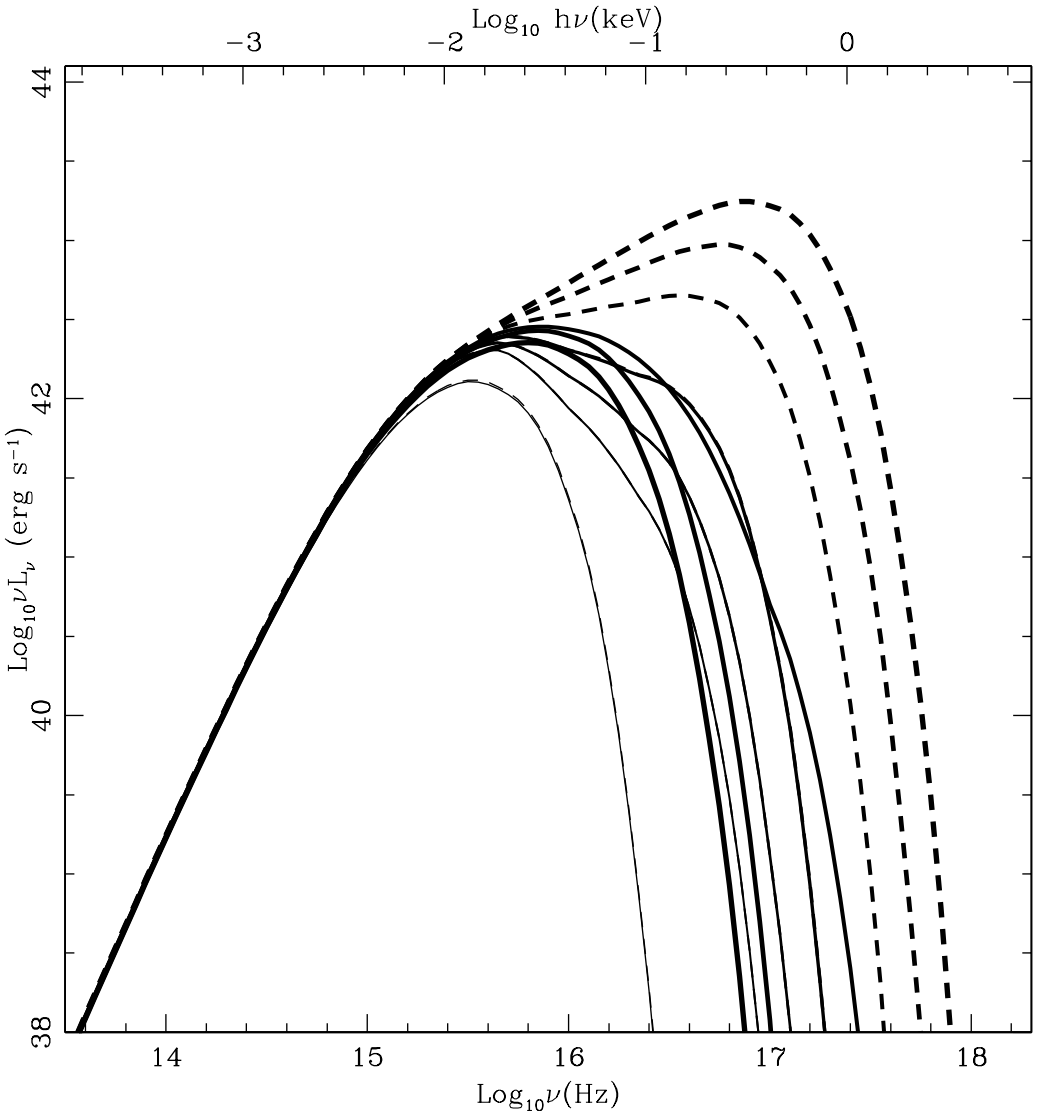

Figure 3.6 Emission spectra for the fiducial circumbinary disk. The dashed lines show the same spectra as calculated in Figure 3.4, from thinnest to thickest curves, at decoupling and at 1 month, 2 yr, 5 yr, 9.2 yr and 20 yr after the merger. The solid lines show the spectra at the same times, but after subtracting the estimated advected flux. The disk is initially geometrically thin and advection is not significant until after the merger. At late times, the inner disk becomes geometrically thick and radiatively inefficient, and this could lead to reduced overall X-ray emission. The disk emits briefly emits soft X-rays at intermediate times, $\sim$ 2 yr after the merger.

is a steeper-than-linear function of the halo mass; Ferrarese 2002). Cosmological merger tree calculations for SMBH mergers consistently show that contributions to the *LISA* data stream from SMBH coalescences will be dominated by moderately low-q events (e.g., Volonteri et al. 2003a; Sesana et al. 2007b; Tanaka & Haiman 2009).

With this new set of parameters, the disk decouples at a time $t_{\mathrm{GW},0}/4 \approx 0.30$ yr before the merger, when the binary has reached a semimajor axis $a_0 \approx 54GM/c^2$. The nominal estimate for the onset of the afterglow emission is $t_{\mathrm{EM}} \approx 0.91$ yr after the merger. The inner edge of the disk has a somewhat higher temperature, $T_0 \approx 8.7 \times 10^6$K. We again use the initial density profile given by equation 3.7. In Figure 3.7 we show the same disk quantities for this thick disk as shown for the fiducial disk in Figure 3.5. This time, the disk remains geometrically thick and radiatively inefficient inside the initial edge radius throughout its evolution. As we did for the fiducial disk model, we estimate the advective suppression of the spectrum for the thick disk and show results in Figure 3.8. This simple calculation suggests that a disk that is sufficiently geometrically thick at decoupling may not, in principle, exhibit any observable spectral evolution.

On the other hand, we also find that the viscously dissipated flux of this thick circumbinary disk becomes super-Eddington at late times (Figure 3.8d). Begelman (2002) proposed that accretion disks with a super-Eddington flux may be able to stay geometrically thin and radiatively efficient. Such accretion flows may also generate powerful outflows and flares. While our simple models do not self-consistently treat advection or the super-Eddington regime, our results underscore the need for further study of this class of accreting systems.

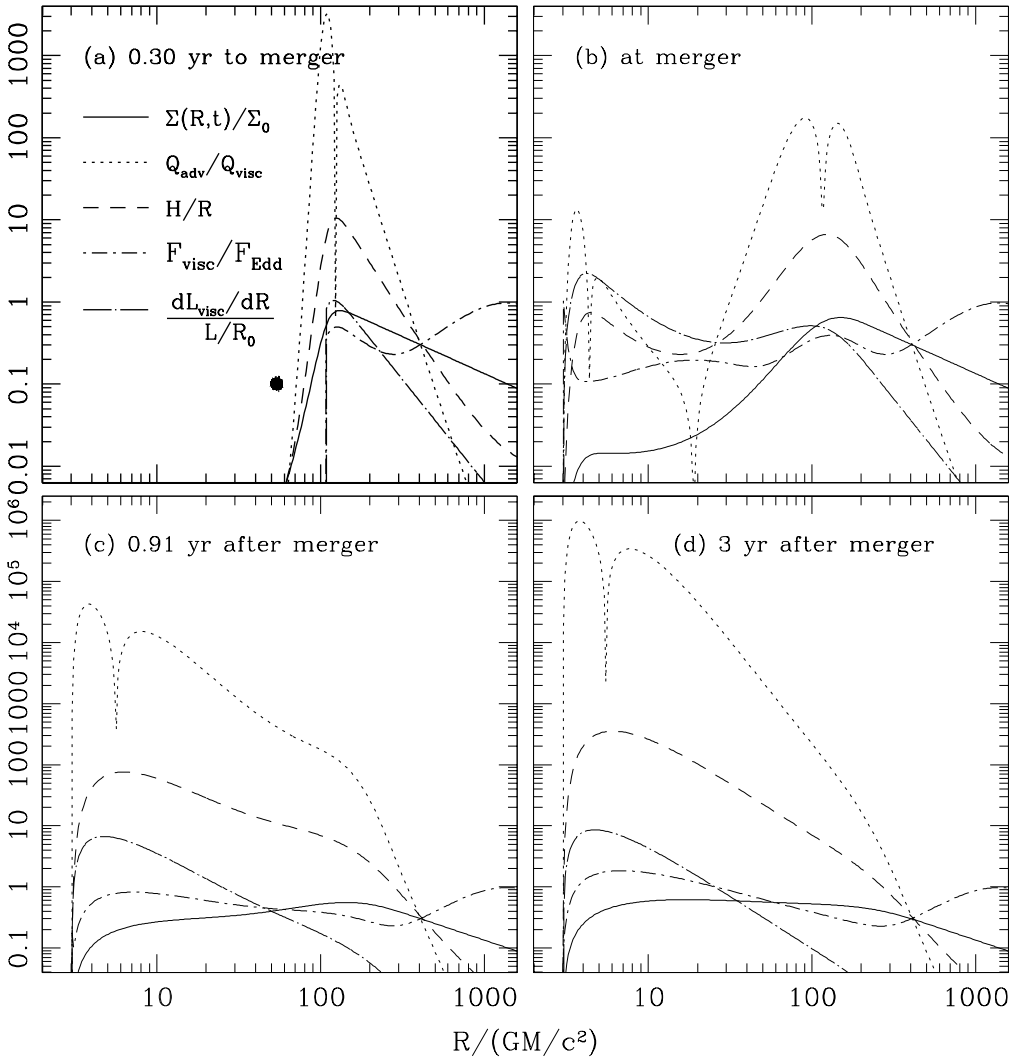

Figure 3.7 Same as Figure 3.5, but for a more massive and geometrically thicker disk with $S = 5$ and $\zeta = 1/3$. The disk is geometrically thick and advection-dominated in the inner regions of interest, from decoupling to several years after the merger. The local flux also exceeds the Eddington limit at late times (Panel d).

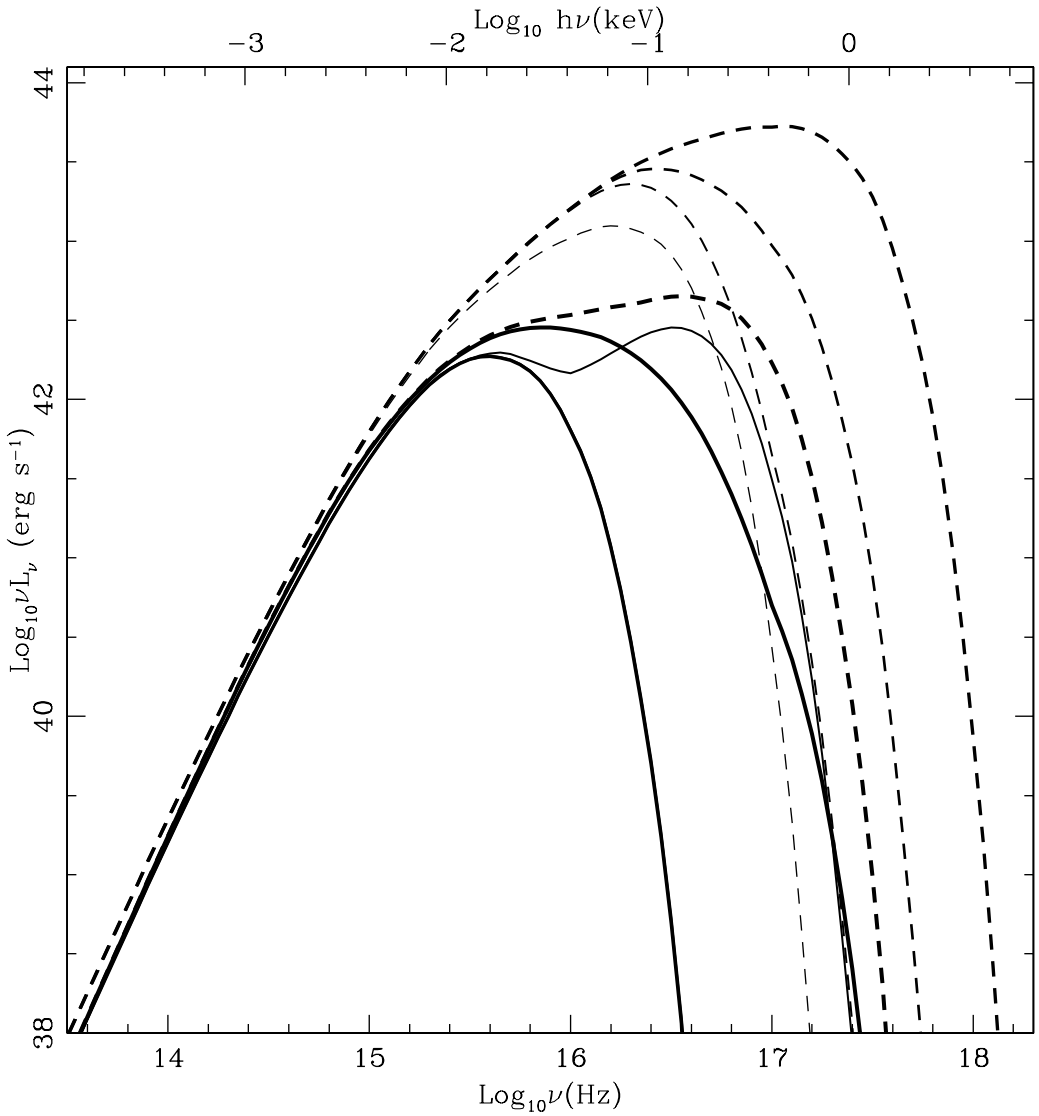

Figure 3.8 The spectra of a more massive and geometrically thick disk, with $S = 5$ and $\zeta = 1/3$, at decoupling and at 1 week, 3 months, 0.91 yr = $t_{\rm EM}$ and 3 yr after the merger. As in Figure 3.6, later spectra are shown with thicker lines. The dashed lines show spectra ignoring advective effects, while the solid lines show spectra under the assumption that advection suppresses the emitted flux. In the latter case, the disk is geometrically thick and radiatively inefficient across the entire region inside the initial gap radius, and thus no obvious afterglow may be observable (all curves overlap except the one 1 week after merger, when our calculations show a brief afterglow that is quickly suppressed by advection).

3.4 Conclusion

We have presented a simple semianalytic model for the viscous evolution of a thin circumbinary disk around a SMBH binary, in the final stages of the binary's evolution. Using this model, we have estimated the time dependence of the thermal spectrum immediately after the binary merger. In what may be the most optimistic scenario, a rapidly evolving soft X-ray signature may be observed soon after the GW event, perhaps years earlier than previously estimated. It is worth emphasizing that the bolometric luminosity is several orders of magnitude brighter than many of the EM counterpart candidates proposed in the literature, and in extreme cases may exceed the Eddington luminosity of the central SMBH. An important feature of the afterglow mechanism discussed in this paper is that it is not strongly limited by the mass of the circumbinary disk. This is in stark contrast to recoil-powered afterglows whose luminosities and observational prospects are generally limited by the mass of the gas bound to the recoiling remnant (e.g., Bode & Phinney 2007; Schnittman & Krolik 2008; Lippai et al. 2008; Shields & Bonning 2008; O'Neill et al. 2009; Megevand et al. 2009; Corrales et al. 2010; Rossi et al. 2010; Anderson et al. 2010).

The reason why an afterglow may be observable so early is because the power per unit area dissipated by viscous dissipation scales as $F \propto \nu\Sigma\Omega^2$. The potential is sufficiently deep close to the SMBH remnant that in this region even low surface densities can generate more power than elsewhere in the disk. A central conclusion of this paper is that enough gas may be able to viscously follow the binary to merger. The presence of gas in the vicinity of the binary at merger may have significant implications for other proposed EM counterpart mechanisms, as well. Some of this gas can reasonably be expected to

contribute to circumprimary or circumsecondary accretion disks which could generate additional observable signatures even if the total gas mass captured in this way is small (Armitage & Natarajan 2005; Lodato et al. 2009; Chang et al. 2010).

In addition to the possibility that the accretion afterglow proposed by MP05 may be observable earlier than previously estimated, our results raise interesting possibilities about the observable features of the post-merger disk at later times. Most of the luminosity is generated in the innermost disk ($R \lesssim 20GM/c^2$), and may be partially reprocessed to IR/optical frequencies by the outer disk ($R \gtrsim 100GM/c^2$). We have considered two mechanisms that could reasonably reprocess a significant fraction of the inner-disk emission: a geometrical warp in the disk just outside the emitting region; and the steep geometrical flaring of the vertical thickness of the inner disk as it evolves We find that the bolometric luminosity of the time-dependent afterglow can exceed the Eddington luminosity of the binary, without violating the local Eddington flux limit. We conclude that the EM signature — the unobstructed signature in soft X-rays, as well as the reprocessed signature in IR/optical — could become comparable to the Eddington luminosity of the binary.

Our calculations also suggest that as it spreads, the innermost disk may become geometrically thick and thus radiatively inefficient, even if it is neither geometrically thick nor radiatively inefficient at decoupling. If the disk is initially geometrically thin at decoupling, then it may emit an afterglow that is powered by the newly formed deep central potential before becoming geometrically thick near the remnant at later times. In such a scenario, an accretion disk around a recently merged SMBH remnant may provide a unique system where the transition from a radiatively efficient to an inefficient accretion

state can be monitored on humanly tractable timescales.

In the most pessimistic scenario, it seems possible for the disk to become geometrically thick even before the binary merges. In such a situation a disk could behave like an ADAF and it is unclear whether a circumbinary cavity should exist at all. Even if a cavity were kept open until decoupling, advective losses may suppress any observable spectral evolution for the disk emission. As suggested by MP05, horizontal advection could also act to make the disk thinner than we have considered here, as in the "slim disk" models of Abramowicz et al. (1988). A slim disk would remain somewhat radiatively efficient and thus could still exhibit an observable evolution of the spectrum or luminosity. Another possible mechanism to keep the circumbinary disk radiatively efficient is super-Eddington fluxes (Begelman 2002). We find that in the same regions of the parameter space where the disk becomes very geometrically thick, the viscously dissipated flux also becomes very high. The nominally radiated fluxes near the central regions around the SMBH remnant can be super-Eddington for physically reasonable parameter values, and this could also help the disk produce an observable evolving EM signature.

Advection-dominated accretion or a super-Eddington flow may produce powerful outflows (e.g., Blandford & Begelman 1999). The outright viscous dissipation of super-Eddington fluxes near the center could also result in a strong outflow. The spin of the SMBH remnant and its orientation are likely to be well constrained from the preceding *LISA* observation of the merger GWs. As such, the observation of a jet near a recently merged SMBH binary would present an unprecedented opportunity to study a SMBH-powered jet where the spin of the central engine is precisely and independently

constrained.

We conclude that the observation of the EM accretion afterglow of a SMBH merger is likely to provide a windfall of empirical constraints on the physics of gas accretion onto SMBHs. Such an observational opportunity would be unprecedented on three points. First, the "initial condition" for the structure of the evolving flow may be relatively well characterized thanks to models for circumbinary disks and the theoretical understanding of the orbital evolution of the binary leading up to the merger. Second, the accretion flow is likely to evolve on humanly tractable timescales. Such an accretion signature will act as a probe of the viscosity in the flow, in contrast with a steady α-disk in which the emission is independent of the α parameter. Third, the mass, spin and orientation of the central SMBH will have been independently determined by *LISA*. The last point is important in terms of emission geometry, energetics, and particularly significant if the accretion flow fuels an outflow or a jet as it fills the central cavity.

In this work, we have used idealized models with a simple radial power-law prescription for the gas viscosity. However, in the α-viscosity model the viscosity is also a function of surface density and disk thickness. In light of the large range of Σ and H/R values found in the regimes of interest in our solutions, it would not be surprising if a more realistic viscosity prescription led to significantly different results from those presented here. Even in our highly idealized calculations, the richness of the physical problem that these systems represent is readily apparent. Given advection, radiation-pressure dominance, super-Eddington fluxes, issues of disk stability and geometry, and rapid viscous evolution, the circumbinary accretion flows around a merging SMBH binary are likely to

produce intriguing observational signatures that would serve as unprecedented probes of fundamental astrophysical processes. The interesting possibilities raised by our simple models underscore the need for more detailed investigations of this interesting class of objects.

APPENDICES

3.A Properties of the Circumbinary Disk After Decoupling

We review below the properties of the inner circumbinary disk at the time of decoupling. Decoupling is defined as the time when $t_{\rm GW}$, the timescale on which the binary orbit shrinks, becomes shorter than $\beta t_{\rm visc}$, the timescale on which the central cavity refills due to viscous spreading of the disk. The procedure is as follows. First, the decoupling condition $t_{\rm GW} \sim \beta t_{\rm visc}$ relates the gas properties at the inner edge to the orbital parameters of the binary in terms of the disk midplane temperature. Second, the midplane temperature is calculated by making simple assumptions about the thermal structure of the disk. MP05 performed these calculations for the inner edge of the disk; we follow their approach closely and reproduce their results while pointing out a few differences. We also discuss possible ways to set the radial power n of the prescribed viscosity in our time-dependent solutions.

Assumptions underlying the model are detailed in the body of this paper and in MP05. Notation for physical quantities is given in our §3.2. Fiducial parameter values are

$M = 10^6 M_\odot M_6$, $\zeta \equiv 4q/(1+q)^2 = 1$, $\alpha = 0.1\alpha_{-1}$, $\beta = 0.1\beta_{-1}$, $\lambda = 1$, $S = 1$, and $\theta_{0.2} = 1$. Unless otherwise specified, the subscript "0" refers to the value at decoupling when applied to binary properties; when applied to gas properties, it refers to the value at decoupling time and at the inner edge of the disk.

A1. Properties at decoupling as functions of midplane temperature

For circular orbits with semimajor axis a, the timescale of gravitational-wave-driven decay of the binary separation is given by:

$$t_{\rm GW} \equiv \frac{a}{da/dt} = \frac{5}{16}\frac{c^5}{G^3M^3}a^4\zeta^{-1}. \tag{3.26}$$

The viscous time is

$$t_{\rm visc} = \frac{2}{3}\frac{R^2}{\nu} = \frac{R^2\Omega\mu m_{\rm P}}{\alpha kT} \approx 144\ {\rm yr} \times \alpha_{-1}^{-1}\lambda^{1/2}T_6^{-1}\left(\frac{a}{100GM/c^2}\right)^{1/2}, \tag{3.27}$$

where $T = 10^6 {\rm K}\ T_6$ is the midplane temperature.

Applying the decoupling condition $t_{\rm GW,0} \sim \beta t_{\rm visc,0}$ gives the following quantities in terms of the midplane temperature at the inner edge, $T_{0,6}$:

$$a_0 \approx 136\frac{GM}{c^2} \times T_{0,6}^{-2/7}\alpha_{-1}^{-2/7}\lambda^{1/7}\left(\beta_{-1}\zeta\right)^{2/7}, \tag{3.28}$$

$$t_{\rm EM} = \frac{3}{4}\beta t_{\rm visc,0} \approx 13\ {\rm yr} \times T_{0,6}^{-8/7}\alpha_{-1}^{-8/7}\lambda^{4/7}M_6\beta_{-1}^{8/7}\zeta^{1/7}, \tag{3.29}$$

$$\nu_0 \approx 2.0\times10^{17}\ {\rm cm^2\ s^{-1}} \times T_{0,6}^{4/7}\alpha_{-1}^{4/7}\lambda^{12/7}M_6\left(\beta_{-1}\zeta\right)^{3/7}, \tag{3.30}$$

$$\Omega_0 \approx 4.5\times10^{-5}\ {\rm s^{-1}} \times T_{0,6}^{3/7}\alpha_{-1}^{3/7}\lambda^{-12/7}M_6^{-1}\left(\beta_{-1}\zeta\right)^{-3/7}. \tag{3.31}$$

We parametrize the surface density via the arbitrary relation $\Sigma_0 \equiv S\dot{M}_{\rm Edd}/(3\pi\nu_0)$, where the parameter S may be interpreted as the product of the outer mass supply rate in Eddington units and the excess in surface density near the inner edge over the value expected for a standard Shakura-Sunyaev disk, which arises from mass accumulation near the binary (Ivanov et al. 1999; Chang et al. 2010). The latter factor depends on tidal interactions and prior accretion history, and in general is expected to exceed unity. We obtain

$$\Sigma_0 \approx 7.3 \times 10^5 \mathrm{g\ cm^{-2}} \times T_{0,6}^{-4/7} \alpha_{-1}^{-4/7} S \lambda^{-12/7} \left(\beta_{-1}\zeta\right)^{-3/7}. \tag{3.32}$$

The pressure near the edge is dominated by the radiation pressure $P_{\rm rad} = (4\sigma/3c)T^4$, if $T \gtrsim 10^6$K. We calculate the scale-height-to-radius ratio H_0/R_0 and the radiation-to-gas pressure ratio $P_{\rm rad,0}/P_{\rm gas,0}$ at the disk edge:

$$\begin{aligned}
\frac{H_0}{R_0} &\approx \sqrt{\frac{\gamma P_{\rm rad,0} H_0}{\Sigma_0}} \frac{1}{\Omega_0 R_0} = \frac{\gamma P_{\rm rad,0}}{\Sigma_0 \Omega_0^2 R_0} \\
&\approx 0.056 \times T_0^4 S^{-1} \lambda^4 M_6 \beta_{-1} \zeta, \qquad (3.33) \\
P_{\rm gas,0} &= \frac{kT_0\Sigma_0}{\mu m_{\rm p} H_0} = \frac{kT_0\Sigma_0^2\Omega^2}{\gamma\mu m_{\rm p} P_{\rm rad,0}}, \\
\frac{P_{\rm rad,0}}{P_{\rm gas,0}} &= \frac{\gamma\mu m_{\rm p} P_{\rm rad,0}^2}{kT_0\Sigma_0^2\Omega^2} \\
&\approx 57 \times T_{0,6}^{51/7} \alpha_{-1}^{2/7} S^{-2} \lambda^{48/7} M_6^2 \left(\beta_{-1}\zeta\right)^{12/7}. \qquad (3.34)
\end{aligned}$$

Above, $\gamma \approx 4/3$ is the relevant adiabatic index, and $\rho = \Sigma/H$ is the gas density. (Note that in the H/R profiles plotted in Figures 3.5 and 3.7, we solve for the scale height H in terms of the total pressure $P_{\rm gas} + P_{\rm rad}$ and do not assume radiation pressure dominance.)

A2. Calculating the midplane temperature

We now turn to calculating the midplane temperature of the inner edge of the disk at decoupling. The thermal spectrum of the inner gas disk differs from that of a blackbody because scattering is a significant source of opacity. In such a medium, photons travel a shorter effective path before becoming thermalized and the absorption opacity is effectively enhanced as (e.g., Rybicki & Lightman 1986; Blaes 2004):

$$\kappa_{\mathrm{eff},\nu} \approx \kappa_{\mathrm{abs},\nu} \sqrt{1 + \frac{\kappa_{\mathrm{es}}}{\kappa_{\mathrm{abs},\nu}}} \tag{3.35}$$

where $\kappa_{\mathrm{abs},\nu}$ and $\kappa_{\mathrm{es}} \approx 0.40$ cm^{2}g^{-1} are the absorption and scattering opacities, respectively. The second approximation in equation 3.35 is applicable if electron scattering is the dominant source of opacity, which is generally the case for our problem in the spectral range of interest (UV frequencies and higher).

The disk radiates through a photosphere, with photons of different energies being thermalized at different depths. The bottom of the photosphere is defined as the height where the effective optical depth $\tau_{\mathrm{eff},\nu} \approx \rho_{\mathrm{p}} H_{\mathrm{p}} \kappa_{\mathrm{eff},\nu} = 1$, where ρ_{p} is the density there and H_{p} is the scale height of the photosphere. Characterizing H_{p} as the ratio of the sound speed at the bottom of the photosphere to the orbital angular speed leads to

$$\rho_{\mathrm{p}} \approx \frac{3c\Omega^2}{4\gamma\sigma T_{\mathrm{p}}^4 \kappa_{\mathrm{abs},\nu}(\kappa_{\mathrm{abs},\nu} + \kappa_{\mathrm{es}})}, \tag{3.36}$$

where T_{p} is the temperature at the bottom of the photosphere.

The main source of absorption is the bound-free process, which in general is a com-

plicated function of composition and photon frequency due to its dependence on the ionization state of the gas. For consistency and simplicity, we prescribe the absorption opacity in the same way as MP05, so that the frequency dependence has the same functional form as the free-free absorption opacity, viz.

$$\kappa_{\mathrm{abs},\nu} \approx \kappa_{\mathrm{abs,R}}\, \rho_{\mathrm{p}} T_{\mathrm{p}}^{-7/2} [187 f_{\nu}(\xi)]. \tag{3.37}$$

Above, $\xi \equiv h\nu/kT_{\mathrm{p}}$, $f_{\nu}(\xi) \equiv \xi^{-3}(1 - e^{-\xi})$ gives the frequency dependence, and ρ_{p} and T_{p} are in cgs. The coefficient $\kappa_{\mathrm{abs},\nu}$ is scaled so that its Rosseland mean yields the Kramer's opacity scaling at solar metallicity, with $\kappa_{\mathrm{abs,R}} \approx 1.6 \times 10^{24}$ cm^{2}g^{-1} and the factor in square brackets having a Rosseland mean of unity. Substituting for the photospheric density using equation 3.36 gives a cubic equation for the absorption-to-scattering opacity ratio $K_{\nu} \equiv \kappa_{\mathrm{abs},\nu}/\kappa_{\mathrm{es}}$ at the bottom of the photosphere:

$$\kappa_{\mathrm{abs},\nu}^2 \left(1 + \frac{\kappa_{\mathrm{abs},\nu}}{\kappa_{\mathrm{es}}}\right) \approx 140 \frac{c\kappa_{\mathrm{abs,R}}}{\gamma\sigma\kappa_{\mathrm{es}}} \Omega^2 T_{\mathrm{p}}^{-15/2} f_{\nu}(\xi). \tag{3.38}$$

We find that in most cases, the vast majority of the disk emission is emitted from in frequencies and regions that are scattering dominated, i.e. $\kappa_{\mathrm{abs},\nu}/\kappa_{\mathrm{es}} \ll 1$. In this limit, 3.38 may be solved trivially to obtain $\kappa_{\mathrm{abs},\nu} \approx \kappa_{\mathrm{abs},*} f_{\nu}^{1/2}$, where we have defined for convenience the frequency-independent quantity $\kappa_{\mathrm{abs},*} \approx 4.7 \times 10^{20}$ cm^{2}g^{-1} $\times (\Omega$ s$)(T_{\mathrm{p}}/\mathrm{K})^{-15/4}$. However, in some of our solutions there is significant emission from regions with $\kappa_{\mathrm{abs},\nu}/\kappa_{\mathrm{es}} \gtrsim 1$. All of the spectra and radial disk profiles in our figures were computed by solving 3.38 generally.

The angle-integrated emergent spectral flux (e.g., Blaes 2004) is

$$F_\nu = \pi \frac{2\epsilon_\nu^{1/2}}{1+\epsilon_\nu^{1/2}} B_\nu, \tag{3.39}$$

where B_ν is the Planck function and $\epsilon_\nu \equiv \kappa_{\mathrm{abs},\nu}/(\kappa_{\mathrm{abs},\nu}+\kappa_{\mathrm{es}}) = (1+K_\nu^{-1})^{-1} \leq 1$ is the frequency-dependent ratio of the absorption opacity to the total opacity. Both ϵ_ν and B_ν are evaluated at the bottom of the photosphere. Each face of the disk contributes half of the total flux F. Integrating equation 3.39, we obtain:

$$\frac{F}{2} = \int_0^\infty F_\nu \, d\nu = \Xi \sigma T_{\mathrm{p}}^4, \tag{3.40}$$

where Ξ is the deviation of the flux from blackbody,

$$\Xi \equiv \frac{15}{\pi^4} \int_0^\infty \frac{2\epsilon_\nu^{1/2}}{1+\epsilon_\nu^{1/2}(\xi)} \frac{e^{-\xi} d\xi}{f_\nu(\xi)}. \tag{3.41}$$

Ξ may be expressed as a function of the frequency-independent quantity $\epsilon_*(T_{\mathrm{p}}, \Omega) \equiv \kappa_{\mathrm{abs},*}/(\kappa_{\mathrm{abs},*}+\kappa_{\mathrm{es}}) \leq 1$. In the limit $\epsilon_\nu^{1/2} \ll 1$, $\Xi \approx 30\pi^{-4}\epsilon_*^{1/2}\int_0^\infty {f_\nu}^{-3/4} e^{-\xi}\, d\xi \approx 0.873\epsilon_*^{1/2}$. In the opposite limit $\epsilon_* \approx 1$ (i.e. very close to blackbody), $\Xi \approx 1/[1+(\epsilon_*^{-1}-1)^{2/3}]$. We find that the brightest regions of our disk generally span the entire range $0 \leq \epsilon_* \leq 1$. We find a useful fitting formula for Ξ to be

$$\Xi \approx \frac{0.873\epsilon_*^{-1/6}}{1-0.127\epsilon_*^{5/6}} \frac{1}{1+\left(\epsilon_*^{-1}-1\right)^{2/3}}. \tag{3.42}$$

The above prescription evaluates Ξ accurately to within $\sim 1\%$ for the entire range of ϵ_*

while reproducing the asymptotic behavior at the extreme limits $\epsilon_* \ll 1$ and $\epsilon_* \approx 1$. We also find that a much simpler formula, $\Xi \approx (4/5)\epsilon_*^{1/2}$, calculates Ξ accurately to within 10% for $0 \leq \epsilon \lesssim 0.9$.

The optical depth between the bottom of the photosphere and the midplane is dominated by electron scattering, and thus has negligible frequency dependence. Provided that this region is optically thick and can be described by a one-zone formalism, the midplane and photosphere temperatures are related by $T_{\rm p}^4 = (4/3)T^4/\tau$, where τ is the optical thickness between the two heights. Following MP05, we write $\tau = \theta\kappa_{\rm es}\Sigma$, where $\theta = \theta_{0.2}/0.2 \leq 1$ is a porosity correction factor. Ignoring horizontal advection, we equate F with the power viscously dissipated per unit area in the disk:

$$F = 2 \times \Xi(\epsilon_*)\frac{4\sigma T^4}{3\tau} \approx \frac{32\sigma T^4 \epsilon_*^{1/2}}{15\tau} \sim \frac{9}{4}\nu\Sigma\Omega^2. \tag{3.43}$$

We now solve for $T_{0,6}$ by substituting equations 3.30, 3.31 and 3.32 into equation 3.43. A useful intermediate result is

$$\epsilon_*^{1/2} \approx \sqrt{\frac{\kappa_{{\rm abs},*}(\Omega, T)}{\kappa_{\rm es}}} \approx 0.19 T_{0,6}^{-27/14} \times \alpha_{-1}^{-3/56} S^{15/32} \lambda^{-93/56} M_6^{-1/2} \left(\beta_{-1}\zeta\right)^{-93/224} \theta_{0.2}^{15/32}, \tag{3.44}$$

which gives $T_{0,6} \propto (\nu_0 \Sigma_0 \Omega_0^2 \epsilon_*^{-1/2} \tau_0)^{14/25}$. That is, the midplane temperature scales with the flux scaling to the +14/25th power, i.e. more sensitively than in the standard black body relation of $T \propto F_{\rm bb}^{1/4}$. We finally arrive at

$$T_{0,6} \approx 1.3 \times \alpha_{-1}^{19/100} S^{343/400} \lambda^{-39/20} M_6^{-7/25} \left(\beta_{-1}\zeta\right)^{-39/80} \theta_{0.2}^{119/400}. \tag{3.45}$$

We derive an inner-edge temperature that is somewhat lower than the value calculated by MP05, primarily because their prescription for the emitted flux is lower than ours by a factor of two. (Compare their expression for F_ν with our equation 3.39. See also our discussion that follows equation 3.13.) As noted above, the midplane temperature is rather sensitive to the flux. We calculate that MP05 underestimated the flux through the omission of the factor of two, and slightly overestimated the frequency integral. Overall, we calculate a bolometric flux 60% higher than they did, and our disk temperature is lower by a factor $1.6^{14/25} \approx 1.3$. We derive identical power-law dependencies on the system parameters as they did.

The lower temperature suggests a slightly longer timescale for the nominal onset of the X-ray afterglow after merger, $t_{\rm EM} \sim (3/4)\beta t_{\rm visc,0} \propto T_0^{-8/7} \sim 9 \times M_6^{1.3}$ yr for fiducial parameters. We also calculate lower values for the quantities H_0/R_0 and $P_{\rm rad,0}/P_{\rm gas,0}$, which are highly sensitive to the disk temperature (equations [3.33] and [3.34]) and thus significantly reduced when the factor of two correction to the flux is included. We summarize our results below, and in our Table 3.1.

$$a_0 \approx 126\frac{GM}{c^2} \times \alpha_{-1}^{-17/50} S^{-49/200} \lambda^{7/10} M_6^{2/25} (\beta_{-1}\zeta)^{17/40} \theta_{0.2}^{-17/200}, \quad (3.46)$$

$$\Omega_0 \approx 5.1 \times 10^{-5}\ {\rm s}^{-1} \times \alpha_{-1}^{51/100} S^{147/400} \lambda^{-51/20} M_6^{-28/25} (\beta_{-1}\zeta)^{-51/80} \theta_{0.2}^{51/400}, \quad (3.47)$$

$$t_{\rm EM} \approx 9.2\ {\rm yr} \times \alpha_{-1}^{-34/25} S^{-49/50} \lambda^{14/5} M_6^{33/25} \beta_{-1}^{17/10} \zeta^{7/10} \theta_{0.2}^{-17/50}, \quad (3.48)$$

$$\Sigma_0 \approx 6.2 \times 10^5 {\rm g\ cm^{-2}} \times \alpha_{-1}^{-17/25} S^{51/100} \lambda^{-3/5} M_6^{4/25} (\beta_{-1}\zeta)^{-3/20} \theta_{0.2}^{-17/100}, \quad (3.49)$$

$$\frac{H_0}{R_0} \approx 0.17 \times \alpha_{-1}^{19/25} S^{243/100} \lambda^{-19/5} M_6^{-3/25} (\beta_{-1}\zeta)^{-19/20} \theta_{0.2}^{119/100}, \quad (3.50)$$

$$\frac{P_{\rm rad,0}}{P_{\rm gas,0}} \approx 430 \times \alpha_{-1}^{167/100} S^{1699/400} \lambda^{-147/20} M_6^{-1/25} (\beta_{-1}\zeta)^{-147/80} \theta_{0.2}^{867/400}, \quad (3.51)$$

$$\epsilon_*^{1/2} \approx 0.11 \times \alpha_{-1}^{-21/50} S^{-237/200} \lambda^{21/10} M_6^{1/25} (\beta_{-1}\zeta)^{21/40} \theta_{0.2}^{-21/200}. \quad (3.52)$$

We calculate a somewhat shorter scale height for the disk than MP05, who had derived $H_0/R_0 \sim 0.46$ for fiducial parameters. We show in §3.3.4 that the advected flux at decoupling is not likely to be significant in the circumbinary disk (for $S = 1$).

A3. Prescribing a value for the viscosity power-law index n

In our model we prescribe a simple radial power-law $\nu \propto R^n$ for the viscosity. In the $\alpha_{\rm gas}$-disk model, however, ν has a physical definition with $\nu \propto \alpha P_{\rm gas}/(\rho\Omega) \propto T/\Omega$. This suggests that a reasonable value for n should satisfy the relation

$$n \approx \frac{d\ln T}{d\ln r} + 3/2. \quad (3.53)$$

Below, we apply the above relationship between ν and T to evaluate an appropriate value for our viscosity power-law index n.

For fiducial parameters, near the disk edge at decoupling and inside this radius thereafter, $\Xi \approx (4/5)\epsilon_*^{1/2}$ and $\epsilon_* \approx \kappa_{\rm abs,*}/\kappa_{\rm es} \propto \Omega T_{\rm p}^{-15/4}$. Applying these approximations to equation 3.43 leads to $T_{\rm p}^{17/8} \propto \nu\Sigma\Omega^{3/2}$. Since $T \propto \tau^{1/4} T_{\rm p} \propto \Sigma^{1/4} T_{\rm p}$, we obtain

$$T(r,t) \propto \left[\frac{\Sigma(r,t)}{\Sigma_0}\right]^{49/68} r^{(8n-18)/17}. \quad (3.54)$$

One way of characterizing the accretion flow is through the power-law index $m \equiv d\ln(\nu\Sigma)/d\ln R$. Writing $d\ln\Sigma/d\ln R = m - n$ and substituting this into equations 3.53 and

3.54 leads to the solution $n \approx (30 + 49m)/85$. If the disk just outside the inner edge behaves like a steady-state accretion solution with $m \approx 0$, then equation 3.53 suggests that in this region a reasonable value of n is $6/17 \approx 0.35$. If, however, the gas in this region behaves more like a decretion profile with $m \approx -1/2$, then $n = 11/170 \approx 0.065$ would be more appropriate.

The disk surface density profile outside the edge at decoupling, however, is not particularly relevant to the viscous refilling rate of the central cavity. Rather, the pertinent value of n is that where the torque gradient is highest, where the most rapid evolution occurs, i.e. near the inner edge of the disk. However, in this region $d\ln(T/\Omega)/d\ln R > 0$ is a rapidly changing function of time and radius. Thus, there is no single value for n that can fully describe the viscous evolution of an α-disk. At late times ($t \gtrsim t_{\rm visc}(R)$ after the merger), however, the inner disk evolves toward a standard quasi-steady accretion solution with $\nu\Sigma \propto (1 - \sqrt{R_*/R})$, where R_* is the inner boundary radius imposed by the SMBH remnant. In this limit, equation 3.53 leads to $n \approx 6/17 + 49/(170/\sqrt{R/R_* - 1})$, which evaluates to $n \approx 0.4$ at radii $R \lesssim R_0$ for $1 \leq R_* c^2/GM \leq 6$. We thus select $n = 0.4$ as our fiducial viscosity power-law index, as it is consistent with the asymptotic evolution of the inner disk at late times.

Upon experimenting with several values for n, we have found that the choice of n does not qualitatively affect the main conclusions of our study, as long as $n \lesssim 1$. The viscous evolution is driven by the single timescale $\tau \propto t_{\rm visc}/(1 - n/2)^2$, and the bolometric and monochromatic luminosities also have weak direct dependencies on n.

3.B Green's Function for the Viscous Evolution of the Disk Surface Density

Equation 3.3, a second-order diffusion equation, is linear if ν does not depend on Σ. Below, we follow the formalism of Lynden-Bell & Pringle (1974; see also Ogilvie 2005) to derive a Green's function solution for the viscous evolution.

Suppose the solution has the exponentially decaying form $\Sigma(R,t) = \exp(-\Lambda t)R^p\varsigma(R)$, where $\varsigma(R)$ is an arbitrary function of radius and p is an arbitrary real number. With the additional assumption that ν behaves as a power-law, $\nu = \nu_0(R/R_0)^n$, equation 3.3 may be rewritten as a modified Bessel equation:

$$R^2\frac{\partial^2\varsigma}{\partial R^2} + \left(2p + 2n + \frac{3}{2}\right)R\frac{\partial\varsigma}{\partial R} + \left[(p+n)\left(\frac{\Lambda}{3\nu_0}R^{2-n} + p + n + \frac{1}{2}\right)\right]\varsigma = 0. \tag{3.55}$$

Selecting $p = n - 1/4$ for convenience and making the substitution $\Lambda = 3\nu_0 k^2 R_0^{-n}$, the equation above has a set of solutions

$$\varsigma_k(R) = R^{-2n}\left[A(k)J_{1/(4-2n)}\left(\frac{kR^{1-n/2}}{1-n/2}\right) + B(k)Y_{1/(4-2n)}\left(\frac{kR^{1-n/2}}{1-n/2}\right)\right], \tag{3.56}$$

where $n < 2$, J and Y are the ordinary Bessel functions of the first and second kind, respectively, $k > 0$ is their mode, and $A(k)$ and $B(k)$ are arbitrary coefficients for each mode.

We note here that even with a simple prescription for ν, the general problem of the viscous evolution of a circumbinary disk around a binary is not easily treated analytically.

For a case where the Green's function is particularly simple, Pringle (1991) showed that one may account for the binary torques by applying a zero-torque boundary condition $\partial(\nu\Sigma R^{1/2})/\partial R = 0$ at the gap opening radius. For binaries of interest in this paper, the gap-opening radius itself is a function of time, which greatly adds to the "extreme algebraic complexity" (Pringle 1991) that is generally involved with applying such a boundary condition at nonzero radius. Our problem, however, is greatly simplified by the decoupling condition, which allows us to approximate the early evolution as if the circumbinary disk experiences no torques from the binary. We thus proceed as if the potential is due to a single central point mass, and approximate the gas orbits as being circular and Keplerian. We discuss the justifications and the modest evolutionary effects associated with this model simplification in §3.2.2.

The boundary condition of interest, then, is that the torques vanish at the center. (At late times, it will be necessary to consider the inner boundary condition imposed at finite radius by the central SMBH remnant. We shall address this point shortly.) Because $Y_{1/(4-2n)}(ky)$ diverges as $y \to 0$, our physical solution requires all $B(k)$ to vanish. The general solution is a sum of all possible modes over all $k \geq 0$ weighted by a set of coefficients $A(k)$, viz.:

$$\Sigma(R,t) = R^{-n-1/4} \int_0^\infty A(k) J_{1/(4-2n)}(ky) \exp\left[-3\nu_0 R_0^{-n} k^2 t\right] \, dk, \tag{3.57}$$

where we have defined $y \equiv R^{1-n/2}/(1-n/2)$ for convenience.

The function $A(k)$ can be evaluated through the use of Hankel transforms. For a given

order of the Bessel function ℓ, Hankel transform pairs satisfy:

$$\Phi(R) = \int_0^\infty \phi(k) J_\ell(kR)\, k\, dk \tag{3.58}$$

$$\phi(k) = \int_0^\infty \Phi(R) J_\ell(kR)\, R\, dR \tag{3.59}$$

For our problem, we may construct the Hankel transform pair

$$\begin{aligned}
\Phi(y) &= R^{n+1/4}\Sigma(y,0) = \int_0^\infty \left[k^{-1}A(k)\right] J_{1/(4-2n)}(ky)\, k\, dk \\
\phi(k) &= k^{-1}A(k) = \int_0^\infty \left[R'^{n+1/4}\Sigma(y',0)\right] J_{1/(4-2n)}(ky')\, y'\, dy',
\end{aligned}$$

from which we obtain

$$\begin{aligned}
A(k) &= \int_0^\infty R'^{n+1/4}\Sigma(y',0)\, J_{1/(4-2n)}(ky')\, k\, y'\, dy' \\
&= \left(1-\frac{n}{2}\right)^{-1} \int_0^\infty \Sigma(y',0)\, J_{1/(4-2n)}(ky')\, k\, R'^{5/4+n/2}\, dR'
\end{aligned} \tag{3.60}$$

Combining equations 3.57 and (3.60), we may write the solution $\Sigma(y,t)$ as an integral function of the initial condition $\Sigma(R',0)$,

$$\begin{aligned}
\Sigma(R,t) &= \left(1-\frac{n}{2}\right)^{-1} R^{-n-1/4} \int_0^\infty R'^{5/4+n/2} \int_0^\infty \Sigma(R', t=0)\, \exp\left(-3\nu_0 R_0^{-n} k^2 t\right) \\
&\qquad \times J_{1/(4-2n)}(ky') J_{1/(4-2n)}(ky)\, k\, dk\, dR' \\
&\equiv \int_0^\infty G_0(R,R',t)\Sigma(R',t=0)\frac{dR'}{R_0},
\end{aligned} \tag{3.61}$$

with the Green's function

$$\begin{aligned} G_0 &= \left(1-\frac{n}{2}\right)^{-1} R^{-1/4-n} R'^{5/4} R_0 \int_0^\infty J_{1/(4-2n)}(ky')\, J_{1/(4-2n)}(ky) \exp\left(-3\nu_0 R_0^{-n} k^2 t\right)\, k\, dk \\ &= \frac{R^{-1/4-n} R'^{5/4} R_0^{1+n}}{6(1-n/2)\nu_0 t} I_{1/(4-2n)}\left[\frac{R^{1-n/2} R'^{1-n/2} R_0^n}{6(1-n/2)^2\nu_0 t}\right] \exp\left[-\frac{\left(R^{2-n}+R'^{2-n}+\right) R_0^n}{12(1-n/2)^2\nu_0 t}\right] \\ &= \frac{2-n}{\tau} r^{-1/4-n} r'^{5/4} I_{1/(4-2n)}\left(\frac{2r^{1-n/2} r'^{1-n/2}}{\tau}\right) \exp\left(-\frac{r^{2-n}+r'^{2-n}}{\tau}\right), \end{aligned} \tag{3.62}$$

where I_ℓ is the modified Bessel function of the first kind and order ℓ, and we have substituted the dimensionless variables $r \equiv R/R_0$ and $\tau \equiv 8(1-n/2)^2 t/t_{\rm visc,0}$.

In deriving the Green's function above, we have applied the zero-torque boundary condition at the origin, instead of at some finite boundary radius R_*. At late times and small radii, $\tau \gg r^{2-n}$, the solutions obtained using equation 3.62 have the radial dependence $\partial \ln(\nu\Sigma)/\partial \ln R \approx (n-2) r^{2-n}/\tau$, viz. they converge to a quasi-steady profile with $\nu\Sigma \propto \Sigma r^n$ constant in radius. Thus, Σ diverges at the center at late times if n is positive.

A physically plausible model should account for the boundary condition imposed by the SMBH remnant at late times. Below, we approximate the Green's function for the case where the zero-torque boundary condition is applied at a finite boundary radius $R_* > 0$. In this case the boundary condition determines the relationship between the $A(k)$ and $B(k)$ coefficients in equation 3.56, viz:

$$\varsigma_k(R) = R^{-2n} A(k) \left[J_{1/(4-2n)}(ky) - \frac{J_{1/(4-2n)}(ky_*)}{Y_{1/(4-2n)}(ky_*)} Y_{1/(4-2n)}(ky)\right], \tag{3.63}$$

where $y_* = R_*^{1-n/2}/(1-n/2)$.

Our new Green's function is then

$$G = R^{-1/4-n} R'^{5/4+n/2} \int_0^\infty J_{1/(4-2n)}(ky)\, J_{1/(4-2n)}(ky') \exp\left[-\frac{R_0^{2-n}k^2}{4\,(1-n/2)^2}\tau\right] \times\left[1 - \frac{J_{1/(4-2n)}(ky_*)}{Y_{1/(4-2n)}(ky_*)}\,\frac{Y_{1/(4-2n)}(ky)}{J_{1/(4-2n)}(ky)}\right] k\, dk. \quad (3.64)$$

The integrand above is greatest where k is small, and decays rapidly as k increases beyond $k^2 \sim 4R_0^{n-2}\tau^{-1}$. The boundary condition, however, has a significant effect on the viscous evolution only where $\tau \gg r^{2-n} = y^2 R_0^{n-2}$ and $y \gtrsim y_*$. Thus, the boundary effect is appreciable only where $k^2 \ll 4y_*^{-2} \gtrsim 4y^{-2}$. In the limit of small argument $J_\ell(x)/Y_\ell(x) \propto x^{2m}$, and we obtain

$$\frac{J_{1/(4-2n)}(ky_*)}{Y_{1/(4-2n)}(ky_*)}\,\frac{Y_{1/(4-2n)}(ky)}{J_{1/(4-2n)}(ky)} \approx \left(\frac{y_*}{y}\right)^{1/(2-n)} = \sqrt{\frac{R_*}{R}}. \quad (3.65)$$

To leading order, then, at late times and small radii the factor in square brackets in equation 3.64 evaluates to $1 - \sqrt{R_*/R}$. This factor may be taken outside of the integral, and gives the asymptotic behavior $\nu\Sigma \propto 1 - \sqrt{R_*/R}$, which is precisely the standard solution for a steady thin accretion disk with a no-torque inner boundary at R_* (e.g., Frank et al. 2002).

We do not evaluate the exact Green's function solution at early times or at large radii, where the effect of the boundary condition is minimal. Instead, we approximate the function by ensuring the correct boundary behavior at late times and small radii, as suggested by Lynden-Bell & Pringle (1974). Accounting for the fact that the boundary at R_* exists only after the time of merger, which evaluates via the decoupling condition to

$\tau_{\text{merge}} = 2\beta(1 - n/2)^2$, we approximate our new Green's function as follows:

$$G(r, r', \tau) = \left\{1 - \sqrt{\frac{r_*}{r}} \exp\left[-\frac{(r - r_*)^{2-n}}{\mathcal{R}(\tau - \tau_{\text{merge}})}\right]\right\} G_0(r, r', \tau), \tag{3.66}$$

where G_0 is the Green's function given in equation 3.62 for the case $R_* = 0$, $r_* \equiv R_*/R_0$, and $\mathcal{R}$ is the ramp function, defined as $\mathcal{R}(x \leq 0) = 0$ and $\mathcal{R}(x > 0) = x$.

Chapter 4

Witnessing the Birth of a Quasar[1]

4.1 Introduction

Observational evidence robustly indicates that all or nearly all galaxies harbor a supermassive black hole in their nucleus (SMBH; e.g., Magorrian et al. 1998). Since cosmological structure formation models predict a hierarchy of galaxy mergers, if nuclear SMBHs were indeed common at earlier times, then these mergers should result in the formation of SMBH binaries (SMBHBs; Begelman et al. 1980), and these binaries should then be common throughout cosmic time (Haehnelt 1994; Menou et al. 2001; Volonteri et al. 2003a; Wyithe & Loeb 2003a; Sesana et al. 2007b; Lippai et al. 2009; Tanaka & Haiman 2009).

It has also long been known, both observationally (e.g., Sanders et al. 1988) and theoretically (e.g., Barnes & Hernquist 1991) that galaxy mergers can drive gas to the nucleus of the merger remnant, which could facilitate the merger of the nuclear SMBHs

[1]This chapter is a reformatted version of an article by the same name by T. Tanaka, Z. Haiman and K. Menou that can be found in The Astronomical Journal, Volume 140, Issue 2, pp. 642-651. The abstract for this paper is reproduced in Section 1.6.3.

on one hand, while also providing fuel for quasar activity on the other. Mergers are therefore generically also believed to trigger quasar activity; the rate of major galaxy mergers can indeed provide an explanation for the observed evolution of the quasar population as a whole (Carlberg 1990; for more recent work, see, e.g., Hopkins et al. 2007a and Wyithe & Loeb 2009 and references therein).

Despite their expected ubiquity, observational evidence for SMBHBs is scarce, and the precise timing of any quasar activity, and when it occurs relative to the merger of the nuclear SMBH binary, remains unclear (Kocsis et al. 2006). A handful of pairs of active SMBHs in the same galaxy have been resolved directly, at ~kpc separation in X–ray (Komossa et al. 2003) and optical (Comerford et al. 2009b) images, and at ~10pc separation in the radio (Rodriguez et al. 2006), confirming that gas is present around the SMBH binary, and that quasar activity can, at least in some systems, commence prior to their coalescence. However, there has been at least one suggestion that luminous activity can be occurring later, at the time of the merger, as well – momentarily interrupted by the coalescence of the SMBHs and reactivated after-wards (Liu et al. 2003b). While there are many more observed SMBHB candidates with small separations (e.g., Roos et al. 1993; Schoenmakers et al. 2000; Merritt & Ekers 2002; Sudou et al. 2003; Liu 2004; Boroson & Lauer 2009), the evidence for these tighter binaries is indirect, and each candidate system has alternative explanations. The expectation is that at large separations, the binaries rapidly lose orbital angular momentum through dynamical friction with background stars and through tidal–viscous interaction with the surrounding gas (Ivanov et al. 1999; Armitage & Natarajan 2002; Escala et al. 2005; Merritt & Milosavljević 2005; Dotti et al. 2007; Sesana et al. 2007a;

Cuadra et al. 2009; Callegari et al. 2009; Colpi et al. 2009; Haiman et al. 2009a; Chang et al. 2010). Once sufficiently compact, gravitational wave (GW) emission rapidly shrinks the orbit, culminating in a merger. How long this process lasts, and at what stage(s) the SMBHs light up as luminous quasars, is, however, also poorly understood theoretically.

Apart from the cosmological context, interest in EM signatures of SMBH mergers surged recently (e.g., Milosavljević & Phinney 2005; Bode & Phinney 2007; Lippai et al. 2008; Schnittman & Krolik 2008; Shields & Bonning 2008; O'Neill et al. 2009; Chang et al. 2010; Megevand et al. 2009; Corrales et al. 2010; Rossi et al. 2010; Anderson et al. 2010; Krolik 2010; Tanaka & Menou 2010; Shapiro 2010), driven by (i) the prospect that the *Laser Interferometer Space Antennae* (*LISA*) will detect the mergers in GWs and provide a tractable list of (perhaps as low as a few hundred; e.g., Kocsis et al. 2008) EM candidates for SMBHBs and (ii) the breakthrough in numerical general relativistic calculations of BH mergers (e.g., Pretorius 2005; Campanelli et al. 2006; Baker et al. 2006a), which led to robust predictions of significant mass–loss and recoil that can significantly perturb the ambient gas. A simultaneous observation of the merger in gravitational and EM waves would enable new scientific investigations in cosmology and BH accretion physics (Cutler 1998; Holz & Hughes 2005; Kocsis et al. 2006, 2007; Lang & Hughes 2008; Phinney 2009; Bloom et al. 2009).

In this paper, we focus on one particular signature of SMBHB coalescence, which we will hereafter refer to as the "viscous afterglow". The physics of this model was discussed by Liu et al. (2003b) in the context of the interruption of jets in double-double radio galaxies, and later by Milosavljević & Phinney (2005; hereafter MP05) in the context

of EM counterparts of *LISA* sources. Prior to merger, the SMBHB torques open and maintain a cavity in the center of a thin circumbinary gas disk (Artymowicz & Lubow 1994). When the binary becomes sufficiently compact, GW emission causes the binary orbit to shrink faster than the gas just outside the cavity can viscously respond. The merger takes place inside the cavity, which is subsequently filled as the disk viscously spreads inward. Because the refilling inner disk produces higher-energy photons than the outer regions, the disk is predicted to transition from an X-ray-dim state to an X-ray-bright one, with its bolometric luminosity increasing by a factor of ~ 10 during this time. This transition is expected to take place on humanly tractable timescales, with the cavity filling in $\sim 10(1+z)(M/10^6 M_\odot)^{1.3}$ yr, where M is the total mass of the binary. A study of an optically selected sample by Gibson et al. (2008) found X-ray-dim AGN to be rare ($\lesssim 2\%$ at $z \sim 2$), suggesting that it would be tractable to catalog and monitor such systems for possible observational signatures of a merger afterglow.

In the observational scenario originally proposed by MP05, *LISA* would detect the GWs from the merger and determine its approximate location in the sky to within ~ 0.1 deg, triggering a follow–up search to identify the EM counterpart and host galaxy. A natural question to ask, however – and the subject of the present paper – is *whether the viscous afterglows may be sufficiently bright and numerous to be detectable in EM surveys alone, even before LISA is launched*. The identification of mergers by their EM signatures alone could, in fact, be valuable for several reasons. First, *LISA* will be sensitive to GWs from relatively low–mass SMBHBs, with total masses of $\sim (10^4 - 10^7)/(1+z)M_\odot$. EM studies could, in principle, detect coalescing SMBHBs outside this mass range, and

therefore complement the *LISA* binary population. Second, while many models for the cosmological evolution of SMBHs predict that *LISA* will detect dozens or hundreds of mergers (if "seed" black holes are abundant and merge often; e.g., Sesana et al. 2007b), there are some SMBH assembly scenarios that may result in far fewer *LISA* events (i.e. if seeds are rare and grow primarily through rapid accretion or are very massive already at formation; Tanaka & Haiman 2009; Lippai et al. 2009). It is therefore plausible that EM surveys could deliver a larger SMBH binary sample than available from GWs. Third, several transient EM surveys are already under way, or are planned to be completed before the expected launch date of the *LISA* mission around 2020.

If luminous quasar activity is triggered by major mergers of galaxies, as argued above, then the viscous afterglow could plausibly be interpreted as the signature of the birth of a quasar. In this paper, we estimate the number of identifiable afterglow sources, i.e. birthing quasars, in the sky, by (i) adopting an idealized time-dependent model (Tanaka & Menou 2010, hereafter TM10) of the evolution of the disk structure, to calculate photometric light curve and variability of the afterglow, and (ii) by using the observed luminosity function of quasars as a proxy for the SMBHB merger rate. Our two main goals are:

1. To assess whether there is any hope of detecting and identifying the viscous afterglows with conventional EM telescopes alone.

2. To see how the identifiability of the afterglows depends on theoretical parameters and to delineate the ideal survey attributes (wavelength, angular coverage and depth). We compare the derived attributes to those similar to planned large sur-

veys of the transient sky: a soft-X-ray survey with specs similar to those that were proposed recently, unsuccessfully, for the the *Lobster-Eye Wide-Field X-ray Telescope* [2] (*LWFT*) mission; and the *Large Synoptic Survey Telescope*[3] (*LSST*) in the optical.

We find that the detectability of the afterglow is sensitive to the properties of the circumbinary disk, in particular to the ratio of the viscous stress to the gas pressure, and to the surface density of the disk. We conclude that purely EM identification of the afterglows by the planned surveys are unlikely, unless the surface density and the viscosity in the circumbinary disk are at the high end of the expected range. In this latter, optimistic scenario, several dozen birthing quasars could be identified in a soft X-ray transient survey. We also find that if $\gtrsim 1\%$ of the X-ray radiation emitted in the central regions is reprocessed into the optical frequencies by dust surrounding the source, or by warps or geometric splaying in the disk itself (TM10), several dozen afterglows could be detected in an optical transient survey, such as *LSST*.

This paper is organized as follows. In § 2, we summarize the viscous afterglow model, and describe our methods for estimating the identifiable population of AGN harboring a recently merged SMBHB. In § 3, we present estimates for the number of identifiable afterglow sources in the sky. We summarize our results and offer our conclusions in § 4.

[2]http://www.star.le.ac.uk/lwft/
[3]http://www.lsst.org/lsst

4.2 A Simple Model for the Afterglow Population

In this section, we describe the model and underlying assumptions used to estimate the number of observable afterglow sources in the sky. We use the idealized Newtonian viscous evolution model described by TM10 to calculate the light–curve and spectral evolution of the source, and to obtain light curves in fixed frequency bands. Then, following Haiman et al. (2009a) we assume that every SMBHB merger ultimately leads to a bright quasar episode. Using the bolometric quasar luminosity function of Hopkins et al. (2007b) as a proxy for the underlying SMBH mass function, we estimate the number of afterglow sources in the sky that exhibit identifiable photometric brightening. [4]

Throughout this paper, $M = M_1 + M_2$, $q = M_1/M_2$ and a denote the binary mass, mass ratio and semi-major axis, respectively. The physical constants G, c, and k_B have their usual meanings.

4.2.1 Modeling Afterglow Light Curves

Below, we briefly recapitulate the main features of the viscous afterglow model, which was advanced by MP05 and elaborated upon by TM10 (see also Shapiro 2010). We refer the reader to those earlier works for further details; a derivation of the disk evolutionary equations, in particular, can be found in Appendix A of TM10.

A SMBHB in a circular orbit of semi-major axis a will open an annular gap in a

[4] As discussed by Haiman et al. (2009a), an alternative method to construct the luminosity function is to calculate the SMBH merger rate from the merger rate of dark matter halos, and apply the post-merger light curve $L(t, M, q)$ to each merger event. That method requires a third ingredient, namely a way to prescribe the SMBH masses from the halo masses. Given the approximate nature of our calculation, we choose the simpler method described in the text.

thin circumbinary disk at a radial distance $R_{\rm wall} \sim 2a$ from the binary's center of mass (Artymowicz et al. 1991). The binary will shrink gradually by depositing orbital angular momentum in the disk, maintaining a nearly self-similar geometry with $R_{\rm wall}/(2a) \sim 1$. The kinematic viscosity $\nu_{\rm gas}$ of the gas comprising the disk is a weak function of radius R, and thus the viscous time $t_{\rm visc}(R) = (2/3)R^2/\nu_{\rm gas}$ roughly scales just outside the cavity as $t_{\rm visc}(R_{\rm wall}) \propto a^2$. Because GW emission shrinks the orbit on a timescale $t_{\rm GW} \equiv (d\ln a/dt)^{-1} \propto a^4$, there exists a value of a inside which the binary orbit begins to shrink faster than the circumbinary gas can viscously follow. This critical binary separation is of the order $a \sim 100GM/c^2$; past this point, the evolution of the binary and the disk are decoupled.

After decoupling, the evolution of the disk surface density Σ can be described by the standard equation for a thin, Keplerian accretion disk (e.g., Frank et al. 2002):

$$\frac{\partial}{\partial t}\Sigma(R,t) = \frac{3}{R}\frac{\partial}{\partial R}\left[R^{1/2}\frac{\partial}{\partial R}\left(\nu\Sigma R^{1/2}\right)\right]. \tag{4.1}$$

In the special case when the dependence of the gas viscosity on radius can be approximated as a power–law, $\nu \propto R^n$, the surface density evolution $\Sigma(R,t)$ can be solved semi-analytically with a Green's function formalism (Lynden-Bell & Pringle 1974, TM10), starting from an arbitrary initial distribution $\Sigma(R, t = 0)$.

The subsequent evolution of the circumbinary disk is sensitive to several properties of the binary-plus-disk system. For a given total binary mass $M \equiv M_6 \times 10^6 M_\odot$, the relevant observable quantities can be expressed in terms the following eight system parameters.

- The parameter $\zeta \equiv 4q(1+q)^{-2} \leq 1$ is the symmetric mass ratio of the binary, scaled

to unity for equal-mass binaries.

- The ratio of the diameters of the circumbinary cavity and the binary separation at decoupling is specified by the parameter $\lambda = R_{\rm wall}/(2a) \sim 1$.

- The gas viscosity is parametrized via the ratio of the viscous stress to the gas pressure, $\alpha = \nu_{\rm gas}\Omega k_{\rm B}T/P_{\rm gas}$ (Shakura & Sunyaev 1973). The choice to scale the viscosity with the gas pressure and not the total pressure is motivated by calculations that suggest that radiation pressure-dominated thin disks may be thermally unstable (Shakura & Sunyaev 1976; Pringle 1976).[5]

- The parameter β gives the ratio at decoupling of the timescale $t_{\rm GW}$ on which the binary shrinks due to GW emission to the viscous time $t_{\rm visc}$ of the gas at the cavity wall. MP05 prescribed $\beta = 0.1$, based on the fact that the gas near the cavity wall has a very steep density gradient and will move on a characteristic time–scale $\sim 0.1 t_{\rm visc}$ after the binary torques vanish. However, the calculations by TM10 suggest that the lower value of $\beta \sim 0.05$ is more appropriate, as it gives a closer approximation to the time elapsing between the merger and when the binary torques cease to influence the gas.

- The disk porosity parameter, θ, prescribes how the optical thickness τ between the disk midplane and the height where the emitted photons are thermalized relates to the surface density: $\tau = \theta\Sigma\sigma_{\rm es}$, where $\sigma_{\rm es}$ is the electron-scattering cross section.

- The viscous evolution model assumes a radial power–law for the kinematic viscosity,

[5]Even if thermally stable (Hirose et al. 2009a), such disks may still be viscously unstable (Lightman & Eardley 1974; Piran 1978; Hirose et al. 2009a).

$\nu_{\rm gas} \propto R^n$. Prior to merger, viscosity in a circumbinary decretion disk may be expected to be a weak function of radius, with $n \lesssim 0.1$; after the merger, $n \sim 0.4$ may be expected in the accretion region of interest (TM10).

- Another parameter of interest is the radial power-law index $m \equiv \partial(\ln \nu\Sigma)/\partial \ln R$ of the disk mass profile at decoupling. The standard steady-state solution for a thin accretion disk around a single central object satisfies $m = 0$, with mass flow $\dot{M} = 3\pi\nu_{\rm gas}\Sigma$ constant with radius. However, at the cavity wall the binary torques *decrete* the gas outward; in this regime, the disk would more likely satisfy $m = -1/2$, with torque density $\nu_{\rm gas}\Sigma R^{1/2} =$ constant (e.g., Pringle 1991). We prescribe initial surface density profiles satisfying

$$\Sigma(R \lesssim R_{\rm wall}) \ll \Sigma(R \gtrsim R_{\rm wall}) \propto R^{-n+m},$$

with a steep exponential drop-off in Σ near $R \approx R_{\rm wall}$ (MacFadyen & Milosavljević 2008; TM10).

- The value of the surface density $\Sigma_{\rm wall}$ at the cavity wall at decoupling depends on how much gas has piled up due to accretion of outer gas and decretion of the gas just inside. We parametrize this value as $\Sigma_{\rm wall} = S\dot{M}_{\rm Edd}/(3\pi\nu)$, where $\dot{M}_{\rm Edd}$ is the accretion rate that would produce a luminosity corresponding to the Eddington limit for the binary mass M, assuming a radiative efficiency of $\eta = 10\%$. The parameter S can be thought of as the product of two quantities: the mass supply rate $\dot{M}$ to the disk in Eddington units, and the enhancement of the disk surface density just

outside the cavity due to mass accumulation. The second factor may be expected to exceed unity.

In addition to the above parameters, the cosmological redshift z of the source will also obviously affect the observed spectra, luminosity, and brightening rates of the sources.

For a given set of parameters, we use the corresponding surface density evolution $\Sigma(R, t)$ to calculate the disk temperature profile at depths where the emitted photons are thermalized, which in turn is used to obtain the monochromatic luminosity in the source rest frame:

$$L_\nu(t) \sim 2 \times \int_{R_{\rm ISCO}}^{10R_{\rm wall}} \pi \frac{2\epsilon_\nu}{1+\epsilon_\nu} B_\nu \, 2\pi R \, dR. \tag{4.2}$$

The leading factor of two on the right hand side of equation 4.2 accounts for the fact that the disk radiates from two faces; B_ν is the Planck function; $\epsilon_\nu \equiv \kappa_{{\rm abs},\nu}/(\kappa_{{\rm abs},\nu} + \kappa_{\rm es})$ is the ratio of the absorption to the total opacity; and the fraction $2\epsilon_\nu/(1+\epsilon_\nu)$ is the deviation of the flux from blackbody (e.g., Rybicki & Lightman 1986; Blaes 2004). The effective temperature in the inner region of the disk is higher than a blackbody disk with the same surface density distribution, and thus it produces a harder spectrum. The lower limit of integration, $R_{\rm ISCO}$, is the radius of the innermost stable circular orbit, for which we adopt the value $3GM/c^2$, consistent with the assumption that the binary remnant has moderate spin, $0.6 \lesssim a_{\rm spin} \lesssim 0.9$. The choice for the upper limit of integration is somewhat arbitrary, and does not significantly affect $L_\nu(t)$ as long as it is sufficiently large; at larger radii the flux is significantly lower, and evolves on much longer timescales than the inner region originally occupied by the cavity.

The overall disk evolution timescale is roughly the viscous time at the cavity wall,

evaluated at decoupling:

$$t_{\rm visc} \sim 120 \text{ yr} \times M_6^{1.32} \zeta^{0.70} \lambda^{2.8} \alpha_{-1}^{-1.36} \beta_{-1}^{0.70} S^{-0.98} \theta_{0.2}^{-0.34}. \tag{4.3}$$

Although the disk continues to brighten and spectrally harden for $\sim t_{\rm visc}$ after the merger, the most dramatic evolution takes place in the first $\sim \beta t_{\rm visc}$, which corresponds to the faster viscous spreading of the sharp density edge at $R \lesssim R_{\rm wall}$.

The time–dependent spectrum of the viscously spreading disk can be divided into two parts: a nearly static low–frequency component produced predominantly by gas outside $R_{\rm wall}$; and a rapidly evolving high–frequency component produced predominantly by the gas flowing into the central cavity. The characteristic frequency that marks the boundary between the static and variable parts of the spectrum can be approximated in the source rest frame (TM10) as

$$h\nu_{\rm var} \sim 30 \text{ eV} \times M_6^{-0.32} \zeta^{-0.45} \lambda^{2.1} \alpha_{-1}^{0.36} \beta_{-1}^{-0.45} S^{0.73} \theta_{0.2}^{0.09}, \tag{4.4}$$

or, in terms of the wavelength,

$$\lambda_{\rm var} \sim 40 \text{ nm} \times M_6^{0.32} \zeta^{0.45} \lambda^{-2.1} \alpha_{-1}^{-0.36} \beta_{-1}^{0.45} S^{-0.73} \theta_{0.2}^{-0.09}. \tag{4.5}$$

Above, we have defined $\alpha_{-1} \equiv \alpha/0.1$, $\beta_{-1} \equiv \beta/0.1$, and $\theta_{0.2} \equiv \theta/0.2$. These equations already reveal that significant brightening will occur primarily at photon energies in the hard UV to X-ray bands. While the characteristic frequency in eq. (4.4) can move into

the optical bands for the most massive SMBHs, the overall evolutionary timescale for these sources, eq. (4.3), becomes long. The most rapid evolution takes place as the cavity fills; once it is filled, the system evolves more gradually to approach a standard steady thin accretion disk solution around a single SMBH, with the quantity νL_ν peaking at a frequency of roughly $\nu_{\rm peak} \sim 15\nu_{\rm var}$. The spectrum falls off steeply at higher frequencies, and is likely unobservable above $\nu \gtrsim 3\nu_{\rm peak}$.

Prior to decoupling, the luminosity of the disk at frequencies $\nu \gtrsim \nu_{\rm var}$ is negligible. Once the cavity is filled, the monochromatic luminosity at frequencies $\nu_{\rm var} \lesssim \nu \lesssim \nu_{\rm peak}$ reaches

$$\nu L_\nu \sim 5 - 30 \times 10^{42}\ {\rm erg\ s^{-1}} M_6^{0.92} \zeta^{-0.42} \lambda^{-1.7} \alpha_{-1}^{0.34} \beta_{-1}^{-0.42} S^{1.2} \theta_{0.2}^{0.08}. \quad (4.6)$$

In order to be identifiable in a survey, an afterglow source must exhibit significant brightening, at least comparable to the typical variability of typical AGN, at luminosities and frequencies to which the survey is sensitive.

4.2.2 Modeling the Population of Afterglow Sources

We now turn to estimating the number of identifiable afterglow sources. The approach described below closely follows that described in Haiman et al. (2009a) for estimating the number of pre-merger sources that may be detectable by their periodic variability. We begin by prescribing the quasar luminosity function as a proxy for the SMBH mass function. Specifically, we adopt the bolometric luminosity function of Hopkins et al. (2007b), and suppose that during a typical bright quasar phase, the luminosity and SMBH mass are related via a simple approximate relation, $L(M) \sim f_{\rm Edd} L_{\rm Edd}(M)$, where $L_{\rm Edd}(M)$

is the Eddington luminosity for an object with mass M, and $f_{\rm Edd}$ is a constant. This is an admittedly rough estimate, as $f_{\rm Edd}$ is known to have a non-negligible spread among the population of bright quasars. However, our simple estimate is sufficient for a proof-of-concept; a more precise calculation is not warranted, given the approximate and highly idealized nature of the afterglow model and the uncertainty in the system parameters. We choose the fiducial values $f_{\rm Edd} \sim S \sim 0.3$ (e.g., Kollmeier et al. 2006) and assume a rest-frame quasar lifetime of $t_{\rm Q} \sim 10^7$ yr (e.g., Martini 2004).

We further assume that there is a one-to-one correspondence between a SMBHB merger and a bright quasar episode, i.e. that a SMBHB merger ultimately triggers quasar activity. This assumption is consistent with our post-merger disk evolution model, which naturally leads to a state with a fully extended disk around a single SMBH, as long as fuel remains available to maintain near-Eddington accretion rate at the outer edge of the disk. Given the comoving number density of AGN $dn_{\rm AGN}/dM$, we are interested in the subset of SMBH merger remnants that are producing an observable, brightening afterglow, and have not yet settled to a later, presumably steadier quasar phase. To estimate this fraction of SMBHs, we use the afterglow model described above, and calculate the duration $t_{\rm ag}$ over which the photometric emission from a circumbinary disk brightens at a rate exceeding some threshold value. This threshold should be chosen to correspond to a brightening rate that is not only measurable, but is also distinguishable from other possible sources of time-variability. The number of variable sources $N_{\rm ag}$ in the sky of such SMBHB remnants

in a given mass and redshift range is then

$$N_{\rm ag} \sim \Delta V(z_{\rm lo}, z_{\rm hi}) \int \frac{dn_{\rm AGN}}{d\ln M} \left[\frac{t_{\rm ag}(M)}{t_{\rm Q}}\right] d\ln M, \tag{4.7}$$

where $\Delta V(z_{\rm lo}, z_{\rm hi})$ is the cosmological comoving volume between redshifts $z_{\rm lo}$ and $z_{\rm hi}$ and $n_{\rm AGN}$ is the space density of SMBHs of mass M, inferred from the quasar luminosity function evaluated at $L(M)$. It is worth emphasizing that this expression does not assume that the birthing quasars have a bolometric luminosity of $L(M)$ – rather, $L(M)$ here represents the characteristic luminosity that is produced by SMBHs of mass M in their quasi–steady quasar state; this asymptotic luminosity is reached only $\gtrsim$ 100 yrs after the merger, according to our afterglow models.

4.3 Results and Discussion

4.3.1 Basic Parameter Dependencies

The dependence of the number of detectable variable sources on the various model parameters for the binary-plus-disk population is of obvious interest, and is not trivial, as each parameter affects differently the luminosity, spectral frequency and brightening rate of the afterglow. For example, increasing the binary mass M increases the source luminosity and lowers the characteristic frequency of the source, while making the afterglow evolve more slowly — thus, the brightening and hardening rates both decrease (making identification more difficult) while the total flux and the total afterglow lifetime both increase (making

a detection easier).

To illustrate how each of the parameters and the source redshift affect the detectability of variable afterglow sources, in Figure 4.1 we first plot the basic quantity $t_{\rm ag,obs} = (1+z)t_{\rm ag}$, representing the amount of time sources are observed to spend at or above the required threshold for the brightening rate. The threshold in this figure is set at $d\ln L_{\rm X}/dt_{\rm obs}$ of at least 10% yr^{-1}, in the soft X–ray frequency window of $0.1 - 3.5$ keV (motivated by the proposed all-sky X-ray transience survey *LWFT*; see below). The solid black curve in each panel shows $t_{\rm ag,obs}$ for the fiducial parameter values $q = \alpha_{-1} = \theta_{0.2} = S = \lambda = 1$, $\beta = 0.05$, $n = 0.4$, $m = -1/2$, and $z = 2$. In each panel, we show how the apparent duration of the rapidly brightening phase is affected by changes (dashed and dotted lines) in one of the system parameters.

In Figure 4.2, we plot the corresponding number $dN_{\rm ag}/d\ln M$ of sources that exhibit a band luminosity $L_{\rm X}$ of at least 10^{40} erg s^{-1} and an observed brightening rate $d\ln L_{\rm X}/dt_{\rm obs}$ of at least 10% yr^{-1} in the same $0.1 - 3.5$ keV frequency window. This is given by the product of the duty cycle $t_{\rm ag,obs}/t_Q$ and the space density of SMBHs (eq. 4.7), except that a further cut is imposed in luminosity. This is because the brightening rate initially may exceed the threshold for SMBHs with masses near the low–mass end of the range shown in the figure, but their band fluxes are still below the imposed threshold; these sources are excluded by subtracting the duration of this initial, sub-luminous state from the duty cycle $t_{\rm ag,obs}/t_Q$. The number is computed for the whole sky, and over a redshift range $1 < z < 3$. The line-style scheme is the same as in Figure 4.1: the solid black curves show $dN_{\rm ag}/d\ln M$ for fiducial parameter values, and the dashed and dotted curves show how

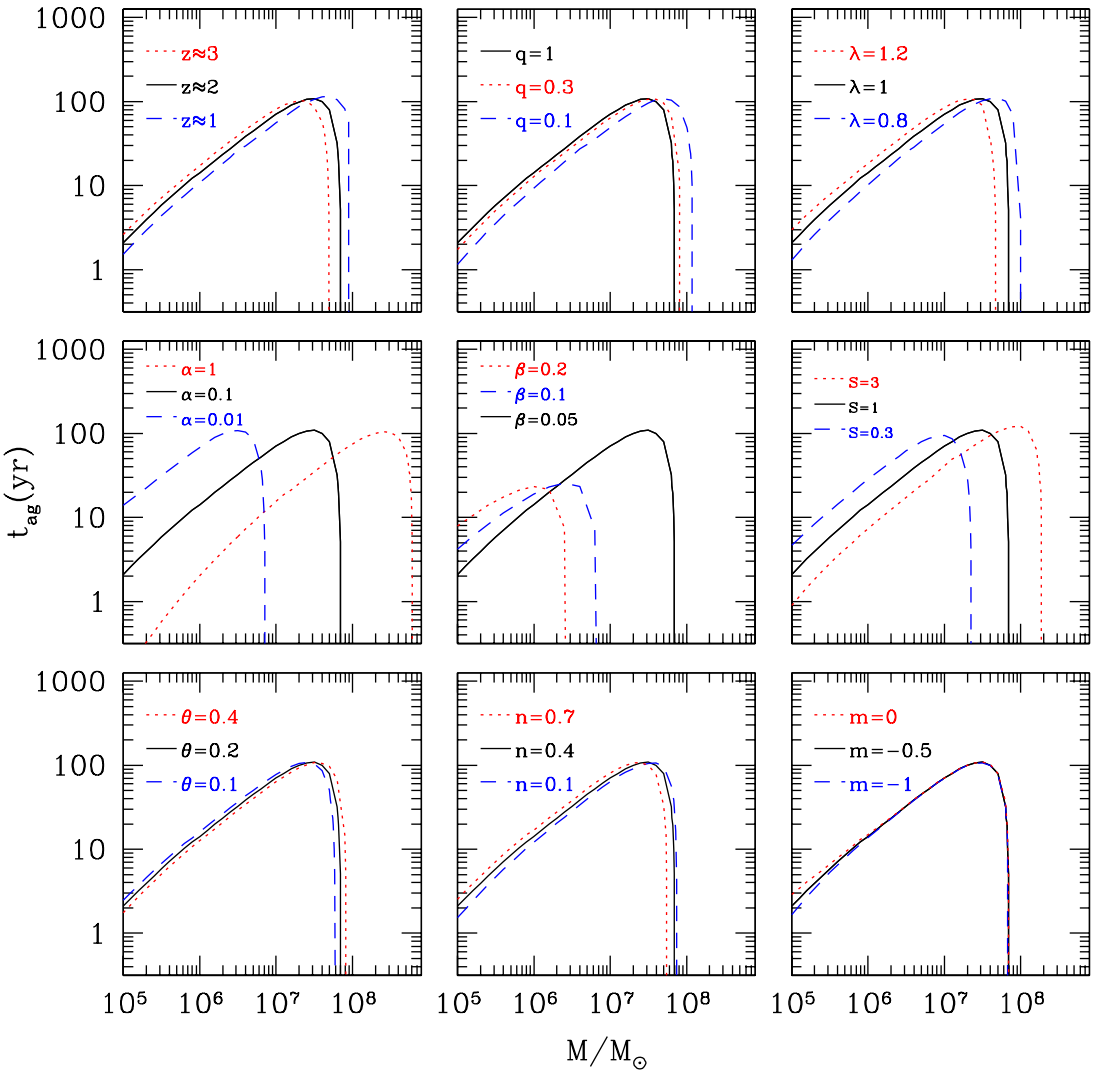

Figure 4.1 The approximate amount of time spent by an afterglow source with an observable band brightening rate of at least 10% yr^{-1} in the 0.1 − 3.5 keV photon energy range. The most luminous such sources spend $t_{\rm ag} \gtrsim 10$ yr in this brightening phase, roughly independently of the model parameter values. The solid curve in each panel shows the fiducial set of parameters, $q = \alpha_{-1} = \lambda = S = \theta_{0.2} = 1, \beta = 0.05, z = 2, n = 0.4$ and $m = -1/2$. The dotted and dashed curves show how the duration of the rapidly brightening phase is affected by changes in each of the model parameters (see § 2.1 for a definition of each parameter).

the number of rapidly brightening sources depends on each parameter. The optimistic luminosity and brightening thresholds in this figure are chosen purely for demonstrative purposes. For reference, a source with $L_X \sim 10^{40}$ erg s^{-1} and $z \sim 1$ would, in fact, have a flux of only $\lesssim 10^{-18}$ erg s^{-1} cm^{-2}, and thus be too faint to be monitored for variability. Also, AGN have been observed to vary in their X-ray brightness by as much as order unity on timescales of years. Although the afterglows in question would exhibit a monotonic increase in X-ray brightness, along with a corresponding monotonic spectral hardening, it is unclear whether a brightening rate of 10% yr^{-1}, even if sustained for several years, and accompanied by a monotonic hardening of the spectrum, would be sufficient to distinguish an afterglow candidate from other X-ray variable sources.

The steep cutoff at high binary masses seen in both t_{ag} and $dN_{ag}/d\ln M$, in Figures 4.1 and 4.2, respectively, has two causes. One is that for sufficiently large M, the emission frequency of the source becomes too low, and falls out of the soft X–ray window. The other reason is that the disk evolution timescale t_{visc} scales as $\propto M^{1.3}$, so that for sufficiently large M the disk evolves so slowly that its brightening rate never reaches the adopted threshold of 10% yr^{-1}.

Figures 4.1 and 4.2 also show that the duration of the brightening phase, and the mass function of the afterglow sources depend strongly only on the parameters α, β and S. This is due to the fact that the afterglow frequency range and evolution timescale scale steeply with these parameters (see equations 4.3 and 4.4). Increasing α, increasing S and decreasing β relative to their fiducial values all have the effect of increasing the afterglow emission frequencies and pushing it further into the survey frequency window, while also

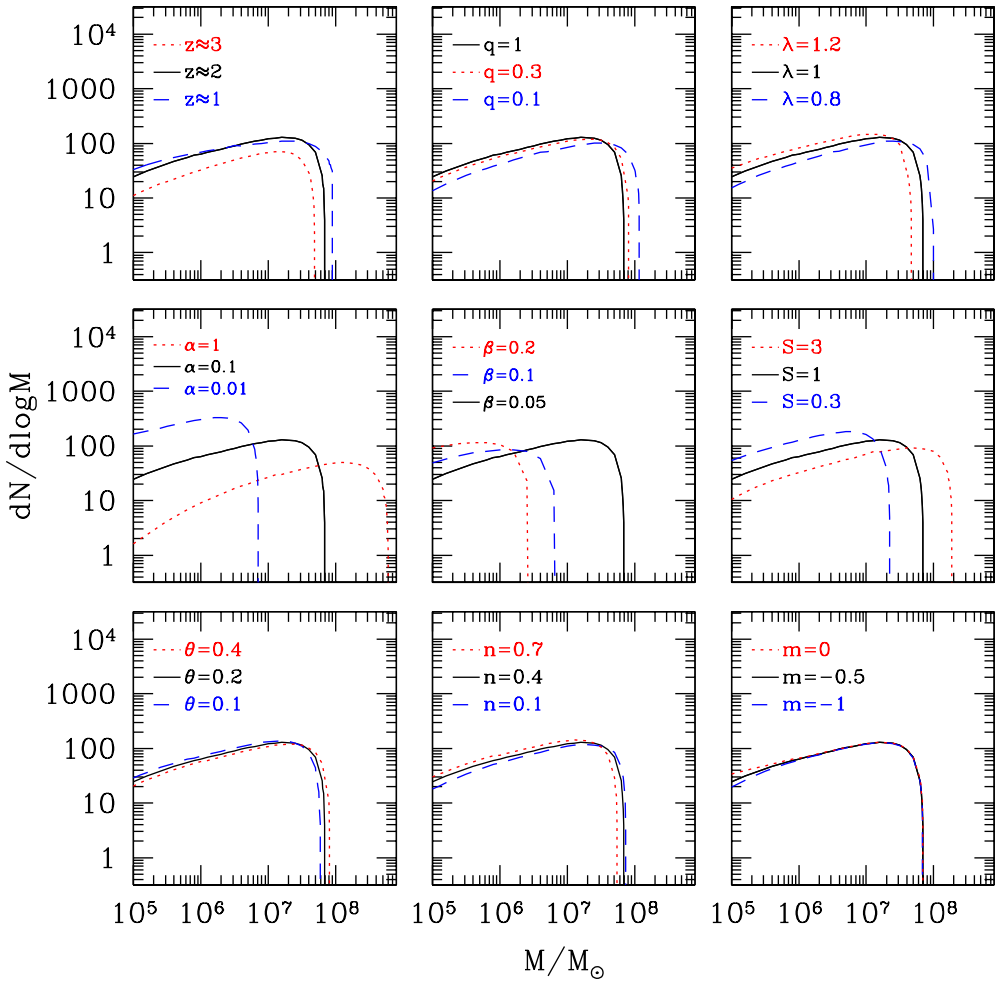

Figure 4.2 The approximate number of afterglow sources in the sky at any given moment whose luminosity in the 0.1 − 3.5 keV energy range is (1) at least 10^{40} erg s^{-1} and (2) increasing by at least 10% yr^{-1}. We use the AGN luminosity function of Hopkins et al. (2007b) as a proxy for the SMBH mass function, and associate each episodic activity of AGN with a SMBH merger whose afterglow light curve is given by the time-dependent model of Tanaka & Menou (2010). The solid curve in each panel shows the fiducial set of parameters, $q = \alpha_{-1} = \lambda = S = \theta_{0.2} = 1$, $\beta = 0.05$, $1.5 < z < 2.5$, $n = 0.4$ and $m = -1/2$. The dotted and dashed curves show how the number of sources are affected by changes in each of the parameters. Note that the most massive sources are not the most luminous in the band of interest (see text). For plausible parameter values, the most luminous sources have masses, band luminosities, and number in the sky of $\gtrsim 10^6 M_\odot$, $L \gtrsim 10^{43}$ erg s^{-1}, and overall number $N_{\rm ag} \lesssim 100$, respectively. The masses of luminous sources of interest coincide with the *LISA* sensitivity window, $10^5 - 10^7 (1+z)^{-1} M_\odot$.

increasing the brightening rate of the afterglow. The quantities $\nu_{\rm var}$ and $t_{\rm visc}$ both depend only weakly on θ, and ζ varies too weakly in the range $0.1 \lesssim q \leq 1$ to have a sizable effect. Increasing (decreasing) the parameter λ results in afterglows that are further inside (outside) the frequency window but evolve much more slowly (quickly) – the two effects tend to cancel out, and yield a relatively weak overall effect on $dN_{\rm ag}/d\ln M$.

It is worth noting that the brightest afterglow sources satisfying a fixed $d\ln L_{\rm X}/dt_{\rm obs} > 10\%$ yr^{-1} are not the most massive ones. This is because disks around more massive BHs evolve more slowly, and the brightening rate is greatest early in the post-decoupling disk evolution when the source is dimmer. For most of the parameter value combinations probed in Figure 4.2 — excepting α, β and S for the moment — the most luminous sources brightening at or above the threshold rate are those with binary masses of $(0.5-2)\times10^6 M_\odot$. Interestingly, this mass range lies in the middle of *LISA*'s sensitivity window.

The maximum band luminosities of these sources are typically $1-4\times10^{43}$ erg s^{-1}, and behave roughly as described in equation 4.6. For $\alpha = 1$ and $S = 3$, the masses and luminosities for the brightest afterglow sources are somewhat greater: $\gtrsim 5\times10^6 M_\odot$ and $\gtrsim 10^{44}$ erg s^{-1}.

The range of luminosities across the parameter combinations probed in Figures 4.1 and 4.2 correspond to an observed soft X-ray flux of $F_{\rm X} \sim 10^{-16} - 10^{-14}\times S^{1.2}$ erg s^{-1} cm^{-2} in the range $1 \lesssim z \lesssim 3$. Thus, if the approximate location of the source is known via a prior GW detection, the afterglow would be observable during the rapidly brightening phase at the sensitivity achieved by existing instruments such as *XMM-Newton*, *ROSAT HRI*, and *Chandra* (at ~ 100 ks exposure; see, e.g., Brandt & Hasinger 2005). For the parameter

combinations probed in Figures 4.1 and 4.2, there are at most ~ 100 sources in the sky with $L_X \gtrsim 10^{43}$ erg s^{-1} and $d \ln L_X / dt_{obs} > 10\%$ yr^{-1}, and they maintain this luminosity and brightening rate for $t_{ag,obs} \gtrsim 10$ yr in the observer frame.

4.3.2 Counts of Birthing Quasars in X-ray and Optical Surveys

Applying the methods described above to estimate the afterglow light-curve and source population, we next calculate the source counts of identifiable afterglows as a function of their apparent flux. Based on the fact that the mergers of massive dark matter halos peak at a redshift $z \sim 2$, we limit our analysis to AGN in the redshift range $1 < z < 3$. We assume that the typical disk-plus-binary system has parameter values $q = 0.1$, $\beta = 0.05$, $\lambda = 1$, $\theta = 0.2$ $n = 0.4$, and $m = -1/2$. In reality, these parameters will vary from system to system, perhaps by a great deal. However, because our calculations above (equation 4.6 and Figure 4.2) suggest that neither the luminosity nor the mass function of afterglow sources are likely to be affected by the values of these parameters by more than an order of magnitude, we limit further exploration of the parameter space to the α-S plane. The value of the viscosity parameter α is highly uncertain; however, numerical simulations of MHD disks suggest in the radiation-dominated regions of an accretion disk $\alpha \sim 1$ is consistent with our type of viscosity prescription $\nu \propto P_{gas}$ (Hirose et al. 2009a). Assuming that the mass supply rate around the afterglow phase is at least comparable to the rate during prolonged AGN activity, we expect S to be at least as great as the ratio of the typical luminosity of the typical AGN episode to the Eddington luminosity limit of the SMBH engine, i.e. $S \geq f_{Edd} = 0.3$. Depending on how much mass the circumbinary disk

accumulates near its inner wall prior to decoupling, S may, however, significantly exceed $f_{\rm Edd}$.

4.3.2.1 X-ray Surveys

We consider a hypothetical survey with observational capabilities similar to those of the proposed *LWFT* mission, sensitive to photon energies of 0.1 – 3.5 keV, down to fluxes of $\sim 3.5 \times 10^{-14}$ erg s^{-1} cm^{-2} (confusion limit for *LWFT*).

In Figure 4.3, we present the X-ray source counts of objects increasing in band luminosity by at least 3% yr^{-1}, 10% yr^{-1}, 30% yr^{-1} and 100% yr^{-1} in the relevant frequency band. (Note that the energy range approximately coincides with the sensitivity window of existing deeper-exposure telescopes such as *Chandra* and *XMM-Newton*, and the planned all-sky survey *eROSITA*.) In the figure, the height of the histogram pillars are the number of sources in each logarithmic flux bin, with each bin having a width $\Delta \log_{10}(F \text{ erg}^{-1} \text{ cm}^2 \text{ s}) = 0.5$. The dashed vertical line in each panel shows the approximate source confusion flux limit. These results suggest that only if S is large, i.e. if there is significant mass accumulation at the inner wall of the circumbinary disk, then rapidly brightening afterglow sources could be detected by the model mission: several sources at $S \sim 3$ and as many as dozens of sources for $S \sim 10$. The brightest of these sources have central SMBH masses of $M \gtrsim 10^6 S^{1.2} M_\odot$, and thus most of the X-ray-detectable sources would also be observable by *LISA*. However, for $S \gtrsim 3$, some would fall outside *LISA*'s sensitivity window. These sources are expected to continue to brighten at a slightly reduced rate at harder frequencies, $h\nu > 3.5$ keV; this is a prediction that could be tested with pointed

follow-up observations. Our calculations indicate that birthing quasars will be difficult to identify with existing and planned wide-angle soft-X-ray surveys. For example, the *eROSITA* all-sky survey is expected to have a semi-annual flux limit of $\gtrsim 10^{-13}$ erg s$^{-1}cm^{-2}$ with a time resolution of $\sim$ months at those flux levels. Intrinsic absorption of soft X-rays by the birthing quasar itself could also be an observational barrier for at least some of the sources (see, e.g., Brandt et al. 2000), especially if galaxies harboring merging SMBHBs tend to be more heavily shrouded in gas and dust than the general population of active galaxies.

Whether the monotonic brightening of the afterglow would be sufficiently distinguishable from the X-ray variability of faint AGN and other sources is an open (and more difficult) question. Luminous X-ray AGN have been known to vary by as much as $\sim 10 - 100\%$ yr^{-1} (e.g., Mushotzky et al. 1993). Any monotonic brightening in the X-ray must be distinguished from other sources of intrinsic variability, in addition to any instrumental error close to the faint-flux detection limit. It would appear likely, however, that a source that monotonically increases its X-ray luminosity by up to order unity per year for several years, while showing a consistent and monotonic hardening of its spectrum for the entire duration, would be fairly unusual. While an increase in the accretion rate through a standard thin accretion disk is also expected to produce a simultaneous brightening and spectral hardening, the spectral evolution will be different in detail. We also note that a variability survey, utilizing the ROSAT all sky survey, as well as pointed ROSAT observations (Grupe et al. 2001) found no correlation between changes in flux and spectral hardening in a sample of luminous soft-X-ray AGN. If the brightening is caused

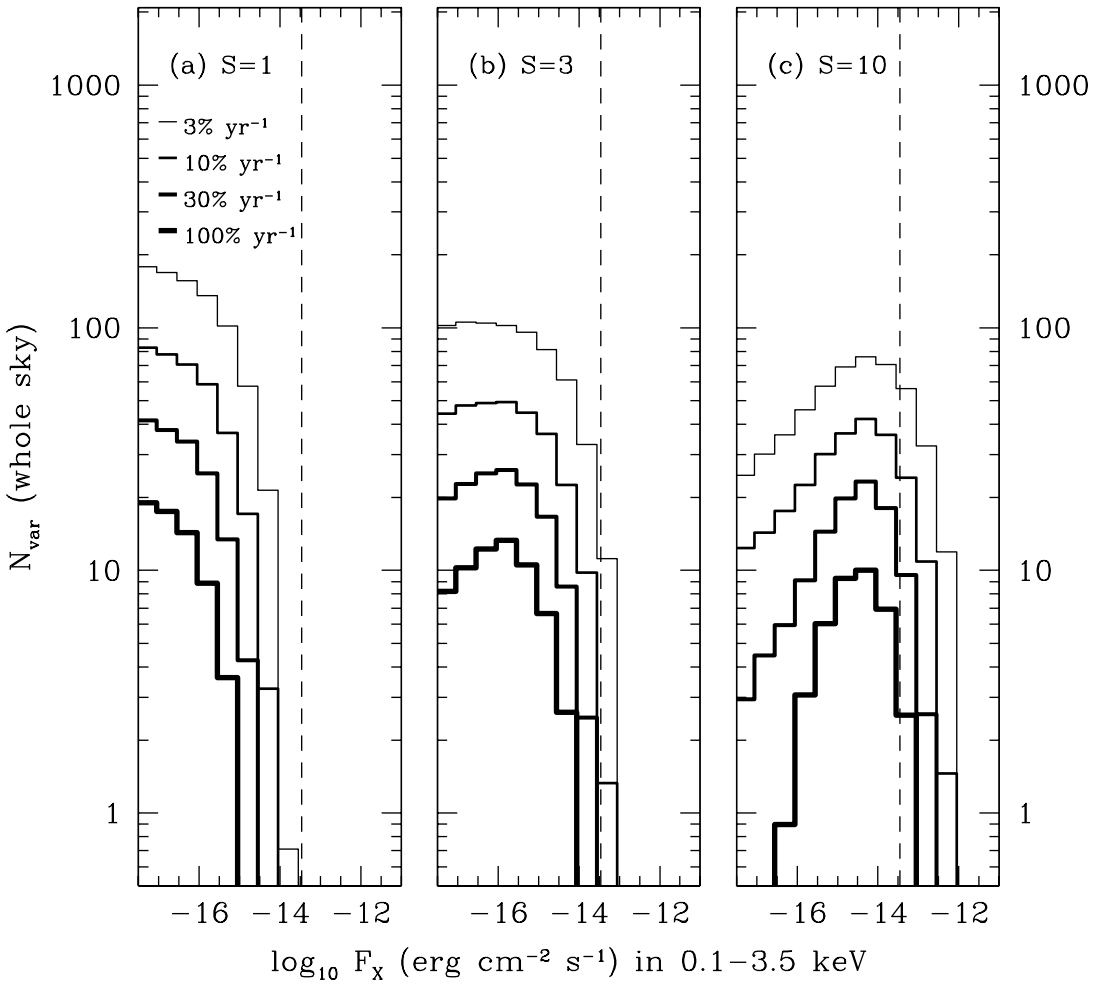

Figure 4.3 The number of afterglow sources as a function of their 0.1 − 3.5 keV flux, in the redshift range $1 < z < 3$. The histogram shows the number of sources in each logarithmic flux bin of width $\Delta \log_{10}(F \text{ erg}^{-1} \text{ cm}^2 \text{ s}) = 0.5$. We consider sources with $10^6 M_\odot < M < 10^9 M_\odot$. The histograms demarcate, in order of decreasing line thickness, the number of sources in each flux bin exhibiting a brightening rate of at least 3% yr^{-1}, 10% yr^{-1}, 30% yr^{-1}, and 100% yr^{-1}. The dashed vertical line in each panel is the detection limit for point sources in our model survey. All panels have parameter values $q = 0.1$, $\alpha_{-1} = \lambda = \theta_{0.2} = 1$, $\beta = 0.05$, $1.5 < z < 2.5$, $n = 0.4$ and $m = -1/2$. We vary the value of S, a measure of the disk mass, in each panel: $S = 1$ in panel (a), $S = 3$ in panel (b), and $S = 10$ in panel (c).

by the viscous afterglow, it would also slow down on humanly tractable timescales and the light curve could be checked at different observational frequencies against the evolutionary models for the viscously spreading disk. This feature would also help distinguish birthing quasars from tidal disruption events.

It is worth emphasizing that a major caveat of the above analysis is that the thin-disk formalism adopted in our afterglow model breaks down for models with $S \gtrsim 3$. Indeed, in this massive regime, the disk midplane temperature becomes sufficiently high that a one-dimensional estimate for the disk scale-height-to-radius ratio H/R, evaluated at the cavity wall, exceeds unity (MP05; TM10). This suggests that horizontal advection would become a significant factor in determining the structure of the inner accretion flow. The evolution and emission properties of the disk in this advective regime is highly uncertain and it remains a subject of active research. Advective disks have generally been associated with radiatively inefficient accretion states (see Narayan & Quataert 2005, for a review), but the situation may be different here, with a high density accretion flow and an intrinsically time-dependent flow. As suggested by MP05, horizontal advection could act to make the disk "slim", as in the models of Abramowicz et al. (1988), and thus relatively radiatively efficient. It is also unclear how well binary torques can open a central cavity in the first place, if the disk becomes geometrically thick. Adding to the various modeling uncertainties is the possibility that a sizable fraction of the viscously dissipated energy could be deposited into a hot disk corona (e.g., Liu et al. 2003a), which would lead to a significant increase of the afterglow high energy emission, well above what is predicted by our strictly thermal afterglow models.

4.3.2.2 Optical Surveys (*LSST*)

Our simple viscous emission model predicts that the same SMBH binary remnant that produces an X-ray afterglow would brighten at lower, optical frequencies at a rate of $\gtrsim$ a few% yr^{-1}, several years prior to the X-ray afterglow (TM10). This less pronounced variability is comparable in magnitude to the intrinsic r.m.s. long-term variability of optical AGN (e.g., Ulrich et al. 1997). However, if the approximate location of the source is known through a GW detection, searching for AGN exhibiting steady optical/ultraviolet brightening may identify the source before the X-ray afterglow. Alternatively, in the absence of a GW signal, a wide-angle variability survey of optical AGN with high time resolution, such as those possible with *LSST*, could possibly still be used to select afterglow candidates for follow-up X-ray observations.

Motivated by these possibilities, we apply our simple source-count estimate to the *LSST* u photometric band (330 − 400 nm). We choose this band because it is the bluest *LSST* filter, and thus the one in which the afterglow brightening is likely to be the most prominent. The results are presented in Figure 4.4. In both panels, the dashed vertical lines demarcate the limiting flux for signal-to-noise of 50 over a $\sim$ 1 yr period ($\sim$ 450 s accumulated exposure). With the fiducial parameter choices in the left panel, there is clearly no hope of a detection. We also find that, unlike in the X-ray bands, raising the value of S does not increase the source counts. This is because while increasing S beyond the fiducial value pushes the evolving portion of the spectrum into the model instrument band, it pushes it further out of the *LSST* u band, reducing the optical variability. Therefore, we instead show, in the right panel, the expected number counts for the larger value $\alpha = 1$,

on which the emission frequency depends less strongly. We find that the sources are still likely to be too dim and too few to be identifiable with high confidence from among the large number of AGN expected to be detected by *LSST*. However, gradually brightening optical AGN could still be cataloged, and their X-ray luminosities could be cross-checked with data from instruments such as *eROSITA* for subsequent X-ray afterglows, which may still prove useful.

More promising for optical surveys is the possibility that the X-ray afterglow may be promptly reprocessed into optical or infrared frequencies, either by dust surrounding the source (MP05) or by warps and vertical splaying in the circumbinary disk (TM10). If a significant portion of the X-ray emission is reprocessed, then the afterglow source will appear as an AGN whose optical/infrared luminosities can brighten by $d \ln L_\nu / dt \gtrsim 10\%\ \ \mathrm{yr}^{-1}$. Such a monotonic variability would exceed the typical long-term r.m.s. variability in AGN brightness at these frequencies, and is likely to be identifiable by *LSST*. For purely demonstrative purposes, we adopt here a simple toy model, in which a fraction $f_{\rm rp}$ of the emitted power above $\nu > 1$ keV in the source's rest frame is reprocessed (thermalized) and re–emitted at frequencies below the ultraviolet energy 10 eV, so that the power below $h\nu \leq 10$ eV is enhanced as

$$L^{(\rm rep)}_{h\nu<0.01\ \rm keV} \sim L_{h\nu<0.01\ \rm keV} + f_{\rm rp} L_{h\nu>1\ \rm keV}. \tag{4.8}$$

We further assume that the fractional energy enhancement is roughly uniform in the optical and infrared — i.e., we ignore line emission from recombination processes — so

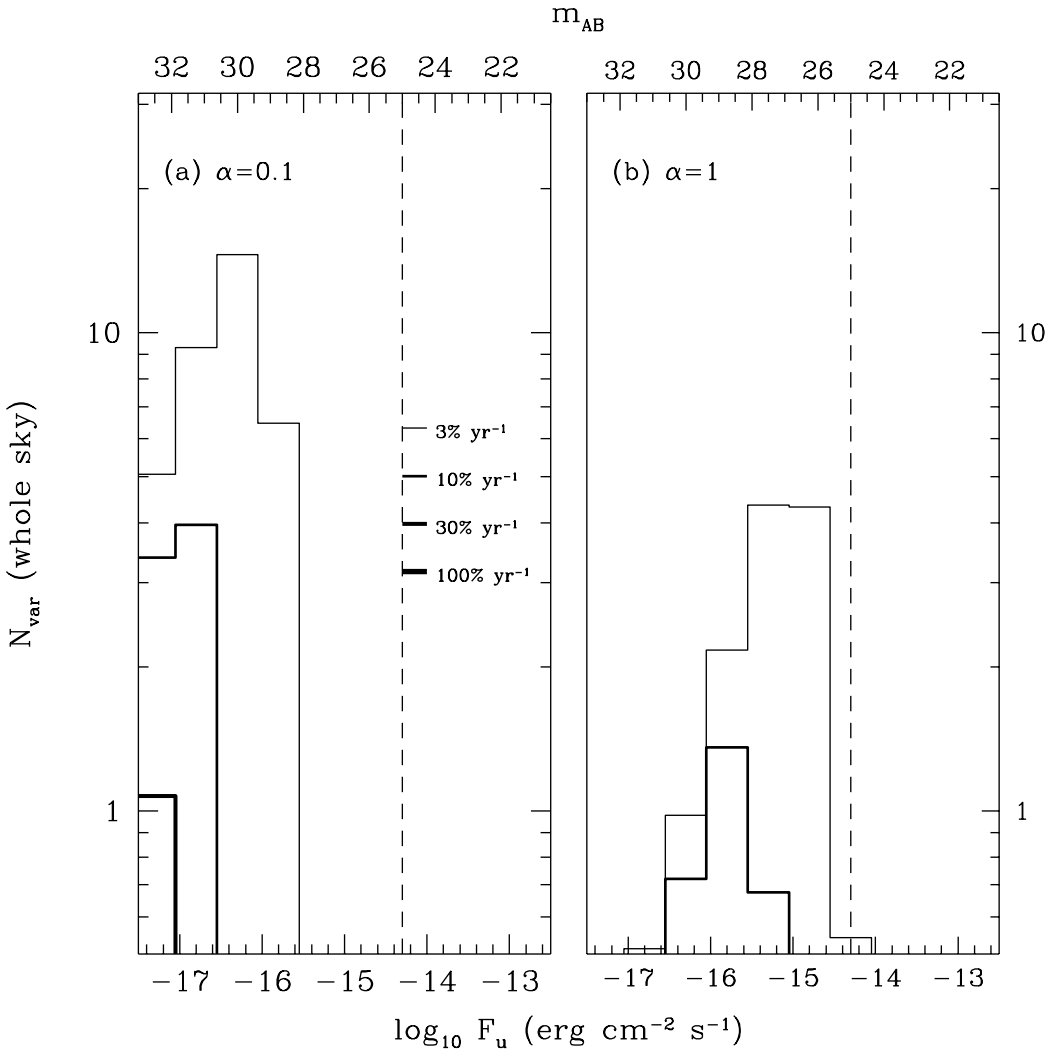

Figure 4.4 Same as Figure 4.3 but in the u photometric band (330 – 400 nm). The dashed vertical line is the limiting flux to achieve a signal-to-noise of at least 50 over ~ 1 yr of operation for *LSST* (~ 30 exposures at 15 s each). Both panels have parameter values $q = 0.1$, $S = \lambda = \theta_{0.2} = 1$, $\beta = 0.05$, $1 < z < 3$, $n = 0.4$ and $m = -1/2$. Panel (a) has $\alpha = 0.1$, and panel (b) has $\alpha = 1$ (more viscous disk).

that for our purposes the optical spectral emission is given by

$$\nu L_\nu^{(\mathrm{rep})} \sim \nu L_\nu \times \left(1 + f_{\mathrm{rp}} \frac{L_{h\nu > 1 \text{ keV}}}{L_{h\nu < 0.01 \text{ keV}}}\right). \tag{4.9}$$

This prescription is similar in spirit to the model of Gierliński et al. (2009), who showed that reprocessing signature of the X-ray outbursts of the stellar-mass black hole system XTE J1817-330 is consistent with reprocessing a constant fraction of the bolometric X-ray luminosity. We also neglect the reprocessing time, which is dominated by the light-travel time and much shorter than the variability timescales of interest here (see, e.g., Peterson & Horne 2004).

The source counts in the *LSST* u band for this model, for a reprocessing fraction $f_{\mathrm{rp}} = 10^{-2}$, are shown in Figure 4.5. Our simple calculations suggest that perhaps dozens of afterglows could be detected if the X-ray emission is reprocessed, for moderately optimistic parameter values, e.g., for $S \gtrsim 1$. It is worth cautioning that the reprocessed fraction f_{rp} is highly dependent on the vertical disk geometry, which itself may be rapidly evolving during the afterglow. For instance, TM10 found in their non-irradiated afterglow models that the scale-height-to-radius ratio H/R can be a steeply increasing function of radius during the period when the cavity is refilling, but not necessarily before or after this phase. Such complications should be included, along with details of radiative transfer, in a more realistic analysis of disk irradiation and reprocessing. Finally, absorption of the reprocessed UV/optical emission by gas and dust, surrounding the nuclear SMBH on larger scales, could be another observational hindrance for at least some of the sources.

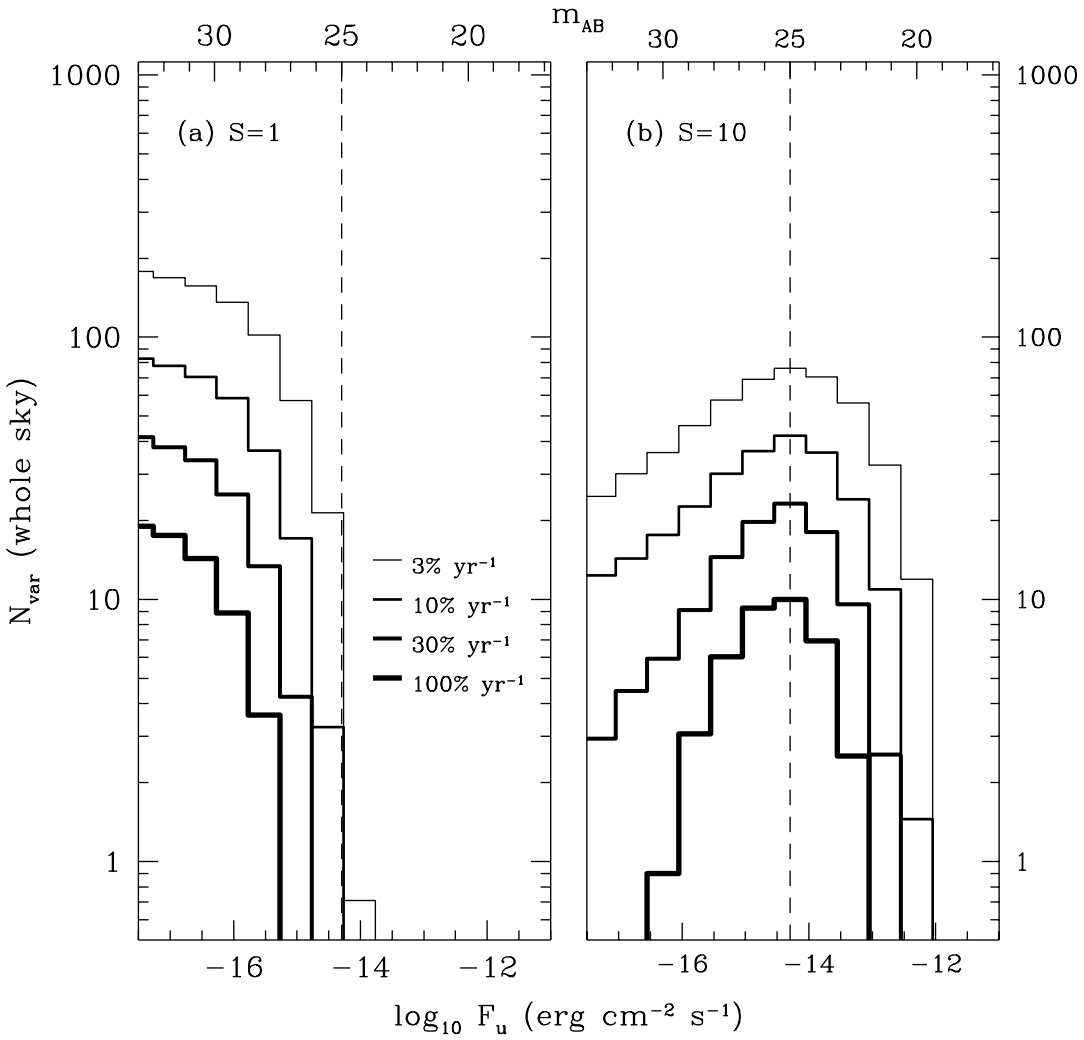

Figure 4.5 Same as Figure 4.4 but assuming that $f_{\rm rp} = 1\%$ of the power emitted above 1 keV is reprocessed into optical and infrared bands (see text for details). The dashed vertical line is the limiting flux as in Figure 4.4. Both panels have parameter values $q = 0.1$, $\alpha_{-1} = \lambda = \theta_{0.2} = 1$, $\beta = 0.05$, $1 < z < 3$, $n = 0.4$ and $m = -1/2$. Panel (a) has $S = 1$, and panel (b) has $S = 3$ (more massive disk).

4.4 Conclusions

Using an idealized model for the population of coalescing SMBHBs, and for the light curve of the afterglow produced by the viscously spreading post-merger circumbinary disk, we have shown that ongoing afterglows of SMBHB mergers may be present in the data sets of wide X-ray and optical surveys. In soft X-ray bands, this requires that the surface density and the viscosity in the circumbinary disk be at the high end of the expected range, while afterglows could only be found in optical surveys if the X-ray emission is promptly and significantly reprocessed into optical frequencies.

Despite the highly approximate nature of our analysis and other model uncertainties, our calculations provide a proof-of-concept for a very general hypothesis: *SMBHB mergers may exhibit identifiable, steady brightening rate for a period of the order of decades, and such afterglows could be detected serendipitously in a large survey that revisits the sky at least every few months for several years.* Our more specific findings can be summarized as follows:

- For optimistic parameter values, several birthing quasars, brightening by at least $d \ln L_X / dt_{obs} > 30\%$ yr^{-1} for several years, could be identified in the 0.1 - 3.5 keV soft X-ray band by an all-sky survey with specifications comparable to those proposed for the *LWFT* mission.

- At any given time, there could be up to $N_{ag} \sim 100$ sources in the sky that exhibit a brightening at or above $d \ln L_X / dt_{obs} > 10\%$ yr^{-1}, with soft X-ray luminosities $L_X \gtrsim 10^{42}$ erg s^{-1}. The most luminous sources typically spend $t_{ag,obs} \gtrsim 10$ yr in this state, and thus can be monitored on humanly tractable timescales. These numbers

depend weakly on most system parameters.

- To have any hope of detecting birthing quasars, a survey has to reach a depth of at least a few $\times 10^{-13}$ erg s^{-1} cm^{-2}. However, the slopes of our calculated $\log N - \log S$ distributions at fluxes just below this threshold are relatively shallow (Figures 4.3, 4.4 and 4.5), implying that surveys should favor large angular sky coverage over depth, once they reach this flux threshold.

- If identified, candidate sources can be followed up by pointed observations at higher frequencies, where they are expected to continue both their monotonic brightening and their spectral hardening.

- Most birthing quasars that are identifiable have, coincidentally, SMBH masses lying in the middle of *LISA*'s sensitivity window ($M \sim 10^6 M_\odot$), and are thus members of the same population that would be probed with GW detections. However, a minority ($\gtrsim$ few % for $S \gtrsim 3$) of the detectable X-ray variables have masses of $\gtrsim 10^7$ $M_\odot$, probing a population above *LISA*'s range.

- These sources may be identifiable by *LSST* if a fraction as low as $\sim 1\%$ of the X-ray flux is promptly reprocessed into the optical frequencies.

Our calculations are contingent on theoretical caveats of the afterglow scenario we have considered. The two primary uncertainties regarding the post-merger evolution of the circumbinary cavity are related to the viscous and advective properties of the disk. As stated in § 4.2.1, the viscosity of accretion flows, including the possibility of viscous instability, are not well understood when radiation pressure dominates gas pressure,

which is the relevant regime for the gas refilling the circumbinary cavity. Additionally, the disk may be geometrically thick (MP05; TM10), either right at decoupling or later during the afterglow phase, suggesting that horizontal advection may play a significant role in determining the surface density evolution and the disk net emission properties. The importance of viscous instabilities in radiation-dominated accretion flows remains a general open question, and the role of advection in a viscously spreading accretion flow remains a largely unexplored regime. More detailed studies of the circumbinary cavity will be needed to address how these effects may affect the emission predicted by simple analyses based on a thin disk formalism such as ours. Another major uncertainty is the validity of our assumption that quasar activity can be associated with SMBH coalescence. In reality, there may not be a one-to-one relation: it is possible that for at least some AGN, gas accretion or changes in radiative efficiency are triggered by mechanisms other than SMBH mergers; conversely, some SMBH mergers may not trigger prolonged quasar activity. If the former is true, our analysis overestimates the number of identifiable afterglow sources; if the latter is true, then our results could in principle be an underestimate.

For completeness, we note that while we focused here on the viscous afterglows, other SMBHB merger–related signatures could also be looked for in EM surveys. For example, the GW–emission–induced mass–loss and recoil can cause strong disturbances in the circumbinary disk, which can produce a detectable afterglow (Lippai et al. 2008; Schnittman & Krolik 2008; Shields & Bonning 2008; O'Neill et al. 2009; Megevand et al. 2009; Corrales et al. 2010; Rossi et al. 2010). For the low SMBH masses of $\sim 10^6 M_\odot$ relevant for *LISA*, these signatures are expected to have a short duration $\sim$ few years (e.g., Corrales

et al. 2010) and would be too rare to be found serendipitously, without a trigger from *LISA*. However Schnittman & Krolik (2008) and Shields & Bonning (2008) focused on these signatures in disks around more massive SMBHs, which occur on longer ($\sim 10^4$yr) time–scales, and proposed detecting a flare by monitoring a population of AGN in the infrared or X-rays bands. Another possibility is that the binary is activated, and produces periodic emission, tracking the orbital frequency, prior to the merger. Haiman et al. (2009a) argued that as long as this emission is at a few percent of the Eddington luminosity, a population of these variable sources, with periods of tens of weeks, may be identifiable in optical or X-ray surveys.

To conclude, the concomitant observation of a SMBHB merger based on GW and EM signals remains by far the most promising scenario for the unambiguous detection of such systems. The precision with which *LISA* would determine the masses, spins, and luminosity distances of coalescing binaries can not be replicated by current or planned EM telescopes. However, detections based on EM signatures alone could still help identify SMBHB mergers before *LISA* is launched, and perhaps more importantly, possibly outside *LISA*'s mass sensitivity window. Detecting the EM signatures from the mergers of the most massive SMBHs would complement the synergistic EM-plus-GW observations of lower-mass systems, and help provide a more complete picture of the accretion physics and cosmological evolution history of SMBHBs.

Chapter 5

Exact Time-dependent Solutions for the Thin Accretion Disk Equation: Boundary Conditions at Finite Radius[1]

5.1 Introduction

Since its emergence in the 1970's (Shakura & Sunyaev 1973; Novikov & Thorne 1973; Lynden-Bell & Pringle 1974), the theory of astrophysical accretion disks has been applied to explain the emission properties of active galactic nuclei, X-ray binaries, cataclysmic binaries, supernovae and gamma-ray bursts; to study planetary and star formation; and to model the evolution of binary and planetary systems. Because accretion disks onto

[1]This chapter is a reformatted version of an article by the same name by T. Tanaka that can be found in the Monthly Notices of the Royal Astronomical Society, Volume 410, Issue 2, pp. 1007-1017. The abstract for this paper is reproduced in Section 1.6.4.

compact objects can dissipate much larger fractions of baryonic rest-mass energies than nuclear reactions, they are often associated with some of the most energetic astrophysical processes in the universe.

If the local gravitational potential is dominated by a central compact object or a compact binary, and if the timescale for the viscous dissipation of energy is longer than the orbital timescale, then the accretion flow near the center of the potential is expected to be nearly axisymmetric. If the gas is able to cool efficiently, then the flow will also be geometrically thin, and one only needs the radial coordinate to describe the mass distribution in the disk (any relevant vertical structure can be integrated or averaged over the disk height). The partial differential equation (Lynden-Bell & Pringle 1974, henceforth LP74)

$$\frac{\partial}{\partial t}\Sigma(R,t) = \frac{1}{R}\frac{\partial}{\partial R}\left[R^{1/2}\frac{\partial}{\partial R}\left(3\nu\Sigma R^{1/2}\right)\right], \tag{5.1}$$

is obtained by combining the equations of mass conservation and angular momentum, and describes the surface density evolution of a thin Keplerian accretion disk due to kinematic viscosity ν.

In general, the viscosity ν depends on the surface density Σ and equation 5.1 is nonlinear. If, however, ν is only a function of radius, then the equation is linear and much more amenable to analytic methods. In particular, a solution that makes use of a Green's function G,

$$\Sigma(R,t) = \int_{R_{\rm in}}^{\infty} G(R,R',t)\,\Sigma(R',t=0)\,dR', \tag{5.2}$$

gives the solution Σ for any $t > 0$ given an arbitrary (and not necessarily differentiable)

profile $\Sigma(R, t = 0)$ and an inner boundary condition imposed at $R_{\rm in}$. For power-law viscosity $\nu \propto R^n$, Lüst (1952) and LP74 derived analytic Green's functions that satisfy a boundary condition of either zero-torque or zero-mass-flux at the coordinate origin, i.e., for the case $R_{\rm in} = 0$. In reality, however, the objects at the center of astrophysical accretion disks have a finite size to which the observational appearance of the disk is sensitive: e.g., the luminosity, spectral hardness, and variability timescales of black hole disks depend strongly on the radius of the innermost stable orbit, and those of circumbinary disks depend on where the inner disk is truncated by the central tidal torques.

Despite the astrophysical relevance of Green's functions to the thin accretion disk equation with boundary conditions imposed at a finite radius, such solutions have not been published. Pringle (1991) derived the Green's function with a zero-flux boundary condition at a nonzero radius in the special case $n = 1$, and noted the "extreme algebraic complexity" involved in calculating a more general solution with $R_{\rm in} > 0$. Time-dependent models of accretion flows have continued to employ solutions that correspond to the central objects having zero physical size (e.g., Metzger et al. 2008; Tanaka & Menou 2010).

In this paper we derive exact Green's functions for equation 5.1 for boundary conditions imposed at a finite radius, for any power-law viscosity $\nu \propto R^n$ with $n < 2$. We show that mathematical difficulties can be minimized with the aid of the appropriate integral transform techniques, namely the Weber transform (Titchmarsh 1923) and the recently proved generalized Weber transform (Zhang & Tong 2007). We present two specific solutions of astrophysical interest: the solution with zero torque at a radius $R_{\rm in} > 0$, which is of interest for accretion disks around black holes and slowly rotating stars; and the solution

with zero mass flow at $R_{\rm in} > 0$, which is applicable to accretion flows that accumulate mass at the disk center due to the injection of angular momentum from the tidal torques of a binary or perhaps the strong magnetic field of the central object.

This paper is organized as follows. In §2, we review the Green's function solutions, derived by Lüst (1952) and LP74, for the thin-disk equation with boundary conditions imposed at the origin. In §3, we derive the new Green's function solutions, which impose boundary conditions at a finite inner boundary radius. We offer our conclusions in §4.

5.2 Green's-Function Solutions with Boundary Conditions at $R = 0$

In the special case where the viscosity is a radial power law, $\nu \propto R^n$, and assuming a separable ansatz of the form $\Sigma(R,t) = R^p\sigma(R)\exp(-\Lambda t)$, where p and Λ are real numbers and σ is an arbitrary function of R, equation 5.1 can be rewritten as the Bessel differential equation:

$$R^2\frac{\partial^2\sigma}{\partial R^2} + \left(2p + 2n + \frac{3}{2}\right)R\frac{\partial\sigma}{\partial R} + \left[(p+n)\left(\frac{\Lambda}{3s}R^{2-n} + p + n + \frac{1}{2}\right)\right]\sigma = 0. \tag{5.3}$$

Above, $s = \nu R^{-n}$ is a constant. With the choices $p = n - 1/4$ and $\Lambda = 3sk^2$, equation 5.3 has the general solution

$$\sigma_k(R) = R^{-2n}\left[A(k)J_\ell(ky) + B(k)Y_\ell(ky)\right]. \tag{5.4}$$

Above, k is an arbitrary mode of the solution; $A(k)$ and $B(k)$ are the mode weights; $\ell = (4-2n)^{-1} > 0$; $y(R) \equiv R^{(1-n/2)}/(1-n/2)$; and J_ℓ and Y_ℓ are the Bessel functions of the first and second kinds, respectively, and of order ℓ. If ℓ is not an integer, then Y_ℓ above may be replaced without loss of generality by $J_{-\ell}$. Integrating the fundamental solution across all possible k-modes gives the solution:

$$\Sigma(R,t) = \int_0^\infty R^{-n-1/4}\left[A(k)J_\ell(ky) + B(k)Y_\ell(ky)\right]\ \exp(-3sk^2t)\,dk. \tag{5.5}$$

The mode-weighting functions $A(k)$ and $B(k)$ are determined by the boundary conditions and the initial surface density profile $\Sigma(R, t=0)$. Our goal is to rewrite equation 5.5 in the Green's function form (equation 5.2) and to write down an explicit symbolic expression for the Green's function $G(R, R', t)$. Throughout this paper, we will employ the following strategy:

1. Using the boundary condition, find an analytic relationship between the mode weights $A(k)$ and $B(k)$.

2. Identify the appropriate integral transform to express the mode weights in terms of the initial profile $\Sigma(R, t=0)$.

3. Insert the time-dependence $\exp(-3sk^2t)$ and integrate over all modes to find the Green's function.

4. Derive analytic expressions for the asymptotic disk behavior at late times and small radii.

Before deriving the solutions with boundary conditions at finite radius, we begin by reviewing the Green's functions of LP74 with boundary conditions at the coordinate origin.

5.2.1 Zero Torque at $R = 0$

An inner boundary condition with zero central torque is of astrophysical interest as it can be used to describe accretion onto a black hole or a slowly rotating star, at radii much larger than the radius of innermost circular orbit or the stellar surface, respectively. The radial torque density g in the disk due to viscous shear is

$$g(R,t) = \nu\Sigma R^2 \frac{\partial\Omega_{\rm K}}{\partial R} \propto \nu\Sigma R^{1/2}, \tag{5.6}$$

where $\Omega_{\rm K}$ is the Keplerian angular velocity of the orbit.

Because the functions J_ℓ and Y_ℓ have the asymptotic behaviors $J_\ell(ky) \propto y^\ell \propto R^{1/4}$ and $Y_\ell(ky) \propto y^{-\ell} \propto R^{-1/4}$ near the origin, at small radii the mode weight $A(k)$ will contribute to the behavior $g \propto R^{1/2}$ while $B(k)$ will contribute to g = constant. Therefore, for the solution to have zero viscous torque at $R = 0$ the function $B(k)$ must be identically zero.

We may relate the surface density distribution at $t = 0$ and the weight $A(k)$ via the integral equation

$$\Sigma(R, t = 0) = R^{-n-1/4} \int_0^\infty A(k) J_\ell(ky)\, dk, \tag{5.7}$$

which may be solved with the use of the Hankel integral transform (e.g., Ogilvie 2005).

A Hankel transform pair of order ℓ satisfies

$$\phi_\ell(x) = \int_0^\infty \Phi_\ell(k)\, J_\ell(kx)\, k\, dk, \tag{5.8}$$

$$\Phi_\ell(k) = \int_0^\infty \phi_\ell(x)\, J_\ell(kx)\, x\, dx. \tag{5.9}$$

For the problem at hand, the suitable transform pair is

$$R^{n+1/4}\Sigma(R, t=0) = \int_0^\infty \left[A(k)k^{-1}\right] J_\ell(ky)\, k\, dk, \tag{5.10}$$

$$A(k)k^{-1} = \int_0^\infty \left[R^{n+1/4}\Sigma(R, t=0)\right] J_\ell(ky)\, y\, dy. \tag{5.11}$$

Combining them gives us $A(k)$:

$$A(k) = \left(1 - \frac{n}{2}\right)^{-1} \int_0^\infty \Sigma(y', 0)\, J_\ell(ky')\, k\, R'^{5/4}\, dR' \tag{5.12}$$

Inserting equation 5.12 and $B(k) = 0$ into equation 5.5, we obtain

$$\Sigma(R, t) = \left(1 - \frac{n}{2}\right)^{-1} R^{-n-1/4} \int_0^\infty R'^{5/4} \int_0^\infty \Sigma(R', t=0)\, J_\ell(ky')\, J_\ell(ky)\, \exp\left(-3sk^2t\right)\, k\, dk\, dR'. \tag{5.13}$$

To pose the solution in terms of a Green's function $G(R, R', t)$ (equation 5.2), we write

$$\begin{aligned} G(R, R', t) &= \left(1 - \tfrac{n}{2}\right)^{-1} R^{-n-1/4} R'^{5/4} \int_0^\infty J_\ell(ky')\, J_\ell(ky)\, \exp\left(-3sk^2t\right)\, k\, dk \\ &= (2-n)\tfrac{R^{-n-1/4}R'^{5/4}}{\tau} I_\ell\left(\tfrac{2R^{1-n/2}R'^{1-n/2}}{\tau}\right) \exp\left(-\tfrac{R^{2-n}+R'^{2-n}}{\tau}\right). \end{aligned} \tag{5.14}$$

Above, I_ℓ is the modified Bessel function of the first kind, and we have substituted

$\tau \equiv 12(1-n/2)^2 st$.

Although the Green's function allows for the calculation of $\Sigma(R,t)$ for arbitrary initial surface density profiles, it is instructive to study the case where the initial surface density is a Dirac delta function,

$$\Sigma(R, t=0) = \Sigma_0 \, \delta(R - R_0) \, R_0, \tag{5.15}$$

for which the solution is (by definition) the Green's function itself. The integral over radius in equation 5.2 becomes trivial and many behaviors of the solution may be expressed analytically. Because any initial surface density profile can be described as a superposition of δ-functions, studying this special case will help illuminate the general behavior of all solutions.

We may evaluate the asymptotic behavior at late times and small radii by noting that for small argument $z \lesssim 0.2\sqrt{1+\ell}$,

$$I_\ell(z) \approx \frac{1}{\Gamma(\ell+1)} \left(\frac{z}{2}\right)^\ell. \tag{5.16}$$

We find

$$\Sigma\left(R \lesssim R_0, t \gtrsim t_\nu(R_0)\right) \approx \frac{2-n}{\Gamma(\ell+1)} \Sigma_0 \left(\frac{R}{R_0}\right)^{-n} \left[8\left(1-\frac{n}{2}\right)^2 \frac{t}{t_\nu(R_0)}\right]^{-1-\ell}, \tag{5.17}$$

where $t_{\nu,0} = (2/3)R_0^2/\nu$ is the local viscous timescale at R_0.

Thus, for these solutions the inward radial mass flow,

$$\dot{M}(R) = -2\pi R \Sigma v_R = 6\pi R^{1/2} \frac{\partial}{\partial R}\left(\nu \Sigma R^{1/2}\right), \tag{5.18}$$

becomes radially constant near the origin and at late times:

$$\dot{M}\left(R \lesssim R_0, t \gtrsim t_\nu(R_0)\right) \approx \frac{2-n}{\Gamma(\ell+1)} 3\pi\nu(R_0)\Sigma_0 \left[8\left(1-\frac{n}{2}\right)^2 \frac{t}{t_\nu(R_0)}\right]^{-1-\ell}. \tag{5.19}$$

In Figure 5.1, we plot the solution $\Sigma(R,t)$ and the radial mass flow $\dot{M}(R,t)$, for the δ-function initial condition (equation 5.15), and for viscosity power-law index values $n = 0.1$ and $n = 1$. In both cases, we see the power-law behavior from equation 5.17 near the origin as the solution approaches $t \sim t_{\nu,0}$. The disk spreads as the gas at inner annuli loses angular momentum to the gas at outer annuli. The gas initially accumulates near the origin, then becomes diffuse as mass is lost into the origin.

5.2.2 Zero Mass flow at $R = 0$

If the accretion flow has a sufficiently strong central source of angular momentum, then the gas will be unable to flow in, and instead accumulate near the origin. Such solutions can be used to describe astrophysical disks around a compact binary (Pringle 1991), and perhaps those around compact objects with strong central magnetic fields (LP74). For circumbinary thin disks, Pringle (1991) demonstrated that such a boundary condition characterizes quite well the effects of an explicit central torque term.

In general, the mass flow has the behavior

$$\dot{M} \propto \int_0^\infty R^{1/2} \frac{\partial}{\partial R}\left[A(k)J_\ell(ky)R^{1/4} + B(k)Y_\ell(ky)R^{1/4}\right] \exp(-3sk^2t)\, dk. \tag{5.20}$$

We have seen above that for solutions with $B(k) = 0$ the mass flow is radially constant

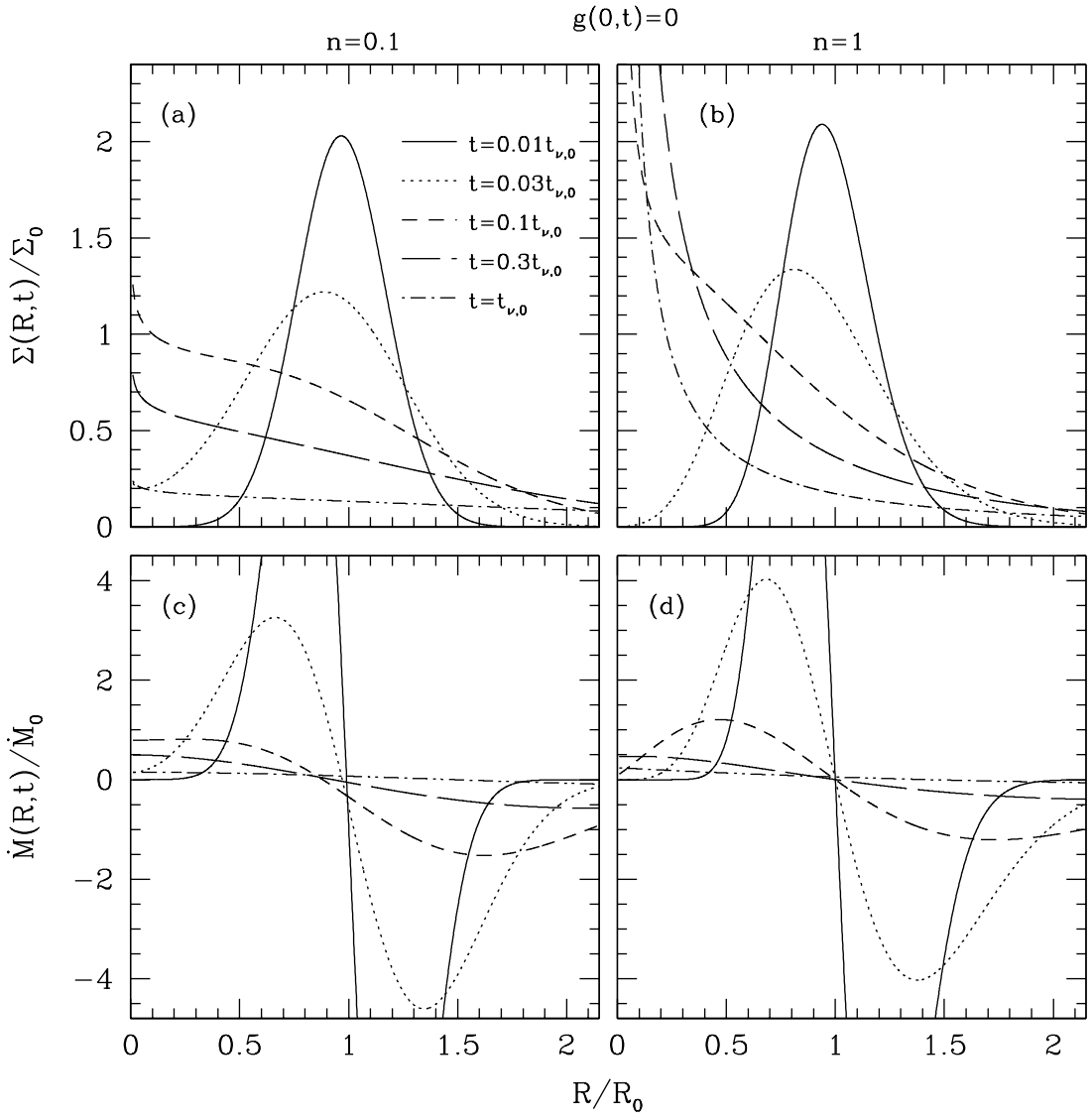

Figure 5.1 The solution $\Sigma(R,t)$, from LP74, and the corresponding radial mass inflow rate $\dot{M}(R,t)$, for a zero-torque boundary condition imposed at $R = 0$ and a δ-function initial profile $\Sigma(R, t = 0) = \Sigma_0 R_0 \delta(R - R_0)$, where the quantities Σ_0 and R_0 are arbitrary. The viscosity is a radial power law with $\nu \propto R^n$. Panels on the left side (a and c) show solutions for $n = 0.1$, and those on the right (b and d) show solutions for $n = 1$. Values for t are in units of the viscous time at R_0, $t_{\nu,0} = (2/3)R_0^2/\nu(R_0)$. We have normalized $\dot{M}$ to the quantity $\dot{M}_0 \equiv 3\pi\nu(R_0)\Sigma_0$. At late times, the solution has the behavior $\Sigma \propto R^{-n}$ and the mass-flow profile $\dot{M}$ becomes flat near the origin.

and finite near the origin. On the other hand, because $Y_\ell(ky) \propto R^{-1/4}$ near the origin, the weights $B(k)$ will all contribute no mass flow there; so for zero mass flow at $R_{\rm in} = 0$, we require $A(k) = 0$.

We note that because the surface density will have a power-law $\Sigma \propto R^{-1/2-n}$ at the origin, for the mass contained in the disk to converge n must be less than 3/2. Thus, for physically realistic solutions with zero mass flow at the origin, ℓ cannot be an integer. It follows that in this case Y_ℓ in equation 5.5 may be replaced by $J_{-\ell}$ without loss of generality. Then the Green's function for this case is derived in exactly the same fashion as in the previous case, the only difference being that the order of the Hankel transforms has the opposite sign. We obtain:

$$G(R, R', t) = (2-n)\frac{R^{-n-1/4}R'^{5/4}}{\tau} I_{-\ell}\left(\frac{2R^{1-n/2}R'^{1-n/2}}{\tau}\right)\exp\left(-\frac{R^{2-n}+R'^{2-n}}{\tau}\right). \tag{5.21}$$

As before, we evaluate the late-time behavior for the δ-function initial condition (equation 5.15):

$$\Sigma\left(R \lesssim R_0, t \gtrsim t_\nu(R_0)\right) \approx \frac{2-n}{\Gamma(1-\ell)}\Sigma_0\left(\frac{R}{R_0}\right)^{-n-1/2}\left[8\left(1-\frac{n}{2}\right)^2\frac{t}{t_\nu(R_0)}\right]^{-1+\ell}. \tag{5.22}$$

Figure 5.2 shows the evolution of the surface density and the radial mass flow for the boundary condition $\dot{M}(R = 0) = 0$. At early times, the behavior is nearly identical to the zero-torque boundary case. At late times, the zero-flux boundary condition causes the gas to accumulate instead of being lost to the origin. The central mass concentration reaches a maximum, then decreases as the disk begins to spread outward.

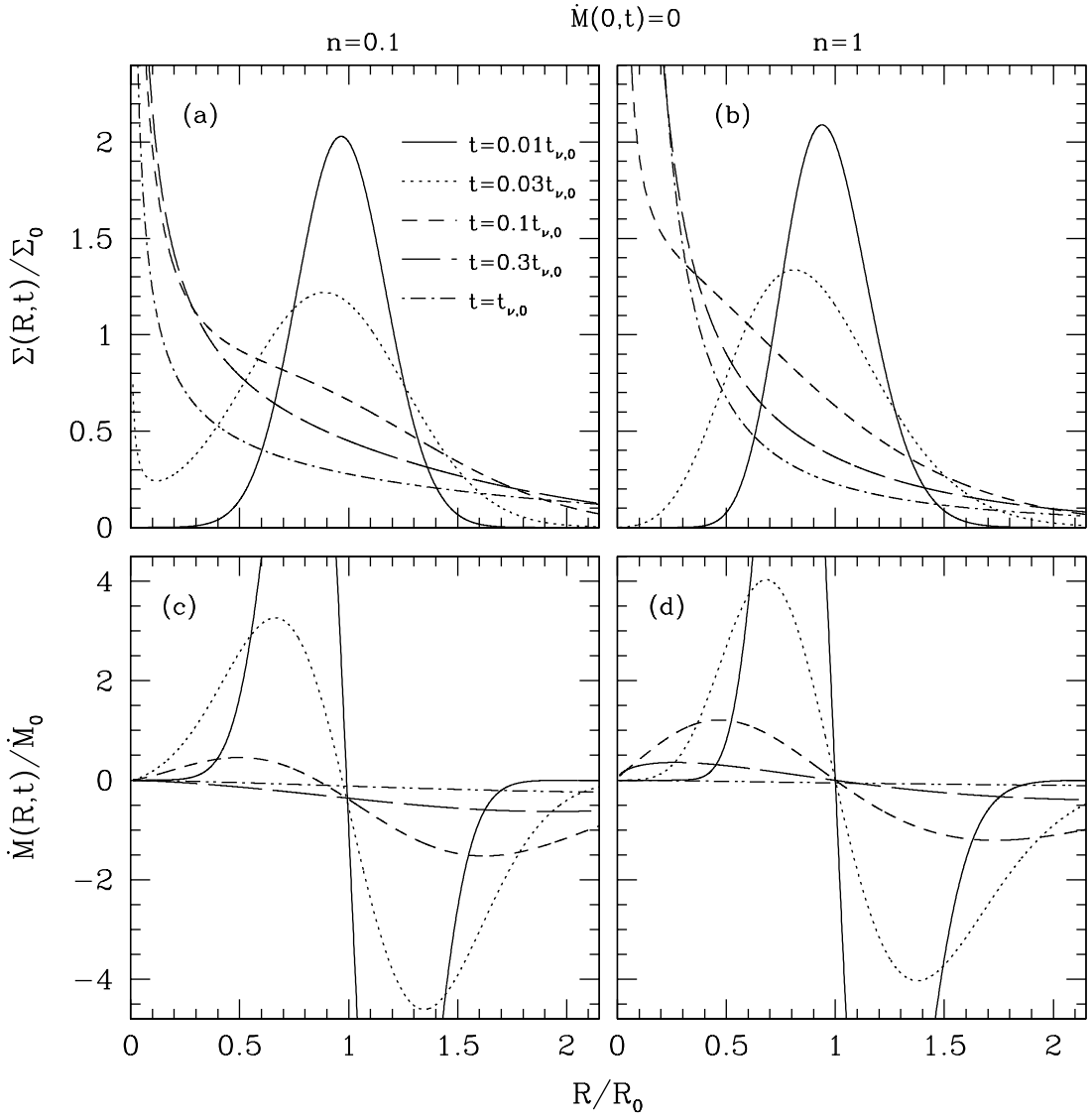

Figure 5.2 Same as Figure 5.1, except that the boundary condition is $\dot{M} = 0$ at $R = 0$. Again, the scales Σ_0 and R_0 are arbitrary. Whereas in the zero-torque case the total mass in the disk monotonically decreases due to mass loss at the origin (onto the black hole or star), the solutions in this figure conserve mass. At late times, the solution has the behavior $\Sigma \propto R^{-n-1/2}$. Gas initially piles up near the origin because of the boundary condition before gradually spreading outward; note that Σ at inner radii decreases from $t = 0.3t_{\nu,0}$ to $t = t_{\nu,0}$.

5.3 Green's-Function Solutions with Boundary Conditions at Finite Radii

Analytic treatment of the case with finite boundary radius was briefly diskussed in LP74 and Pringle (1991), but to the author's knowledge explicit solutions have never before been published. We show below that the Green's functions for finite boundary radii can be derived with the aid of the appropriate integral transform techniques, and that they can be represented as ordinary integrals of analytic functions.

5.3.1 Zero Torque at $R_{\rm in}$

We wish to solve the problem as in §5.2.1, but with $R_{\rm in} > 0$, i.e.

$$g(R_{\rm in}) \propto \Sigma\nu R^{1/2}\Big|_{R=R_{\rm in}} \propto \Sigma(R_{\rm in})R_{\rm in}^{n+1/2} = 0. \tag{5.23}$$

We may relate the mode weights $A(k)$ and $B(k)$ by requiring that every mode of the solution satisfy the boundary condition, i.e.:

$$A(k)J_\ell(ky_{\rm in}) + B(k)Y_\ell(ky_{\rm in}) = 0, \tag{5.24}$$

where $y_{\rm in} \equiv y(R_{\rm in})$. Substituting $C(k) = A(k)/Y_\ell(ky_{\rm in}) = -B(k)/J_\ell(ky_{\rm in})$, we obtain

$$\Sigma(R,t) = \int_0^\infty C(k)R^{-n-1/4}\left[J_\ell(ky)Y_\ell(ky_{\rm in}) - Y_\ell(ky)J_\ell(ky_{\rm in})\right]\exp(-3sk^2t)\,dk. \tag{5.25}$$

The function $C(k)$ may be evaluated with the use of the Weber integral transform (Titchmarsh 1923). A Weber transform pair satisfies

$$\phi_\ell(x) = \int_0^\infty \Phi_\ell(\kappa) \frac{J_\ell(\kappa x)Y_\ell(\kappa) - Y_\ell(\kappa x)J_\ell(\kappa)}{J_\ell^2(\kappa)+Y_\ell^2(\kappa)}\ \kappa\, d\kappa, \tag{5.26}$$

$$\Phi_\ell(\kappa) = \int_1^\infty \phi_\ell(x)\left[J_\ell(\kappa x)Y_\ell(\kappa) - Y_\ell(\kappa x)J_\ell(\kappa)\right]\ x\, dx. \tag{5.27}$$

Proceeding as before, we construct the pair

$$R^{n+1/4}\Sigma(R, t=0) = \int_0^\infty \left[C(\kappa)\kappa^{-1}\right] \frac{J_\ell(\kappa x)Y_\ell(\kappa) - Y_\ell(\kappa x)J_\ell(\kappa)}{J_\ell^2(\kappa)+Y_\ell^2(\kappa)}\ \kappa\, d\kappa, \tag{5.28}$$

$$C(\kappa)\kappa^{-1} = \int_1^\infty \left[R^{n+1/4}\Sigma(R, t=0)\right]\left[J_\ell(\kappa x)Y_\ell(\kappa) - Y_\ell(\kappa x)J_\ell(\kappa)\right]\ x\, dx. \tag{5.29}$$

Above, we have substituted $x = y/y_{\rm in} \geq 1$ and $\kappa = k y_{\rm in}$. Note the lower limit of integration in equation 5.29 is nonzero to account for the finite boundary radius. Combining equations 5.28 and 5.29 to eliminate $C(\kappa)$, and inserting the time-dependence factor $\exp(-3sk^2t) = \exp[-2(1-n/2)^2\kappa^2 t/t_\nu(R_{\rm in})]$, we obtain our new Green's function:

$$G(R, R', t) = \left(1-\tfrac{n}{2}\right) R^{-n-1/4} R'^{5/4} R_{\rm in}^{n-2} \int_0^\infty \frac{[J_\ell(\kappa x)Y_\ell(\kappa) - Y_\ell(\kappa x)J_\ell(\kappa)][J_\ell(\kappa x')Y_\ell(\kappa) - Y_\ell(\kappa x')J_\ell(\kappa)]}{J_\ell^2(\kappa)+Y_\ell^2(\kappa)} \times \exp\left[-2\left(1-\tfrac{n}{2}\right)^2 \kappa^2 \frac{t}{t_\nu(R_{\rm in})}\right]\ \kappa\, d\kappa. \tag{5.30}$$

Whereas the integral over k in equation 5.14 has an analytic solution, to the author's knowledge there is no analytic expression for the integral in equation 5.30. Nonetheless, equation 5.30 gives an exact expression for the Green's function. While it is somewhat more unwieldy than the solutions for $R_{\rm in} = 0$, the additional computational cost of an

ordinary integral is not likely to be a significant practical barrier (e.g., one could tabulate the integral in terms of the three dimensionless variables x, x' and $t/t_\nu(R_{\rm in})$).

The Green's function in equation 5.14 does have a closed-form expression for the special case $n = 1$ (i.e., $\ell = 1/2$). As noted by Pringle (1991), in this case the Bessel functions become easier to handle analytically, with $J_{1/2}(x) = \sqrt{\pi/2}\, x^{-1/2} \sin x$ and $Y_{1/2}(x) = -\sqrt{\pi/2}\, x^{-1/2} \cos x$. For this value of n we obtain for our Green's function

$$\begin{aligned} G(R,R',t) &= \tfrac{1}{\pi R_{\rm in}} \left(\tfrac{R'}{R}\right)^{5/4} (x\,x')^{-1/2} \int_0^\infty \sin\left[\kappa(x-1)\right] \sin\left[\kappa(x'-1)\right] \exp\left[-\tfrac{\kappa^2}{2} \tfrac{t}{t_\nu(R_{\rm in})}\right] d\kappa \\ &= \tfrac{R^{-3/2} R' R_{\rm in}^{-1/2}}{2\sqrt{2\pi}} \sqrt{\tfrac{t_\nu(R_{\rm in})}{t}} \left\{ \exp\left[-\tfrac{(x-x')^2}{2} \tfrac{t}{t_\nu(R_{\rm in})}\right] - \exp\left[-\tfrac{(x+x'-2)^2}{2} \tfrac{t}{t_\nu(R_{\rm in})}\right] \right\}. \end{aligned} \tag{5.31}$$

Note that in the case $n = 1$, x and x' are simply $\sqrt{R/R_{\rm in}}$ and $\sqrt{R'/R_{\rm in}}$, respectively.

For general values of n, we can evaluate the behavior at late times $t \gtrsim t_\nu(R_0) > t_\nu(R_{\rm in})$ by noting that in this regime only the modes $\kappa^2 \lesssim 1$ contribute to the integral in equation 5.30. For the central region $R \lesssim R_0$ at late times, we obtain the following analytic expression for the δ-function initial condition:

$$\begin{aligned} \Sigma\left(R \lesssim R_0, t \gtrsim t_\nu(R_0)\right) &\approx \tfrac{2-n}{2^{1+2\ell}\Gamma^2(1+\ell)} \Sigma_0 \left(\tfrac{R}{R_{\rm in}}\right)^{-n} \left(1 - \sqrt{\tfrac{R_{\rm in}}{R}}\right) \left(\sqrt{\tfrac{R_0}{R_{\rm in}}} - 1\right) \\ &\quad \times \int_0^\infty \exp\left[-2\left(1 - \tfrac{n}{2}\right)^2 \kappa^2 \tfrac{t}{t_\nu(R_{\rm in})}\right] \kappa^{1+2\ell}\, d\kappa \\ &= \tfrac{2-n}{\Gamma(1+\ell)} \Sigma_0 \left(\tfrac{R}{R_{\rm in}}\right)^{-n} \left(1 - \sqrt{\tfrac{R_{\rm in}}{R}}\right) \left(\sqrt{\tfrac{R_0}{R_{\rm in}}} - 1\right) \left[8\left(1 - \tfrac{n}{2}\right)^2 \tfrac{t}{t_\nu(R_{\rm in})}\right]^{-1-\ell}. \end{aligned} \tag{5.32}$$

We see that the Green's function explicitly gives the asymptotic behavior $\Sigma \propto 1 - \sqrt{R_{\rm in}/R}$, which has been used extensively for solutions of accretion disks near zero-torque boundary

surfaces (e.g., LP74, Frank et al. 2002).[2]

We show in Figure 5.3 the exact solutions for the δ-function initial condition, with the no-torque boundary condition imposed at $R_{\rm in} = R_0/5$. The qualitative evolution is as predicted by LP74: at early times, far from the boundary, the disk spreads inward in very much the same manner as the solutions with $R_{\rm in} = 0$, and so the $R_{\rm in} = 0$ Green's function suffices; at late times, once the gas reaches the vicinity of the boundary it exhibits the behavior $\Sigma \propto R^{-n}(1 - \sqrt{R_{\rm in}/R})$ in that neighborhood.

5.3.2 Zero Mass Flux at $R_{\rm in}$

We now consider the boundary condition of zero mass flow at a particular radius,

$$\dot{M}(R_{\rm in}) \propto \frac{\partial}{\partial R}\left(\nu\Sigma R^{1/2}\right)\Big|_{R=R_{\rm in}} \propto \frac{\partial}{\partial y}\left\{y^\ell\left[A(k)J_\ell(ky) + B(k)Y_\ell(ky)\right]\right\}\Big|_{y=y_{\rm in}} = 0. \tag{5.33}$$

From the relations $\partial[x^\ell J_\ell(x)]/\partial x = x^\ell J_{\ell-1}(x)$ and $\partial[x^\ell Y_\ell(x)]/\partial x = x^\ell Y_{\ell-1}(x)$, we obtain the relationship between A and B:

$$\frac{A(\kappa)}{B(\kappa)} = -\frac{Y_{\ell-1}(\kappa)}{J_{\ell-1}(\kappa)}. \tag{5.34}$$

The solution is then

$$\Sigma(R,t) = \int_0^\infty C(\kappa)R^{-n-1/4}\left[J_\ell(\kappa x)\,Y_{\ell-1}(\kappa) - Y_\ell(\kappa x)\,J_{\ell-1}(\kappa)\right]\exp\left[-2\left(1-\frac{n}{2}\right)^2\kappa^2\frac{t}{t_\nu(R_{\rm in})}\right]\,\kappa\,d\kappa \tag{5.35}$$

[2]The factor arises from assuming that Ω is nearly Keplerian at the radius where the torque $g \propto \partial\Omega/\partial R = 0$ (Frank et al. 2002).

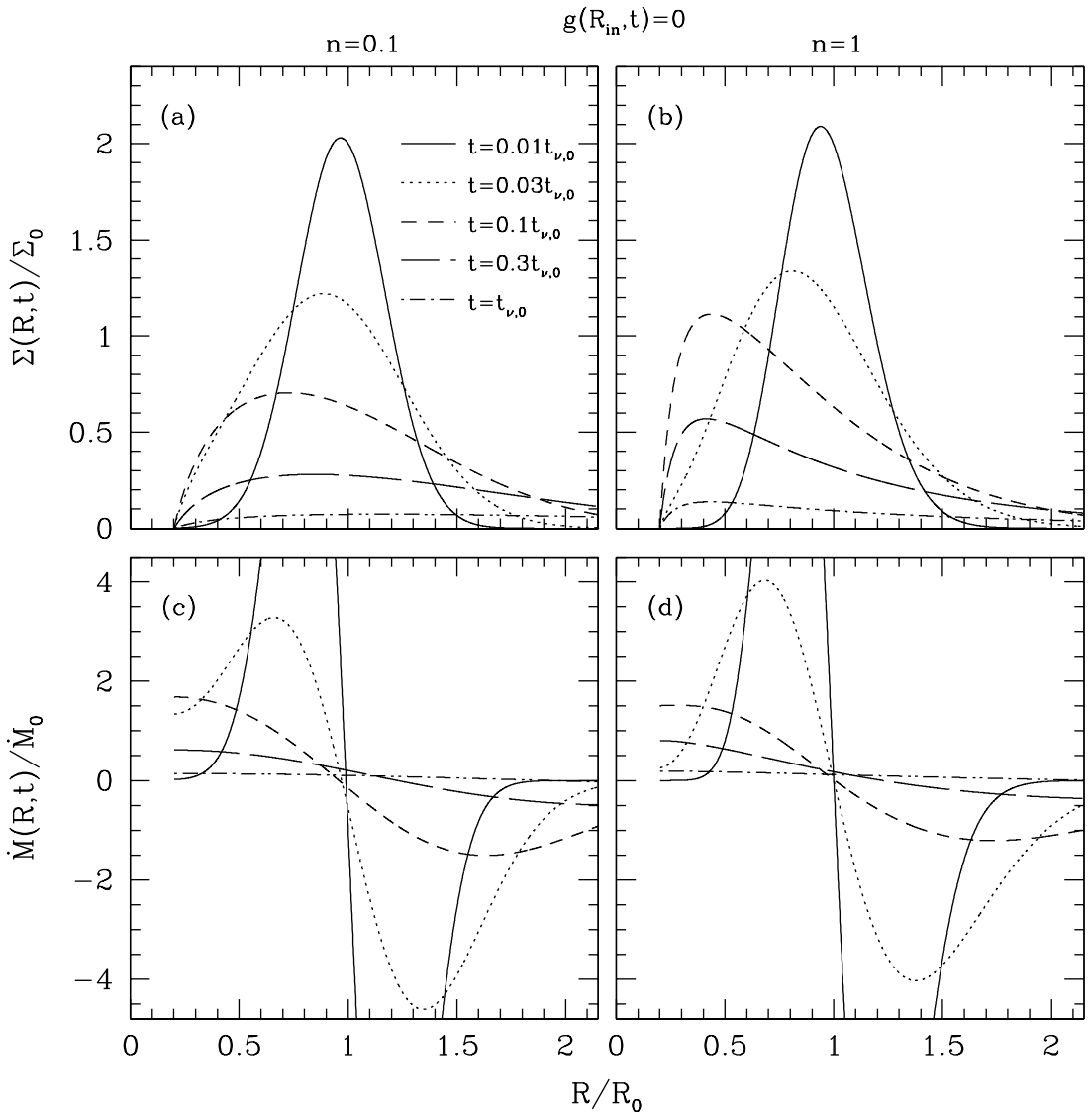

Figure 5.3 Same as as Figure 5.1, except that the zero-torque boundary condition is applied at a finite radius $R_{\rm in} = R_0/5$. As gas flows near the inner boundary, it exhibits the well-known behavior $\Sigma \propto R^{-n}(1 - \sqrt{R_{\rm in}/R})$ of LP74.

Pringle (1991) solved the special case $n = 1$ analytically, but noted the mathematical difficulty in deriving a solution for a more general case. We note that the mode weight $C(\kappa)$ can be obtained with the use of the recently proved generalized Weber transform (Zhang & Tong 2007),

$$\phi_\ell(x) = \int_0^\infty \frac{W_\ell(\kappa,x;a,b)}{Q_\ell^2(\kappa;a,b)}\,\Phi_\ell(\kappa)\,\kappa\,d\kappa, \tag{5.36}$$

$$\Phi_\ell(\kappa) = \int_1^\infty W_\ell(\kappa,x;a,b)\,\phi_\ell(x)\,x\,dx. \tag{5.37}$$

The functions $W_\ell(\kappa,x;a,b)$ and $Q_\ell^2(\kappa;a,b)$ are defined as follows:

$$\begin{aligned} W_\ell(\kappa,x;a,b) &\equiv J_\ell(\kappa x)\left[aY_\ell(\kappa)+b\kappa\,Y'_\ell(\kappa)\right]-Y_\ell(\kappa x)\left[aJ_\ell(\kappa)+b\kappa\,J'_\ell(\kappa)\right] \\ &= J_\ell(\kappa x)\left[(a-\ell b)\,Y_\ell(\kappa)+b\kappa\,Y_{\ell-1}(\kappa)\right]-Y_\ell(\kappa x)\left[(a-\ell b)\,J_\ell(\kappa)+b\kappa\,J_{\ell-1}(\kappa)\right] \end{aligned} \tag{5.38}$$

$$\begin{aligned} Q_\ell^2(\kappa;a,b) &\equiv \left[aY_\ell(\kappa)+b\kappa\,Y'_\ell(\kappa)\right]^2+\left[aJ_\ell(\kappa)+b\kappa\,J'_\ell(\kappa)\right]^2 \\ &= \left[(a-\ell b)\,Y_\ell(\kappa)+b\kappa\,Y_{\ell-1}(\kappa)\right]^2+\left[(a-\ell b)\,J_\ell(\kappa)+b\kappa\,J_{\ell-1}(\kappa)\right]^2. \end{aligned} \tag{5.39}$$

Above, J'_ℓ and Y'_ℓ are the ordinary derivatives of the Bessel functions. If $a = 1$ and $b = 0$, the pair is identical to the ordinary Weber transform (equations 5.26 and 5.27).

The choice $a = \ell$ and $b = 1$ corresponds to the desired boundary condition $\dot{M}(R_{\rm in},t) = 0$. The Green's function is then:

$$\begin{aligned} G(R,R',t) &= \left(1-\tfrac{n}{2}\right)R^{-n-1/4}R'^{5/4}R_{\rm in}^{n-2}\int_0^\infty \frac{[J_\ell(\kappa x)Y_{\ell-1}(\kappa)-Y_\ell(\kappa x)J_{\ell-1}(\kappa)][J_\ell(\kappa x')Y_{\ell-1}(\kappa)-Y_\ell(\kappa x')J_{\ell-1}(\kappa)]}{J_{\ell-1}^2(\kappa)+Y_{\ell-1}^2(\kappa)} \\ &\quad\times\exp\left[-2\left(1-\tfrac{n}{2}\right)^2\kappa^2\frac{t}{t_\nu(R_{\rm in})}\right]\,\kappa\,d\kappa. \end{aligned} \tag{5.40}$$

A specific instance of the above Green's function was derived by Pringle (1991) for the case $n = 1$. We can use equation 5.40 to reproduce that previous solution by noting that $J_{-1/2}(x) = -Y_{1/2}(x) = \sqrt{\pi/2}\, x^{-1/2} \cos x$ and $Y_{-1/2}(x) = J_{1/2}(x) = \sqrt{\pi/2}\, x^{-1/2} \sin x$. We obtain:

$$\begin{aligned} G(R, R', t) &= \frac{1}{\pi R_{\rm in}} \left(\frac{R'}{R}\right)^{5/4} (x\,x')^{-1/2} \int_0^\infty \cos\left[\kappa(x-1)\right] \cos\left[\kappa(x'-1)\right] \exp\left[-\frac{\kappa^2}{2} \frac{t}{t_\nu(R_{\rm in})}\right] d\kappa \\ &= \frac{R^{-3/2} R' R_{\rm in}^{-1/2}}{2\sqrt{2\pi}} \sqrt{\frac{t_\nu(R_{\rm in})}{t}} \left\{ \exp\left[-\frac{(x-x')^2}{2} \frac{t}{t_\nu(R_{\rm in})}\right] + \exp\left[-\frac{(x+x'-2)^2}{2} \frac{t}{t_\nu(R_{\rm in})}\right] \right\}. \end{aligned} \tag{5.41}$$

The only difference between this Green's function and the one for $n = 1$ and zero torque at $R_{\rm in}$ (equation 5.31) is the sign in between the exponential functions.

For general values of n, the analytic late-time behavior of equation 5.40 turns out to be identical to that for the case $R_{\rm in} = 0$ (equation 5.22). This can be confirmed by observing that for $\ell < 1$ and small arguments $\kappa \ll 1$ and $\kappa x \ll 1$, $W_\ell(\kappa, x; \ell, 1) \approx \csc(\ell\pi) J_{-\ell}(\kappa x) J_{\ell-1}(\kappa)$ and $Q_\ell^2(\kappa; \ell, 1) \approx \csc^2(\ell\pi) J_{\ell-1}^2(\kappa)$, and therefore the large fraction in equation 5.40 is approximately equal to $J_{-\ell}(\kappa x) J_{-\ell}(\kappa x')$.

Figure 5.4 shows the solution for the δ-function initial condition and the zero-flux boundary condition at $R_{\rm in} = R_0/5$. The panels showing the mass flow clearly exhibit the desired boundary condition. Note that the case $n = 1$ (panels b and d) is the case solved analytically by Pringle (1991). The $n = 1$ case, however, leads to a more rapid evolution and steeper late-time profiles than solutions with lower values for n; e.g., for the innermost regions of circumbinary disks around supermassive black holes, the viscosity is believed to be roughly constant with radius (Milosavljević & Phinney 2005; Tanaka & Menou 2010).

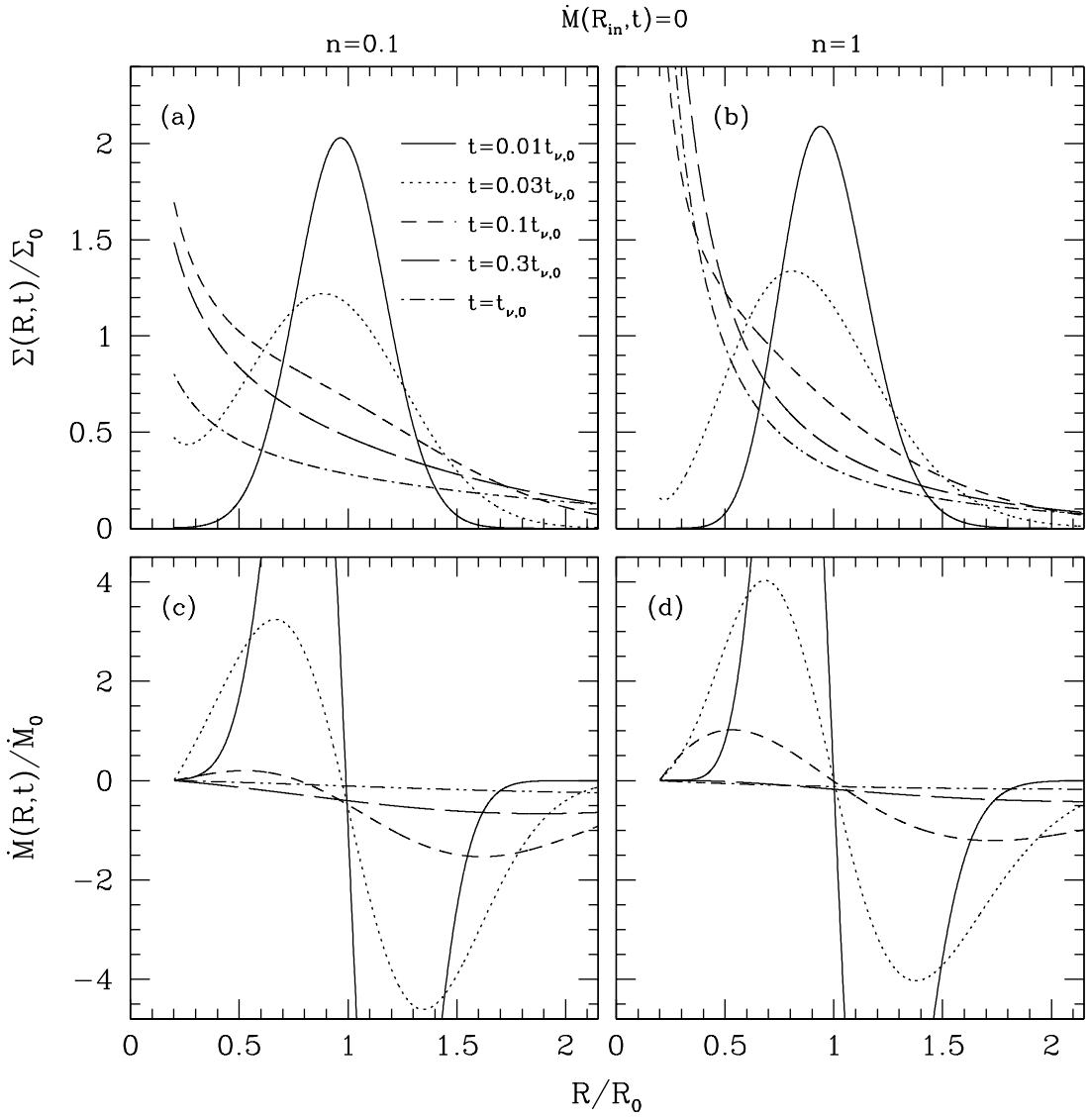

Figure 5.4 Same as as Figure 5.2, except that the zero-flux boundary condition is applied at a finite radius $R_{\rm in} = R_0/5$. Note that the $n = 1$ case was solved analytically by Pringle (1991).

5.4 Conclusion

We have presented Green's functions to the thin accretion disk equation, in the special case of a power-law viscosity profile $\nu \propto R^n$, for two different types of boundary conditions of either zero viscous torque or zero mass flow imposed at a finite inner radius $R_{\rm in} > 0$. They are extensions of the elegant analytic solutions derived by Lüst (1952) and LP74 for the same boundary conditions applied at $R_{\rm in} = 0$. While the problem of the finite-radius boundary had been mentioned previously in the literature, to the author's knowledge these solutions have not been explicitly pursued, and are presented here for the first time. The new solutions can be used to model the time-dependent behavior of the innermost regions of accretion disks, where the finite physical size of the central objects can significantly affect the observable characteristics of the disk. They complement the numerous approximate solutions and numerical treatments in the literature.[3]

The integral transforms used to derive the solutions are applicable to a wide class of boundary conditions, and may be applicable to astrophysical thin-disk systems and configurations not considered here. Because the generalized Weber transform by its nature is applicable to many second-order differential equations with intrinsic cylindrical symmetry, they may also prove to be useful in solving other mathematical equations in astrophysics and other fields.

[3]For example, Cannizzo et al. (1990) studied the accretion of a tidally disrupted star onto a black hole via numerical solutions and analytic self-similar solutions. The problem of a thin disk with $\dot{M} = 0$ at a finite radius was diskussed for the non-linear case $\nu \propto \Sigma^m \nu^n$ by Pringle (1991) and Ivanov et al. (1999), with both papers providing numerical solutions as well as analytic approximations.

Chapter 6

Electromagnetic Counterparts of Supermassive Black Hole Binaries Resolved by Pulsar Timing Arrays[1]

6.1 Introduction

Over the last several years, the possibility of observing both the gravitational-wave (GW) and electromagnetic (EM) emission signatures of coalescing supermassive black hole (SMBH) binaries has received intense attention (Holz & Hughes 2005; Kocsis et al. 2006, 2007, 2008; Dotti et al. 2006; for an overview of proposed mechanisms for EM signatures, see Haiman et al. 2009b; Schnittman 2011). The bursts of GWs emitted by such systems

[1]This chapter is a reformatted version of an article by the same name by T.Tanaka, K.Menou and Z.Haiman that was submitted in July to the Monthly Notices of the Royal Astronomical Society. The abstract for this article can be found in Section 1.6.5.

can now be predicted by numerical general relativity (Pretorius 2005; Baker et al. 2006a; Campanelli et al. 2006), and are expected to be observed by current and future detectors. Of particular interest is the temporal evolution of the gravitational waveform, which can be used to extract the luminosity distance, help constrain the location of the source on the sky, and determine the masses and spins of the SMBHs. If an EM signature of the coalescence can also be identified, this would allow for a determination of the source redshift, turning merging black holes into "standard sirens" for probing cosmic expansion.[2] Such multi-messenger observations would also enable astronomical investigations of SMBHs whose masses, spins and orbital parameters are already known, presenting ideal laboratories for investigating accretion physics in active galactic nuclei (AGN). Furthermore, if major mergers of galaxies trigger luminous AGN activity (e.g., Sanders et al. 1988; Hernquist 1989; Carlberg 1990; Barnes & Hernquist 1991; Hernquist & Mihos 1995; Mihos & Hernquist 1996; Kauffmann & Haehnelt 2000; Hopkins et al. 2007a, 2008), then the characteristic EM emission promptly following the SMBH coalescence may herald the birth of a quasar (Tanaka et al. 2010).

To date, theoretical studies of EM signatures of GW-emitting SMBH binaries have largely centred on systems expected to be detected by the planned space-borne observatory *LISA*, i.e. systems with total mass $\sim 10^{5-7}(1+z)^{-1}M_{\odot}$ out to $z \gtrsim 10$. A paramount feature of *LISA* for multi-messenger astronomy is the precision with which it is expected to determine the sky position of SMBH sources (Kocsis et al. 2006): to $\lesssim 1$ deg (Vecchio 2004; Lang & Hughes 2008), or perhaps even to $\lesssim 1'$ when spin-induced precession (Lang

[2]The importance of such GW+EM observations for cosmography was first diskussed by Schutz (1986) in the context of merging neutron star binaries.

& Hughes 2006) or higher-order harmonics (McWilliams et al. 2010) are included in the analysis of the waveform.

In this paper, we evaluate the prospects of electromagnetically identifying GW-emitting SMBH binaries that are individually resolved by pulsar timing arrays (PTAs). PTAs will detect GWs from compact SMBH binaries that are more massive (total mass $M \gtrsim 10^7 M_\odot$) and less compact (observer's-frame period of $P_{\rm obs} \sim 1$ yr) than those detectable by *LISA*. PTA detections will be comprised mostly of the collective background of the SMBH binary population at low redshift ($z \lesssim 1.5$). However, theoretical population-synthesis studies (Sesana et al. 2009; Sesana & Vecchio 2010; Kocsis & Sesana 2011) predict that the GW signals from the most massive and/or most nearby PTA sources will stick out above the background, and be individually resolved. According to those studies, resolved sources have chirp masses

$$\mathcal{M} \equiv M_1^{3/5} M_2^{3/5} M^{-1/5} = \eta^{3/5} M, \tag{6.1}$$

typically around $\mathcal{M} \sim 10^{8.5} M_\odot$. Above, $M_2 \leq M_1$ are the masses of each member of the binary, and $\eta \equiv M_1 M_2 / M^2 \leq 1/4$ is the symmetric mass ratio. The redshift probability distribution is poorly known, but is expected to decline steeply outside the range $0.1 \lesssim z \lesssim 1.5$, owing to the small volume at lower z and to the attenuation of the GW signal, as well as to the decline of the intrinsic SMBH merger rate at higher z. The orbits of PTA sources decay slowly, and most will not coalesce within a human lifetime – determining the masses and luminosity distances for the majority of sources will present a particularly difficult challenge. Another major challenge is the relatively poor sky-localization accuracy, which

is expected to be $\Delta\Omega \lesssim 3\,\mathrm{deg}^2$ (if the contributions to the signal from individual pulsars are known; Corbin & Cornish 2010; hereafter CC10) to as large as $\Delta\Omega \sim 40\,\mathrm{deg}^2$ (if individual pulsar contributions cannot be extracted from the data; Sesana & Vecchio 2010; hereafter SV10).

In this paper, we investigate whether individually resolved PTA sources may be viable targets for EM identification. In particular, we address the following two questions:

1. What is the average number N_g of candidate host galaxies — that is, interloping galaxies that could plausibly harbor the GW source — in a typical error box of a PTA detection? Of particular interest is whether there are plausible scenarios for detecting individually resolved sources with $N_g < 1$, i.e., cases where the source may be uniquely identified with an EM search of the three–dimensional PTA error box.

2. In cases where $N_g > 1$, what can be done to distinguish the true host galaxy of the source from the other interlopers? Motivated by the hypothesis that galaxy mergers can fuel AGN activity, we will consider the differences in thermal emission properties predicted by disk models of AGN powered by a compact SMBH binary as opposed to one powered by a solitary SMBH of the same total mass.

It is important to note that because most PTA sources will not merge within a human lifetime, many of the mechanisms hypothesized to elicit EM signatures for SMBH coalescences (e.g., gravitational recoil effect, loss of binary mass via GW emission, post-merger tidal response of a circumbinary accretion disk) are not applicable for PTA sources. Periodic emission modulated at the orbital frequency of the PTA source (e.g., Haiman et al. 2009b and references therein) and double-peaked broad emission lines (e.g., Gaskell 1996;

Zhou et al. 2004; Bogdanović et al. 2008; Blecha & Loeb 2008; Boroson & Lauer 2009; Bogdanović et al. 2009; Dotti et al. 2009; Tang & Grindlay 2009; Shen & Loeb 2010) have been previously considered in the literature as possible tell-tale EM features of compact SMBH binaries.

This paper is organized as follows. In §6.2, we provide a brief overview of the expected population of PTA-resolved SMBH binaries as well as the anticipated detection error box of such objects. We consider the error box prescriptions of SV10 and CC10, characterize the types of astronomical objects that are plausible hosts of a PTA-resolved binary, and estimate the number of such objects. In §3, we describe a toy model to calculate the dynamical state and thermal emission features of gas accreting onto a resolved SMBH binary. We diskuss several features predicted by the model which, if observed, could help the EM identification of individually resolved PTA sources. We summarize our findings and offer our conclusions in §4.

Throughout this paper, c denotes the speed of light; G is the gravitational constant; h is Planck's constant; $k_{\rm B}$ is the Boltzmann constant; $m_{\rm p}$ is the mass of the proton; and $\sigma_{\rm SB}$ is the Stefan-Boltzmann constant.

6.2 Plausible Hosts of PTA-resolved Binaries

As stated in §6.1, theoretical models predict that SMBH binaries individually resolved by PTAs are most likely to have masses of $M \gtrsim 10^8 M_\odot$, observed periods of $P_{\rm obs} \sim 1$ yr, and lie in a redshift range $0.1 \lesssim z \lesssim 1.5$. Given the PTA detection of such a GW-source binary, we wish to evaluate the number N_g of candidate galaxies that could plausibly host it. To

this end, we will first review the volume of the error box in which we must look for the source, based on previous work on the source localization capability of PTAs. Then, we will evaluate the number of interloping host galaxies in the error box by estimating the number of (i) sufficiently massive dark matter halos, (ii) sufficiently luminous luminous galaxies, and (iii) AGN. We adopt a standard ΛCDM cosmology with $h = 0.70$, $\Omega_{\rm m} = 0.27$, $\Omega_\Lambda = 1 - \Omega_{\rm m}$, $\Omega_{\rm b} = 0.046$, and $\sigma_8 = 0.81$ (WMAP 7-year results, Jarosik et al. 2011).

6.2.1 The PTA Error Box

We consider two different estimates of the size of the error box of PTA-resolved sources. The first is based on the calculations by SV10, who assumed that the contributions to the signal from GW perturbations at the individual pulsars (the so-called "pulsar term") cannot be extracted from the PTA data. The location on the sky of a resolved source can then be determined within an error of $\Delta\Omega \sim 40\,{\rm deg}^2({\rm SNR}/10)^{-2}$, where SNR is the signal-to-noise ratio. The error of the signal amplitude $A \propto \mathcal{M}^{5/3} D_L^{-1}$ will be of order $\sim 30\%$. Given the wide spread in chirp mass distribution predicted by population synthesis models of resolved sources, in the absence of an independent measurement of $\mathcal{M}$ the only constraint on D_L comes from the maximum distance at which PTAs are expected to detect individually resolved sources. The population synthesis models (Sesana et al. 2009; Kocsis & Sesana 2011) predict that the majority of resolved sources will lie below a maximum redshift $z_{\rm max} \sim 1.5$, or a luminosity distance below $D_{L,{\rm max}} \lesssim 10^4$ Mpc. This "worst-case"

error box has a comoving volume of

$$\Delta V^{(\mathrm{SV})} \sim 3 \times 10^8 \left(\frac{\Delta\Omega}{40\,\mathrm{deg}^2} \right) \mathrm{Mpc}^3. \tag{6.2}$$

More optimistic numbers are obtained by CC10, who suggested that utilizing information on the distances to individual pulsars in the array can greatly enhance the measurement capabilities of PTAs. They concluded that if the individual pulsar term can be extracted from the signal, then this would double the signal power and enable direct measurement of the chirp mass. They estimate that for a system with SNR= 20 (corresponding to a detection of SNR= 10 without pulsar distance information), a resolved source can be localized with distance and angular errors of $\Delta D_L/D_L < 20\%$ and $\Delta\Omega < 3\,\mathrm{deg}^2$, respectively. Noting that the comoving distance $D(z)$ in the relevant redshift range can be analytically approximated[3] as $D \approx cH_0^{-1} z\,(1-0.2z)$, we may estimate the error box in the CC10 scenario as

$$\Delta V^{(\mathrm{CC})} \approx 1.2 \times 10^6\, z^3 (1-0.2z)^3 \left(\frac{\Delta\Omega}{3\,\mathrm{deg}^2} \right) \left(\frac{\Delta D_L/D_L}{20\%} \right) \mathrm{Mpc}^3. \tag{6.3}$$

6.2.2 Interloper Counts

Where does an individually resolved PTA source live? The mass M of a nuclear SMBH is known to correlate with the velocity dispersion σ of the host galaxy (the "$M-\sigma$ relation"; Ferrarese & Merritt 2000; Gebhardt et al. 2000; Tremaine et al. 2002), as well as with

[3]This fitting formula has an error of less than 1% in D at $z \leq 1.4$ and roughly 5% at $z = 1.9$. It is provided for the reader's convenience; all distance and volume calculations in this paper are performed using exact expressions.

the stellar luminosity of the host (the "$M - L$ relation"; Kormendy & Richstone 1995; Magorrian et al. 1998; Häring & Rix 2004; Lauer et al. 2007); with more massive halos and luminous galaxies hosting more massive SMBHs. That resolved PTA sources are expected to be exceptionally massive ($M \gtrsim 10^8 M_\odot$) implies that the host should be a giant elliptical galaxy or be among the most massive spiral galaxies (with velocity dispersion $\sigma \gtrsim 200$ km s^{-1} of the spheroid component; e.g.Gültekin et al. 2009).

Assuming that SMBH binaries are able to overcome the "final parsec" problem (e.g., Escala et al. 2005; Mayer et al. 2007; Callegari et al. 2009; Colpi et al. 2009; Hayasaki 2009; see, however, Lodato et al. 2009), we expect the PTA host galaxy to be the product of a relatively recent merger. A natural question to ask is whether such galaxies typically lie in the field, or in the centre of a cluster. We can answer this question qualitatively by considering the dependence of the major merger rates of the most massive dark matter halos on their environments. Analyses by Fakhouri & Ma (2009) and Bonoli et al. (2010) of the Millennium simulation results (Springel et al. 2005) indicate that while the rate of major mergers is enhanced in over-dense environments, this effect is weak: for halo masses and redshifts of interest ($M \gtrsim 10^{13} M_\odot$ and at $z \lesssim 1.5$), the ratio of merger rates between the most and least over-dense regions is of order unity. We interpret this result to mean that there is no strong reason to search for PTA sources in galaxy clusters as opposed to those in the field.

6.2.2.1 The most massive halos

One conservative way to estimate the number N_g of host galaxy candidates in the error box is to simply count the dark matter halos that are massive enough to plausibly harbour the source SMBH binary. We use the observational results of Dutton et al. (2010), who infer a double-power-law fit for the relation between the SMBH mass M and the host halo mass $M_{\rm halo}$ for local elliptical galaxies. We extrapolate their results to higher redshifts by postulating the canonical z-dependence based on the theory for spherically collapsing halos (see, e.g., Wyithe & Loeb 2003a),

$$M_{\rm halo}(M,z) \propto F(z) \equiv \sqrt{\frac{d(z)}{d(0)}\frac{\Delta_{\rm c}(0)}{\Delta_{\rm c}(z)}}, \tag{6.4}$$

where $d(z) = -[(\Omega_{\rm m}/\Omega_\Lambda)(1+z)^3 + 1]^{-1}$ and $\Delta_{\rm c}(z) = 18\pi^2 + 82d(z) - 39d^2(z)$. We obtain

$$M_{\rm halo} \approx 2.3\times10^{13} M_9^{0.75}\left[\frac{1}{2} + 13M_9^{1.77}\right]^{0.50} F(z)\, M_\odot, \tag{6.5}$$

where $M_9 \equiv M/(10^9 M_\odot)$. We estimate the number of candidate host halos inside the three-dimensional PTA error box by integrating the halo mass function of Jenkins et al. (2001; their equation 9) above $M_{\rm halo}$.

The most massive halos with $M_{\rm halo} \gtrsim$ few $\times 10^{14} M_\odot$, which are associated with galaxy clusters, may be expected to contain more than one plausible host galaxy. Since the halo mass function at $M_{\rm halo} \gtrsim$ few $\times 10^{14} M_\odot$ drops much more steeply than linear with mass, whereas the sub-halo mass function increases less steeply than linear (Giocoli et al. 2010), most galaxies with halos masses $\sim 10^{13} M_\odot$ will reside in the field, rather than in groups

and clusters. The multiple occupancy of massive galaxies in the most massive halos will then represent only a small increase in our total counts of interlopers. As the purpose of the exercise in this section is to give order-of-magnitude estimates for interlopers, we will neglect sub-halos in our analysis.

6.2.2.2 The brightest galaxies

A second way to estimate the number of candidate host galaxies is through the $M - L$ relation, where L is the luminosity of the host galaxy. Of particular interest is the fact that the $M - L$ relation and the $M - \sigma$ relations are diskrepant at the high-mass end (here σ denotes the velocity dispersion of the host). The former predicts higher masses for the most massive SMBHs, and higher number densities for fixed BH mass (Lauer et al. 2007, and references within). This therefore results in a greater number of individually resolvable PTA sources (Sesana et al. 2009).

Since σ is used to infer $M_{\rm halo}$, we expect that for a fixed SMBH mass, the number of expected interloping host galaxies, inferred from the $M - L$ relation, would also be greater than the number of halos, inferred from the $M - \sigma$ and $\sigma - M_{\rm halo}$ relations.

To evaluate this different estimate quantitatively, we adopt the $M - L$ relation found by (Lauer et al. 2007) for the most luminous core galaxies in their sample,

$$M_V \approx -22.0 - 1.8 \log_{10} M_9, \tag{6.6}$$

where M_V is the V-band magnitude of the host galaxy.

To compute the number of sufficiently luminous galaxies, we use the results of

Gabasch et al. (2004, 2006), who measured the luminosity function in multiple wavelength bands between $150-900$ nm, and studied the redshift evolution in each band out to $z \gtrsim 2$. The luminosity function is given in the form of a standard Schechter function (Schechter 1976),

$$\phi(M_V) = \frac{2}{5}(\ln 10)\phi^* \left[10^{(2/5)(M_V^*-M_V)}\right]^{\alpha_0+1} \exp\left[-10^{(2/5)(M_V^*-M_V)}\right]. \tag{6.7}$$

They set a constant value for the parameter α_0 while fitting M_V^* and ϕ_* to a power-law redshift dependence of the form

$$M_V^*(z) = M_{V,0}^* + A\ln(1+z), \tag{6.8}$$

$$\phi^*(z) = \phi_0^*(1+z)^B. \tag{6.9}$$

The five fitting parameters $(\alpha_0, M_{V,0}^*, \phi_0^*, A, B)$ vary with the wavelength band of the luminosity function. Because the Gabasch et al. results do not have fits for the V-band, we interpolate the parameters between neighboring bands to obtain the following values: $\alpha_0 \approx -1.3$, $M_{V,0}^* \approx -21.1$, $\phi_0^* \approx 6.2\times10^{-3}$ Mpc$^{-3}M_V^{-1}$, $A \approx -1.18$, and $B \approx -1.05$.

6.2.2.3 The brightest AGN

Finally, a third method to identify plausible hosts is to search for AGN that are luminous enough to be plausibly powered by a $M \sim 10^9 M_\odot$ SMBH. AGN activity is an ideal scenario for identifying the EM counterparts of PTA sources, as the interaction between a compact SMBH binary and its accretion flow provide a natural physical mechanism for eliciting a smoking-gun EM signature. However, a significant uncertainty with this approach is

whether the host of a resolved PTA source is likely to be undergoing an observable AGN episode. While multiple studies have suggested that galaxy mergers trigger AGN activity (refs. in §6.1) whether the two phenomena are causally related remains an open question.

Recently, Schawinski et al. (2011) suggested that the low Sersic indices in most X-ray-selected AGN hosts at $1.5 < z < 3$ indicate that they are disk galaxies, and therefore unrelated to mergers (see, however, Governato et al. 2009, who suggest mergers can result in disk galaxies). Further, even if one accepts that there exists a direct causal connection between galaxy mergers and luminous AGN activity, it is uncertain whether such a trend extends to the most massive galaxies at $z < 1.5$. The mass fraction of cold gas in massive galaxies tend to decrease toward lower redshift, and gas-poor "dry" mergers are thought to play an important (if not dominant) role in the assembly of giant elliptical galaxies at $z < 1$, in the field as well as in clusters (van Dokkum 2005; Lin et al. 2008, 2010). On the other hand, the amount of gas required to fuel a luminous AGN episode is a small fraction of the total gas content of even very gas-poor galaxies; the most luminous known AGN are situated in giant elliptical galaxies; a plurality of mergers of massive galaxies at $z < 1$ are gas-rich (Lin et al. 2008); many early-type galaxies identified as undergoing a dry merger have been found to contain detectable amounts of gas in followup HI observations (e.g., Donovan et al. 2007; Sánchez-Blázquez et al. 2009); and even though the hosts of the most luminous quasars tend to be ellipticals, they are not exclusively so in the SMBH mass regime of interest ($M \gtrsim 10^8 M_\odot$; e.g., Percival et al. 2001; Floyd et al. 2004; Zakamska et al. 2006, and references therein). We conclude that PTA sources powering luminous AGN activity is a plausible scenario, and not merely an expedient assumption.

We parameterize the minimum luminosity for the AGN counterpart in terms of the Eddington luminosity $L_{\rm Edd}(M) = 4\pi GM\,\mu_e m_{\rm p} c/\sigma_{\rm T}$, where μ_e is the mean molecular weight per electron and $\sigma_{\rm T}$ is the Thomson cross section:

$$L_{\rm min}(M) = f_{\rm min} L_{\rm Edd}\,(M)\,. \tag{6.10}$$

We choose $f_{\rm min} = 10^{-2}$ for our minimum Eddington ratio $L/L_{\rm Edd}$, motivated by the fact that this quantity is observed to peak at $L/L_{\rm Edd} \sim 0.1 - 0.3$ (e.g., Kollmeier et al. 2006).

In order to estimate the number of AGN that are bright enough to correspond to a PTA source with mass M, we adopt the observationally motivated fits to the AGN luminosity function given by Hopkins et al. (2007b; their equations 6, 18 and 20, and Table 3).

Note that although the Eddington ratio distribution and the luminosity function cited above are expressed in terms of the bolometric luminosity, they are actually proxies for the optical luminosity. Hopkins et al. (2007b) noted that their bolometric luminosity function is effectively equivalent to the optical luminosity function, and Kollmeier et al. (2006) uses the flux at 510 nm to estimate the bolometric luminosity. Our AGN interlopers are therefore optically luminous AGN, and we assume nothing a priori about the X-ray and UV emission of accreting PTA sources. We will diskuss the importance of searching for the EM counterpart at optical wavelengths in §6.3.3.

6.2.3 Expected Counts of Interloping Galaxies

If the individual pulsar contributions to the signal cannot be extracted, as in the SV10 scenario, then the chirp mass and luminosity distance of the resolved source cannot be independently known, and the source can only be localized within a solid angle $\Delta\Omega$. The only constraints on $\mathcal{M}$ and D_L are then model-dependent, and come from theoretical expectations for the population of resolvable sources, given the detection threshold of the array. The upper end of the chirp mass distribution of SMBH binaries, along with the detector sensitivity, sets a maximum luminosity distance $D_{L,\max}$ (equivalently, $z_{\max}$). Similarly, the chirp mass distribution in the local Universe determines a minimum chirp mass $\mathcal{M}_{\min}$ required for a PTA source to be resolved. Note that since $M = \eta^{-3/5}\mathcal{M} \geq 2^{6/5}\mathcal{M}$, the quantity $\mathcal{M}_{\min}$ also sets a lower limit $M_{\min} \approx 2.3\mathcal{M}_{\min}$ on the gravitational mass.

The number of interloping halos can be expressed as

$$N_{\rm halo}^{\rm (SV)} = \frac{\Delta\Omega}{4\pi} \int_0^{z_{\max}} \int_{M_{\rm halo}(M_{\min},z)}^{\infty} \frac{dn_{\rm halo}}{dM_{\rm halo}}\, dM_{\rm halo}\, \frac{dV}{dz}\, dz, \tag{6.11}$$

where n is the comoving number density of dark matter halos, and $dV/dz = 4\pi D_L^2\, dD_L/dz$ is the comoving volume element. The lower limit of the integral over halo mass is given by equation 6.5. Similarly, the numbers of interloping galaxies and AGN are given by the expressions

$$N_{\rm gal}^{\rm (SV)} = \frac{\Delta\Omega}{4\pi} \int_0^{z_{\max}} \int_{-\infty}^{M_V(M_{\min},z)} \phi\, dM_V\, \frac{dV}{dz}\, dz, \tag{6.12}$$

$$N_{\rm AGN}^{\rm (SV)} = \frac{\Delta\Omega}{4\pi} \int_0^{z_{\max}} \int_{L_{\min}(M_{\min})}^{\infty} \frac{dn_{\rm AGN}}{dL}\, dL\, \frac{dV}{dz}\, dz. \tag{6.13}$$

The limits of integration over luminosity are taken from equations 6.6 and 6.10.

We show in Figure 6.1 the estimated number of interlopers for the worst-case error box in the SV10 scenario, assuming $\Delta\Omega = 40\,\mathrm{deg}^2$, as a function of the maximum redshift $z_{\max}$ and minimum BH binary mass $M_{\min}$. Panels (a), (b) and (c) show the isonumber contours of the expected number of interloping massive halos, luminous galaxies and luminous AGN, respectively. All three methods to estimate the number of interlopers yield on the order of $N_g \gtrsim 10^2$ for PTA sources with $M > 10^9 M_\odot$, if the redshift range is restricted to $z_{\max} \sim 1$.

The difference in the number of interlopers between the top two panels (halos vs. galaxies) for the most massive SMBHs arises because the observed SMBH samples yield an internally inconsistent set of $M-\sigma$, $L-\sigma$ and $M-L$ relations, as mentioned above. While the interpretation of this inconsistency is beyond the scope of our paper, we note that Tundo et al. (2007) diskussed this issue, and concluded that the intrinsic scatter in the relations produces a selection bias: using the observed BH samples yields a biased $L-\sigma$ relation (too low L for given σ). This suggests that the $M-L$ galaxy relation we adopted may also be biased and it under-predicts L; correcting this bias would decrease the number of galaxy interlopers.

If the GW signal can be used to constrain $\mathcal{M}$ and D_L of the source via statistical inference, as suggested by CC10, then the numbers of interloping halos, luminous galaxies and AGN are given by

$$N_{\mathrm{halo}}^{(\mathrm{CC})}(z) = \frac{\Delta\Omega}{4\pi}\int_{z_-}^{z_+}\int_{M_{\mathrm{halo}}(M_{\min},z)}^{\infty}\frac{dn_{\mathrm{halo}}}{dM_{\mathrm{halo}}}\,dM_{\mathrm{halo}}, \tag{6.14}$$

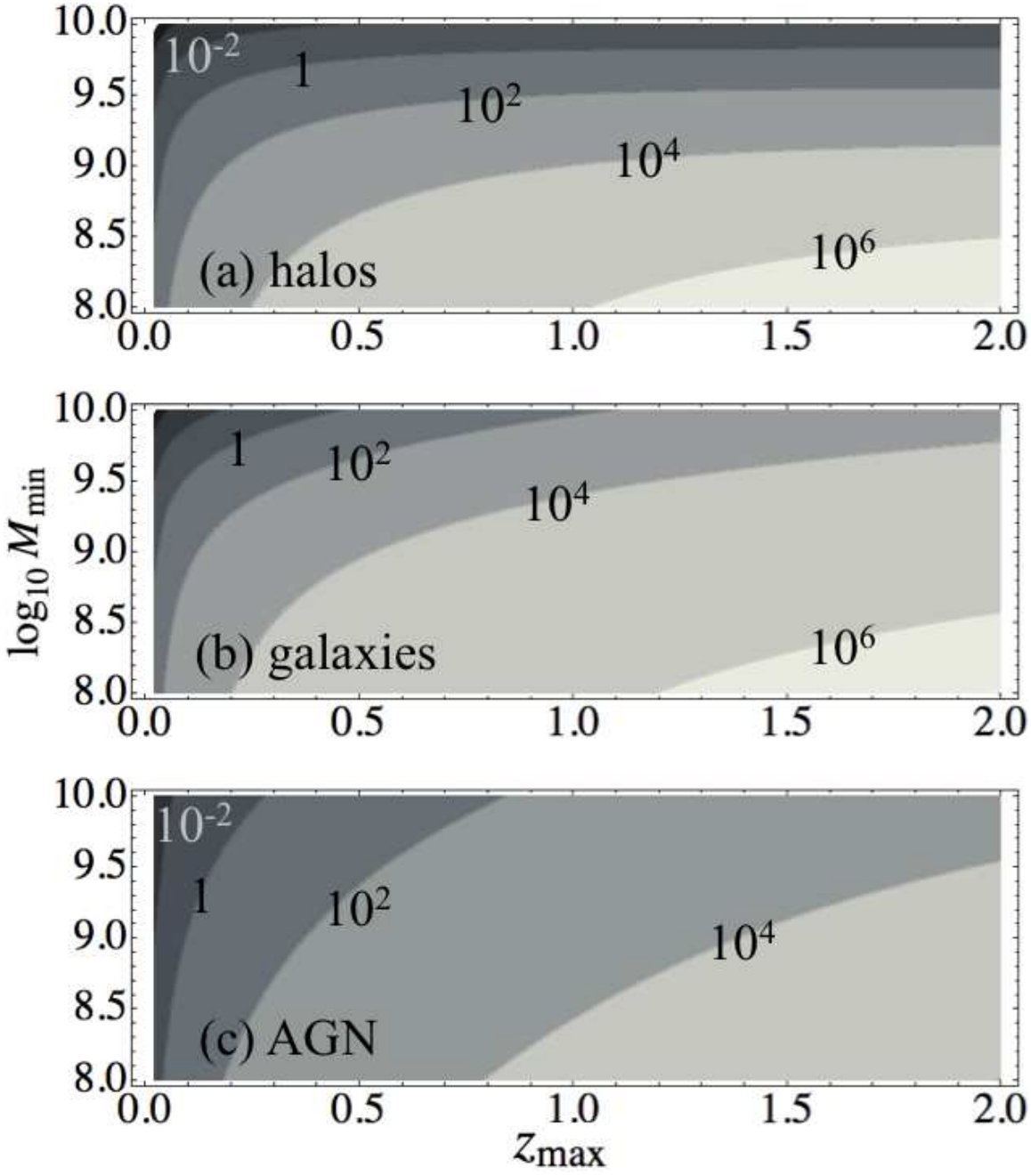

Figure 6.1 Estimates of the number of interloping host objects — (a) massive dark matter halos, (b) luminous galaxies, and (c) luminous AGN — in the conical error volumes suggested by SV10. The extent of the error volume is limited by $z_{\rm max}$, the maximum redshift at which PTAs can resolve an individual source, and the angular localization $\Delta\Omega = 40\,{\rm deg}^2$. The number of interlopers is calculated by assuming a minimum SMBH mass $M_{\rm min}$, which then sets the minimum host mass/luminosity through equations 6.5, 6.6, and 6.10.

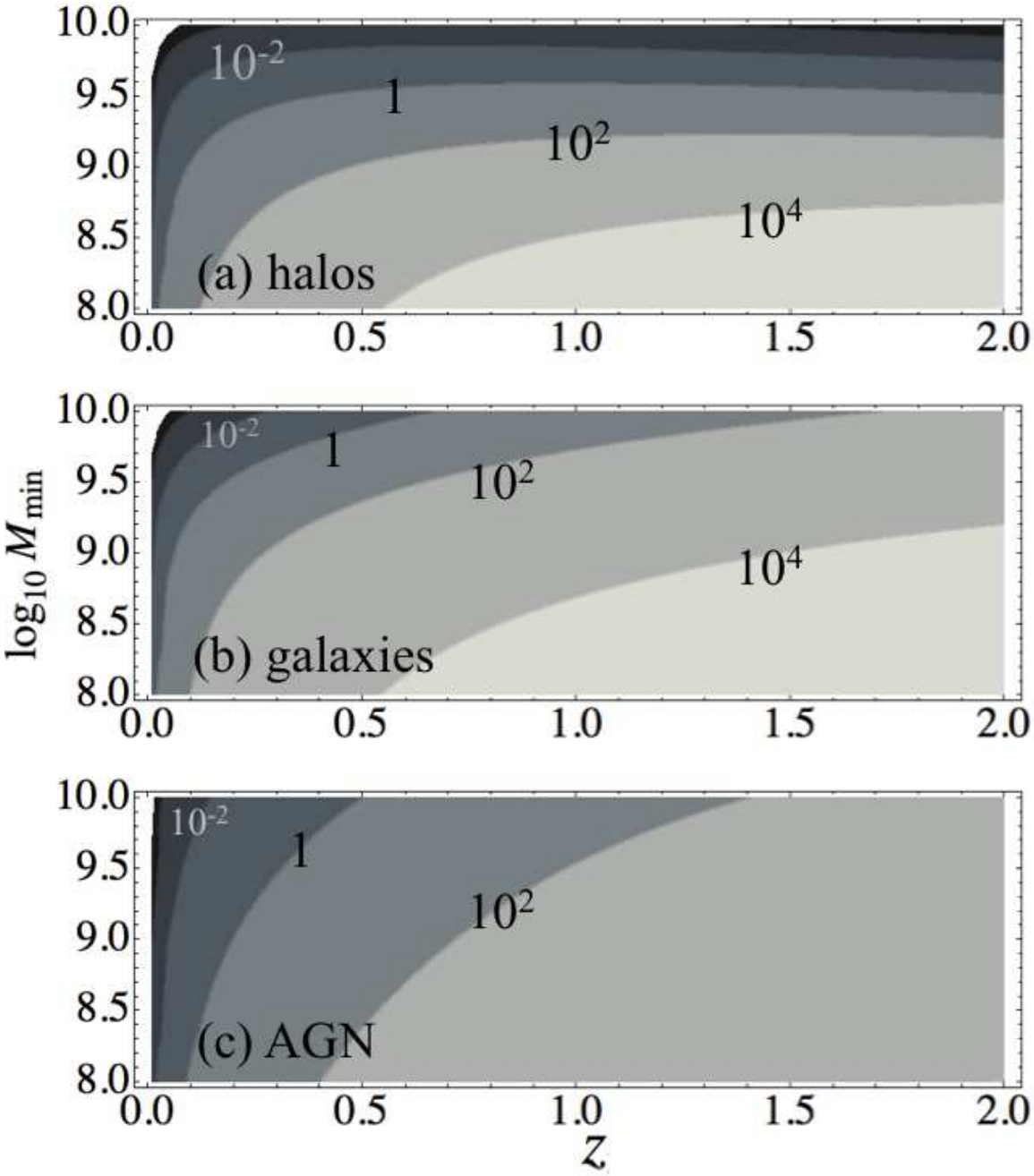

Figure 6.2 Same as Figure 6.1, except that the error volume is calculated from the results of CC10, who assumed that the pulsar term of the GW signal can be used to infer the luminosity distance to the source binary. The error box is limited by the uncertainty $\Delta D_L/D_L = 20\%$ in the luminosity distance to the source, and the angular localization $\Delta\Omega = 3\,\mathrm{deg}^2$. Note that whereas the horizontal axis in Figure 6.1 showed the maximal PTA detection range $z_{\rm max}$, here it denotes the actual redshift z of the source.

$$N_{\rm gal}^{\rm (CC)}(z) = \frac{\Delta\Omega}{4\pi}\int_{z_-}^{z_+}\int_{-\infty}^{M_V(M_{\rm min},z)}\phi\, dM_V\,\frac{dV}{dz}\,dz, \quad (6.15)$$

$$N_{\rm AGN}^{\rm (CC)}(z) = \frac{\Delta\Omega}{4\pi}\int_{z_-}^{z_+}\int_{L_{\rm min}(M_{\rm min})}^{\infty}\frac{dn_{\rm AGN}}{dL}\,dL, \quad (6.16)$$

respectively. Above, the redshifts $z_\pm = z(D_L \pm \Delta D_L)$ bound the radial extent of the error box. We adopt $\Delta\Omega = 3\,{\rm deg}^2$ and $\Delta D_L/D_L = 20\%$. We ignore errors due to weak lensing, which are expected to be on the order of several percent for sources with $z \lesssim 1.5$ (Kocsis et al. 2006; Hirata et al. 2010; Shang & Haiman 2011). We do not place an upper limit on the host halo mass (or on the host galaxy luminosity). In principle, such an upper limit could be computed, given PTA's observational error on the chirp mass and the spread in the ratio between the chirp mass and the gravitational mass of the binary (i.e., from the model-dependent mass ratio distribution of resolved sources). For example, CC10 provide a chirp mass error estimate of $\Delta\mathcal{M} \sim 5\%$. Converting the chirp mass to the gravitational mass, however, can introduce a large uncertainty, e.g. a factor of ~ 2 depending on whether the mass ratio is 0.1 or 1. Since the number density of interlopers decrease rapidly with increasing halo mass (or luminosity), this simplification should not affect our estimates.

In Figure 6.2, we plot the number of interloping host candidates against the source redshift z. Not surprisingly, the prospects for EM identification improve dramatically in the CC10 scenario. For massive ($M \gtrsim 10^9 M_\odot$) resolved PTA sources, we anticipate that the error box will contain a single host candidate at $z \lesssim 0.2$, and several hundred at $z \lesssim 0.7$. We expect only a single group-sized halo ($M \gtrsim$ few $\times\, 10^{13} M_\odot$) in the error box at any redshift in the CC10 scenario. Note that the number N_g of interlopers is not

necessarily a monotonically increasing function of z, as the decline in the number densities of the interloping objects (in particular massive halos) competes with the increase in the comoving size of the error boxes.

Our simple calculations show that in the scenario of CC10, resolved PTA sources with $M \gtrsim 10^9 M_\odot$ and $z \lesssim 0.5$ are likely to have at worst dozens of interlopers in the error box. With this low number, one could conceivably perform follow-up observations of each individual candidate. If, on the other hand, luminosity distances to the source cannot be determined, this number increases to $\sim 10^3$, suggesting that it will become extremely difficult to electromagnetically identify the source in the absence of an obvious, tell-tale EM signature.

In practice, the number of interloping galaxies may be somewhat larger than the value computed by equations 6.11−6.16. The halo mass of any given candidate host system will not be known a priori, and the intrinsic scatter in the $M_{\rm SMBH} - M_{\rm halo}$ ($M_{\rm SMBH} - \sigma_{\rm host}$) relation will lower the minimum halo mass threshold for candidacy. On the other hand, the simple calculations presented here do not consider detailed demographic properties of resolved PTA sources and plausible hosts, such as the presence of a nuclear stellar core (Makino 1997; Ravindranath et al. 2002; Milosavljević et al. 2002; Volonteri et al. 2003b) or galaxy morphology. Including such factors in the analysis will narrow the field of candidate hosts.

As we argue in 6.3.3, candidate AGN counterparts may be further vetted by examining their UV and X-ray emission for features indicative of a central SMBH binary (see also Sesana et al. 2011 for an in-depth diskussion of possible high-energy signatures

for pre-decoupling — i.e., $t_{\rm GW} > t_\nu$ — PTA sources). In addition, PTA sources are sufficiently nearby that it should be possible to observe an interloping AGN together with its host galaxy. It should therefore be possible to combine the AGN emission, the galaxy luminosity and the inferred SMBH mass to cross-check candidate counterparts.

6.3 Accretion disks Around PTA-source Binaries

Motivated by the results of the previous section that the number of plausible host galaxies in the PTA error box may be tractable for follow-up EM searches, we next model the EM emission properties of SMBH binaries detectable by PTAs. We focus our attention on SMBH binaries that are undergoing luminous accretion, as these are the most promising class of objects for EM identification.

Normalizing the binary mass M and rest-frame period P to the typical orders of magnitude expected of resolved PTA sources, $M = 10^9 M_\odot M_9$ and $P = 1$ yr P_1, we write the semi-major axis for the source binary as

$$a(M,P) = 101\, M_9^{-2/3} P_1^{2/3} \frac{GM}{c^2} = 2.23 \times 10^{-2} M_9^{1/3} P_1^{2/3} \text{ pc}. \tag{6.17}$$

Binaries detectable by PTAs have long overcome the so-called "final parsec" problem. The rest-frame time to merger for a binary with mass M and semi-major axis a, driven by GW emission alone, is

$$t_{\rm merge} \leq \frac{5}{256} \frac{c^5}{G^3 M^3} \frac{a^4}{\eta}$$

$$= 1.9 \times 10^3 M_9 \eta_{1:4}^{-1} \left(\frac{a}{10^2 GM/c^2} \right)^4 \text{ yr}$$
$$= 2.0 \times 10^3 M_9^{-5/3} \eta_{1:4}^{-1} P_1^{8/3} \text{ yr} \quad (6.18)$$

(Peters 1964). Because typical resolved sources have $M/\mathcal{M} = \eta^{-3/5} \sim 3$, we normalize the symmetric mass ratio $\eta \equiv (M_2/M_1)/[1+M_2/M_1]^2$ to the value $\eta_{1:4} \equiv \eta(M_2/M_1 = 0.25) = 0.16$. Note that our ad hoc translation between M and $\mathcal{M}$ is not very sensitive to the value of q; the ratio $M/\mathcal{M}$ varies by less than a factor of two in the range $0.1 \leq M_2/M_1 \leq 1$. The upper bound in equation 6.18 corresponds to binaries in circular orbits, with eccentric orbits merging faster. Recent work has shown that binaries may have eccentricities as high as ~ 0.6 at decoupling (Roedig et al. 2011; see also Armitage & Natarajan 2005; Cuadra et al. 2009). Thus, typical PTA-resolved sources will coalesce on scales of $\sim 10^3$ years. However, exceptionally compact sources will coalesce on scales of several years; for example, a binary with $P = 0.1$ yr — approximately the lowest binary period that is expected to be observable with PTAs — will merge in $t_{\rm merge} \sim 4$ yr.

The tidal torques of the compact SMBH binary provide a particularly promising mechanism for producing a tell-tale observable feature. Theoretical calculations (Goldreich & Tremaine 1980; Artymowicz et al. 1991; Artymowicz & Lubow 1994; Armitage & Natarajan 2002; Bate et al. 2003; Hayasaki et al. 2007; MacFadyen & Milosavljević 2008; Cuadra et al. 2009; Chang et al. 2010) robustly predict that in geometrically thin circumbinary accretion disks, binary torques can open an annular, low-density gap around the orbit of the secondary. The gas inside the gap accretes onto the individual SMBHs while the gas outside is pushed outward by the tidal torques. The binary's tidal torques trans-

fer orbital angular momentum into the outer disk, causing the binary's orbit to shrink gradually while maintaining a roughly axisymmetric circumbinary gap.

Following Milosavljević & Phinney (2005), we parametrize the location of the outer edge of the gap, $R \approx 3^{2/3}a \approx 2.08a$ (Artymowicz et al. 1991), as $R_\lambda \equiv 2\lambda a$, where $\lambda \sim 1$ is a dimensionless parameter. We are interested in circumbinary disks that are truncated inside $R_\lambda \sim 200 M_9^{-2/3} P_1^{2/3} GM/c^2$. Below, we model surface density profiles and thermal emission spectra of such disks, and consider thermal emission due to leakage of gas into the cavity and onto individual SMBHs.

6.3.1 Disk Properties and Binary Decay

6.3.1.1 Disk around a solitary SMBH

Adopting a geometrically thin, thermal gray-body disk model (e.g., Blaes 2004; Milosavljević & Phinney 2005), we estimate the properties of circumbinary disks around resolved PTA sources. As a reference model, let us consider a disk around a solitary SMBH.

Disks whose kinematic viscosities ν scale with the total pressure $p = p_{\rm gas} + p_{\rm rad}$ are known to be thermally unstable (Shakura & Sunyaev 1976; Pringle 1976; see, however, Hirose et al. 2009b). We therefore adopt a prescription in which the kinematic viscosity scales with gas pressure (a.k.a. the "β-disk" model), which is known to be thermally stable:

$$\nu = \frac{2}{3}\frac{\alpha P_{\rm gas}}{\rho\Omega} = \frac{2}{3}\frac{\alpha k_{\rm B} T}{\mu m_{\rm p}\Omega}. \tag{6.19}$$

This viscosity prescription is also consistent with previous analyses of thin circumbinary

disks (Milosavljević & Phinney 2005; Tanaka & Menou 2010). However, an important caveat is that radiation pressure-dominated disks may be *viscously* unstable (Lightman & Eardley 1974; Piran 1978; Hirose et al. 2009a); our results are conditional on this theoretical uncertainty. We choose $\alpha = 0.3$ as the fiducial value, and write $\alpha_{0.3} = \alpha/0.3$. We assume that shear viscosity is the dominant mechanism for the transport of angular momentum.

The surface density Σ and the mid-plane temperature T of the disk are obtained through the following equations:

$$\Xi(\Omega, T_{\rm p})\sigma_{\rm SB}T_{\rm p}^4 = \frac{9}{8}\nu\Sigma\Omega^2 \tag{6.20}$$

$$T_{\rm p}^4 = \frac{4}{3\tau}T^4 \tag{6.21}$$

$$\tau = \theta\kappa\Sigma \tag{6.22}$$

$$\Sigma = \frac{\dot{M}}{3\pi\nu}. \tag{6.23}$$

Above, Ξ is the deviation of the bolometric flux from blackbody due to the photons being thermalized above the mid-plane (see, e.g., Blaes 2004), $T_{\rm p}$ is the temperature of the thermalization photosphere, and τ is the optical depth between the mid-plane and the thermalization photosphere. The dimensionless parameter $\dot{m} \equiv \dot{M}/\dot{M}_{\rm Edd}$ describes the accretion rate in units of the Eddington rate, assuming a radiative efficiency of 0.1, i.e. $L_{\rm Edd} = 0.1\dot{M}_{\rm Edd}c^2$. The quantity θ is a porosity factor that relates the surface density to the optical depth. We set it to 0.2 following Turner (2004) and express our results in terms of $\theta_{0.2} = \theta/0.2$; however, most disk properties are not very sensitive to this parameter.

The disk scale height H is evaluated in the usual way:

$$c_s^2 \equiv \frac{P}{\rho} = H^2\Omega^2 + 4\pi G\Sigma H, \tag{6.24}$$

where $c_s \equiv \sqrt{P/\rho}$ is the isothermal sound speed calculated from the total pressure $P = P_{\rm gas} + P_{\rm rad}$, and the volume density ρ of the disk is given by $\rho = \Sigma/H$. The second term on the right-hand side of equation 6.24 is due to the disk's self-gravity (e.g., Paczynski 1978).

In the regions of interest, the dominant source of opacity is electron scattering, and vertical pressure in the disk is due primarily to radiation pressure. In this regime, the gray-body factor Ξ can be approximated as $\Xi \approx 0.17(\Omega\ {\rm yr})^{1/2}[T_{\rm p}/(10^4{\rm K})]^{-15/8}$ (Tanaka & Menou 2010), and with a little algebra we obtain the surface density profile in the disk:

$$\Sigma(R) \approx 1.3\times10^6{\rm g\ cm^{-2}}\left(\frac{R}{100GM/c^2}\right)^{-6/17} M_9^{16/85}\dot{m}^{36/85}\alpha_{0.3}^{-4/5}\theta_{0.2}^{-1/5}. \tag{6.25}$$

A steady-state disk far from the central object satisfies $\dot{M} = 3\pi\nu\Sigma$ = constant, and so we have $\nu \propto \Sigma^{-1} \propto R^{6/17}$.

6.3.1.2 Circumbinary disks around orbit-decaying binaries

After a circumbinary gap is opened, the SMBH binary undergoes several stages of orbital decay. Let us briefly examine the different stages, and the orbital evolution timescale (or residence time) $t_{\rm res} \equiv a/|da/dt|$ for each. Our goal here is to describe the structure of a dense gaseous annulus, extending at least a factor of few in radius, that is created around the PTA source. The annulus results from inward migration of the binary from larger radii

in a more extended accretion disk. For a more thorough diskussion of the orbital decay of SMBHB binaries, through various physical regimes in a thin disk, see, e.g. Haiman et al. (2009a).

We begin with disk-driven orbital decay, in which the binary's tidal torques transfer its orbital angular momentum to the surrounding gas. At large orbital separations, the mass of the gas at the edge of the cavity far exceeds the mass of the secondary. In this regime, analogous to disk-dominated Type II migration for proto-planets, the binary's orbital evolution is limited only by the rate at which the nearby gas can transport away angular momentum, i.e.

$$t_{\rm res}^{\rm (disk)} = t_\nu(R_\lambda) = \frac{2R_\lambda^2}{3\nu(R_\lambda)}. \tag{6.26}$$

The tidal torques prevent the gas from flowing inward of R_λ, and so the region inside the gap is starved. Any gas that is initially present will be depleted on the local viscous timescale (Chang et al. 2010). In standard steady-state thin-disk models the viscosity is an increasing function of radius, so this drainage occurs on timescales shorter than that of the binary's orbital decay.

When the mass of the secondary becomes comparable to the local disk mass, the orbital decay slows down with respect to the local viscous time. The gas piles up immediately outside the cavity, forming a decretion region in which the viscous torque $\mathcal{T}_\nu = 3\pi\nu\Sigma\Omega R^2$ is nearly constant with radius (Pringle 1991). We apply the analytic model of Ivanov et al. (1999) to calculate the residency time for this secondary-dominated migration stage:

$$t_{\rm res}^{\rm (sec)} = \frac{\eta M}{4\pi R_\lambda^2 \Sigma(R_\lambda)} t_\nu(R_\lambda). \tag{6.27}$$

Note that there are two competing effects influencing $t_{\rm res}^{\rm (sec)}$: the decay slows down as the local disk mass decreases with respect to the secondary, but this is mitigated to a small extent by the fact that Σ outside the cavity increases due to pile-up. The enhancement of Σ relative to that of a disk around a solitary SMBH of the same mass as the binary (equation 6.25) has the functional form (Ivanov et al. 1999)

$$\frac{\Sigma^{\rm (binary)}}{\Sigma^{\rm (solitary)}} = \left\{1 + A\left[1 - \left(\frac{R_\lambda}{R_\lambda^{\rm (disk/sec)}}\right)^{1/2}\right]\right\}^B \left(\frac{R}{R_\lambda}\right)^{-1/2} \tag{6.28}$$

in the neighborhood $R \gtrsim R_\lambda$. Above, $R_\lambda^{\rm (disk/sec)}$ is the radius of the cavity when the transition from disk-dominated to secondary-dominated migration occurs, i.e. when $\eta M = 4\pi R_\lambda^2 \Sigma(R_\lambda)$. For reasonable parameter values, $R_\lambda^{\rm (disk/sec)} > 10^3 GM/c^2$. The dimensionless quantities A and B in equation 6.28 depend on the viscosity and mass profiles of the disk (see Ivanov et al. 1999, for details). We typically find that $A \sim 4$ and $B \sim 0.2$ in our disk models; i.e., the fractional surface density enhancement during secondary-dominated migration is no greater than $(1 + A)^B \sim 1.4$.

At yet smaller separations, the binary's orbital evolution begins to be driven by GW emission. Since binaries of interest here are far from merging, GW emission can be approximated by the leading term in the Newtonian quadrupole. For circular orbits, the residence time is given by Peters (1964)

$$t_{\rm res}^{\rm (GW)} = 4t_{\rm merge} = \frac{5}{64}\frac{c^5}{G^3M^3}\frac{a^4}{\eta}. \tag{6.29}$$

As the binary's orbital decay accelerates due to GW emission, the pileup caused by

secondary-dominated migration spreads out. Past the point where $t_{\rm res}^{\rm (GW)} \approx t_\nu(R_\lambda)$, the binary begins to outrun the disk, as the decay timescale for a becomes rapidly shorter than that on which the disk can viscously spread.

Let us now diskuss the gravitational stability of the disk, based on the stability criteria of a radiation-pressure dominated fluid summarized by Thompson (2008). If the radiative diffusion timescale is much shorter than the dynamical timescale, then the radiation pressure does not stabilize the fluid and gravitational fragmentation occurs on the same length scales as it would in the absence of radiation pressure. If the radiative diffusion timescale is much longer than the dynamical time, which we find to be the case for our disk models, then radiation pressure acts to make the fluid more Jeans-stable.[4]. We use the Toomre criterion, and assume that the disk is gravitationally stable when $Q(R) \equiv c_{\rm s}\Omega/(\pi G\Sigma) > 1$. Note that the only effect of radiation pressure is that the sound speed $c_{\rm s}$ is computed from the total pressure, not just the gas pressure.

In the radiation pressure-dominated regime of our gray-body disk, the gas volume density *decreases* with increasing surface density, because the disk scale height is a stronger-than-linear function of Σ. This leads to the somewhat counterintuitive behavior that increasing the surface density in the inner regions of the disk makes it more gravitationally stable. We find

$$R_{\rm Q} \equiv R(Q=1) = 550\frac{GM}{c^2}\alpha_{0.3}^{34/165}\dot{m}^{62/165}M_9^{-16/55}\theta_{0.2}^{17/55}. \tag{6.30}$$

[4]The gas is susceptible to an additional weak diffusive instability that grows on the Kelvin-Helmholtz timescale, $t_{\rm KH} \sim \kappa c_s^2/(\pi G c)$. We find that the viscous timescale is shorter than the Kelvin-Helmholtz timescale — i.e., the diffusive instability is irrelevant — in all but the outermost annulus of the radiation-dominated region of our fiducial circumbinary disks (e.g., at $R = 400GM/c^2$, $t_{\rm KH} \sim 2\times10^6$ yr and $t_\nu \sim 10^5$ yr). Even in the small region where the disk is formally unstable to the diffusive instability, it is plausible that local turbulence can quench its growth. Hence we assume the diffusive instability is unimportant.

Equation 6.30 is valid for radiation pressure-dominated regions only; if the surface density is sufficiently low, then gas pressure dominates and decreasing the surface density further increases R_Q. That is, R_Q has a minimum value as a function of $\dot{m} \approx 0.1$ for fiducial disk parameters.

The gas density profile in the outer regions $R > R_Q$, where classical thin-disk models predict $Q < 1$, is uncertain. One possibility that has been explored by Sirko & Goodman (2003) and others (Thompson et al. 2005; Levin 2007; Lodato et al. 2009) is that feedback mechanisms (such as nuclear fusion from stars that formed in the disk or their supernovae) inject sufficient energy as to maintain marginal gravitational stability with $Q \approx 1$ in the outer regions. However, the profile of the outer disk is not central to this study, as we are interested in radiation from the central regions of the disk, where the presence of a compact binary is most likely to produce characteristic features that may distinguish them from accretion disks around single SMBHs. To keep our analysis as simple as possible, we simply neglect the thermal radiation of the disk outside R_Q.

What is the accretion rate in the disk? Uncertainty regarding the outer gas distribution notwithstanding, quasars are able to efficiently supply SMBHs of mass $> 10^8 M_\odot$ with enough fuel to maintain luminosities of $0.1 - 1 L_{\rm Edd}$ for periods of 10^{6-8} yr. That most quasars radiate at just under the Eddington limit while few exceed $L_{\rm Edd}$ suggests that the accretion rate is limited by radiative feedback, rather than by the availability of fuel. In principle, therefore, the surface density in a circumbinary disk can be significantly greater than in a disk around a solitary SMBH of the same total mass. This is not because of the mass accumulation of gas outside the binary's orbit, but because a circumbinary disk has

a much lower luminosity owing to its low-density central cavity. If binary torques inhibit gas from accessing the centre of the potential, then this effectively reduces the radiative efficiency of the system, i.e. $L_{\rm disk} \ll \dot{M}c^2$. We find that even at $\dot{m} \sim 10$, the locally viscously dissipated flux in our disks does not provide sufficient radiation pressure to unbind gas from the local gravitational field.[5]

In Figure 6.3, we plot, for a circumbinary disk around a binary with $M_9 = 1, M_2/M_1 = 1/4$, several transition radii as a function of the mass accretion parameter $\dot{m}$. We show the radii at which the disk transitions from being radiation pressure-dominated to gas pressure-dominated; from where the opacity is dominated by electron scattering to free-free absorption; and from Jeans-stable to unstable. We note that in general, radiation pressure acts to stabilize the disk against Jeans collapse, and that the radius R_Q closely corresponds to the radius where the disk becomes radiation pressure-dominated. We also plot the size of the cavity $R_\lambda \sim 2a$ at which the binary's orbital evolution transitions from being gas-driven to GW-driven, and the value of R_λ where $t_{\rm res}^{\rm GW} = t_\nu(R_\lambda)$. We find that the disk is geometrically thin ($H/R \ll 1$) or marginally thin ($H/R \lesssim 1$) for $\dot{m} < 10$.

In conclusion, Figure 6.3 suggests that a Jeans-stable circumbinary annulus could exist, instantaneously, around an individually resolved PTA source, for any value of the supply rate, extending at least by a factor of two in radius (from the inner radius shown by the [red] dashed line, to the outer radius shown by the [black] dotted curve). However, in order for this annulus to be created through the in-ward migration of the secondary BH from larger radii in a more extended disk, we require that the disk is stable to radii

[5]We remind the reader that $\dot{m}$ is defined with respect to the Eddington limit assuming a radiative efficiency of ~ 0.1. Strictly speaking, our circumbinary disks are not super-Eddington, even when the parameter $\dot{m}$ exceeds unity.

that extend beyond the gas/GW-driven transition. This latter requirement (i.e., the dotted [black] curve must lie above the thick [blue] solid curve) means that in practice, the gaseous annulus exists in PTA sources only if the mass supply rate is $\dot{m} \gtrsim 1$. As argued above, while radiative feedback may disallow such high (super-Eddington) rates in thin disks around a single BH, they can naturally be maintained in disks around binaries, owing to their low radiative efficiency.

6.3.2 Surface Density Evolution of the Circumbinary Gas

Let us now address the surface density profile of the outer disk at the time when the binary SMBH becomes observable by PTAs.

The tidal torque density $dT_{\rm tide}/dR$ is sharply peaked in a narrow region that roughly coincides with the edge of the cavity R_λ, preventing the gas from accreting inward. Everywhere else in the disk, the tidal torques are negligible compared to the viscous torques. The effect of the tidal torques in the region $R \approx R_\lambda$ can thus be approximated as a boundary condition prohibiting mass flow across R_λ (Pringle 1991; Ivanov et al. 1999):

$$\dot{M}(R_\lambda, t) = 6\pi\nu\Sigma \frac{\partial \ln(\Sigma\nu R^{1/2})}{\partial \ln R}\Big|_{R=R_\lambda} = 0, \tag{6.31}$$

Note that our disk is not steady-state, and the local mass flow rate $\dot{M}$ need not be radially constant.

The surface density evolution of the circumbinary disk R_λ is governed by the standard equation for viscous disks (e.g., Pringle 1981; Frank et al. 2002) without including an

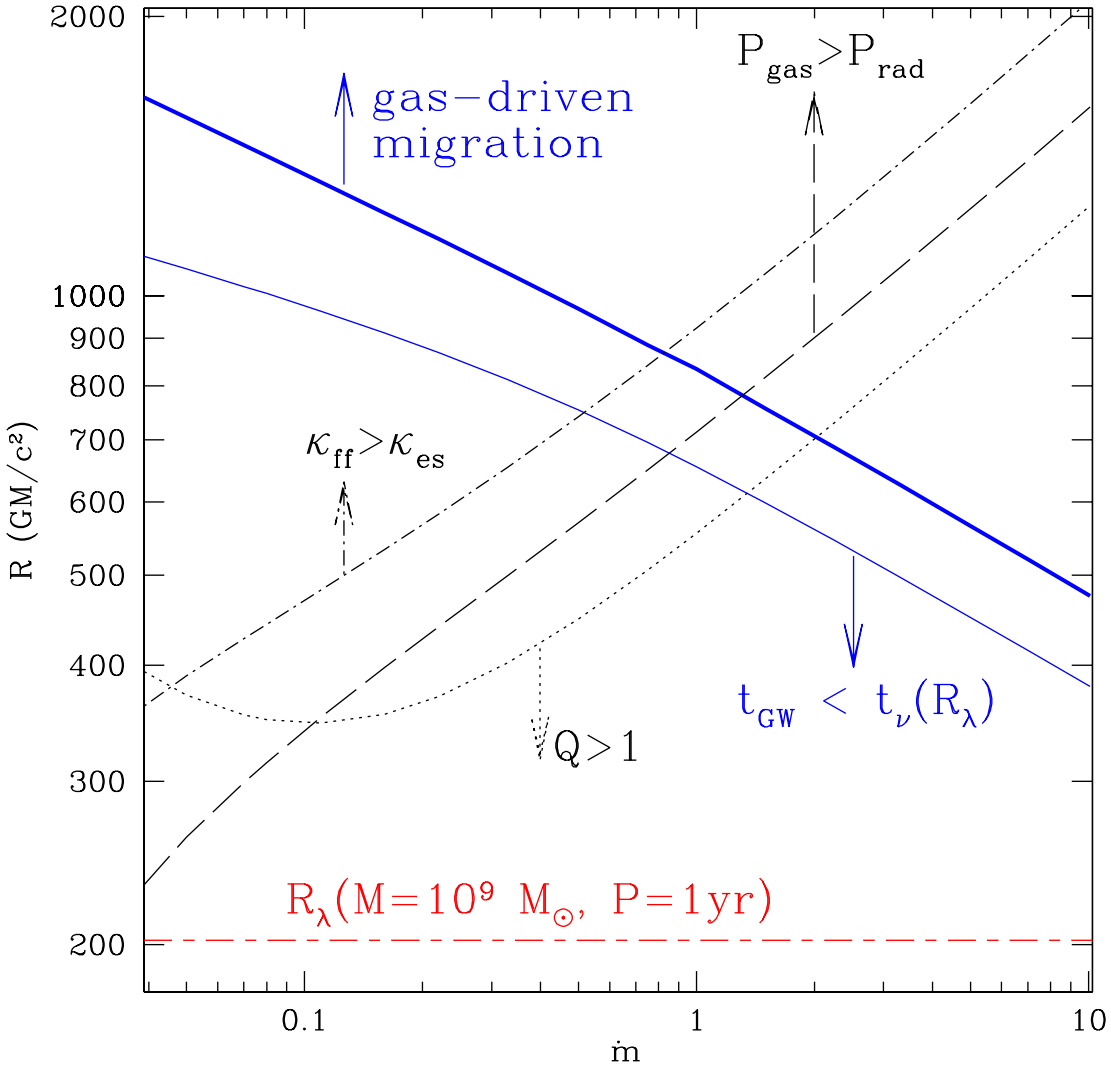

Figure 6.3 In black lines, several transition boundaries within a steady-state, thin accretion disk are plotted as a function of the mass supply rate $\dot{m} = \dot{M}/(0.1L_{\rm Edd}/c^2)$. Binary parameters are $M = 10^9 M_\odot$, $M_2/M_2 = 1/4$; disk parameters are $\alpha = 0.3$, $\lambda = 1$, $\theta = 0.2$. Plotted are the radii in the disk where the disk is marginally stable to gravitational fragmentation (R_Q; dotted lines); where the radiation pressure equals the gas pressure (dashed); and where free-free opacity equals electron-scattering opacity (dash-dot). In blue lines, we plot the size of the circumbinary cavity $R_\lambda = 2\lambda a$ where the binary's orbital decay transitions from being gas-driven to GW-driven (thick lines), and when the orbital decay timescale becomes shorter than the viscous timescale at the inner edge of the disk (thin lines). For reference, the radius of the cavity for a $10^9 M_\odot$ binary with a rest-frame period of 1 yr is plotted as a horizontal red dashed line.

explicit term for the tidal torques:

$$\frac{\partial}{\partial t}\Sigma(R,t) = \frac{1}{R}\frac{\partial}{\partial R}\left[R^{1/2}\frac{\partial}{\partial R}\left(3\nu\Sigma R^{1/2}\right)\right]. \tag{6.32}$$

A semi-analytic solution for the thin-disk equation 6.32 with the boundary condition in equation 6.31 was derived by Tanaka (2011), for a finite boundary $R_\lambda > 0$ and a special viscosity prescription $\nu \propto R^n$. The solution can be written in the form

$$\Sigma(R,t) = \int_{R_\lambda}^{\infty} G(R,R',t;R_\lambda)\,\Sigma_{\rm init}(R')\,dR', \tag{6.33}$$

where $G(R,R',t;R_\lambda)$ is the Green's function specific to the boundary condition and the chosen value of the viscosity power-law index n; and $\Sigma_{\rm init}(R)$ is an arbitrary initial density profile. We find that our disks satisfy $\nu \propto R^{0.4}$ inside $R < 10^3 GM/c^2$, and thus adopt $n = 0.4$.

In order to model a thin accretion disk around a GW-driven SMBH binary, we modify the solution of Tanaka (2011) in two ways. First, we derive a more general Green's function to allow for a boundary condition with nonzero mass flux across the inner boundary:

$$\dot{M}(R_\lambda,t) = f_{\rm leak}\dot{M}_{\rm ss}(R_\lambda,t). \tag{6.34}$$

Above, $\dot{M}_{\rm ss} = 3\pi\nu\Sigma$ is the standard accretion rate for expected of a steady-state disk, and $0 \leq f_{\rm leak} < 1$ is a numerical factor representing the incomplete suppression of gas inflow into the cavity. The case $f_{\rm leak} = 0$ corresponds to total suppression of accretion by the binary's tidal torques.

The boundary condition in equation 6.34 is motivated by results from numerical simulations, which show that in general the binary torques do not completely prevent accretion into the gap, but rather allow some gas to leak into the centre of the disk with a suppressed mass flux $f_{\rm leak} \lesssim 0.1$ (e.g., Artymowicz & Lubow 1996; Günther et al. 2004; Ochi et al. 2005; MacFadyen & Milosavljević 2008). We choose $f_{\rm leak} = 0.1$ as our fiducial value.

The long-term behavior of the gas is to pile up near the cavity and satisfy the power-law $\nu\Sigma \propto R^{(f_{\rm leak}-1)/2}$ in the vicinity of the boundary. In comparison, a steady-state around a solitary central mass disk satisfies $\nu\Sigma =$ constant, and a boundary condition imposing zero inward mass flux satisfies $\nu\Sigma \propto R^{-1/2}$ at the boundary. Note, however, that the value of $f_{\rm leak}$ does not have a strong effect on the mass profile (and hence the luminosity produced) outside the cavity-opening radius R_λ; the fractional surface density enhancement due to secondary-dominated migration is typically of order unity.

The gas that enters the cavity does so in nearly radial orbits MacFadyen & Milosavljević (2008), and so is dynamically decoupled from the circumbinary disk. Thus, the surface density and mass flux inside R_λ can consistently be disregarded in the Green's function formalism. The leaked gas can presumably form accretion disks around one or both SMBHs (Hayasaki et al. 2007), presumably at the usual AGN radiative efficiency ~ 0.1. The mass supply rate of such circum-secondary (or -primary) disks will be modulated by $f_{\rm leak}\dot{M}_{\rm ss}$, which decreases as the binary outruns the circumbinary gas. Because the viscous time at the outer edge of such disks are shorter than at R_λ, they will be nearly steady-state, with the surface density profile at any given time being determined by the instantaneous

mass flux into the cavity. Thus, the bolometric luminosities of the disks around each disk may be roughly expressed as $L < 0.1\ f_{\rm leak}\dot{M}_{\rm ss}(R_\lambda)c^2$, and would be Eddington-limited by the potential of the individual black holes they orbit. This suggests that if the quantity $f_{\rm leak}\dot{m}$ exceeds unity, the region inside the cavity would develop radiation-driven outflow winds. Thus, the parameter $f_{\rm leak}$ affects the energetic output due to accretion inside the cavity far more than it does that of the circumbinary disk.

The Green's function for the boundary condition in equation 6.34 is given by

$$\begin{aligned} G(R,R',t;R_\lambda) &= \left(1-\frac{n}{2}\right)R^{-n-1/4}R'^{5/4} \\ &\quad\times\int_0^\infty \left[J_\ell(ky)\widetilde{Y}_\ell(ky_\lambda)-Y_\ell(ky)\widetilde{J}_\ell(ky_\lambda)\right] \\ &\quad\times\left[J_\ell(ky')\widetilde{Y}_\ell(ky_\lambda)-Y_\ell(ky')\widetilde{J}_\ell(ky_\lambda)\right] \\ &\quad\times\left[\widetilde{J}^2_\ell(ky_\lambda)+\widetilde{Y}^2_\ell(ky_\lambda)\right]^{-1} \\ &\quad\times\exp\left[-3\Lambda k^2 t\right]\ k\,dk. \end{aligned} \tag{6.35}$$

Above, $\ell = 1/(4-2n)$ and $\Lambda = (1-n/2)^2\nu R^{-n}$ are constants. We have introduced the variables $y = R^{1-n/2}$, $y' = R'^{1-n/2}$ and $y_\lambda = R_\lambda^{1-n/2}$, as well as the functions

$$\widetilde{J}_\ell(x) = x\,J_{\ell-1}(x) - \frac{f_{\rm leak}}{2-n}J_\ell(x) \qquad \text{and} \tag{6.36}$$

$$\widetilde{Y}_\ell(x) = x\,Y_{\ell-1}(x) - \frac{f_{\rm leak}}{2-n}Y_\ell(x). \tag{6.37}$$

Taking $f_{\rm leak} \to 0$ leads to the solution given in Tanaka (2011; his equation 42) imposing $\dot{M}(R_\lambda) = 0$.

The second modification to the Green's function formalism is to allow the inner boundary to move inward as an explicitly known function of time, i.e., $R_\lambda(t) = 2\lambda a(t)$. This is done through a time-weighted superposition of different Green's functions at intermediate values of R_λ (see Appendix).

We take our initial condition as the disk profile when $t_{\rm res}^{\rm GW} = t_\nu(R_\lambda)$, just as the binary is just beginning to outrun the circumbinary gas. We note that the torques exerted by the disk are not entirely negligible at this stage. For simplicity, we approximate the contribution of the disk torques to the orbital decay rate, $a/(da/dt)_{\rm disk} = t_{\rm res}^{\rm (sec)}$, as being constant. This is justified as follows. The disk torques are a weak function of R_λ, at least as long as the quantity $\nu(R_\lambda)\Sigma(R_\lambda)$ is comparable to the steady-state value. Once GW emission dominates the orbital decay, the contribution of disk torques becomes quickly negligible regardless of the value of $\nu(R_\lambda)\Sigma(R_\lambda)$.

The orbital decay is then given by

$$\frac{da}{dt} = \left(\frac{da}{dt}\right)_{\rm (GW)} + \left(\frac{da}{dt}\right)_{\rm disk} = \frac{a}{t_{\rm res}^{\rm (sec)}} + \frac{64}{5}\frac{c^5}{G^3M^3}\frac{\eta}{a^3}. \tag{6.38}$$

With the assumption that $t_{\rm res}$ is roughly constant, this has the analytic solution

$$t = \frac{t_{\rm res}^{\rm (sec)}}{4}\ln\left[\frac{a_0^4 + 64c^5\eta t_{\rm res}^{\rm (sec)}/(5G^3M^3)}{a^4(t) + 64c^5\eta t_{\rm res}^{\rm (sec)}/(5G^3M^3)}\right]. \tag{6.39}$$

In Figure 6.4, we plot a model surface density profile of the circumbinary disk around a binary with $M_9 = 1$ and $M_2/M_1 = 1/4$. As the initial condition, we take a steady-state surface density profile with $\dot{m} = 3$, at the time when $t_{\rm (GW)} = t_\nu(R_\lambda)$ (solid black

curve).[6] Initially, the disk has a cavity radius of $R_\lambda = 510GM/c^2$ and a rest-frame period of $P = 11$ yr. We then evolve the profile using our Green's function to when the orbital period is $P = 1$ yr (short-dashed blue curve) and $P = 0.1$ yr (long-dashed red curve). We denote with a green dot-dashed line the radius where $Q = 1$, beyond which the disk is expected to be susceptible to the Jeans instability. All the disk profiles are truncated at twice the binary's semi-major axis, and lose mass across this radius through the boundary condition in equation 6.34, with $f_{\rm leak} = 0.1$. Note that the boundary radius R_λ moves inward faster than the gas can pile up. We see that a small amount of gas is able to follow the binary's orbital decay, even though the bulk of the circumbinary disk is getting left behind by the inspiraling binary.

6.3.3 Thermal Emission of Accreting PTA Sources

Since the innermost gas is missing from the accretion disks around PTA binaries, it is probable that their accretion flows will emit less UV and thermal X-rays compared to ordinary AGN powered by solitary BHs. The question then is how UV- and X-ray-deficient these objects are; the answer depends on how much gas is able to follow the binary's decaying orbit, and how much of this gas is further able to leak into the cavity and accrete onto individual SMBHs. This is a complex problem characterized by dynamical

[6]Strictly speaking, the surface density profile at this time should deviate somewhat from the steady-state one. During secondary-dominated migration, the circumbinary surface density profile $\Sigma(R \gtrsim R_\lambda)$ can become greater than the steady-state profile by at most a factor of ~ 1.4 (equation 6.28; note that the pileup must be smaller if one accounts for the fact that $f_{\rm leak} > 0$). Prior to our initial condition, GW emission accelerates the binary's orbital evolution, and the circumbinary pileup spreads out; however, the surface density does not decrease below the steady-state profile, since $t_{\rm res} > t_\nu(R_\lambda)$. A difference in Σ of less than 40% is insignificant compared to the other theoretical uncertainties, and we employ the steady-state profile for simplicity.

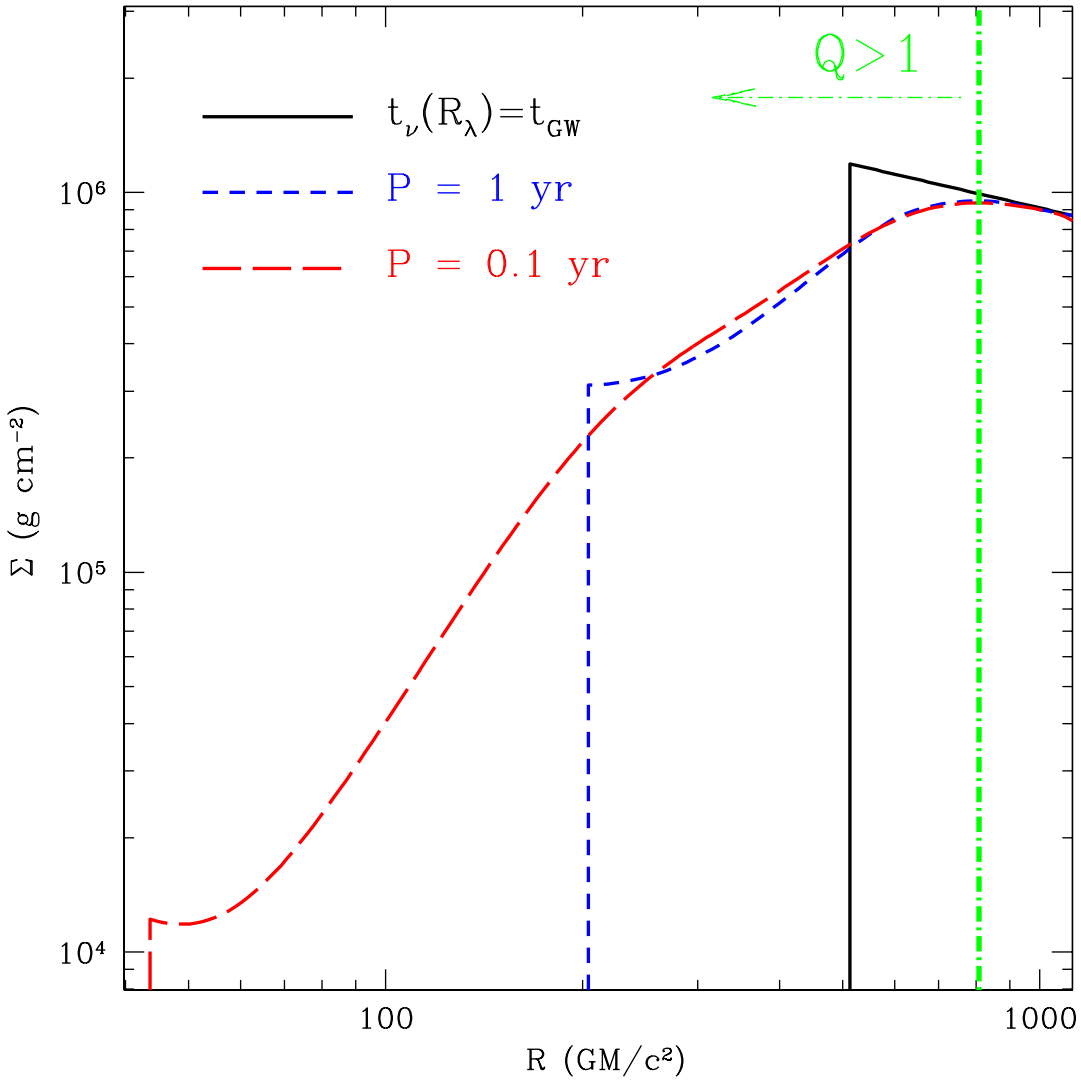

Figure 6.4 The surface density profiles Σ for a circumbinary disk around a PTA source. The binary's mass is $10^9 M_\odot$ and its mass ratio is $M_2/M_1 = 1/4$. We adopt a moderately high value of the accretion parameter $\dot{m} = 3$, and assume that the circumbinary gas can leak into the cavity at the rate given in equation 6.34, with $f_{\rm leak} = 0.1$. The solid black curve shows the surface density profile when GW emission begins to dominate the binary's orbital decay ($t_\nu(R_\lambda) = t_{\rm GW}$, $R_\lambda = 510GM/c^2$, $P = 11$ yr). Using the semi-analytic method described in the text, we solve for the surface density profile in the disk at later times, when $P = 1$ yr (short-dashed blue line) and $P = 0.1$ yr (long-dashed red line). The dot-dash green line denotes the radius inside which the circumbinary disk is stable against Jeans collapse.

richness in 3D, and considerable theoretical uncertainty of the underlying fluid physics. With the above caveat in mind, we will use as a first approximation the toy surface density evolution model introduced in §6.3.2 to estimate the thermal emission from an accreting PTA source.

In Figure 6.5, we show the thermal spectrum for a $M_9 = 1$, $M_2/M_1 = 1/4$ PTA source, calculated from the circumbinary gas surface density profiles in Figure 6.4. We have plotted (i) the circumbinary disk with $\dot{m} = 3$ inside the radius where $Q = 1$ (dotted, left hump); (ii) an accretion disk around the secondary SMBH (dotted, right hump) fueled by leakage into the cavity and truncated at the Hill radius, for which we use $R_{\rm H} \sim 0.5\eta^{1/3}a$; and (iii) the combined emission of the two disks (solid thick line). For comparison, we also show the spectrum for an Eddington-limited thin disk (dashed lines) around a solitary SMBH with the same mass as the binary. For simplicity, we have assumed that all of the gas leaked into the cavity fuels a circum-secondary disk. We have plotted spectra when the source has a binary orbital period of $P = 1$ yr and when $P = 0.1$ yr.

The result is what would be expected intuitively. The infrared and optical flux, which is produced almost exclusively in the circumbinary disk, does not vary greatly from what is expected from a standard thin disk. However, the flux drops precipitously below wavelengths of $\lambda \lesssim 3000$Å ($\nu > 10^{15}$ Hz in the figure). This is in stark contrast to most unobscured quasars thought to be powered by $\sim 10^{8-9}M_\odot$ SMBHs, which have their *brightest* emission in the rest-frame near-UV near their Lyman-α line. The bolometric luminosity of the accreting PTA source is roughly $\sim 0.03L_{\rm Edd}$ for $P = 1$ yr, and $\sim 10^{-3}L_{\rm Edd}$ for $P = 0.1$ yr. The optical and infrared emission is dominated by the circumbinary disk,

whereas the UV and X-rays are produced by circum-secondary accretion fueled by leakage of circumbinary gas into the cavity. As the binary evolves toward shorter periods, the circum-secondary disk is depleted — the viscous time at the Hill radius is typically a few hundred years, shorter than the time to binary merger — and as a result, less gas is able to leak into the cavity, decreasing the UV and X-ray emission. The degree to which the UV and X-ray emission is suppressed depends on the model parameters (in particular $f_{\rm leak}$) and on the binary period.

We also note that the downturn in the near-UV flux at $\lambda \lesssim 300$ nm could help distinguish PTA sources from single-SMBH AGN. This feature will be observable in the optical if the source redshift is high; e.g., at $z = 1$ it will be in the V band. Hence, even in the optical, this source will have an unusual color: it will appear fainter in the U and B bands than a typical AGN. The downturn could be distinguished from reddening due to dust obscuration through the deviation from the power-law spectral shape of dust reddening.

We propose that once an individually resolved PTA source is detected and its error box determined, searching for AGN with weak UV emission lines (e.g., Ly α) and/or weak soft X-ray emission is a promising method to narrow the field of interlopers. AGN whose soft X-ray fluxes are weaker by more than a factor of 10 compared to the average have indeed been detected, and are estimated to constitute at most $\sim 1\%$ of the general AGN population (e.g., Brandt et al. 2000; Leighly et al. 2007; Gibson et al. 2008; Wu et al. 2011, and refs. therein). There have also been observations of quasars with exceptionally weak lines (Diamond-Stanic et al. 2009); these objects have infrared and optical emission consistent with those of typical luminous AGN, and also tend to be X-ray weak (Shemmer

et al. 2009). That X-ray weak AGN are so rare suggests that it will be possible to narrow the number of interlopers in a typical PTA error box by a factor of ≈ 100, i.e. either to a handful of objects, or yielding a unique EM counterpart candidate. It is possible, furthermore, that some of these rare X-ray weak AGN are in fact the SMBH binaries PTAs will be detecting.

Our results also strongly suggest that AGN counterparts to PTA sources should draw from optically selected surveys, as their nature makes them likely to be missed by X-ray searches (see, however, Sesana et al. 2011, who investigate the possible X-ray searches of PTA source binaries that have not yet decoupled). The *Large Synoptic Survey Telescope*[7] should be able to detect all of the optically luminous AGN in the PTA error box within $z \sim 1$. It may be possible to follow up candidates individually, but comparing the optical data with that of wide-field X-ray surveys such as *MAXI*[8], *eROSITA*[9] or *Wide Field X-ray Telescope*[10] would greatly facilitate the multi-wavelength search for counterpart candidates inside the error box.

Additional follow-up studies of candidates may further corroborate the identification of a counterpart. For example, the gas that leaks radially into the cavity can shock-heat the outer edge of the circum-secondary (or circum-primary) disk and produce hot spots. The viscously dissipated luminosity of a circum-secondary disk is roughly $L_{\rm disk2} \lesssim (1/2)GM_2\dot{M}_2/R_{\rm ISCO,2}$, where $\dot{M}_2 \leq f_{\rm leak}\dot{M}(R_\lambda)$ is the mass supply rate of the circum-secondary disk and $R_{\rm ISCO,2}$ is the radius of innermost stable circular orbit around the

[7] http://www.lsst.org/lsst
[8] http://maxi.riken.jp/top/
[9] http://www.mpe.mpg.de/erosita/
[10] http://www.wfxt.eu/home/Overview.html

secondary. The time-averaged power per unit mass of the hot spots is limited by the amount of kinetic energy the flow can deposit at the outer edge of the circum-secondary disk, i.e. $L_{\rm hot} \lesssim GM_2\dot{M}_2/R_H$. It follows directly that the time-averaged ratio between between the hot-spots and the intrinsic luminosity of the circum-secondary disk is

$$\frac{L_{\rm hot}}{L_{\rm disk2}} \lesssim \frac{R_{\rm ISCO,2}}{R_H} \sim \eta^{-1/3}\frac{R_{\rm ISCO,2}}{a} \gtrsim \left(\frac{a}{GM/c^2}\right)^{-1}. \tag{6.40}$$

In other words, the time-averaged power of a hot spot is of order $\gtrsim 1-10\%$ of the circum-secondary disk luminosity for resolved PTA sources. In principle, the luminosity of any single flare could be much greater. Because streaming into the cavity is expected to be modulated quasi-periodically by the binary's orbital period (e.g., Hayasaki et al. 2007; MacFadyen & Milosavljević 2008), EM counterparts of resolved PTA sources may be characterized by periodic UV flares.

In the same vein, if the orbital plane lies close to the line of sight, the UV lines would display strong periodic Doppler shifting with respect to the optical emission, modulated at the binary's orbital period (e.g., Halpern & Filippenko 1988). Thus, monitoring candidate counterparts for periodic or quasi-periodic variability on orbital timescales may prove a fruitful route for identification Haiman et al. (2009a). As a proof of this concept, we note that Boroson & Lauer (2009) recently reported a candidate SMBH binary, with two sets of broad emission lines separated by $3,500 {\rm km\ s^{-1}}$, with inferred component masses of $M_1 = 10^{8.9}\ {\rm M}_\odot$ and $M_2 = 10^{7.3}\ {\rm M}_\odot$. The binary interpretation, however, could be ruled out by the lack of any change in the velocity offset between two spectra taken ≈ 1 year apart (Chornock et al. 2009).

Lastly, we consider the scenario of Chang et al. (2010), in which the circum-primary disk brightens prior to merger due to tidal excitation by the shrinking binary. The power generated by this process can be approximated as (see their equation 15)

$$L_{\rm tide} \sim \frac{GMM_{\rm in}}{2at_{\rm merge}} \sim 1.4\times10^{43} M_9^{7/3}\eta_{1:4}P_1^{-10/3}\frac{M_{\rm in}}{100M_\odot}{\rm erg\ s^{-1}}, \quad (6.41)$$

where $M_{\rm in}$ is the mass of the circum-primary disk. Extrapolation of the Chang et al. (2010) results to binaries with mass $M \sim 10^9 M_\odot$ (their calculations only considered binaries up to $M = 10^8 M_\odot$) suggests a value of $M_{\rm in} \sim 100 M_\odot$. Similar values are obtained by estimating the disk mass that can be fueled by gas leaking into the cavity with $f_{\rm leak} \sim 0.1$, when the time to merger is comparable to the viscous timescale at the outer edge of the circum-primary disk.

Equation 6.41 suggests that the power produced by tidal excitation of the circum-primary disk is negligible compared to the thermal disk emission if the binary period is $P \sim 1$ yr. However, for sources with $P \sim 0.1$ yr, the tidally excited emission would rival the bolometric output of the thermal emission. The tidal component, which would have a peak frequency in the UV and soft X-rays, will brighten dramatically prior to merger on timescales of several years to decades. Even though $P \sim 0.1$ yr sources are predicted to be much rarer and also much more difficult to resolve individually with PTAs, they present tantalizing possibilities for observing EM signatures that are directly related to binary coalescence.

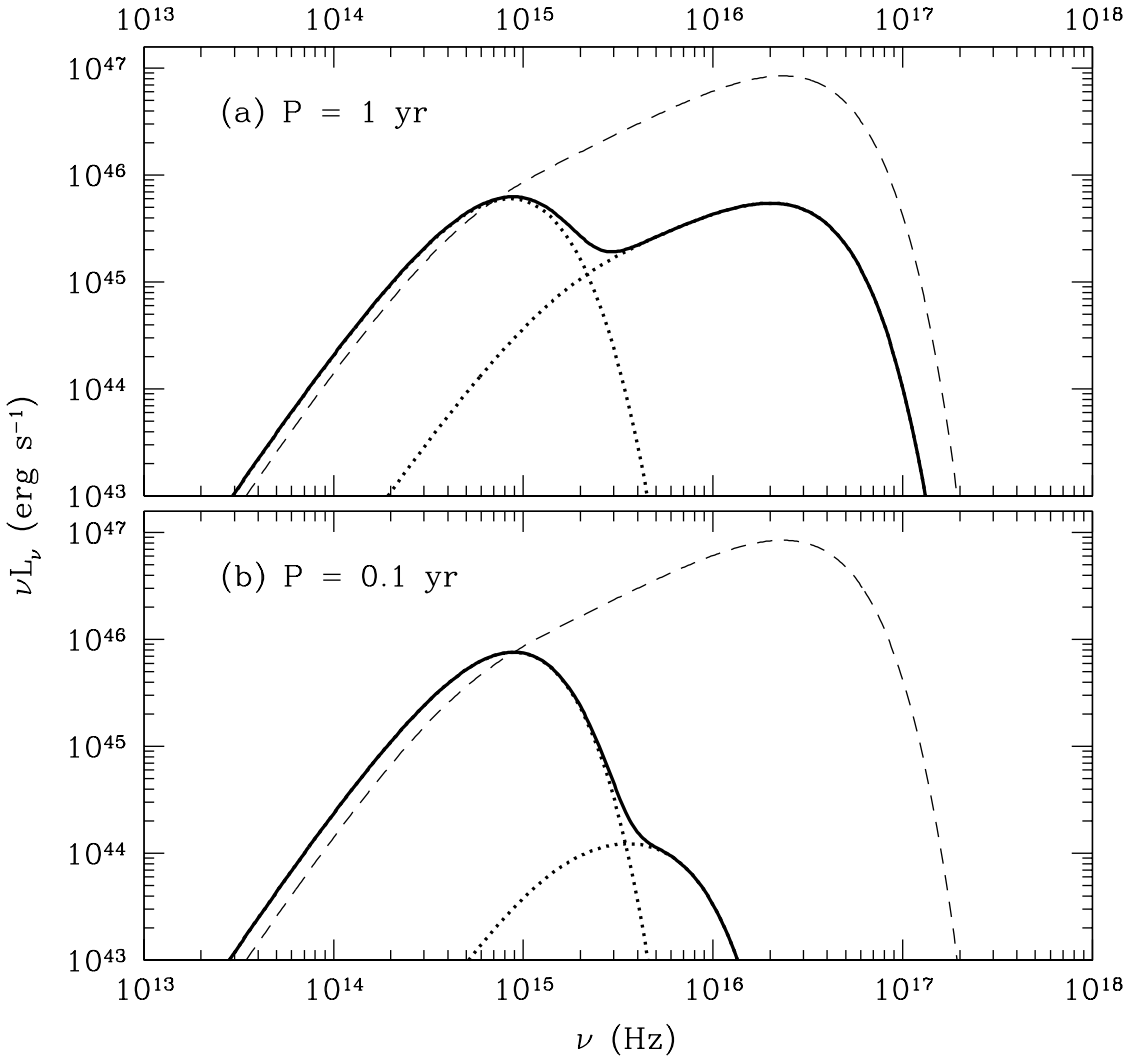

Figure 6.5 We plot the spectral energy distribution from the thin, viscous circumbinary accretion disk from Figure 6.4 (red curve; $M = 10^9 M_\odot$, $P = 1$ yr), when the source has a period (a) $P = 1$ yr and (b) $P = 0.1$ yr. We assume that some of the gas at the cavity/disk boundary is able to accrete into the cavity at a suppressed rate and fuel a small accretion disk around the secondary SMBH (see text for details). The spectra from individual disks are shown in dotted lines, with the circumbinary cavity emitting at lower frequencies. The combined spectrum from the circumbinary and circum-secondary accretion disks is shown in solid thick lines. We plot for reference the model AGN spectrum (thin dashed) for an Eddington-limited thin disk around a single SMBH with the same mass as the binary. The spectra for the PTA source are UV and X-ray weak compared to a standard thin disk model around a single SMBH. However, the infrared and optical emission, mostly produced in the circumbinary disk, is similar to what would be expected for a single-SMBH disk.

6.4 Conclusions

In this paper we considered the possibility that individually resolved PTA sources — SMBH binaries with $M \sim 10^9 M_\odot$, $M_2/M_1 \sim 1/4$, $P \sim 0.1 - 1$ yr and $z \lesssim 1.5$ — may be identified with EM observations if they reside in gas-rich environments. Multi-wavelength observations of such systems would allow for studies of an AGN in a system that is known to harbor a compact SMBH binary, thus providing a unique window into gas accretion in a rapidly time-varying gravitational potential. If, as suggested by Corbin & Cornish (2010), PTAs can constrain the luminosity distance to individually resolved sources, these SMBHs can be used as "standard sirens" to measure the cosmic expansion history. Interestingly, the predicted redshift distribution of these sources lies between $0.1 \lesssim z \lesssim 1.5$, comparable to that of the deepest Type-Ia supernovae surveys, and a range where *LISA* detection rates are expected to be low (Sesana et al. 2007b).

Our findings can be summarized as follows:

- The number of interloping massive halos and AGN that may be confused with the PTA source is typically $N_g \sim 10^4$ for a $10^9 M_\odot$ binary if the contributions to the signal from individual pulsars are not identifiable in the PTA data.

- In the more optimistic case, the pulsar term can be constrained and utilized to better determine the sky location of the source as well as constrain its redshift and mass. The number of interlopers can then drop to $N_g \sim 10 - 100$ at $z \sim 0.5$, and perhaps to $N < 1$ for close sources at redshifts as close as $z \lesssim 0.2$.

- By considering the orbital evolution history of an accreting PTA source, first by tidal

interactions with the circumbinary gas and then by GW emission, we showed that a gaseous accretion disk around the source can be expected to be gas-poor both inside and immediately outside its orbit. They would thus have optical and infrared luminosities comparable with typical quasars, while exhibiting low soft X-ray luminosities and weak UV emission lines. The downturn in flux below 300 nm could be diskerned by optical observations if the source redshift is $z \gtrsim 1$. Searching for AGN in the PTA error box with one or more of these atypical characteristics could lead to the identification of a single EM counterpart. Further monitoring candidate counterparts for periodicity and other theoretically predicted pre-coalescence signatures may also aid identification.

APPENDIX

6.A Green's-function Solution for the Thin-disk Equation with Moving Inner Boundaries

Consider a disk at time t which has surface density profile $\Sigma(R,t)$ and moving inner boundary $R_\lambda(t)$. During an infinitesimal time-step Δt, the boundary moves to $R_\lambda + \Delta R_\lambda = R_\lambda(t+\Delta t) < R_\lambda$, and the surface density evolves to

$$\Sigma(R,t+\Delta t) = \int_{R_\lambda+\Delta R_\lambda}^{\infty} G(R,R',t+\Delta t; R_\lambda+\Delta R_\lambda)\Sigma_{\rm init}(R')\,dR'. \tag{6.42}$$

From equations 6.33 and 6.42, we may write the derivative $d\Sigma/dt$ through its definition.

$$\begin{aligned}
\frac{d}{dt}\Sigma(R,t) &\equiv \lim_{\Delta t\to 0}\frac{\Sigma(R,t+\Delta t)-\Sigma(R,t)}{\Delta t} \\
&= \lim_{\Delta t\to 0}\Big[\int_{R_\lambda+\Delta R_\lambda}^{\infty}\frac{G(R,R',t+\Delta t;R_\lambda+\Delta R_\lambda)}{\Delta t}\Sigma_{\rm init}(R')\,dR' \\
&\qquad -\int_{R_\lambda}^{\infty}\frac{G(R,R',t;R_\lambda)}{\Delta t}\Sigma_{\rm init}(R')\,dR'\Big] \\
&= \lim_{\Delta t\to 0}\int_{R_\lambda}^{\infty}\left[\frac{G(R,R',t+\Delta t;R_\lambda+\Delta R_\lambda)}{\Delta t}-\frac{G(R,R',t;R_\lambda)}{\Delta t}\right]\Sigma_{\rm init}(R')\,dR' \\
&\qquad +\lim_{\Delta t\to 0}\int_{R_\lambda+\Delta R_\lambda}^{R_\lambda}\frac{G(t+\Delta t;R_\lambda+\Delta R_\lambda)}{\Delta t}\Sigma_{\rm init}(R')\,dR'. \qquad (6.43)
\end{aligned}$$

The last term evaluates to

$$\lim_{\Delta t\to 0}\frac{R_\lambda(t)-R_\lambda(t+\Delta t)}{\Delta t}\,G(R,R_\lambda,t;R_\lambda)\,\Sigma_{\rm init}(R_\lambda)=-\frac{dR_\lambda}{dt}\,G(R,R_\lambda,t;R_\lambda)\,\Sigma_{\rm init}(R_\lambda), \qquad (6.44)$$

which must be zero for all $t > 0$ since the initial surface density profile is zero inside $R_\lambda(t=0)$. We are left with

$$\begin{aligned}
\frac{d}{dt}\Sigma(R,t) &= \lim_{\Delta t\to 0}\int_{R_\lambda}^{\infty}\left[\frac{G\left(R,R',t+\Delta t;R_\lambda(t+\Delta t)\right)}{\Delta t}-\frac{G(R,R',t;R_\lambda)}{\Delta t}\right]\Sigma_{\rm init}(R')\,dR' \\
&= \int_{R_\lambda(t)}^{\infty}\left[\frac{\partial G(R,R',t;R_\lambda)}{\partial t}+\frac{\partial G(R,R',t;R_\lambda)}{\partial R_\lambda}\frac{dR_\lambda}{dt}\right]\Sigma_{\rm init}(R')\,dR'. \qquad (6.45)
\end{aligned}$$

Equation 6.45 is an ordinary differential equation with respect to the variable t. We wish to begin with an initial condition $\Sigma_{\rm init}(R)$ and integrate to a subsequent time t^*, with $R^*_\lambda = R_\lambda(t^*)$. In the context of this paper, the initial condition is the point when the binary's orbital decay becomes GW-driven, and t^* is the time when the system is observed as an

individually resolved PTA source and $R_\lambda(t) \geq R^*_\lambda$ describes the (GW-driven) evolutionary history of the gap-opening radius prior to t^*.

Direct integration gives the expression

$$\Sigma(R, t^*) = \Sigma_{\rm init}(R') + \int_0^{t^*} \int_{R_\lambda(t)}^{\infty} \left[\frac{\partial G(R, R', t; R_\lambda)}{\partial t} + \frac{\partial G(R, R', t; R_\lambda)}{\partial R_\lambda} \frac{dR_\lambda}{dt} \right] \Sigma_{\rm init}(R')\, dR'\, dt. \quad (6.46)$$

We may write this in the form of a Green's function,

$$\Sigma(R, t) = \int_0^{\infty} \mathcal{G}(R, R', t^*)\, \Sigma_{\rm init}(R')\, dR', \quad (6.47)$$

where

$$\mathcal{G}(R, R', t^*) = \delta(R - R') + \int_0^{t^*} \left[\frac{\partial G(R, R', t; R_\lambda)}{\partial t} + \frac{\partial G(R, R', t; R_\lambda)}{\partial R_\lambda} \frac{dR_\lambda}{dt} \right] dt. \quad (6.48)$$

Note that $G = 0$ for all $R < R_\lambda$ and $R' < R_\lambda$. The Dirac δ-function can be awkward to implement in a numerical integration scheme. We rewrite it in terms of the known Green's function with a fixed boundary by using the fact that any Green's function evaluated at $t = 0$ is the δ-function, i.e.,

$$G(R, R', t^*; R^*_\lambda) = \delta(R - R') + \int_0^{t^*} \frac{\partial G(R, R', t; R^*_\lambda)}{\partial t}\, dt. \quad (6.49)$$

We thus obtain for our "master" Green's function:

$$\mathcal{G}(R,R',t^*) = G(R,R',t^*;R^*_\lambda) + \int_0^{t^*} \left[\frac{\partial G(R,R',t;R_\lambda)}{\partial t} - \frac{\partial G(R,R',t;R^*_\lambda)}{\partial t} + \frac{\partial G(R,R',t;R_\lambda)}{\partial R_\lambda} \frac{dR_\lambda}{dt} \right] dt. \tag{6.50}$$

We see that if the boundary is stationary, the second term above vanishes and $\mathcal{G} = G$. Note the similarity of the mathematical form of our solution to DuHamel's theorem (e.g., Carslaw & Jaeger 1959) for time-dependent boundary conditions. This is not surprising, given that both are based on the superposition principle. For our thin accretion disk problem, the function $\mathcal{G}(R,R',t)$ is explicitly known and easily tabulated, given a specific combination of: (i) the parameter $f_{\rm leak}$; (ii) boundary evolution $R_\lambda(t)$; and (iii) viscous power-law index $\nu(R) \propto R^n$, with $n < 2$.

Chapter 7

Conclusion

The last five years have seen remarkable advances in the theory of coalescing SMBHs. 2005 saw the solution to the longstanding technical problem of simulating the gravitational waveform of coalescence (Pretorius 2005; Campanelli et al. 2006; Baker et al. 2006a). This breakthrough was followed by an explosion in numerical studies of the recoil effect, spin evolution and orbital eccentricity. Attention has now turned to the problem of extracting physical information from the waveform.

Starting at around the same time, the possibility that merging SMBH binaries could be used as standard sirens for cosmology became a major focal point (Holz & Hughes 2005; cf. Schutz 1986). As might be expected of any new field with many exciting scientific prospects and few empirical constraints, the theoretical modeling of EM signatures of merging SMBH binaries has been very much a theoretician's playground. Numerous mechanisms for eliciting an EM signature of compact and coalescing SMBH binaries and recoiling binary remnants have already been proposed (Chapter 1).

The excitement accompanying the ongoing and rapid theoretical developments is matched by the eager anticipation of forthcoming observations by planned GW detectors and EM telescopes. I conclude my thesis with an overview of the future prospects for this nascent field and some speculative thoughts on its direction.

7.1 Gravitational-wave Astronomy

In March 2011, NASA announced that it is canceling its participation in the *LISA* mission — as well as in the *International X-ray Observatory* and in *EJSM-Laplace*, a mission that would send a spacecraft to Jupiter and its moons Ganymede and Europa — as a joint-level project. Although NASA has not ruled out the possibility of support or participation in these missions on lesser scales, it now falls on the ESA to take total ownership of them. ESA is now evaluating all three missions to determine whether they can be redefined and redesigned to fit the new financial constraints while still providing a sufficiently rewarding scientific return. Before the announcement, the *LISA Pathfinder* mission — a test mission of a single prototype *LISA*-type module — had been planned for launch as early as 2013.

Despite the cloud of uncertainty cast by NASA's withdrawal, the overall outlook for *LISA*, or a mission similar to it, remains positive. It could be possible to build a less sensitive observatory comparable to the original *LISA* design with an ESA-only budget. The goal sensitivity levels for *LISA* were so high, and SMBH coalescences are believed to be such robustly detectable events (Figure 1.5), that a lesser version of the mission could still detect them. The most significant loss in a redefined space detector mission is likely

to be the detectability of EMRI events.[1]

The performance of ground-based GW experiments such as *Advanced LIGO* over the next decade will be paramount in determining the future of *LISA*, and of GW Astronomy in general. The scientific yield of successful detections will be enormous — to name just a few: constraints on the GW speed and graviton mass, probe of the population of black-hole and neutron-star binaries in the Galaxy, and tests of alternative gravity theories which predict waveforms different from those of general relativity. On a practical level the unequivocal confirmation of the existence of GWs will further strengthen the scientific case for future missions such as *LISA*, *DECIGO* and the proposed *Einstein Telescope*[2]. Ground-based detections will also serve as the first real test for the highly intricate and difficult task of extracting physical quantities from the waveforms of coalescence. Because some sources of ground-based GW interferometers are expected to trigger luminous EM counterparts such as gamma-ray bursts, these experiments will lay the ground work for time-sensitive multi-messenger astronomy (Kanner et al. 2008). Two of the main challenges are the rapid localization of the source from the GW signal and a highly synergistic coordination of the GW detectors with EM telescopes.

It is not impossible that ground-based observatories will fail to detect GWs — although this would be a truly baffling result. The existence of GWs is not just a postulate, but rather a fundamental prediction of general relativity. It is more than a hypothesis, as it has been indirectly observed, for more than three decades, in the orbital decay of the Hulse-Taylor binary pulsar. Considering that general relativity has survived virtually

[1]Oral presentation by Oliver Jennrich at *Bridging Electromagnetic Astrophysics and Cosmology with Gravitational Waves* workshop in Milan, Italy, March 2011.

[2]`http://www.et-gw.eu/`

every empirical test (see Will 2006 for a comprehensive review), a total lack of direct detections would suggest that there is something fundamentally incomplete about our understanding of gravity on macroscopic scales. This would be a tremendous scientific result in its own right; unexpected results are the greatest catalysts of scientific discovery. However, non-detections of expected sources by *Advanced LIGO* and its sister missions could have devastating effects on the future allocation of resources toward future GW interferometers.

Pulsar timing arrays offer a promising way to detect the stochastic nano-Hz GW background, but the ability to resolve individual SMBH binaries will be limited to the precision with which radio telescopes can measure timing residuals and the number of the most precisely timed pulsars. The *Square Kilometer Array* offers by far the most long-term promise for pulsar timing. The sensitivity of the array will allow for the most precise timing of pulsars to date; the large field of view, which will permit the instrument to time as many as thousands of pulsars per day, will enable the discovery of numerous pulsars suitable for inclusion in PTAs (Cordes et al. 2004; Sesana et al. 2008).

7.2 The Prospects for EM Observations of SMBH Binaries

Roughly two decades of deep, wide-angle EM surveys have revolutionized the study of AGN. A significant focus of EM astronomy is the search for transient activity. Current and forthcoming high-cadence, wide-field surveys — such as *Swift*, *LSST* and *WFXT* — will provide, for the first time, a comprehensive picture of the transient EM sky. The timing of these missions is fortuitous for the purposes of multi-messenger astronomy of GW

sources: within the next decade, both EM and GW observatories will be actively scanning the sky for transient and variable events.

Ongoing and future time-domain EM surveys will detect conspicuous AGN variability, such as rapid and persistent evolution, sudden transitions, and periodic variability. They should also detect many tidal disruption events, a significant fraction of which could be caused by SMBH binaries (Chen et al. 2011). Very recently, *Swift* detected the sudden onset of persistent blazar-like emission from a SMBH that was previously inactive (Burrows et al. 2011) — the first event of its kind to be observed. If SMBH binaries do elicit dramatic AGN signatures, then these signatures should be discovered by EM surveys. On the other hand, it may also be the case that "extraordinary" time-varying behavior is actually quite ordinary for AGN, and that SMBH binaries account for only a small fraction of them. This underscores the possibility that synergy with GW observatories may be the only way to unambiguously identify SMBH binaries. New studies of the transient EM sky will almost certainly find new puzzles as well as clues to understanding the interactions SMBHs can have with their environment.

The time-domain surveys will be complemented by the next generation of high-resolution, high-senstivity space telescopes, including the *James Webb Space Telescope* and the *International X-ray Observatory*. These missions will be able to probe the quasar population out to $z \sim 10$ and perform high-precision spectroscopy and high-resolution imaging of AGN. These telescopes also offer the best chance at observing the EM counterparts to the most distant merger events detectable by a detector like *LISA*.

7.3 Future Modeling of EM Signatures of SMBH Binaries

Gas accretion is the only mechanism that powers persistent, luminous emission from SMBHs in the Universe, so it is logical to make it a focal point of efforts to model EM emission from SMBH binaries. As with accretion onto a single SMBH, the problem here consists of a simple and well-characterized engine and a complicated and poorly understood fueling mechanism. The primary difference between the solitary and binary problems is also the strongest physical reason to expect an EM signature; in the latter case, the engine is able to interact with the fuel in a coherent way, on timescales determined by its orbital parameters.

The central question, then, is: What is this interaction? A central cavity, as explored at length in this thesis, could produce a signature, as could radial leakage of gas into the cavity. Other possibilities include large, periodic fluctuations in the radial (azimuthally averaged) structure of the inner disk, and axially asymmetric structures that result in variations in the damping of spiral density waves.

Most studies of gas-binary dynamics have focused on α-prescription disks in the gas pressure-dominated regime, where the disk is geometrically thin and radiative transport is unimportant. Simulations of SMBH circumbinary disks (e.g., MacFadyen & Milosavljević 2008; Cuadra et al. 2009) have also focused largely on the gas-dominated regime. However, for merging binaries, the relevant regime is the radiation pressure-dominated one, where gas-binary dynamics have not been extensively studied. To fully consider the problem numerically could be extremely prohibitive, as it requires treatments of vertical structure and radiation transport. Analytic and semianalytic arguments will be useful in

outlining the possible dynamical differences between different regimes. In the same vein, understanding the orbital dynamics of fluids for different viscosity prescriptions, as well as in other disk models such as slim disks or magnetohydrodynamic disks, may provide predictive differences between accretion mechanisms.

Even in the absence of knowledge regarding the detailed (and likely quite complicated) gas dynamics, it may be possible to characterize the physical characteristics of the gas inside the binary's orbit. For example, the suppression of accretion into the cavity may naturally cause the central accretion flow to be radiatively inefficient, as in an ADAF, even if the circumbinary gas is a gas pressure-dominated thin disk. The tidal precursor model of Chang et al. (2010) relies on a geometrically thin circumprimary disk, while the analytic argument for a prompt flare by Krolik (2010) requires that the gas be optically thick.

On the other hand, it is possible that future EM surveys will detect an unambiguous EM signature of a coalescing or compact SMBH binary without the benefit of a detailed theoretical model — e.g., with the aid of GW localization or perhaps through an unmistakable periodic emission feature. As discussed in Chapter 1, if this is a SMBH whose coalescence has been observed in GWs, the knowledge of its mass and spin can be used to confirm and calibrate AGN theory. A more exciting scenario would be the discovery of a "live" binary that has not yet merged. Even if such a system is discovered only in EM emission, it will provide an unprecedented opportunity to observe an accretion flow that is being continuously and regularly stirred and prodded by the central engine. Detailed follow-up studies of the response of the central gas could be used to infer the gas

dynamics and properties, as well as to derive knowledge about the underlying accretion mechanism.

Bibliography

Abbott, B., Abbott, R., Adhikari, R., Ageev, A., Allen, B., Amin, R., Anderson, S. B., Anderson, W. G., Araya, M., Armandula, H., & et al. 2005, Physical Review Letters, 94, 181103

Abel, T., Bryan, G. L., & Norman, M. L. 2000, ApJ, 540, 39

—. 2002, Science, 295, 93

Abramowicz, M. A., Czerny, B., Lasota, J. P., & Szuszkiewicz, E. 1988, ApJ, 332, 646

Alexander, D. M., Brandt, W. N., Smail, I., Swinbank, A. M., Bauer, F. E., Blain, A. W., Chapman, S. C., Coppin, K. E. K., Ivison, R. J., & Menéndez-Delmestre, K. 2008, AJ, 135, 1968

Alvarez, M. A., Wise, J. H., & Abel, T. 2009, ApJ, 701, L133

Amaro-Seoane, P., Gair, J. R., Freitag, M., Miller, M. C., Mandel, I., Cutler, C. J., & Babak, S. 2007, Classical and Quantum Gravity, 24, 113

Amaro-Seoane, P., Sesana, A., Hoffman, L., Benacquista, M., Eichhorn, C., Makino, J., & Spurzem, R. 2010, MNRAS, 402, 2308

Anderson, M., Lehner, L., Megevand, M., & Neilsen, D. 2010, Phys. Rev. D, 81, 044004

Armitage, P. J. & Natarajan, P. 2002, ApJ, 567, L9

—. 2005, ApJ, 634, 921

Artemova, I. V., Bjoernsson, G., & Novikov, I. D. 1996, ApJ, 461, 565

Artymowicz, P., Clarke, C. J., Lubow, S. H., & Pringle, J. E. 1991, ApJ, 370, L35

Artymowicz, P. & Lubow, S. H. 1994, ApJ, 421, 651

—. 1996, ApJ, 467, L77+

Arun, K. G., Babak, S., Berti, E., Cornish, N., Cutler, C., Gair, J., Hughes, S. A., Iyer, B. R., Lang, R. N., Mandel, I., Porter, E. K., Sathyaprakash, B. S., Sinha, S., Sintes, A. M., Trias, M., Van Den Broeck, C., & Volonteri, M. 2009, Classical and Quantum Gravity, 26, 094027

Ayal, S., Livio, M., & Piran, T. 2000, ApJ, 545, 772

Bachiller, R. 1996, ARA&A, 34, 111

Backer, D. C., Jaffe, A. H., & Lommen, A. N. 2004, Coevolution of Black Holes and Galaxies, 438

Bahcall, J. N., Kirhakos, S., Saxe, D. H., & Schneider, D. P. 1997, ApJ, 479, 642

Baker, J., Campanelli, M., Lousto, C. O., & Takahashi, R. 2004, Phys. Rev. D, 69, 027505

Baker, J. G., Boggs, W. D., Centrella, J., Kelly, B. J., McWilliams, S. T., Miller, M. C., & van Meter, J. R. 2008, ApJ, 682, L29

Baker, J. G., Centrella, J., Choi, D., Koppitz, M., & van Meter, J. 2006a, Phys. Rev. Lett., 96, 111102

Baker, J. G., Centrella, J., Choi, D., Koppitz, M., van Meter, J. R., & Miller, M. C. 2006b, ApJ, 653, L93

Baker, J. G., McWilliams, S. T., van Meter, J. R., Centrella, J., Choi, D., Kelly, B. J., & Koppitz, M. 2007, Phys. Rev. D, 75, 124024

Balbus, S. A. & Hawley, J. F. 1991, ApJ, 376, 214

Barack, L. & Cutler, C. 2004, Phys. Rev. D, 69, 082005

Bardeen, J. M. 1970, Nature, 226, 64

Barkana, R. 2004, MNRAS, 347, 59

Barkana, R. & Loeb, A. 2001, Phys. Rep., 349, 125

Barnes, J. E. & Hernquist, L. E. 1991, ApJ, 370, L65

Barth, A. J., Bentz, M. C., Greene, J. E., & Ho, L. C. 2008, ApJ, 683, L119

Batcheldor, D., Robinson, A., Axon, D. J., Perlman, E. S., & Merritt, D. 2010, ApJ, 717, L6

Bate, M. R., Lubow, S. H., Ogilvie, G. I., & Miller, K. A. 2003, MNRAS, 341, 213

Begelman, M. C. 2002, ApJ, 568, L97

Begelman, M. C., Blandford, R. D., & Rees, M. J. 1980, Nature, 287, 307

—. 1984, Reviews of Modern Physics, 56, 255

Begelman, M. C., Volonteri, M., & Rees, M. J. 2006, MNRAS, 370, 289

Bekenstein, J. D. 1973, ApJ, 183, 657

Berczik, P., Merritt, D., Spurzem, R., & Bischof, H. 2006, ApJ, 642, L21

Berti, E., Buonanno, A., & Will, C. M. 2005, Phys. Rev. D, 71, 084025

Berti, E. & Volonteri, M. 2008, ApJ, 684, 822

Bianchi, S., Chiaberge, M., Piconcelli, E., Guainazzi, M., & Matt, G. 2008, MNRAS, 386, 105

Binney, J. & Tremaine, S. 1987, Galactic dynamics (Princeton University Press, Princeton, NJ)

Blaes, O., Lee, M. H., & Socrates, A. 2002, ApJ, 578, 775

Blaes, O. M. 2004, in Accretion Discs, Jets and High Energy Phenomena in Astrophysics, ed. V. Beskin, G. Henri, F. Menard, & et al., 137–185

Blanchet, L. 2006, Living Reviews in Relativity, 9, 4

Blandford, R. D. & Begelman, M. C. 1999, MNRAS, 303, L1

Blandford, R. D. & Konigl, A. 1979, ApJ, 232, 34

Blandford, R. D. & McKee, C. F. 1982, ApJ, 255, 419

Blandford, R. D. & Payne, D. G. 1982, MNRAS, 199, 883

Blandford, R. D. & Znajek, R. L. 1977, MNRAS, 179, 433

Blecha, L. & Loeb, A. 2008, MNRAS, 390, 1311

Bloom, J. S., Holz, D. E., Hughes, S. A., & Menou, K. 2009, in Decadal Survey White Paper, Vol. 2010, No. 20

Bode, N. & Phinney, S. 2007, APS Meeting Abstracts, 1010

Bogdanović, T., Eracleous, M., Mahadevan, S., Sigurdsson, S., & Laguna, P. 2004, ApJ, 610, 707

Bogdanović, T., Eracleous, M., & Sigurdsson, S. 2009, ApJ, 697, 288

Bogdanović, T., Reynolds, C. S., & Miller, M. C. 2007, ApJ, 661, L147

Bogdanović, T., Smith, B. D., Sigurdsson, S., & Eracleous, M. 2008, ApJS, 174, 455

Bondi, H. 1952, MNRAS, 112, 195

Bondi, H. & Hoyle, F. 1944, MNRAS, 104, 273

Bonning, E. W., Shields, G. A., & Salviander, S. 2007, ApJ, 666, L13

Bonoli, S., Shankar, F., White, S. D. M., Springel, V., & Wyithe, J. S. B. 2010, MNRAS, 404, 399

Boroson, T. A. & Lauer, T. R. 2009, Nature, 458, 53

Boylan-Kolchin, M., Ma, C., & Quataert, E. 2004, ApJ, 613, L37

—. 2006, MNRAS, 369, 1081

—. 2008, MNRAS, 383, 93

Brandt, W. N. & Hasinger, G. 2005, ARA&A, 43, 827

Brandt, W. N., Laor, A., & Wills, B. J. 2000, ApJ, 528, 637

Bromley, J. M., Somerville, R. S., & Fabian, A. C. 2004, MNRAS, 350, 456

Bromm, V., Coppi, P. S., & Larson, R. B. 2002, ApJ, 564, 23

Bromm, V. & Loeb, A. 2003, ApJ, 596, 34

Burrows, D. N., Kennea, J. A., Ghisellini, G., Mangano, V., Zhang, B., Page, K. L., Eracleous, M., Romano, P., Sakamoto, T., Falcone, A. D., Osborne, J. P., Campana, S., Beardmore, A. P., Breeveld, A. A., Chester, M. M., Corbet, R., Covino, S., Cummings, J. R., D'Avanzo, P., D'Elia, V., Esposito, P., Evans, P. A., Fugazza, D., Gelbord, J. M., Hiroi, K., Holland, S. T., Huang, K. Y., Im, M., Israel, G., Jeon, Y., Jeon, Y. ., Kawai, N., Krimm, H. A., Mészáros, P., Negoro, H., Omodei, N., Park, W. ., Perkins, J. S., Sugizaki, M., Sung, H. ., Tagliaferri, G., Troja, E., Ueda, Y., Urata, Y., Usui, R., Antonelli, L. A., Barthelmy, S. D., Cusumano, G., Giommi, P., Marshall, F. E., Melandri, A., Perri, M., Racusin, J. L., Sbarufatti, B., Siegel, M. H., & Gehrels, N. 2011, arXiv e-prints, 1104.4787

Callegari, S., Mayer, L., Kazantzidis, S., Colpi, M., Governato, F., Quinn, T., & Wadsley, J. 2009, ApJ, 696, L89

Campanelli, M., Lousto, C., Zlochower, Y., & Merritt, D. 2007a, ApJ, 659, L5

Campanelli, M., Lousto, C. O., Marronetti, P., & Zlochower, Y. 2006, Phys. Rev. Lett., 96, 111101

Campanelli, M., Lousto, C. O., Zlochower, Y., & Merritt, D. 2007b, Physical Review Letters, 98, 231102

Cannizzo, J. K., Lee, H. M., & Goodman, J. 1990, ApJ, 351, 38

Carlberg, R. G. 1990, ApJ, 350, 505

Carslaw, H. S. & Jaeger, J. C. 1959, Conduction of heat in solids

Chandrasekhar, S. 1943, ApJ, 97, 255

Chang, P., Strubbe, L. E., Menou, K., & Quataert, E. 2010, MNRAS, 1003

Chen, X., Madau, P., Sesana, A., & Liu, F. K. 2009, ApJ, 697, L149

Chen, X., Sesana, A., Madau, P., & Liu, F. K. 2011, ApJ, 729, 13

Chornock, R., Bloom, J. S., Cenko, S. B., Silverman, J. M., Filippenko, A. V., Hicks, M. D., Lawrence, K. J., Chang, P., Comerford, J. M., George, M. R., Modjaz, M., Oishi, J. S., Quataert, E., & Strubbe, L. E. 2009, The Astronomer's Telegram, 1955, 1

Cisternas, M., Jahnke, K., Inskip, K. J., Kartaltepe, J., Koekemoer, A. M., Lisker, T., Robaina, A. R., Scodeggio, M., Sheth, K., Trump, J. R., Andrae, R., Miyaji, T., Lusso, E., Brusa, M.,

Capak, P., Cappelluti, N., Civano, F., Ilbert, O., Impey, C. D., Leauthaud, A., Lilly, S. J., Salvato, M., Scoville, N. Z., & Taniguchi, Y. 2011, ApJ, 726, 57

Civano, F., Elvis, M., Lanzuisi, G., Jahnke, K., Zamorani, G., Blecha, L., Bongiorno, A., Brusa, M., Comastri, A., Hao, H., Leauthaud, A., Loeb, A., Mainieri, V., Piconcelli, E., Salvato, M., Scoville, N., Trump, J., Vignali, C., Aldcroft, T., Bolzonella, M., Bressert, E., Finoguenov, A., Fruscione, A., Koekemoer, A. M., Cappelluti, N., Fiore, F., Giodini, S., Gilli, R., Impey, C. D., Lilly, S. J., Lusso, E., Puccetti, S., Silverman, J. D., Aussel, H., Capak, P., Frayer, D., Le Floch, E., McCracken, H. J., Sanders, D. B., Schiminovich, D., & Taniguchi, Y. 2010, ApJ, 717, 209

Cole, S., Lacey, C. G., Baugh, C. M., & Frenk, C. S. 2000, MNRAS, 319, 168

Collin, S. & Zahn, J.-P. 1999, A&A, 344, 433

Colpi, M., Callegari, S., Dotti, M., & Mayer, L. 2009, Class. Quantum Grav., 26, 094029

Comerford, J. M., Gerke, B. F., Newman, J. A., Davis, M., Yan, R., Cooper, M. C., Faber, S. M., Koo, D. C., Coil, A. L., Rosario, D. J., & Dutton, A. A. 2009a, ApJ, 698, 956

Comerford, J. M., Griffith, R. L., Gerke, B. F., Cooper, M. C., Newman, J. A., Davis, M., & Stern, D. 2009b, ApJ, 702, L82

Corbin, V. & Cornish, N. J. 2010, arXiv e-prints, 1008.1782

Cordes, J. M., Kramer, M., Lazio, T. J. W., Stappers, B. W., Backer, D. C., & Johnston, S. 2004, New A Rev., 48, 1413

Cornish, N. J. & Crowder, J. 2005, Phys. Rev. D, 72, 043005

Cornish, N. J. & Porter, E. K. 2007, Phys. Rev. D, 75, 021301

Corrales, L. R., Haiman, Z., & MacFadyen, A. 2010, MNRAS, 404, 947

Cuadra, J., Armitage, P. J., Alexander, R. D., & Begelman, M. C. 2009, MNRAS, 393, 1423

Cutler, C. 1998, Phys. Rev. D, 57, 7089

Cutler, C. & Flanagan, É. E. 1994, Phys. Rev. D, 49, 2658

Damour, T., Gopakumar, A., & Iyer, B. R. 2004, Phys. Rev. D, 70, 064028

Damour, T. & Vilenkin, A. 2005, Phys. Rev. D, 71, 063510

Deffayet, C. & Menou, K. 2007, ApJ, 668, L143

Detweiler, S. 1979, ApJ, 234, 1100

Di Matteo, T., Springel, V., & Hernquist, L. 2005, Nature, 433, 604

Diamond-Stanic, A. M., Fan, X., Brandt, W. N., Shemmer, O., Strauss, M. A., Anderson, S. F., Carilli, C. L., Gibson, R. R., Jiang, L., Kim, J. S., Richards, G. T., Schmidt, G. D., Schneider, D. P., Shen, Y., Smith, P. S., Vestergaard, M., & Young, J. E. 2009, ApJ, 699, 782

Dijkstra, M., Haiman, Z., Rees, M. J., & Weinberg, D. H. 2004, ApJ, 601, 666

Djorgovski, S. G., Courbin, F., Meylan, G., Sluse, D., Thompson, D., Mahabal, A., & Glikman, E. 2007, ApJ, 662, L1

Doeleman, S. S., Weintroub, J., Rogers, A. E. E., Plambeck, R., Freund, R., Tilanus, R. P. J., Friberg, P., Ziurys, L. M., Moran, J. M., Corey, B., Young, K. H., Smythe, D. L., Titus, M., Marrone, D. P., Cappallo, R. J., Bock, D., Bower, G. C., Chamberlin, R., Davis, G. R., Krichbaum, T. P., Lamb, J., Maness, H., Niell, A. E., Roy, A., Strittmatter, P., Werthimer, D., Whitney, A. R., & Woody, D. 2008, Nature, 455, 78

Donovan, J. L., Hibbard, J. E., & van Gorkom, J. H. 2007, AJ, 134, 1118

Dotti, M., Colpi, M., Haardt, F., & Mayer, L. 2007, MNRAS, 379, 956

Dotti, M., Montuori, C., Decarli, R., Volonteri, M., Colpi, M., & Haardt, F. 2009, MNRAS, 398, L73

Dotti, M., Salvaterra, R., Sesana, A., Colpi, M., & Haardt, F. 2006, MNRAS, 372, 869

Dutton, A. A., Conroy, C., van den Bosch, F. C., Prada, F., & More, S. 2010, MNRAS, 407, 2

Ebisuzaki, T., Makino, J., & Okumura, S. K. 1991, Nature, 354, 212

Echeverria, F. 1989, Phys. Rev. D, 40, 3194

Edgar, R. 2004, New A Rev., 48, 843

Ehlers, J., Rosenblum, A., Goldberg, J. N., & Havas, P. 1976, ApJ, 208, L77

Eisenstein, D. J. & Hu, W. 1999, ApJ, 511, 5

Eracleous, M. & Halpern, J. P. 2003, ApJ, 599, 886

Eracleous, M., Halpern, J. P., Gilbert, A. M., Newman, J. A., & Filippenko, A. V. 1997, ApJ, 490, 216

Escala, A., Larson, R. B., Coppi, P. S., & Mardones, D. 2004, ApJ, 607, 765

—. 2005, ApJ, 630, 152

Esposito, F. P. 1971, ApJ, 165, 165

Fabbiano, G., Elvis, M., Markoff, S., Siemiginowska, A., Pellegrini, S., Zezas, A., Nicastro, F., Trinchieri, G., & McDowell, J. 2003, ApJ, 588, 175

Fabian, A. C., Iwasawa, K., Reynolds, C. S., & Young, A. J. 2000, PASP, 112, 1145

Fakhouri, O. & Ma, C. 2009, MNRAS, 394, 1825

Fan, X. 2006, New A Rev., 50, 665

Fan, X., Narayanan, V. K., Lupton, R. H., Strauss, M. A., Knapp, G. R., Becker, R. H., White, R. L., Pentericci, L., Leggett, S. K., Haiman, Z., Gunn, J. E., Ivezić, Ž., Schneider, D. P., Anderson, S. F., Brinkmann, J., Bahcall, N. A., Connolly, A. J., Csabai, I., Doi, M., Fukugita, M., Geballe, T., Grebel, E. K., Harbeck, D., Hennessy, G., Lamb, D. Q., Miknaitis, G., Munn, J. A., Nichol, R., Okamura, S., Pier, J. R., Prada, F., Richards, G. T., Szalay, A., & York, D. G. 2001, AJ, 122, 2833

Fan, X., Strauss, M. A., Schneider, D. P., Becker, R. H., White, R. L., Haiman, Z., Gregg, M., Pentericci, L., Grebel, E. K., Narayanan, V. K., Loh, Y., Richards, G. T., Gunn, J. E., Lupton, R. H., Knapp, G. R., Ivezić, Ž., Brandt, W. N., Collinge, M., Hao, L., Harbeck, D., Prada, F., Schaye, J., Strateva, I., Zakamska, N., Anderson, S., Brinkmann, J., Bahcall, N. A., Lamb, D. Q., Okamura, S., Szalay, A., & York, D. G. 2003, AJ, 125, 1649

Favata, M., Hughes, S. A., & Holz, D. E. 2004, ApJ, 607, L5

Ferrarese, L. 2002, ApJ, 578, 90

Ferrarese, L. & Ford, H. 2005, Space Sci. Rev., 116, 523

Ferrarese, L. & Merritt, D. 2000, ApJ, 539, L9

Ferrarese, L., van den Bosch, F. C., Ford, H. C., Jaffe, W., & O'Connell, R. W. 1994, AJ, 108, 1598

Fitchett, M. J. 1983, MNRAS, 203, 1049

Flanagan, É. É. & Hughes, S. A. 1998, Phys. Rev. D, 57, 4566

—. 2005, New Journal of Physics, 7, 204

Floyd, D. J. E., Kukula, M. J., Dunlop, J. S., McLure, R. J., Miller, L., Percival, W. J., Baum, S. A., & O'Dea, C. P. 2004, MNRAS, 355, 196

Foster, R. S. & Backer, D. C. 1990, ApJ, 361, 300

Frank, J., King, A., & Raine, D. J. 2002, Accretion Power in Astrophysics: Third Edition (Cambridge University Press)

Gabasch, A., Bender, R., Seitz, S., Hopp, U., Saglia, R. P., Feulner, G., Snigula, J., Drory, N., Appenzeller, I., Heidt, J., Mehlert, D., Noll, S., Böhm, A., Jäger, K., Ziegler, B., & Fricke, K. J. 2004, A&A, 421, 41

Gabasch, A., Hopp, U., Feulner, G., Bender, R., Seitz, S., Saglia, R. P., Snigula, J., Drory, N., Appenzeller, I., Heidt, J., Mehlert, D., Noll, S., Böhm, A., Jäger, K., & Ziegler, B. 2006, A&A, 448, 101

Gair, J. R., Barack, L., Creighton, T., Cutler, C., Larson, S. L., Phinney, E. S., & Vallisneri, M. 2004, Classical and Quantum Gravity, 21, 1595

Galloway, D. K. & Sokoloski, J. L. 2004, ApJ, 613, L61

Gammie, C. F. 2001, ApJ, 553, 174

Gaskell, C. M. 1983, in Liege International Astrophysical Colloquia, Vol. 24, Liege International Astrophysical Colloquia, ed. J.-P. Swings, 473–477

Gaskell, C. M. 1985, Nature, 315, 386

—. 1996, ApJ, 464, L107+

Gebhardt, K., Bender, R., Bower, G., Dressler, A., Faber, S. M., Filippenko, A. V., Green, R., Grillmair, C., Ho, L. C., Kormendy, J., Lauer, T. R., Magorrian, J., Pinkney, J., Richstone, D., & Tremaine, S. 2000, ApJ, 539, L13

Gezari, S., Heckman, T., Cenko, S. B., Eracleous, M., Forster, K., Gonçalves, T. S., Martin, D. C., Morrissey, P., Neff, S. G., Seibert, M., Schiminovich, D., & Wyder, T. K. 2009, ApJ, 698, 1367

Gezari, S., Martin, D. C., Milliard, B., Basa, S., Halpern, J. P., Forster, K., Friedman, P. G., Morrissey, P., Neff, S. G., Schiminovich, D., Seibert, M., Small, T., & Wyder, T. K. 2006, ApJ, 653, L25

Ghez, A. M., Salim, S., Hornstein, S. D., Tanner, A., Lu, J. R., Morris, M., Becklin, E. E., & Duchêne, G. 2005, ApJ, 620, 744

Gibson, R. R., Brandt, W. N., & Schneider, D. P. 2008, ApJ, 685, 773

Gierliński, M., Done, C., & Page, K. 2009, MNRAS, 392, 1106

Giocoli, C., Tormen, G., Sheth, R. K., & van den Bosch, F. C. 2010, MNRAS, 404, 502

Goldreich, P. & Tremaine, S. 1980, ApJ, 241, 425

González, J. A., Hannam, M., Sperhake, U., Brügmann, B., & Husa, S. 2007, Physical Review Letters, 98, 231101

Goodman, J. 2003, MNRAS, 339, 937

Gopal-Krishna, Biermann, P. L., & Wiita, P. J. 2003, ApJ, 594, L103

Gould, A. & Rix, H. 2000, ApJ, 532, L29

Governato, F., Brook, C. B., Brooks, A. M., Mayer, L., Willman, B., Jonsson, P., Stilp, A. M., Pope, L., Christensen, C., Wadsley, J., & Quinn, T. 2009, MNRAS, 398, 312

Granato, G. L., De Zotti, G., Silva, L., Bressan, A., & Danese, L. 2004, ApJ, 600, 580

Green, P. J., Myers, A. D., Barkhouse, W. A., Mulchaey, J. S., Bennert, V. N., Cox, T. J., & Aldcroft, T. L. 2010, ApJ, 710, 1578

Grupe, D., Thomas, H., & Beuermann, K. 2001, A&A, 367, 470

Guedes, J., Diemand, J., Zemp, M., Kuhlen, M., Madau, P., & Mayer, L. 2008, Astronomische Nachrichten, 329, 1004

Gültekin, K., Richstone, D. O., Gebhardt, K., Lauer, T. R., Tremaine, S., Aller, M. C., Bender, R., Dressler, A., Faber, S. M., Filippenko, A. V., Green, R., Ho, L. C., Kormendy, J., Magorrian, J., Pinkney, J., & Siopis, C. 2009, ApJ, 698, 198

Günther, R., Schäfer, C., & Kley, W. 2004, A&A, 423, 559

Haehnelt, M. G. 1994, MNRAS, 269, 199

—. 2003, Classical and Quantum Gravity, 20, 31

Haehnelt, M. G., Natarajan, P., & Rees, M. J. 1998, MNRAS, 300, 817

Haiman, Z. 2004, ApJ, 613, 36

Haiman, Z., Abel, T., & Rees, M. J. 2000, ApJ, 534, 11

Haiman, Z. & Bryan, G. L. 2006, ApJ, 650, 7

Haiman, Z., Ciotti, L., & Ostriker, J. P. 2004, ApJ, 606, 763

Haiman, Z., Jimenez, R., & Bernardi, M. 2007, ApJ, 658, 721

Haiman, Z., Kocsis, B., & Menou, K. 2009a, ApJ, 700, 1952

Haiman, Z., Kocsis, B., Menou, K., Lippai, Z., & Frei, Z. 2009b, Classical and Quantum Gravity, 26, 094032

Haiman, Z. & Loeb, A. 2001, ApJ, 552, 459

Haiman, Z., Rees, M. J., & Loeb, A. 1997, ApJ, 476, 458

Haiman, Z., Thoul, A. A., & Loeb, A. 1996, ApJ, 464, 523

Halpern, J. P. & Eracleous, M. 2000, ApJ, 531, 647

Halpern, J. P. & Filippenko, A. V. 1988, Nature, 331, 46

Häring, N. & Rix, H.-W. 2004, ApJ, 604, L89

Hawking, S. W. 1966, ApJ, 145, 544

Hayasaki, K. 2009, PASJ, 61, 65

Hayasaki, K., Mineshige, S., & Ho, L. C. 2008, ApJ, 682, 1134

Hayasaki, K., Mineshige, S., & Sudou, H. 2007, PASJ, 59, 427

Hayasaki, K. & Okazaki, A. T. 2009, ApJ, 691, L5

Heckman, T. M., Krolik, J. H., Moran, S. M., Schnittman, J., & Gezari, S. 2009, ApJ, 695, 363

Heger, A., Fryer, C. L., Woosley, S. E., Langer, N., & Hartmann, D. H. 2003, ApJ, 591, 288

Heggie, D. C. 1975, MNRAS, 173, 729

Heller, C. H. & Shlosman, I. 1994, ApJ, 424, 84

Hernquist, L. 1989, Nature, 340, 687

Hernquist, L. & Mihos, J. C. 1995, ApJ, 448, 41

Hickox, R. C., Jones, C., Forman, W. R., Murray, S. S., Kochanek, C. S., Eisenstein, D., Jannuzi, B. T., Dey, A., Brown, M. J. I., Stern, D., Eisenhardt, P. R., Gorjian, V., Brodwin, M., Narayan, R., Cool, R. J., Kenter, A., Caldwell, N., & Anderson, M. E. 2009, ApJ, 696, 891

Hirata, C. M., Holz, D. E., & Cutler, C. 2010, Phys. Rev. D, 81, 124046

Hirose, S., Blaes, O., & Krolik, J. H. 2009a, ApJ, 704, 781

Hirose, S., Krolik, J. H., & Blaes, O. 2009b, ApJ, 691, 16

Hobbs, G., Archibald, A., Arzoumanian, Z., Backer, D., Bailes, M., Bhat, N. D. R., Burgay, M., Burke-Spolaor, S., Champion, D., Cognard, I., Coles, W., Cordes, J., Demorest, P., Desvignes, G., Ferdman, R. D., Finn, L., Freire, P., Gonzalez, M., Hessels, J., Hotan, A., Janssen, G., Jenet, F., Jessner, A., Jordan, C., Kaspi, V., Kramer, M., Kondratiev, V., Lazio, J., Lazaridis, K., Lee, K. J., Levin, Y., Lommen, A., Lorimer, D., Lynch, R., Lyne, A., Manchester, R., McLaughlin, M., Nice, D., Oslowski, S., Pilia, M., Possenti, A., Purver, M., Ransom, S., Reynolds, J., Sanidas, S., Sarkissian, J., Sesana, A., Shannon, R., Siemens, X., Stairs, I., Stappers, B., Stinebring, D., Theureau, G., van Haasteren, R., van Straten, W., Verbiest, J. P. W., Yardley, D. R. B., & You, X. P. 2010, Classical and Quantum Gravity, 27, 084013

Hoffman, L. & Loeb, A. 2006, ApJ, 638, L75

—. 2007, MNRAS, 377, 957

Holley-Bockelmann, K., Micic, M., Sigurdsson, S., & Rubbo, L. J. 2010, ApJ, 713, 1016

Holz, D. E. & Hughes, S. A. 2005, ApJ, 629, 15

Hopkins, P. F., Bundy, K., Hernquist, L., & Ellis, R. S. 2007a, ApJ, 659, 976

Hopkins, P. F. & Hernquist, L. 2009, ApJ, 698, 1550

—. 2010, MNRAS, 407, 447

Hopkins, P. F., Hernquist, L., Cox, T. J., & Kereš, D. 2008, ApJS, 175, 356

Hopkins, P. F., Richards, G. T., & Hernquist, L. 2007b, ApJ, 654, 731

Hoyle, F. & Lyttleton, R. A. 1939, in Proceedings of the Cambridge Philosophical Society, Vol. 35, Proceedings of the Cambridge Philosophical Society, 405–+

Hughes, S. A. 2002, MNRAS, 331, 805

Hughes, S. A. & Blandford, R. D. 2003, ApJ, 585, L101

Ichimaru, S. 1977, ApJ, 214, 840

Islam, R. R., Taylor, J. E., & Silk, J. 2004, MNRAS, 354, 629

Ivanov, P. B., Papaloizou, J. C. B., & Polnarev, A. G. 1999, MNRAS, 307, 79

Iwasawa, M., Funato, Y., & Makino, J. 2006, ApJ, 651, 1059

Jarosik, N., Bennett, C. L., Dunkley, J., Gold, B., Greason, M. R., Halpern, M., Hill, R. S., Hinshaw, G., Kogut, A., Komatsu, E., Larson, D., Limon, M., Meyer, S. S., Nolta, M. R., Odegard, N., Page, L., Smith, K. M., Spergel, D. N., Tucker, G. S., Weiland, J. L., Wollack, E., & Wright, E. L. 2011, ApJS, 192, 14

Jenet, F., Finn, L. S., Lazio, J., Lommen, A., McLaughlin, M., Stairs, I., Stinebring, D., Verbiest, J., Archibald, A., Arzoumanian, Z., Backer, D., Cordes, J., Demorest, P., Ferdman, R., Freire, P., Gonzalez, M., Kaspi, V., Kondratiev, V., Lorimer, D., Lynch, R., Nice, D., Ransom, S., Shannon, R., & Siemens, X. 2009, arXiv e-prints, 0909.1058

Jenet, F. A., Hobbs, G. B., van Straten, W., Manchester, R. N., Bailes, M., Verbiest, J. P. W., Edwards, R. T., Hotan, A. W., Sarkissian, J. M., & Ord, S. M. 2006, ApJ, 653, 1571

Jenkins, A., Frenk, C. S., White, S. D. M., Colberg, J. M., Cole, S., Evrard, A. E., Couchman, H. M. P., & Yoshida, N. 2001, MNRAS, 321, 372

Johnson, B. M. & Quataert, E. 2007, ApJ, 660, 1273

Kanner, J., Huard, T. L., Márka, S., Murphy, D. C., Piscionere, J., Reed, M., & Shawhan, P. 2008, Classical and Quantum Gravity, 25, 184034

Kauffmann, G. & Haehnelt, M. 2000, MNRAS, 311, 576

Kauffmann, G., Heckman, T. M., Tremonti, C., Brinchmann, J., Charlot, S., White, S. D. M., Ridgway, S. E., Brinkmann, J., Fukugita, M., Hall, P. B., Ivezić, Ž., Richards, G. T., & Schneider, D. P. 2003, MNRAS, 346, 1055

Kawamura, S., Nakamura, T., Ando, M., Seto, N., Tsubono, K., Numata, K., Takahashi, R., Nagano, S., Ishikawa, T., Musha, M., Ueda, K., Sato, T., Hosokawa, M., Agatsuma, K., Akutsu, T., Aoyanagi, K., Arai, K., Araya, A., Asada, H., Aso, Y., Chiba, T., Ebisuzaki, T., Eriguchi, Y., Fujimoto, M., Fukushima, M., Futamase, T., Ganzu, K., Harada, T., Hashimoto, T., Hayama, K., Hikida, W., Himemoto, Y., Hirabayashi, H., Hiramatsu, T., Ichiki, K., Ikegami, T., Inoue, K. T., Ioka, K., Ishidoshiro, K., Itoh, Y., Kamagasako, S., Kanda, N., Kawashima, N., Kirihara, H., Kiuchi, K., Kobayashi, S., Kohri, K., Kojima, Y., Kokeyama, K., Kozai, Y., Kudoh, H., Kunimori, H., Kuroda, K., Maeda, K., Matsuhara, H., Mino, Y., Miyakawa, O., Miyoki, S., Mizusawa, H., Morisawa, T., Mukohyama, S., Naito, I., Nakagawa, N., Nakamura, K., Nakano, H., Nakao, K., Nishizawa, A., Niwa, Y., Nozawa, C., Ohashi, M., Ohishi, N., Ohkawa, M., Okutomi, A., Oohara, K., Sago, N., Saijo, M., Sakagami, M., Sakata, S., Sasaki, M., Sato, S., Shibata, M., Shinkai, H., Somiya, K., Sotani, H., Sugiyama, N., Tagoshi, H., Takahashi, T., Takahashi, H., Takahashi, R., Takano, T., Tanaka, T., Taniguchi, K., Taruya, A., Tashiro, H., Tokunari, M., Tsujikawa, S., Tsunesada, Y., Yamamoto, K., Yamazaki, T., Yokoyama, J., Yoo, C., Yoshida, S., & Yoshino, T. 2006, Classical and Quantum Gravity, 23, 125

Kazantzidis, S., Mayer, L., Colpi, M., Madau, P., Debattista, V. P., Wadsley, J., Stadel, J., Quinn, T., & Moore, B. 2005, ApJ, 623, L67

Keeton, C. R., Kuhlen, M., & Haiman, Z. 2005, ApJ, 621, 559

Kidder, L. E. 1995, Phys. Rev. D, 52, 821

King, A. R. & Pringle, J. E. 2006, MNRAS, 373, L90

Kocsis, B., Frei, Z., Haiman, Z., & Menou, K. 2006, ApJ, 637, 27

Kocsis, B., Haiman, Z., & Menou, K. 2008, ApJ, 684, 870

Kocsis, B., Haiman, Z., Menou, K., & Frei, Z. 2007, Phys. Rev. D, 76, 022003

Kocsis, B. & Loeb, A. 2008, Phys. Rev. Lett., 101, 041101

Kocsis, B. & Sesana, A. 2011, MNRAS, 411, 1467

Kollmeier, J. A., Onken, C. A., Kochanek, C. S., Gould, A., Weinberg, D. H., Dietrich, M., Cool, R., Dey, A., Eisenstein, D. J., Jannuzi, B. T., Le Floc'h, E., & Stern, D. 2006, ApJ, 648, 128

Komatsu, E., Dunkley, J., Nolta, M. R., Bennett, C. L., Gold, B., Hinshaw, G., Jarosik, N., Larson, D., Limon, M., Page, L., Spergel, D. N., Halpern, M., Hill, R. S., Kogut, A., Meyer, S. S., Tucker, G. S., Weiland, J. L., Wollack, E., & Wright, E. L. 2009, ApJS, 180, 330

Komossa, S. 2006, Mem. Soc. Astron. Italiana, 77, 733

Komossa, S., Burwitz, V., Hasinger, G., Predehl, P., Kaastra, J. S., & Ikebe, Y. 2003, ApJ, 582, L15

Komossa, S., Zhou, H., & Lu, H. 2008, ApJ, 678, L81

Kong, A. K. H., McClintock, J. E., Garcia, M. R., Murray, S. S., & Barret, D. 2002, ApJ, 570, 277

Konigl, A. 1982, ApJ, 261, 115

Konigl, A. & Pudritz, R. E. 2000, Protostars and Planets IV, 759

Königsdörffer, C. & Gopakumar, A. 2006, Phys. Rev. D, 73, 124012

Kormendy, J. & Bender, R. 2009, ApJ, 691, L142

Kormendy, J. & Gebhardt, K. 2001, in American Institute of Physics Conference Series, Vol. 586, 20th Texas Symposium on relativistic astrophysics, ed. J. C. Wheeler & H. Martel, 363–381

Kormendy, J. & Richstone, D. 1995, ARA&A, 33, 581

Koushiappas, S. M., Bullock, J. S., & Dekel, A. 2004, MNRAS, 354, 292

Koushiappas, S. M. & Zentner, A. R. 2006, ApJ, 639, 7

Kraft, R. P., Hardcastle, M. J., Worrall, D. M., & Murray, S. S. 2005, ApJ, 622, 149

Krolik, J. H. 2010, ApJ, 709, 774

Krolik, J. H. & Hawley, J. F. 2002, ApJ, 573, 754

Lacey, C. & Cole, S. 1993, MNRAS, 262, 627

Lada, C. J. 1985, ARA&A, 23, 267

Lang, R. N. & Hughes, S. A. 2006, Phys. Rev. D, 74, 122001

—. 2008, ApJ, 677, 1184

Lauer, T. R., Ajhar, E. A., Byun, Y., Dressler, A., Faber, S. M., Grillmair, C., Kormendy, J., Richstone, D., & Tremaine, S. 1995, AJ, 110, 2622

Lauer, T. R., Faber, S. M., Richstone, D., Gebhardt, K., Tremaine, S., Postman, M., Dressler, A., Aller, M. C., Filippenko, A. V., Green, R., Ho, L. C., Kormendy, J., Magorrian, J., & Pinkney, J. 2007, ApJ, 662, 808

Leahy, J. P. & Williams, A. G. 1984, MNRAS, 210, 929

Lehto, H. J. & Valtonen, M. J. 1996, ApJ, 460, 207

Leighly, K. M., Halpern, J. P., Jenkins, E. B., Grupe, D., Choi, J., & Prescott, K. B. 2007, ApJ, 663, 103

Levin, Y. 2007, MNRAS, 374, 515

Lightman, A. P. & Eardley, D. M. 1974, ApJ, 187, L1+

Lin, D. N. C. & Pringle, J. E. 1987, MNRAS, 225, 607

Lin, L., Cooper, M. C., Jian, H., Koo, D. C., Patton, D. R., Yan, R., Willmer, C. N. A., Coil, A. L., Chiueh, T., Croton, D. J., Gerke, B. F., Lotz, J., Guhathakurta, P., & Newman, J. A. 2010, ApJ, 718, 1158

Lin, L., Patton, D. R., Koo, D. C., Casteels, K., Conselice, C. J., Faber, S. M., Lotz, J., Willmer, C. N. A., Hsieh, B. C., Chiueh, T., Newman, J. A., Novak, G. S., Weiner, B. J., & Cooper, M. C. 2008, ApJ, 681, 232

Lippai, Z., Frei, Z., & Haiman, Z. 2008, ApJ, 676, L5

—. 2009, ApJ, 701, 360

Liu, B. F., Mineshige, S., & Ohsuga, K. 2003a, ApJ, 587, 571

Liu, F. K. 2004, MNRAS, 347, 1357

Liu, F. K., Wu, X., & Cao, S. L. 2003b, MNRAS, 340, 411

Liu, X., Greene, J. E., Shen, Y., & Strauss, M. A. 2010, ApJ, 715, L30

Liu, X., Shen, Y., & Strauss, M. A. 2011a, arXiv e-prints, 1104.0951

—. 2011b, arXiv e-prints, 1104.3391

Liu, X., Shen, Y., Strauss, M. A., & Hao, L. 2011c, arXiv e-prints, 1104.0950

Lobanov, A. P. & Roland, J. 2005, A&A, 431, 831

Lodato, G. & Natarajan, P. 2006, MNRAS, 371, 1813

Lodato, G., Nayakshin, S., King, A. R., & Pringle, J. E. 2009, MNRAS, 398, 1392

Lubow, S. H., Seibert, M., & Artymowicz, P. 1999, ApJ, 526, 1001

Lüst, R. 1952, Z.Naturforsch, 7a, 87

Lynden-Bell, D. 1969, Nature, 223, 690

—. 1978, Phys. Scr, 17, 185

Lynden-Bell, D. & Pringle, J. E. 1974, MNRAS, 168, 603

MacFadyen, A. I. & Milosavljević, M. 2008, ApJ, 672, 83

Machacek, M. E., Bryan, G. L., & Abel, T. 2001, ApJ, 548, 509

Madau, P. & Quataert, E. 2004, ApJ, 606, L17

Magain, P., Letawe, G., Courbin, F., Jablonka, P., Jahnke, K., Meylan, G., & Wisotzki, L. 2005, Nature, 437, 381

Maggiore, M. 2000, Phys. Rep., 331, 283

Magorrian, J., Tremaine, S., Richstone, D., Bender, R., Bower, G., Dressler, A., Faber, S. M., Gebhardt, K., Green, R., Grillmair, C., Kormendy, J., & Lauer, T. 1998, AJ, 115, 2285

Makino, J. 1997, ApJ, 478, 58

Makino, J. & Ebisuzaki, T. 1996, ApJ, 465, 527

Makino, J. & Funato, Y. 2004, ApJ, 602, 93

Manchester, R. N. 2008, in American Institute of Physics Conference Series, Vol. 983, 40 Years of Pulsars: Millisecond Pulsars, Magnetars and More, ed. C. Bassa, Z. Wang, A. Cumming, & V. M. Kaspi, 584–592

Marconi, A., Risaliti, G., Gilli, R., Hunt, L. K., Maiolino, R., & Salvati, M. 2004, MNRAS, 351, 169

Martini, P. 2004, in Coevolution of Black Holes and Galaxies, ed. L. C. Ho, 169–+

Mayer, L., Kazantzidis, S., Madau, P., Colpi, M., Quinn, T., & Wadsley, J. 2007, Science, 316, 1874

McClintock, J. E., Narayan, R., Davis, S. W., Gou, L., Kulkarni, A., Orosz, J. A., Penna, R. F., Remillard, R. A., & Steiner, J. F. 2011, arXiv e-prints, 1101.0811; to appear in Classical and Quantum Gravity; Special Volume for GR19

McWilliams, S. T., Thorpe, J. I., Baker, J. G., & Kelly, B. J. 2010, Phys. Rev. D, 81, 064014

Megevand, M., Anderson, M., Frank, J., Hirschmann, E. W., Lehner, L., Liebling, S. L., Motl, P. M., & Neilsen, D. 2009, Phys. Rev. D, 80, 024012

Menou, K. 2003, Classical and Quantum Gravity, 20, 37

Menou, K., Esin, A. A., Narayan, R., Garcia, M. R., Lasota, J., & McClintock, J. E. 1999, ApJ, 520, 276

Menou, K. & Haiman, Z. 2004, ApJ, 615, 130

Menou, K., Haiman, Z., & Narayanan, V. K. 2001, ApJ, 558, 535

Merritt, D. 2002, ApJ, 568, 998

Merritt, D. & Ekers, R. D. 2002, Science, 297, 1310

Merritt, D. & Milosavljević, M. 2005, Liv. Rev. Rel., 8, 8

Merritt, D., Milosavljević, M., Favata, M., Hughes, S. A., & Holz, D. E. 2004, ApJ, 607, L9

Merritt, D. & Poon, M. Y. 2004, ApJ, 606, 788

Merritt, D., Storchi-Bergmann, T., Robinson, A., Batcheldor, D., Axon, D., & Cid Fernandes, R. 2006, MNRAS, 367, 1746

Mesinger, A., Bryan, G. L., & Haiman, Z. 2006, ApJ, 648, 835

Metzger, B. D., Piro, A. L., & Quataert, E. 2008, MNRAS, 390, 781

Micic, M., Holley-Bockelmann, K., Sigurdsson, S., & Abel, T. 2007, MNRAS, 380, 1533

Mihos, J. C. & Hernquist, L. 1996, ApJ, 464, 641

Miller, J. M. 2007, ARA&A, 45, 441

Milosavljević, M. & Merritt, D. 2001, ApJ, 563, 34

—. 2003, ApJ, 596, 860

Milosavljević, M., Merritt, D., Rest, A., & van den Bosch, F. C. 2002, MNRAS, 331, L51

Milosavljević, M. & Phinney, E. S. 2005, ApJ, 622, L93

Mineshige, S., Kawaguchi, T., Takeuchi, M., & Hayashida, K. 2000, PASJ, 52, 499

Misner, C. W., Thorne, K. S., & Wheeler, J. A. 1973, Gravitation (W. H. Freeman and Company)

Mösta, P., Palenzuela, C., Rezzolla, L., Lehner, L., Yoshida, S., & Pollney, D. 2010, Phys. Rev. D, 81, 064017

Murray, N., Quataert, E., & Thompson, T. A. 2005, ApJ, 618, 569

Mushotzky, R. F., Done, C., & Pounds, K. A. 1993, ARA&A, 31, 717

Narayan, R. 2005, New Journal of Physics, 7, 199

Narayan, R. & Quataert, E. 2005, Science, 307, 77

Narayan, R. & Raymond, J. 1999, ApJ, 515, L69

Narayan, R. & Yi, I. 1994, ApJ, 428, L13

—. 1995a, ApJ, 444, 231

—. 1995b, ApJ, 452, 710

Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, ApJ, 490, 493

Novikov, I. D. & Thorne, K. S. 1973, in Black Holes (Les Astres Occlus), 343–450

Ochi, Y., Sugimoto, K., & Hanawa, T. 2005, ApJ, 623, 922

Ogilvie, G. I. 2005, Unpublished lecture notes on "Accretion Discs", http://www.damtp.cam.ac.uk/user/gio10/lecture5.pdf

Oh, S. P. & Haiman, Z. 2002, ApJ, 569, 558

Omukai, K., Schneider, R., & Haiman, Z. 2008, ApJ, 686, 801

O'Neill, S. M., Miller, M. C., Bogdanović, T., Reynolds, C. S., & Schnittman, J. D. 2009, ApJ, 700, 859

O'Shea, B. W. & Norman, M. L. 2007, ApJ, 654, 66

Ostriker, E. C. 1999, ApJ, 513, 252

Ouyed, R., Pudritz, R. E., & Stone, J. M. 1997, Nature, 385, 409

Paczynski, B. 1978, Acta Astron., 28, 91

Palenzuela, C., Anderson, M., Lehner, L., Liebling, S. L., & Neilsen, D. 2009, Phys. Rev. Lett., 103, 081101

Palenzuela, C., Lehner, L., & Liebling, S. L. 2010a, Science, 329, 927

Palenzuela, C., Lehner, L., & Yoshida, S. 2010b, Phys. Rev. D, 81, 084007

Percival, W. J., Miller, L., McLure, R. J., & Dunlop, J. S. 2001, MNRAS, 322, 843

Peres, A. 1962, Physical Review, 128, 2471

Peters, P. C. 1964, Physical Review, 136, 1224

Peters, P. C. & Mathews, J. 1963, Physical Review, 131, 435

Peterson, B. M. 1993, PASP, 105, 247

Peterson, B. M., Ferrarese, L., Gilbert, K. M., Kaspi, S., Malkan, M. A., Maoz, D., Merritt, D., Netzer, H., Onken, C. A., Pogge, R. W., Vestergaard, M., & Wandel, A. 2004, ApJ, 613, 682

Peterson, B. M. & Horne, K. 2004, Astronomische Nachrichten, 325, 248

Peterson, B. M., Korista, K. T., & Cota, S. A. 1987, ApJ, 312, L1

Phinney, E. S. 2009, in Decadal Survey White Paper, Vol. 2010, No. 235

Piran, T. 1978, ApJ, 221, 652

Press, W. H. & Schechter, P. 1974, ApJ, 187, 425

Pretorius, F. 2005, Phys. Rev. Lett., 95, 121101

Pringle, J. E. 1976, MNRAS, 177, 65

—. 1981, ARA&A, 19, 137

—. 1991, MNRAS, 248, 754

Proga, D. & Begelman, M. C. 2003, ApJ, 592, 767

Ptak, A., Terashima, Y., Ho, L. C., & Quataert, E. 2004, ApJ, 606, 173

Pursimo, T., Takalo, L. O., Sillanpää, A., Kidger, M., Lehto, H. J., Heidt, J., Charles, P. A., Aller, H., Aller, M., Beckmann, V., Benítez, E., Bock, H., Boltwood, P., Borgeest, U., de Diego, J. A., De Francesco, G., Dietrich, M., Dultzin-Hacyan, D., Efimov, Y., Fiorucci, M., Ghisellini, G., González-Pérez, N., Hanski, M., Heinämäki, P., Honeycutt, R. K., Hughes, P., Karlamaa, K., Katajainen, S., Knee, L. B. G., Kurtanidze, O. M., Kümmel, M., Kühl, D., Lainela, M., Lanteri, L., Linde, J. V., Lähteenmäki, A., Maesano, M., Mahoney, T., Marchenko, S., Marscher, A., Massaro, E., Montagni, F., Nesci, R., Nikolashvili, M., Nilsson, K., Nurmi, P., Pietilä, H., Poyner, G., Raiteri, C. M., Rekola, R., Richter, G. M., Riehokainen, A., Robertson, J. W., Rodríguez-Espinoza, J., Sadun, A., Shakhovskoy, N., Schramm, K. J., Schramm, T., Sobrito, G., Teerikorpi, P., Teräsranta, H., Tornikoski, M., Tosti, G., Turner, G. W., Valtaoja, E., Valtonen, M., Villata, M., Wagner, S. J., Webb, J., Weneit, W., & Wiren, S. 2000, A&AS, 146, 141

Quataert, E. & Narayan, R. 1999, ApJ, 520, 298

Quinlan, G. D. 1996, New A, 1, 35

Quinlan, G. D. & Hernquist, L. 1997, New A, 2, 533

Raiteri, C. M., Villata, M., Aller, H. D., Aller, M. F., Heidt, J., Kurtanidze, O. M., Lanteri, L., Maesano, M., Massaro, E., Montagni, F., Nesci, R., Nilsson, K., Nikolashvili, M. G., Nurmi, P., Ostorero, L., Pursimo, T., Rekola, R., Sillanpää, A., Takalo, L. O., Teräsranta, H., Tosti, G., Balonek, T. J., Feldt, M., Heines, A., Heisler, C., Hu, J., Kidger, M., Mattox,

J. R., McGrath, E. J., Pati, A., Robb, R., Sadun, A. C., Shastri, P., Wagner, S. J., Wei, J., & Wu, X. 2001, A&A, 377, 396

Ravindranath, S., Ho, L. C., & Filippenko, A. V. 2002, ApJ, 566, 801

Rees, M. J. 1978, Nature, 275, 516

—. 1984, ARA&A, 22, 471

—. 1988, Nature, 333, 523

Rees, M. J., Begelman, M. C., Blandford, R. D., & Phinney, E. S. 1982, Nature, 295, 17

Reisswig, C., Husa, S., Rezzolla, L., Dorband, E. N., Pollney, D., & Seiler, J. 2009, Phys. Rev. D, 80, 124026

Reynolds, C. S. & Nowak, M. A. 2003, Phys. Rep., 377, 389

Richards, G. T., Strauss, M. A., Pindor, B., Haiman, Z., Fan, X., Eisenstein, D., Schneider, D. P., Bahcall, N. A., Brinkmann, J., & Brunner, R. 2004, AJ, 127, 1305

Richstone, D., Ajhar, E. A., Bender, R., Bower, G., Dressler, A., Faber, S. M., Filippenko, A. V., Gebhardt, K., Green, R., Ho, L. C., Kormendy, J., Lauer, T. R., Magorrian, J., & Tremaine, S. 1998, Nature, 395, A14+

Rieger, F. M. & Mannheim, K. 2000, A&A, 359, 948

Robertson, B., Hernquist, L., Cox, T. J., Di Matteo, T., Hopkins, P. F., Martini, P., & Springel, V. 2006, ApJ, 641, 90

Rodriguez, C., Taylor, G. B., Zavala, R. T., Peck, A. B., Pollack, L. K., & Romani, R. W. 2006, ApJ, 646, 49

Roedig, C., Dotti, M., Sesana, A., Cuadra, J., & Colpi, M. 2011, MNRAS, 979

Roos, N. 1981, A&A, 104, 218

Roos, N., Kaastra, J. S., & Hummel, C. A. 1993, ApJ, 409, 130

Rossi, E. M., Lodato, G., Armitage, P. J., Pringle, J. E., & King, A. R. 2010, MNRAS, 401, 2021

Rudak, B. & Paczynski, B. 1981, Acta Astronomica, 31, 13

Rybicki, G. B. & Lightman, A. P. 1986, Radiative Processes in Astrophysics (Wiley-VCH)

Salpeter, E. E. 1964, ApJ, 140, 796

Sánchez-Blázquez, P., Gibson, B. K., Kawata, D., Cardiel, N., & Balcells, M. 2009, MNRAS, 400, 1264

Sanders, D. B., Soifer, B. T., Elias, J. H., Madore, B. F., Matthews, K., Neugebauer, G., & Scoville, N. Z. 1988, ApJ, 325, 74

Sano, T., Inutsuka, S., Turner, N. J., & Stone, J. M. 2004, ApJ, 605, 321

Saripalli, L. & Subrahmanyan, R. 2009, ApJ, 695, 156

Sazhin, M. V. 1978, Soviet Ast., 22, 36

Schawinski, K., Treister, E., Urry, C. M., Cardamone, C. N., Simmons, B., & Yi, S. K. 2011, ApJ, 727, L31+

Schechter, P. 1976, ApJ, 203, 297

Schmidt, M. 1963, Nature, 197, 1040

Schnittman, J. D. 2011, Classical and Quantum Gravity, 28, 094021

Schnittman, J. D. & Buonanno, A. 2007, ApJ, 662, L63

Schnittman, J. D., Buonanno, A., van Meter, J. R., Baker, J. G., Boggs, W. D., Centrella, J., Kelly, B. J., & McWilliams, S. T. 2008, Phys. Rev. D, 77, 044031

Schnittman, J. D. & Krolik, J. H. 2008, ApJ, 684, 835

Schoenmakers, A. P., de Bruyn, A. G., Röttgering, H. J. A., van der Laan, H., & Kaiser, C. R. 2000, MNRAS, 315, 371

Schutz, B. F. 1986, Nature, 323, 310

—. 1999, Classical and Quantum Gravity, 16, A131

—. 2009, Class. Quantum Grav., 26, 094020

Sesana, A. 2010, ApJ, 719, 851

Sesana, A., Haardt, F., & Madau, P. 2007a, ApJ, 660, 546

Sesana, A., Haardt, F., Madau, P., & Volonteri, M. 2004, ApJ, 611, 623

—. 2005, ApJ, 623, 23

Sesana, A., Roediig, C., Reynolds, M. T., & Dotti, M. 2011, arXiv e-prints

Sesana, A. & Vecchio, A. 2010, Phys. Rev. D, 81, 104008

Sesana, A., Vecchio, A., & Colacino, C. N. 2008, MNRAS, 390, 192

Sesana, A., Vecchio, A., & Volonteri, M. 2009, MNRAS, 394, 2255

Sesana, A., Volonteri, M., & Haardt, F. 2007b, MNRAS, 377, 1711

Shakura, N. I. & Sunyaev, R. A. 1973, A&A, 24, 337

—. 1976, MNRAS, 175, 613

Shang, C. & Haiman, Z. 2011, MNRAS, 411, 9

Shankar, F., Bernardi, M., & Haiman, Z. 2009a, ApJ, 694, 867

Shankar, F., Crocce, M., Miralda-Escudé, J., Fosalba, P., & Weinberg, D. H. 2010, ApJ, 718, 231

Shankar, F., Weinberg, D. H., & Miralda-Escudé, J. 2009b, ApJ, 690, 20

Shapiro, P. R., Iliev, I. T., & Raga, A. C. 1999, MNRAS, 307, 203

Shapiro, S. L. 2005, ApJ, 620, 59

—. 2010, Phys. Rev. D, 81, 024019

Shapiro, S. L., Lightman, A. P., & Eardley, D. M. 1976, ApJ, 204, 187

Shapiro, S. L. & Teukolsky, S. A. 1986, Black Holes, White Dwarfs and Neutron Stars: The Physics of Compact Objects (John Wiley & Sons, New York, NY)

Shemmer, O., Brandt, W. N., Anderson, S. F., Diamond-Stanic, A. M., Fan, X., Richards, G. T., Schneider, D. P., & Strauss, M. A. 2009, ApJ, 696, 580

Shen, Y. & Loeb, A. 2010, ApJ, 725, 249

Shen, Y., Strauss, M. A., Oguri, M., Hennawi, J. F., Fan, X., Richards, G. T., Hall, P. B., Gunn, J. E., Schneider, D. P., Szalay, A. S., Thakar, A. R., Vanden Berk, D. E., Anderson, S. F., Bahcall, N. A., Connolly, A. J., & Knapp, G. R. 2007, AJ, 133, 2222

Shen, Z., Lo, K. Y., Liang, M., Ho, P. T. P., & Zhao, J. 2005, Nature, 438, 62

Shields, G. A. & Bonning, E. W. 2008, ApJ, 682, 758

Shields, G. A., Bonning, E. W., & Salviander, S. 2009a, ApJ, 696, 1367

Shields, G. A., Rosario, D. J., Smith, K. L., Bonning, E. W., Salviander, S., Kalirai, J. S., Strickler, R., Ramirez-Ruiz, E., Dutton, A. A., Treu, T., & Marshall, P. J. 2009b, ApJ, 707, 936

Shlosman, I. & Begelman, M. C. 1987, Nature, 329, 810

—. 1989, ApJ, 341, 685

Siemens, X., Mandic, V., & Creighton, J. 2007, Physical Review Letters, 98, 111101

Sillanpaa, A., Haarala, S., Valtonen, M. J., Sundelius, B., & Byrd, G. G. 1988, ApJ, 325, 628

Sirko, E. & Goodman, J. 2003, MNRAS, 341, 501

Soltan, A. 1982, MNRAS, 200, 115

Somerville, R. S. & Kolatt, T. S. 1999, MNRAS, 305, 1

Spaans, M. & Silk, J. 2006, ApJ, 652, 902

Springel, V., White, S. D. M., Jenkins, A., Frenk, C. S., Yoshida, N., Gao, L., Navarro, J., Thacker, R., Croton, D., Helly, J., Peacock, J. A., Cole, S., Thomas, P., Couchman, H., Evrard, A., Colberg, J., & Pearce, F. 2005, Nature, 435, 629

Stone, N. & Loeb, A. 2011, MNRAS, 412, 75

Stubbs, C. W. 2008, Classical and Quantum Gravity, 25, 184033

Sudou, H., Iguchi, S., Murata, Y., & Taniguchi, Y. 2003, Science, 300, 1263

Takeuchi, S., Mineshige, S., & Ohsuga, K. 2009, PASJ, 61, 783

Tanaka, T. 2011, MNRAS, 410, 1007

Tanaka, T. & Haiman, Z. 2009, ApJ, 696, 1798

Tanaka, T., Haiman, Z., & Menou, K. 2010, AJ, 140, 642

Tanaka, T. & Menou, K. 2006, ApJ, 649, 345

—. 2010, ApJ, 714, 404

Tanaka, T., Menou, K., & Haiman, Z. in prep., to be submitted to MNRAS

Tang, S. & Grindlay, J. 2009, ApJ, 704, 1189

Taniguchi, Y. & Wada, K. 1996, ApJ, 469, 581

Taylor, J. H., Fowler, L. A., & McCulloch, P. M. 1979, Nature, 277, 437

Taylor, J. H. & Weisberg, J. M. 1989, ApJ, 345, 434

Tegmark, M., Silk, J., Rees, M. J., Blanchard, A., Abel, T., & Palla, F. 1997, ApJ, 474, 1

Thompson, T. A. 2008, ApJ, 684, 212

Thompson, T. A., Quataert, E., & Murray, N. 2005, ApJ, 630, 167

Thorne, K. S. 1980, Reviews of Modern Physics, 52, 285

Tichy, W. & Marronetti, P. 2008, Phys. Rev. D, 78, 081501

Titchmarsh, E. C. 1923, Proc. London Math. Soc. (2), 22, 15

Tremaine, S., Gebhardt, K., Bender, R., Bower, G., Dressler, A., Faber, S. M., Filippenko, A. V., Green, R., Grillmair, C., Ho, L. C., Kormendy, J., Lauer, T. R., Magorrian, J., Pinkney, J., & Richstone, D. 2002, ApJ, 574, 740

Tundo, E., Bernardi, M., Hyde, J. B., Sheth, R. K., & Pizzella, A. 2007, ApJ, 663, 53

Turner, E. L. 1991, AJ, 101, 5

Turner, N. J. 2004, ApJ, 605, L45

Uchida, Y. & Shibata, K. 1985, PASJ, 37, 515

Ulrich, M., Maraschi, L., & Urry, C. M. 1997, ARA&A, 35, 445

Valtonen, M. J. 1996, MNRAS, 278, 186

Valtonen, M. J., Lehto, H. J., Nilsson, K., Heidt, J., Takalo, L. O., Sillanpää, A., Villforth, C., Kidger, M., Poyner, G., Pursimo, T., Zola, S., Wu, J., Zhou, X., Sadakane, K., Drozdz, M., Koziel, D., Marchev, D., Ogloza, W., Porowski, C., Siwak, M., Stachowski, G., Winiarski, M., Hentunen, V., Nissinen, M., Liakos, A., & Dogru, S. 2008, Nature, 452, 851

Valtonen, M. J., Lehto, H. J., Sillanpää, A., Nilsson, K., Mikkola, S., Hudec, R., Basta, M., Teräsranta, H., Haque, S., & Rampadarath, H. 2006, ApJ, 646, 36

Valtonen, M. J., Mikkola, S., Merritt, D., Gopakumar, A., Lehto, H. J., Hyvönen, T., Rampadarath, H., Saunders, R., Basta, M., & Hudec, R. 2010, ApJ, 709, 725

van Dokkum, P. G. 2005, AJ, 130, 2647

Vecchio, A. 2004, Phys. Rev. D, 70, 042001

Volonteri, M., Haardt, F., & Madau, P. 2003a, ApJ, 582, 559

Volonteri, M., Madau, P., & Haardt, F. 2003b, ApJ, 593, 661

Volonteri, M. & Rees, M. J. 2005, ApJ, 633, 624

—. 2006, ApJ, 650, 669

Wandel, A., Peterson, B. M., & Malkan, M. A. 1999, ApJ, 526, 579

Wang, J. & Merritt, D. 2004, ApJ, 600, 149

Weisberg, J. M., Nice, D. J., & Taylor, J. H. 2010, ApJ, 722, 1030

Wen, L. 2003, ApJ, 598, 419

Whalen, D., Abel, T., & Norman, M. L. 2004, ApJ, 610, 14

Will, C. M. 1998, Phys. Rev. D, 57, 2061

—. 2006, Living Reviews in Relativity, 9, 3

Will, C. M. & Yunes, N. 2004, Classical and Quantum Gravity, 21, 4367

Willott, C. J., McLure, R. J., & Jarvis, M. J. 2003, ApJ, 587, L15

Wu, J., Brandt, W. N., Hall, P. B., Gibson, R. R., Richards, G. T., Schneider, D. P., Shemmer, O., Just, D. W., & Schmidt, S. J. 2011, arXiv e-prints, 1104.3861

Wyithe, J. S. B. & Loeb, A. 2003a, ApJ, 590, 691

—. 2003b, ApJ, 595, 614

—. 2009, MNRAS, 395, 1607

Yoo, J. & Miralda-Escudé, J. 2004, ApJ, 614, L25

Yoshida, N., Abel, T., Hernquist, L., & Sugiyama, N. 2003, ApJ, 592, 645

Yoshida, N., Omukai, K., & Hernquist, L. 2008, Science, 321, 669

Yu, Q. 2002, MNRAS, 331, 935

Yu, Q. & Tremaine, S. 2002, MNRAS, 335, 965

Yuan, F., Markoff, S., Falcke, H., & Biermann, P. L. 2002, A&A, 391, 139

Zakamska, N. L., Strauss, M. A., Krolik, J. H., Ridgway, S. E., Schmidt, G. D., Smith, P. S., Heckman, T. M., Schneider, D. P., Hao, L., & Brinkmann, J. 2006, AJ, 132, 1496

Zel'Dovich, Y. B. 1964, Soviet Physics Doklady, 9, 195

Zhang, J., Fakhouri, O., & Ma, C. 2008, MNRAS, 389, 1521

Zhang, X. & Tong, D. 2007, Applied Mathematics and Computation, 193, 116

Zhao, H., Haehnelt, M. G., & Rees, M. J. 2002, New A, 7, 385

Zhou, H., Wang, T., Zhang, X., Dong, X., & Li, C. 2004, ApJ, 604, L33

Zier, C. & Biermann, P. L. 2001, A&A, 377, 23