From seed to supermassive: simulating the origin, evolution and impact of massive black holes

Ricarda S. Beckmann
Balliol College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Trinity 2017

"The evolution of the Universe can be likened to a display of fireworks that has just ended: some few red wisps, ashes, and smoke. Standing on a well-chilled cinder, we see the fading of the suns and try to recall the vanished brilliance of the origin of the worlds."

— Georges Lemaître, 1931

Abstract

First observed as early as redshift $z = 7$ and now thought to be found at the centre of every massive galaxy in the local Universe, the evolution history of supermassive black holes (SMBHs) spans over 13 billion years. In this thesis, the coevolution between SMBHs and their host galaxies is studied using a set of hydrodynamical simulations to isolate different components of the interaction between black holes and cosmic gas. The simulations range from black hole accretion in an idealised context to the impact of feedback in the cosmological simulations of the HORIZON suite.

The origin of SMBHs during the first billion years of the Universe is a highly non-linear problem, where small-scale behaviour influences large-scale behaviour and vice versa. Gas fuelling a black hole flows from the cosmic web, through its host galaxy and into the black hole's gravitational potential, before eventually reaching its event horizon. Even discounting the complex physical processes at play, resolving the 19 orders of magnitude in spatial scale involved is beyond the capabilities of current simulations. Some of the length scales therefore have to be covered by sub-grid algorithms which need to be able to handle a wide range of environments.

Idealised accretion simulations presented in this thesis show that the Bondi-Hoyle-Lyttleton (BHL) accretion algorithm is sufficiently versatile. It automatically determines the accretion rate onto the black hole by the mass flux into its accretion region when the black hole's gravitational potential becomes resolved. The accretion rate onto the black hole therefore naturally converges to the correct solution once the size of the accretion region approaches the physical size of the black hole. A drag force algorithm that compensates for unresolved dynamical friction, on the other hand, produces a force on the black hole that can unphysically accelerate it relative to the bulk flow of the gas. It needs to be switched off when gas properties are measured within the black hole's gravitational potential.

A study of black hole accretion within an isolated cooling halo confirms that the accretion algorithm is able to handle the flow configurations encountered within an evolving galaxy. To ensure gas is always accreted within the black hole's gravitational potential, a refinement algorithm called "zoom-within-zoom" is introduced in this thesis. It allows the black hole environment to be resolved by orders of magnitude above that of its host galaxy. A low mass seed black hole with a strong drag force early on takes advantage of this extra information during the black hole's early evolution. In the longer term, resolving gas clouds in the black hole vicinity to sub-pc scales has a lasting impact on both the mass evolution and duty cycle of massive black holes.

Sub-pc size clumps also play a deciding role in the first 200 Myr of evolution of a SMBH progenitor in a full cosmological context: 90% of its mass is gained through interactions with dense clumps, which fuel super-Eddington accretion bursts. Once the gas within the host galaxy settles into a rotationally supported disc, star formation and black hole accretion slow down. As both primarily occur within the central 30 pc of the compact host galaxy, star formation in proto-galaxies has a major impact on black hole accretion even in the absence of feedback.

At low redshift, on the other hand, feedback becomes the crucial link between a SMBH and its host galaxy. A comparison of two simulations from the HORIZON suite, run with and without active galactic nuclei (AGN) feedback respectively, shows that AGN feedback is able to prevent as much as 90% of the stellar mass from forming in the most massive galaxies. Quenching proceeds via a combination of AGN driven outflows and reduced inflows and evolves with redshift as the $M_{\rm SMBH} - M_*$ relation flattens from $z = 5$ to $z = 0$.

In conclusion, neither the evolution of galaxies nor that of black holes can be understood without the context of the other. At high redshift, the competition between star formation and black hole accretion inside the compact host galaxy intrinsically links the origin of SMBHs to the early evolution of galaxies. At low redshift, AGN feedback modulates the gas supply of the host galaxy, which has a lasting impact on star formation. The coevolution of black holes and galaxies therefore spans their entire history.

Statement of Originality

I hereby declare that no part of this thesis has been accepted, or is currently being submitted, for any degree or diploma or certificate or any other qualification at the University of Oxford or elsewhere. Except where otherwise stated, the work in this thesis is entirely my own, and has been carried out under the supervision of Professor Julien Devriendt and Professor Adrianne Slyz.

The work presented in this thesis is part of the Horizon-UK project, which uses the DiRAC Complexity system, operated by the University of Leicester IT Services as part of the STFC DiRAC HPC Facility (www.dirac.ac.uk). This equipment is funded by BIS National E-Infrastructure capital grant ST/K000373/1 and STFC DiRAC Operations grant ST/K0003259/1.

Chapter 3 was submitted on the 14.7.2017 to The Monthly Notices of the Royal Astronomical Society: Beckmann, R. S., Slyz, A. & Devriendt, J., **'Bondi or not Bondi: the impact of resolution on accretion and drag force modelling for Supermassive Black Holes'**

Chapter 4 will be submitted to The Monthly Notices of the Royal Astronomical Society: Beckmann, R. S., Devriendt, J. & Slyz, A., **'A zoom-in on the art of planting supermassive black holes in galaxies'**

Chapter 6 was accepted for publication on the 18.7.2017 by The Monthly Notices of the Royal Astronomical Society: Beckmann, R. S., Devriendt, J., Slyz, A., Peirani, S., Richardson, M. L. A., Dubois, Y., Pichon, C., Chisari, N. E., Kaviraj, S., Laigle, C. & Volonteri, M., **'Cosmic evolution of stellar quenching by AGN feedback: clues from the Horizon-AGN simulation'**

Ricarda Beckmann, *July 2017*

Acknowledgements

First and foremost I thank my supervisors, Julien Devriendt and Adrianne Slyz, for all their input, support, time and patience over the last four years. It has been a privilege to be your student and to work with you. Thank you, I could not have done it without you.

I would like to thank all my collaborators, Elisa Chisari, Yohan Dubois, Marta Volonteri, Rebekka Bieri, Christophe Pichon, Sugata Kaviraj, Sebastien Peirani and Clotilde Laigle, for being so great to work with. I would also liked to thank Chris Lintott, Pedro Ferreira and Jo Dunkley for support along the way. A special thanks goes to Mark Richardson, for being both a collaborator and a friend over the last few years. I have learned a huge amount from you, both professionally and outside astrophysics (I'm sure the obscure popular culture references will come in handy) and having you here has been great. Thank you.

There have been a great many people who have made my time here in Oxford some of the best years of my life, both in Balliol College and in the department. Danielle, Ed, Shravan, Holly, Nina, Ben, Andre, Vera, Laura, Sarah, Helen, Rahul, and Ellen, thank you. Becky, a special thanks goes to you. Thank you for being such a great friend, putting up with me as a housemate and reading my entire thesis. My time here would not have been the same without you. Simon, thank you for being part of my life, and in particular for your support over the last few months. Having you with me has made everything a little easier. I would also like to thank midnight bowls of Cheerios, the society coffee shop on St Michael's street, and green tea in general.

And finally, I would like to thank my parents, Barbara and Christoph. Your unwavering support and belief in me are the foundations on which this DPhil was built. Thank you for everything.

Contents

List of Figures

Glossary

c	speed of light
χ	Eddington ratio
c_s	speed of sound
Δx	cell spacing
ϵ_{f}	feedback efficiency parameter in the AGN model
ϵ_{r}	radiative efficiency of an AGN
ϵ_*	star formation efficiency
G	gravitational constant
H_0	Hubble constant
k_B	Boltzman constant
L	luminosity
l_{J}	Jeans length
M	mass
$\mathcal{M}$	Mach number
$\dot{M}$	accretion rate
M_{J}	Jeans mass
m_p	proton mass

M_{seed} formation mass of a black hole

M_* stellar mass

$\mathrm{M}_\odot$ solar mass

Ω_{b} cosmic baryon density parameter

Ω_{DM} cosmic dark matter density parameter

Ω_Λ cosmic dark energy density parameter

Ω_{m} cosmic matter density parameter

P pressure

ρ density

ϕ_{g} gravitational potential

R^{A} accretion radius

R^{B} Bondi radius

R^{BHL} Bondi-Hoyle-Lyttleton Radius

R_{Sch} Schwarzschild radius

σ velocity dispersion of the galaxy

σ_8 amplitude of mass fluctuations in the cosmic microwave background

σ_T Thomson cross section

T temperature

t_{ff} free fall time

V volume

v velocity

z redshift

Acronyms

ΛCDM Lambda cold dark matter

ADAF advection-dominated accretion flow

AGN active galactic nuclei

AMR adaptive mesh refinement

BH black hole

BHL Bondi-Hoyle-Lyttleton

CGM circumgalactic medium

CIC cloud-in-cell

CMB cosmic microwave background

DM dark matter

Edd Eddington

FLA flux limited accretion

GSMF galaxy stellar mass function

H-AGN Horizon-AGN

H-noAGN Horizon-noAGN

HMF halo mass function

IGM intergalactic medium

IMBH	intermediate mass black hole
IMF	initial mass function
ISM	interstellar medium
JWST	James-Webb Space Telescope
MPI	message passing interface
NFW profile	Navarro-Frenk-White profile
PHEW	parallel hierarchical watershed
Pop II	population II star
Pop III	population III star
QSO	quasi-stellar-object
SDSS	Sloan Digital Sky Survey
SFR	star formation rate
SgrA*	Sagittarius A*
SMBH	supermassive black hole
SN	supernova
ULIRG	ultra-luminous infrared galaxy
ULX	ultra-luminous X-ray source

Chapter 1

Introduction

A black hole is a massive object that has collapsed to a singularity of infinite density (Schwarzschild, 1916; Kerr, 1963). They arise as a solution to the Einstein field equations in general relativity (Einstein, 1916) and can be fully described by only three variables: mass, spin and electric charge (Misner et al., 1973). Each black hole is enveloped by an event horizon, within which gravity is so strong that the escape velocity exceeds the speed of light, thereby defining a compact region in spacetime from which nothing can escape. A black hole forms when a sufficiently massive object, such as a star, collapses under its own gravity (Oppenheimer & Snyder, 1939) and gains mass by material entering its event horizon, or by merging with other compact objects such as other black holes or neutron stars.

First treated as merely a mathematical curiosity, supermassive black holes (SMBHs) are now expected to be found in the centre of every large galaxy (Kormendy & Ho, 2013; Heckman & Best, 2014). When active, their accretion energy can far exceed the binding energy of their host galaxy (Fabian, 2012), despite containing much less than one percent of the baryonic mass of the galaxy and being outnumbered by stars by a factor of 10^{12}. If only a small percentage of that energy is able to effectively couple to the gas of the galaxy, it could explain the correlations between the properties of SMBHs and those of their host (Ferrarese & Merritt, 2000; Gebhardt et al., 2000) and offer a possible solution to the overcooling problem (Binney & Tabor, 1995) by preventing the overproduction of galaxies at the massive end of the galaxy stellar mass function (GSMF) (White & Frenk, 1991).

Black holes themselves are fairly simple objects that are well understood in isolation. Their origin and evolution in the context of a galaxy, however, is a complex problem with many open questions, particularly if the host galaxy is evolving on comparable timescales. Where do these black holes come from, and how did they grow so massive, so early? How is their growth influenced by their host galaxy and how do SMBHs in turn influence their host?

With the discovery of the first quasars at redshift $z = 6$ (Fan et al., 2004, 2006), powered by black holes with masses already in excess of 10^9 $\mathrm{M}_\odot$, the origin of SMBHs is firmly a question for the first billion years of the Universe. This is an epoch of the Universe for which we have few observations. Even with the advent of the James-Webb Space Telescope (JWST), our ability to study the progenitors of SMBHs, which are expected to form as early as $z = 20$ with masses as low as a few hundred solar masses, will remain limited. The work presented here tackles these questions using a series of hydrodynamical simulations, from idealised, isolated accretion flows onto black holes to the full sample of 100,000 galaxies and their SMBHs found in the suite of HORIZON simulations (Dubois et al., 2014b).

Hydrodynamical simulations, combined with sub-grid models to include a range of unresolved physics, have become a powerful tool to study the origin and evolution of galaxies and their components, including SMBHs. The rapidly increasing computational power of modern supercomputers has allowed larger simulations to be run, but the computational cost is still a limiting factor. A trade-off has to made between the number of galaxies studied, the number of years covered and the smallest length-scales resolved. On the one hand, there are large-scale cosmological simulations that provide large samples of $> 100,000$ galaxies (Dubois et al., 2014b; Genel et al., 2014; Schaye et al., 2014), but at a comparatively low resolution of 1 kpc. On the other hand, isolated simulations of black holes provide detailed descriptions of their accretion discs (Jiao et al., 2015; Sądowski & Gaspari, 2017) but fail to capture the complex changing environment encountered by early SMBHs. A successful numerical experiment has to be carefully designed to include all relevant processes and length scales, while minimising numerical costs.

The origin of SMBHs is a particularly challenging problem, because the host galaxy forms and evolves on similar timescales to the black hole itself. It is therefore necessary to capture the cosmological context of the galaxy, from the Mpc scales of cosmic filaments and voids through the kpc and pc scales of the emerging galaxy, down to the sub-pc scales of the black hole environment. Cosmological zoom simulations, where a high resolution region containing all progenitor halos of a particular halo is embedded in a large, low-resolution box, are a type of simulation able to cover much of the requisite range of scales. Modern zoom simulations reach resolutions of 1 pc, which allows for plenty of detail within the galaxy itself but still far exceeds the event horizon of a SMBH. For example a seed black hole with a mass of 10^3 $\mathrm{M}_\odot$ has a

Schwarzschild radius of only $R_{\rm Sch} = \frac{2GM_{\rm BH}}{c^2} \approx 10^{-10}$ pc, where G is the gravitational constant and c is the speed of light. No single simulation can currently, or in the forseeable future, cover 19 orders of magnitude to span scales from 10^{-10} to 10^9 pc .

The largest scale on which the black hole can be expected to influence its local environment is given by the extent of its gravitational potential, quantified by its accretion radius (see Section 3.5.2 for details). For the same seed black hole mass of $M_{\rm BH} = 10^3$ M$_\odot$, the accretion radius at rest, $\frac{GM_{\rm BH}}{c_s^2} = 0.04$ pc, assuming a sound speed of $c_s = 10$ km/s. The challenge, then, is to extend the current range of length scales covered by zoom simulations by at least two orders of magnitude to study the origin and evolution of SMBHs in a cosmological context. While resolving the gravitational length scales of the black hole will far improve our understanding of the black hole environment, much of the black hole physics itself will remain unresolved, and require so-called sub-grid algorithms to be included in the simulations. Sub-grid algorithms combine information available in the simulation with an analytic or empirical model to include the impact of physical processes that occur on scales below the resolution limit of the simulation. To be successful, they need to reproduce all relevant effects on a given length scale, using information from that length scale. Extending the range of scales resolved in a simulation means that existing algorithms will have to be carefully examined to asses their suitability for the new regime. Work on this subject is presented in Chapters 4, 5 and 6.

Studying the impact of SMBH feedback on galaxy evolution requires a very different approach. As the scaling relations between SMBHs and galaxy properties are a statistical measure, a large population of objects is required to study their origin. However, there are many processes shaping the evolution of galaxies (See Benson, 2010, for a thorough review), so isolating the impact of a particular contribution, such as the feedback from active SMBHs, can be challenging. The work on this subject presented in this thesis uses the HORIZON suite of simulations (Dubois et al., 2014b), a suite of large scale cosmological simulations following the evolution of more than 100,000 galaxies. To isolate the impact of active galactic nuclei (AGN), HORIZON offers two simulations. The two simulations were run from identical initial conditions, and rely on identical numerical recipes, except that Horizon-AGN (H-AGN) includes a prescription for SMBH and their feedback, whereas Horizon-noAGN (H-noAGN) does not. This means one can identify the same galaxy in both simulations and directly compare its evolution with and without feedback from $z = 5$ to $z = 0$. The work on this subject is presented in Chapter 7.

This thesis is structured as follows: Chapter 2 gives a brief overview of galaxy evolution for context and summarises our current understanding of SMBHs. Chapter 3 introduces RAMSES, the code used for the hydrodynamical simulations presented in this thesis, and includes an overview of numerical methods. Chapter 4 presents a detailed study of the impact of resolution on two current sub-grid algorithms for black holes, Bondi-Hoyle-Lyttleton (BHL) accretion and an analytic drag force. Chapter 5 presents the zoom-within-zoom algorithm used to extend the range of length scales in cosmological simulations, as well as a study on the impact of this extra resolution on the growth of a SMBH in an isolated galaxy. Chapter 6 investigates the origin and growth of a SMBH in a full cosmological context. The penultimate chapter, Chapter 7, is based on the HORIZON simulations and details work on the impact of AGN feedback on star formation. Chapter 8 summarises the work.

Chapter 2

Theory and observation

Found in galaxies as far away as the first quasars at redshift $z = 7$ (Mortlock et al., 2011) and as close as in the centre of our own galaxy (Gillessen et al., 2009), the origin, evolution and impact of supermassive black holes (SMBHs) spans the entire evolution history of the Universe. As one cannot understand the long-term evolution of SMBHs without understanding their environment inside their host galaxies, which are themselves evolving, this chapter includes a brief review of galaxy evolution to provide context for the work on SMBHs, before delving deeper into our current knowledge of SMBHs and their origin.

2.1 Galaxy formation and evolution

2.1.1 Cosmology and dark matter

The energy density of the Universe is shared between a few distinct components, including dark energy, dark matter, radiation, neutrinos and baryonic matter. Dark energy drives the expansion of the Universe, dark matter dominates the large scale structure and interactions with baryonic matter are responsible for our everyday experience. Baryons form stars, gas, galaxies, planets, coffee cups and life. A variety of models have been proposed for dark matter and energy, but the so called Lambda cold dark matter (ΛCDM) model, which uses cold dark matter plus a cosmological constant, is favoured by strong observational constraints (Blumenthal et al., 1984; Bertone et al., 2005) and will be assumed to hold for the rest of this work. Using this model, observations confirm a cosmic dark matter to baryon ratio of $\Omega_{\mathrm{DM}}/\Omega_{\mathrm{b}} = 5$ (Bennett et al., 2013; Ade et al., 2014).

The Universe is thought to have started as a singularity of infinite density and temperature, the Big Bang, and has been expanding ever since. Cooling rapidly, the first second of its evolution saw inflation, a period of accelerating expansion, and the formation of elementary particles. All large scale structure, including galaxies, originates in the Gaussian density fluctuations in the distribution of matter at the end of inflation, which act as the seeds for gravitational collapse. The power spectrum of these fluctuations is still visible in the cosmic microwave background (CMB) (Penzias & Wilson, 1965; Bennett et al., 2013; Ade et al., 2014), the redshifted field of photons freestreaming since photons decoupled from matter when the universe was approximately 379,000 years old. The smallest structures collapse first, following a hierarchical structure formation model, with progressively larger structures forming as the evolution continues. Simulations of this process based on ΛCDM show the emergence of a three-dimensional dark matter structure composed of filaments, sheets and voids: the cosmic web. This process leads to a characteristic mass function for dark matter halos, the halo mass function (HMF) (Press & Schechter, 1974; Shaw et al., 2006). Halos are approximately stable, gravitationally bound dark matter objects that can be characterised by empirical profiles. While the purely gravitational collapse of dark matter is generally well understood, observations indicate an approximately constant dark matter density in the inner parts of halos (called “cores”), while simulations, and empirical models like the Navarro-Frenk-White profile (NFW profile) or the Einasto profile (Einasto, 1965; Navarro et al., 1997; Dutton & Maccio, 2014) indicate a steepening, power-law like behaviour (called “cusps”). This is known as the “cusp-core” problem (de Blok, 2010).

2.1.2 The origin of galaxies

Galaxies are gravitationally bound systems of gas, stars, dust, dark matter and a variety of stellar remnants, including neutron stars, stellar mass black holes and SMBHs. The current population of galaxies shows a large morphological and colour diversity, from blue spirals to red and dead ellipticals. Galaxies also span a large range in mass, from supermassive galaxies reaching a stellar mass of up to $M_* > 10^{12}$ M$_\odot$ at the centre of clusters to ultra-faint dwarfs with stellar masses as low as $M_* \sim 10^9$ M$_\odot$ (Benson, 2010; Naab & Ostriker, 2017).

The first galaxies are thought to have collapsed around redshift $z = 50 - 20$, when the first mini-halos developed a sufficiently deep gravitational potential. On the largest scales, baryonic matter in the form of gas follows the flow of dark matter along the cosmic web and into halos. How it then enters a galaxy is believed to depend on both the galaxy’s mass and on redshift (Dekel & Birnboim, 2006). At high redshift, gaseous filaments form at the intersection of voids and sheets in the intergalactic medium (IGM), and feed as cold gas into proto-galaxies. At lower redshifts, a virial shock develops in massive halos. This heats infalling gas to the potential's virial temperature until that gas cools radiatively, loses pressure support and settles at the bottom of the potential well (Rees & Ostriker, 1977; White & Rees, 1978; Birnboim & Dekel, 2003). Density increases until certain regions exceed the local Jeans length $l_{\rm J}$ and fragment (see Section 3.2), eventually forming the first, metal-free stars, so called population III stars (Pop IIIs) (Bromm et al., 1999; Abel et al., 2000; Abel, 2002). Winds and supernovae (SNe) from these massive and short-lived stars enrich the interstellar medium (ISM), allowing it to cool more efficiently via metal line transitions (Heger et al., 2001; Bromm & Loeb, 2003; Whalen et al., 2010). As the gas continues to fragment and form further generation of stars, the basic processes of galaxy evolution are in place.

2.1.3 Galaxy evolution

Observations show that galaxies today have a characteristic GSMF (see the black symbols in Figure 2.1), but that predictions from the ΛCDM model far overproduce both its massive and the low-mass ends (White & Frenk, 1991; Kauffmann et al., 1993) (see the dashed line in Figure 2.1). Some mechanism is required to supply extra energy and prevent star formation in both low mass and high mass galaxies to produce the observed relation. Current results suggest that stellar feedback prevents star formation in galaxies with $M_* < 10^{11}$ M$_\odot$ (Dekel & Silk, 1986; Heckman et al., 2000; Benson et al., 2003; Hopkins et al., 2014; Keller & Wadsley, 2017) and feedback from active galactic nuclei (AGN) for more massive ones (King, 2003; Murray et al., 2005; Fabian et al., 2006a; Rafferty et al., 2006). Simulations that include both have successfully reproduced the characteristic shape of the GSMF (see coloured lines in Figure 2.1). The hierarchical structure formation model also predicts that the largest objects form last, so the most massive galaxies should have assembled most recently. Observations show that the oldest stars are concentrated in the centres of the most massive galaxies (Hopkins et al., 2009; Carrasco et al., 2010; van Dokkum et al., 2010), suggesting that most of the stars in large galaxies were formed in smaller progenitors. The big challenge of galaxy evolution is to explain the origin of such a diverse population of objects from comparatively simple initial conditions.

There are two main types of galaxy morphologies: disc dominated and elliptical. They have been found to have different formation histories (Kormendy & Bender, 2012). Current models, built on a mixture of simulations and observations, suggest that all galaxies undergo an early collapse phase, during which they form a compact

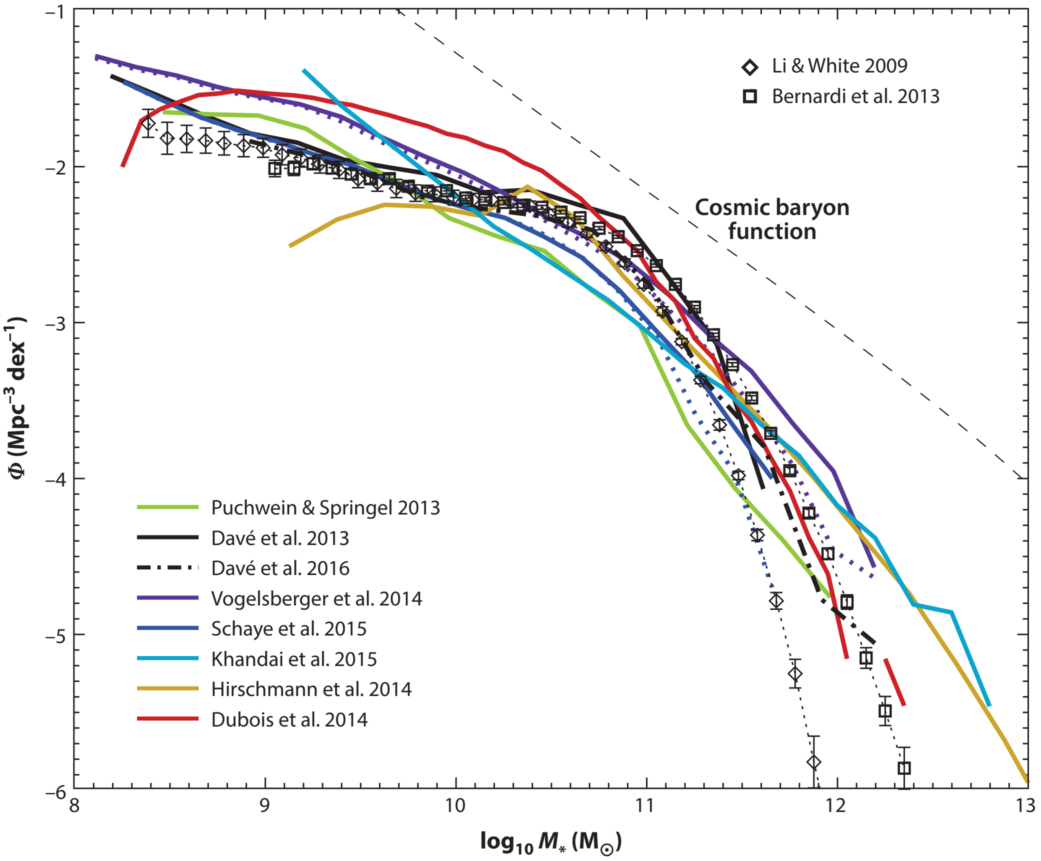

Figure 2.1: Galaxy stellar mass functions at $z = 0$ comparing results from simulations (coloured lines) to two observational measures (black symbols). The black dashed line shows the GSMF expected from the HMF assuming a universal baryon fraction of 15 %. The simulation marked 'Dubois et al. 2014' is the HORIZON simulation, which is investigated further in Chapter 7. (Reprint of Figure 4 from Naab & Ostriker, 2017).

disc with a comparatively homogenous stellar population (Eggen et al., 1962). The formation histories for disc and elliptical galaxies then diverge due to differences in their environments.

Disc galaxies are rotation dominated. They consist of a stellar and a gas disc, and typically have a spherical central bulge or pseudo-bulge (Kormendy & Bender, 2012), although bulgeless galaxies have also been reported (Simmons et al., 2013; Smethurst et al., 2017). A bulge is the dispersion dominated, tightly packed central group of stars in a galaxy. If it is rotation dominated, it is referred to as a “pseudo-bulge” (Kormendy & Bender, 2012). Disc galaxies tend to have continuous star formation histories that include young stellar populations, giving them their characteristic blue colour, and can have a variety of instability driven features, such as bars or spiral arms (Caroll & Ostlie, 2008; Buta, 2013). They are thought to undergo strong stellar feedback early on, in the form of SN and stellar winds that effectively heat and eject gas at high redshift (Weil et al., 1998; Naab & Ostriker, 2017). This gas is pre-enriched with metals from the first generations of stars, and circulates in the halo. It is later gradually re-accreted with higher angular momentum, and so settles into the disc where it feeds continuous star formation (Fall, 1979; Oppenheimer et al., 2010; Brook et al., 2014; Christensen et al., 2016). In general, disc galaxies show an “inside-out” assembly history, with the oldest stars found in the central bulge or bar and the youngest, and most metal rich, stars near the edge of the disc (Kent, 1986; MacArthur et al., 2004; Conselice, 2014; Sánchez-Blázquez et al., 2014).

Most elliptical galaxies by contrast are dispersion dominated and contain more homogeneous populations of older stars, giving them a red colour (Caroll & Ostlie, 2008). They grow early by in-situ star formation in deep potential wells where feedback is comparatively weak, and they form compact but massive systems (Feldmann et al., 2010; Rodriguez-Gomez et al., 2016; Galametz et al., 2009). After the cosmic peak of star formation around $z \sim 2$, in-situ star formation ceases but elliptical galaxies continue to grow by accreting stars from smaller systems (Oser et al., 2010; Gabor & Davé, 2012; Ceverino et al., 2015). This leads to a significant evolution in size, as well as gravitational heating and results in dispersion dominated galaxies with old stellar populations; the observed “red and dead” ellipticals (Baldry et al., 2004, 2006; Franx et al., 2008). However, as gas continues to accrete onto the galaxy even at $z < 2$ from the circumgalactic medium (CGM), an effective mechanism is required to prevent the galaxy from continuing to form stars. Research has shown that stellar feedback is unable to gravitationally unbind gas from galaxies with $M_* > 10^{11}$ $\mathrm{M}_\odot$ (Dekel & Silk, 1986; Benson et al., 2003; Heckman & Kauffmann, 2006; Heckman &

Thompson, 2016), but that the energy liberated by AGN, powered by SMBHs, might be able to (King, 2003; Murray et al., 2005; Fabian et al., 2006a). This question is explored further in Chapter 7.

All galaxies are embedded in much more extended dark matter halos, with a typical mass ratio of $M_{\rm DM}/M_* = 10$. Within galaxies, the ratio of dark matter to baryonic matter varies but is usually in the range of $10 - 20\%$, with dwarf galaxies found to be particularly DM dominated (Kormendy & Freeman, 2004).

During their lifetime, galaxies undergo a variety of interactions, including minor and major mergers, tidal stripping of smaller companions and harassment during close encounters, all of which influence their structure (Conselice, 2014). They collect in larger structures such as clusters and groups and it has been shown that environment correlates with a variety of galaxy properties (Dressler, 1980; Pimbblet et al., 2002; Yang et al., 2007; Pasquali & Nachname, 2015; Smethurst et al., 2017). Even in the absence of interactions, galaxies undergo secular evolution due to instabilities (Kormendy & Freeman, 2004), evolving magnetic fields (Moss et al., 2012; Beck & Wielebinski, 2013; Schober et al., 2013), radiation emitted by stars, cosmic rays (Recchia et al., 2017), stellar winds, SNe (Hopkins & Quataert, 2011), and feedback from black holes (Ciotti & Ostriker, 2007; Dubois et al., 2012a; DeGraf et al., 2014; Taylor & Kobayashi, 2015, e.g). Galactic winds are a ubiquitous feature for both morphological types, and they can carry an amount of gas comparable to a galaxy's star formation rate (Veilleux et al., 2005).

This short summary presented here barely scrapes the surface of the wealth of ongoing work on galaxy evolution, and it is only intended as a brief overview to provide context for the origin and evolution of SMBHs. The reader is advised that many of the models presented in this section are still contested, and is referred to Benson (2010); Conselice (2014); Naab & Ostriker (2017) and Cappellari (2016) for thorough reviews on the subject. Current unresolved questions in galaxy evolution include the role of major mergers, the form of gas accretion from the IGM, the overcooling problem, the "missing satellite problem", the source of reionization, the impact of baryons on halos, the star formation histories of galaxies, and the coevolution of SMBHs and galaxies, among many others.

2.2 Supermassive black holes and galaxies

2.2.1 Observational evidence for black holes

Black holes are typically split into three mass categories, a convention adopted for the remainder of this work: stellar mass black holes, with masses up to 10^3 $M_\odot$, intermediate mass black holes (IMBHs) with masses in the range $10^3 - 10^6$ $M_\odot$, and SMBHs with masses above 10^6 $M_\odot$.

Observationally, it is difficult to conclusively prove the existence of a black hole. Nevertheless, black holes have been reported in a variety of contexts, and over a large range of mass scales, using a range of observational techniques. These include gravitational wave signals during mergers (Abbott et al., 2016a), micro-lensing (Bennett et al., 2002; Minniti et al., 2015), emission from infalling material (Hu et al., 2008), emission from an accretion disc (Agis-Gonzalez et al., 2014), and the impact of a black hole's gravitational field on the orbits of nearby stars (Gillessen et al., 2007). All techniques only detect black holes interacting with other matter, and cannot detect isolated, quiescent black holes. Any attempt at studying the entire sample of black holes is therefore always an extrapolation from the accreting subsample.

Stellar mass black holes with masses below < 10 $M_\odot$ are expected to be most common, with some estimates suggesting 10^8 objects in the Milky Way alone (Fender & Belloni, 2012). They can be observed as X-ray binaries, when matter tidally stripped from a companion star is accreted by the black hole. During infall, a black hole converts its potential energy into photons, particularly X-rays. Isolated stellar mass black holes in the Milky Way can sometimes be found in micro-lensing events, when passing in front of a background star (Bennett et al., 2002). Finding stellar mass black holes outside the Milky Way is challenging. It has been suggested that ultra-luminous X-ray sources (ULXs), observed in nearby galaxies, are powered by stellar mass black holes, or possibly IMBHs (Fender & Belloni, 2012). Using the gravitational wave chirp emitted during the final stage of a merger, the LIGO consortium recently announced the detection of two mergers, involving black holes in the mass range $7.5 - 36$ $M_\odot$, leaving remnants with masses of $20^{+6.1}_{-1.7}$ $M_\odot$ and 62^{+4}_{-4} $M_\odot$ respectively (Abbott et al., 2016a,b).

The observational evidence for IMBHs is scarce, with some objects recently identified in nearby dwarf galaxies (Paragi et al., 2014; Baldassare et al., 2015; Ahn et al., 2017). It has also been suggested the ULXs could be powered by IMBHs (Fender & Belloni, 2012). The observational dearth of IMBHs can partially be explained by the difficulty in detecting them, as their low masses mean their region of gravitational influence is small (Greene, 2012). They are most easily found when accreting so observations are limited by accretion duty cycles. As IMBHs are thought to be the progenitors of more massive black holes, their comparative scarcity in observations is one of the unsolved questions in black hole physics. With the advent of the new space-based gravitational wave observatiory, eLISA, we will be able to observe black hole mergers in the mass range $10^3 - 10^7$ $\mathrm{M}_\odot$ out to redshift $z = 10$ (Barausse et al., 2015) by the 2030s.

The closest SMBH, Sagittarius A* (SgrA*), is postulated to be located at the centre of the Milky Way. Long-term observations of stellar orbits show the presence of a massive dark object, with a mass of $4.31 \pm 0.42 \times 10^6$ $\mathrm{M}_\odot$ (Gillessen et al., 2009), concentrated in a region of less than 125 au, which coincides with a radio source with a diameter of < 1 au. These two observations, together with the observed near-infrared and X-ray flares that can be explained with material accreting onto a black hole, provide compelling evidence for SgrA*'s nature as a black hole, embedded in a dense nuclear star cluster (Genzel et al., 2010). The new Event Horizon telescope, a very long baseline interferometry array, will be able to directly image the "black hole shadow", i.e. the strong gravitational lensing around a black hole, testing the predictions of general relativity and giving conclusive proof as to their nature (Doeleman et al., 2008; Broderick et al., 2014; Lu et al., 2014).

There are three other ways to find and measure the mass of SMBHs, which all rely on measuring the rotational velocity of an object in the gravitational potential of the black hole: maser emission from the accretion disc, broad line emission from cold material near the black hole, and dynamical modelling of spatially resolved kinematics. A combination of all three techniques has yielded a sample of about a hundred SMBHs (Kormendy & Ho, 2013; McConnell & Ma, 2013) that suggest most nearby galaxies house a SMBH (see Figure 2.2). The most massive nearby object to date, which has a mass of 9.7×10^9 $\mathrm{M}_\odot$, was identified by McConnell et al. (2011) in NGC 3842.

Powerful radio emitters, first found in the 1960s and dubbed quasi-stellar radio source (quasar) at the time, turned out to be part of a large population of active SMBHs, now called AGN. In recent years, large scale surveys have been revealing an ever growing sample of AGN (Bower et al., 2006; Shen et al., 2008; Guo et al., 2010; Delvecchio et al., 2014; Jiang et al., 2016, among others) covering the entire range of redshifts accessible with telescopes today (Heckman & Best, 2014). Despite being a varied population with a range of luminosities, spectral shapes and duty cycles, they are all thought to be powered by a central SMBH (Haiman & Loeb, 2001; Volonteri

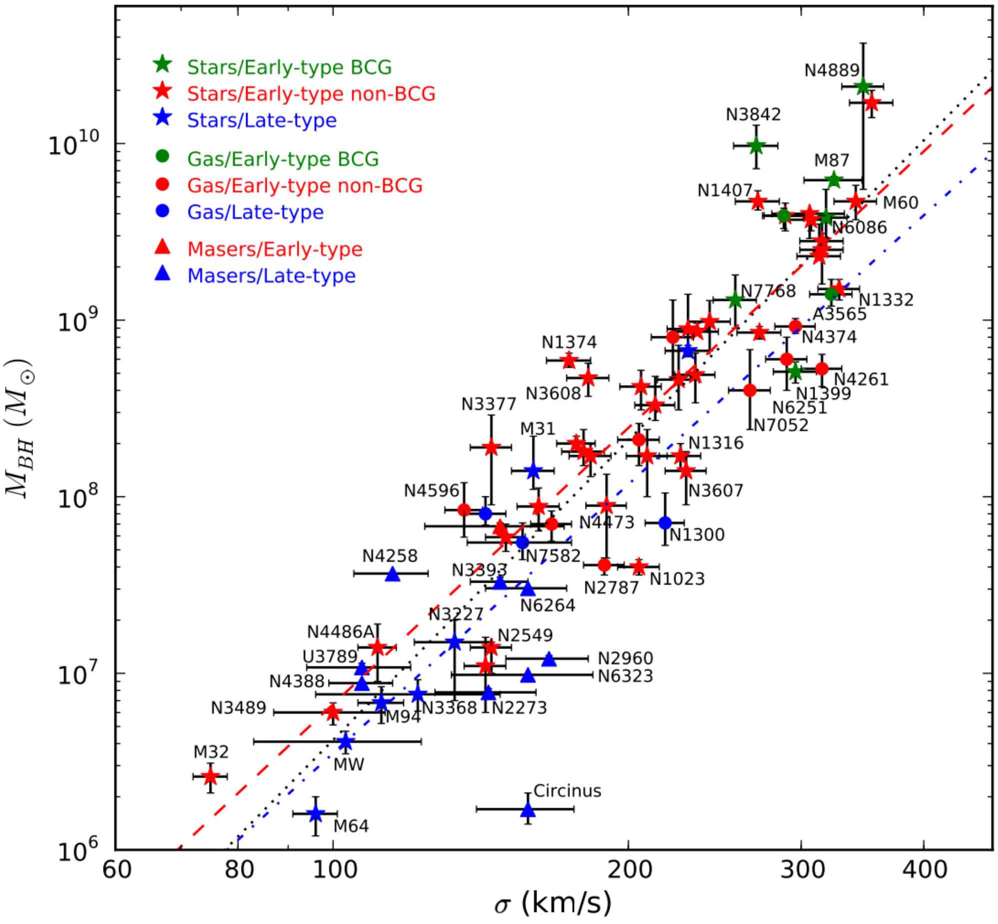

Figure 2.2: $M_{\text{SMBH}} - \sigma$ relation for 72 local SMBHs. The tight correlation between black hole and galaxy properties suggests a link between black hole accretion and the star formation history in the host galaxy (Reprinted of Figure 1 from McConnell & Ma, 2013).

& Rees, 2005), embedded in a range of gas configurations and seen at a range of inclination angles (Urry, 2003). The oldest, and some of the most massive, SMBHs have been found at very high redshift in the form of quasars, mostly in the Sloan Digital Sky Survey (SDSS). Using a colour selection criteria and spectroscopic follow-up, Fan et al. identified a sample of about 20 quasars at redshift $z \geq 6$ (Fan et al., 2001, 2004, 2006). The highest redshift object identified to date is a 2×10^9 $M_\odot$ black hole at redshift $z = 7.085$, found by Mortlock et al. (2011).

This large sample of SMBHs, found at the centres of massive galaxies over a large range of redshifts, begs the question where these objects come from? What role do they play in galaxy evolution? How did they grow so massive in such a limited amount of time? If the first massive objects already had more than 10^9 $M_\odot$ at $z = 7$, why are the most massive objects we see today not significantly more massive? How do they influence their host galaxy? How does their host galaxy influence them?

2.2.2 Evidence for the coevolution of supermassive black holes and galaxies

The interest in SMBHs and AGN increased sharply once it was realised that there exists a remarkably tight correlation between the mass of a central SMBH and the stellar mass of the host galaxy's bulge (Magorrian et al., 1998; Marconi & Hunt, 2003). Since then, further correlations have been identified, such as between the mass of the SMBH and the velocity dispersion, σ, of its host galaxy (Gebhardt et al., 2000; Ferrarese & Merritt, 2000) (see Figure 2.2, reprinted from McConnell & Ma, 2013). This correlation, known as the M-σ relation, hints at strong coevolution between galaxies and SMBHs, but the question of how they influence each other is not yet fully understood.

Further support for the link between star formation history and SMBH accretion came from the discovery that star formation in galaxies and accretion onto SMBHs follow a remarkably similar trend over cosmic time, with a slow increase throughout the history of the Universe up to redshift $z = 3$, a broad peak around $z = 3 - 2$ and sharp fall-off to $z = 0$ (See Figure 2.3, reprinted from Shankar et al. 2009).

The obvious common denominator between star formation and black hole accretion rate is the gas supply of the galaxy, in particular the cold, dense gas (Benson et al., 2003; McKee & Ostriker, 2007), which fuels both processes. Any processes that impacts the cold dense gas, whether by reducing the total gas mass or by heating the existing supply in the galaxy, is expected to impact both star formation and black hole accretion. Both stars and active SMBHs produce energetic feedback that

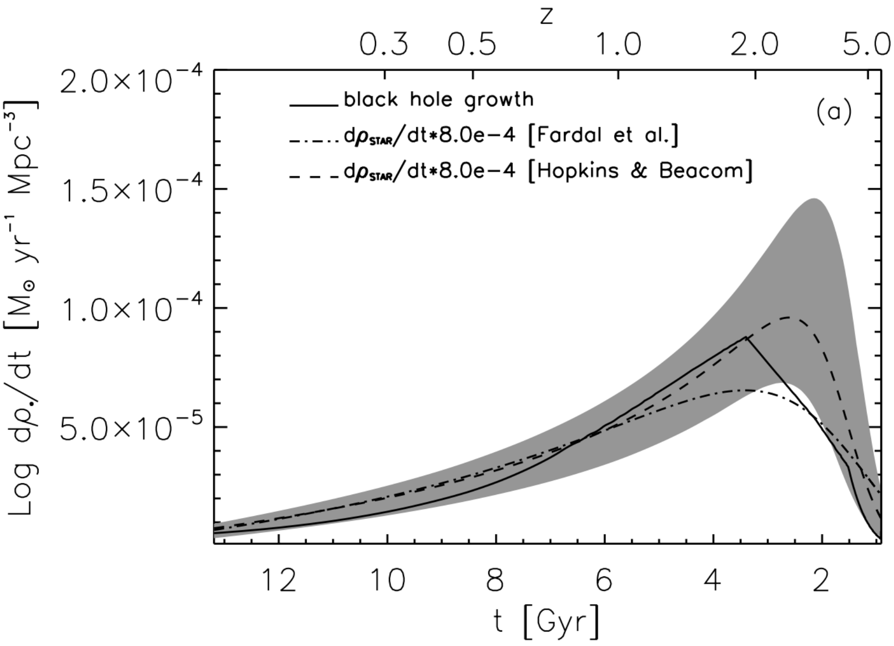

Figure 2.3: Redshift evolution of the star formation density and black hole growth. Their similar global trends provide further hints at a co-evolution between black hole accretion and star formation (Reprint of Figure 6 from Shankar et al., 2009).

transfers energy and momentum to the gas in their vicinity, but the gas in galaxies is subject to many competing processes at the same time and their relative importance is difficult to disentangle.

2.2.3 Supermassive black holes as active galactic nuclei

Observationally, AGN display a wide variety of spectral shapes and luminosities, forming a veritable zoo of objects with various emission and absorption features. Much work, particularly in the 1990s, led to a unified AGN model (Urry, 2003; Beckmann & Shrader, 2013) in which a SMBH at the centre of a galaxy is surrounded by an accretion disc and a corona that radiates in the optical through to the soft X-ray as it is heated by magnetic and viscous processes.

Based on extensive work on X-ray binaries, the accretion disc is thought to be in one of two states. In one state, characterised by a soft X-ray spectrum, it appears to be geometrically thin and optically thick (Shakura & Sunyaev, 1973). In the second state, the disc puffs up and enters an advection-dominated accretion flow (ADAF), which makes it geometrically thick but optically thin (Narayan & Yi, 1994). This pressure supported disc is radiatively inefficient and characterised by hard X-rays. X-ray binaries frequently cycle between the modes, tracing out a characteristic path

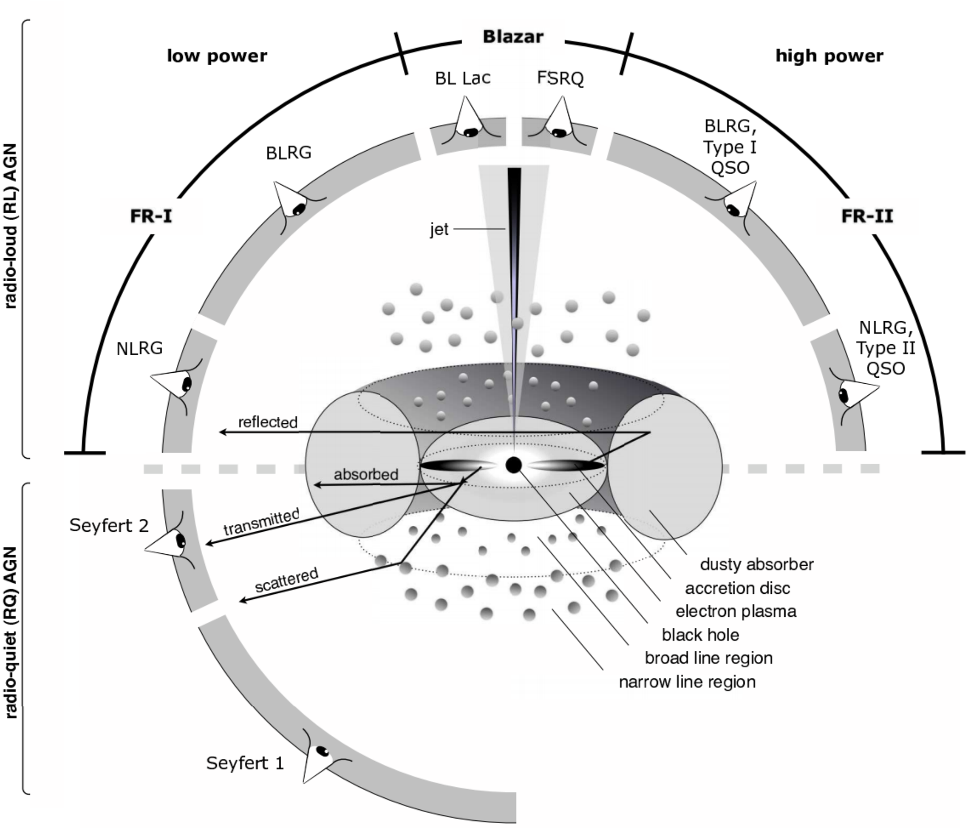

Figure 2.4: Characteristic features of the engine powering an AGN. Their spectral type depends on the features present as well as the inclination (Reprint of Figure 1 from Beckmann & Shrader, 2013).

on the Hardness-vs-Intensity diagram that suggests the cycle is not reversible (Fender & Belloni, 2012). The details of this evolution in accretion disc state are not well understood, as the presence of instabilities make it difficult to tackle analytically and the importance of magnetic fields and radiation hamper numerical efforts (Abramowicz & Fragile, 2013). Even less is known about the accretion discs of AGN, as the range of outer boundary conditions is much greater than in X-ray binaries (Begelman, 2014). Observational evidence exists for both disc types, but few instances of a dramatic change in accretion state have been observed (Brandt & Alexander, 2010).

Near the SMBH, gas can be found at both high velocities in the broad line region, and low velocities in the narrow line region. A relativistic jet is formed near the black hole, perpendicular to the accretion disc, and the disc is surrounded by a dusty torus that leads to significant obscuration from some angles (see Figure 2.4). Depending on the state of the AGN's accretion disc and the local environment, not all components of the AGN are necessarily present. This central engine powering AGN are stipulated to be found in one of two modes: a radiative mode or a jet mode, which are discussed in more detail in the following two sections. The variation in spectral shape in AGN arises because of intrinsic variations in black hole mass, density, size, ionisation parameters, accretion disc state, luminosity, and most importantly, inclination and the resulting obscuration.

2.2.3.1 Active galactic nuclei in radiative mode

AGN in a radiative mode are more likely to be found in high-density pseudo-bulges. Thought to be fed by high density cold gas funnelled inwards due to secular processes, they accrete with relatively high accretion rates, with Eddington ratios of $\chi > 0.01$ (Fabian, 2012; Heckman & Best, 2014). Accretion probably proceeds via a geometrically thin, optically thick accretion disc shielded by a dusty torus. Depending on the presence of a jet, they can have strong or weak radio emission.

The Eddington ratio, $\chi = \dot{M}_{\mathrm{SMBH}}/\dot{M}_{\mathrm{Edd}}$ is a measure of how efficiently a black hole of a given mass is accreting. It is based on the idea that for a luminous, massive object, here taken to have mass M_{SMBH}, there is a maximum luminosity, L_{Edd}, at which the outwards radiation pressure balances the inward gravitational attraction, and the gas is held in hydrostatic equilibrium (Caroll & Ostlie, 2008). Assuming that the luminosity of an AGN, L_{AGN} is a function of its accretion rate, such that

$$L_{\mathrm{AGN}} = \epsilon_{\mathrm{r}} \dot{M}_{\mathrm{SMBH}} c^2 \tag{2.1}$$

where $\epsilon_{\rm r}$ is the percentage of the accretion energy radiated away and c is the speed of light, the radiative force per unit area, $P_{\rm rad}$, at a distance r from the AGN can be written as

$$P_{\rm rad} = \frac{L_{\rm AGN}}{4\pi r^2 c}. \tag{2.2}$$

For the forces to balance, i.e. when $F_{\rm grav} = F_{\rm rad}$,

$$\frac{GM_{\rm SMBH} m_p}{r^2} = P_{\rm rad} \times m_e \kappa, \tag{2.3}$$

where m_p is the mass of a proton and m_e is the mass of an electron. κ is its unit cross-section per unit mass, and $P_{\rm rad}$ is the radiation pressure acting on electrons so $\kappa = \frac{\sigma_T}{m_p}$ where σ_T is the Thomson cross section. It is assumed that the gas is monoatomic.

Equation 2.3 can be solved for the Eddington luminosity of the object using Equation 2.2,

$$L_{\rm Edd} = \frac{4\pi G M_{\rm SMBH} c m_p}{\sigma_T}, \tag{2.4}$$

which in turn can be solved for a maximum accretion rate, $\dot{M}_{\rm Edd}$ onto the SMBH

$$\dot{M}_{\rm Edd} = \frac{4\pi G M_{\rm SMBH} m_p}{\epsilon_{\rm r} c \sigma_T}. \tag{2.5}$$

$\dot{M}_{\rm Edd}$ is the accretion rate at which the radiation pressure from the AGN exactly balances the gravitational attraction on the infalling gas. For higher accretion rates, the gas may be driven out as a wind. The Eddington accretion rate can therefore be seen as a maximum accretion rate onto the AGN, assuming that the gas in the ISM consists mostly of ionised hydrogen, and the radiation is emitted isotropically. Given the relationship between accretion rates and luminosities, the Eddington ratio can be expressed as

$$\chi = \frac{\dot{M}_{\rm SMBH}}{\dot{M}_{\rm Edd}} = \frac{L_{\rm AGN}}{L_{\rm Edd}}. \tag{2.6}$$

Observations show that quasars, by and large, respect their Eddington limits (Begelman, 2014) but it has been suggested that the radiation is effectively trapped in the flow and advected into the black hole at very high accretion rates. This means accretion rates onto the sink could exceed the Eddington limit, while the emergent luminosity is still Eddington limited (Begelman, 1979; Abramowicz et al., 1988). There is a small observed population of super-Eddington objects, which claim to have $\chi > 1$ (Du et al., 2015; Zuo et al., 2015; Kara et al., 2016). At $z > 2$, accretion rates of $\chi > 0.01$ are more commonly observed, whereas at lower redshift, objects with $\chi < 0.01$ are more common (Kollmeier et al., 2006; Shen et al., 2008; Simmons et al., 2012; Trakhtenbrot et al., 2017). This evolution in accretion rates, combined with the fact that AGN above redshift $z > 2$ are on average more massive than AGN at lower redshift leads to an observational effect known as "cosmic downsizing" (Hasinger et al., 2005; Barger et al., 2005; Heckman & Kauffmann, 2006).

However, a quasar whose radiative pressure on ionised hydrogen locally balances the gravitational acceleration near the SMBH will not be powerful enough to arrest the infall of gas into the much deeper potential of the entire galaxy. It has been suggested that the radiation pressure primarily acts on dust, which has a much larger cross section than ionised hydrogen. Intriguingly, the ratio of cross sections is approximately equal to the ratio of black hole mass to galaxy stellar mass, as inferred from the $M_{\rm SMBH} - \sigma$ in Figure 2.2, such that $\frac{\sigma_T}{\sigma_{\rm dust}} \approx \frac{M_{\rm SMBH}}{M_*}$. So an AGN that is at its local Eddington limit for ionised hydrogen is also approximately at the Eddington limit on dust for the entire galaxy. This has been taken as tentative observational evidence for the importance of radiative AGN feedback (Murray et al., 2005; Fabian, 2012). Alternatively, it has been suggested that the AGN drives a local wind, which then sweeps up gas as it moves through the galaxy. A range of models have been suggested for the AGN to transfer its energy to the surrounding gas, using either energy driven or momentum driven mechanisms (See comprehensive reviews by Fabian, 2012; King & Pounds, 2015).

2.2.3.2 Active galactic nuclei in jet mode

By contrast, AGN with less efficient accretion, at $\chi < 0.01$, enter a jet mode where a significant share of their accretion energy is extracted by powerful jets. They are found in classical bulges and ellipticals, and are associated with strong radio emission. Their accretion discs are truncated by an advection-dominated, geometrically thick and therefore radiatively inefficient inner accretion flow and are thought to be fuelled by gas slowly cooling in a self-regulating fashion. The observed "fundamental plane of black hole accretion" shows that the jet power scales with black hole mass as well as with the AGN's X-ray luminosity (Merloni et al., 2003; Gültekin et al., 2014), accretion rate (Sbarrato et al., 2014) and bolometric disc luminosity (Inoue et al., 2017). The mechanisms by which these jets form and how they stay collimated on kpc scales are areas of ongoing research. Suggestions involve some variation of the Blandford-Znajek mechanism, where magnetic field lines threading the accretion disc are wound up as the disc rotates (Blandford & Znajek, 1977; Koide, 2002; Penna, 2014).

Strong observational evidence for the impact of jets on the star formation of their host galaxies can be found in brightest cluster galaxies of nearby cool core clusters (Fabian, 2012). These clusters have a short cooling time, as observed in X-rays, and should appear as bright starbursts but show little ongoing star formation. The AGN in these galaxies generate powerful jets that inflate bubbles of relativistic plasma, visible in the radio, which then rise buoyantly through the CGM and IGM (McNamara & Nulsen, 2007). The fact that their mechanical power is sufficiently high to offset the cooling rate of the host galaxy, and that the fraction of galaxies observed to contain long-lived bubbles is in excess of 95%, is strong circumstantial evidence for the impact of AGN on host galaxy star formation (Fabian, 2012). The exact mechanism through which the mechanical energy of the jet is dissipated into the CGM is currently not well understood.

2.2.4 Observations of outflows from active galactic nuclei

Frequent and fast outflows, with a range of outflow velocities, covering angles and mass loading factors, seem to be a ubiquitous feature of AGN in massive galaxies (Veilleux et al., 2005; Fabian, 2012; King & Pounds, 2015). Observed outflows range from ultrafast collimated jets (Baade & Minkowski, 1954; Bridle & Perley, 1984; Forman et al., 2007; McNamara et al., 2009), associated with jet mode AGN, to winds with much larger covering fractions and a range of velocities (Weiner et al., 2009; Reeves et al., 2009; Tombesi et al., 2010; Parker et al., 2017; Genzel et al., 2014; Harrison et al., 2014). These outflows are able to drive out a significant gas mass (Heckman et al., 2000; Veilleux et al., 2005; Weiner et al., 2009; Sturm et al., 2011) using only $5 - 10\%$ of the accretion power of the AGN (Moe et al., 2009; Saez et al., 2009; Dunn et al., 2010) and have been observed even at high redshifts of $z > 6$ (Maiolino et al., 2012; De Rosa et al., 2014).

While there is strong observational evidence that outflows are associated with AGN, the impact of these galactic scale winds on the star formation of the host galaxy is much more difficult to prove observationally. Originally it was reported that AGN are preferentially found in so called “green valley galaxies”, which are thought to be transitioning from actively star forming to quenched (Schawinski et al., 2007; Nandra et al., 2007; Smethurst et al., 2017), but recent results point out that these galaxies are instead actively star forming but dust reddened (Cardamone et al., 2010), and that AGN are preferentially hosted in star forming galaxies (Rafferty et al., 2011; Rosario et al., 2013; Azadi et al., 2015).

For massive galaxies, there is evidence for a correlation between the existence of outflows and star formation (Carniani et al., 2016), AGN luminosity and outflow velocity (Sturm et al., 2011), star formation rate (SFR) and wind velocity (Weiner et al., 2009; Cano-Díaz et al., 2012), and SFR and AGN luminosity (Lanzuisi et al., 2017; Page et al., 2012). This suggests that the three processes are intrinsically linked. However, the picture rapidly becomes more complex. Förster Schreiber et al. (2014) report evidence for 8 massive galaxies around $z \approx 2$ that have both a nuclear, fast outflow associated with an AGN, as well as a weaker outflow probably driven by stellar feedback, so galaxies may be subject to a range of different feedback mechanisms simultaneously. Rosario et al. (2012) and Ruschel-Dutra et al. (2016) find that high luminosity AGN are correlated with star formation rate, but low luminosity AGN are not, suggesting low luminosity AGN have little impact on their host galaxy's star formation. By contrast Cheung et al. (2016) present evidence for the presence of a centrally driven wind in typical quiescent galaxies that host low-luminosity AGN, suggesting that low luminosity AGN play a vital role in reducing the star formation of their host galaxy. If AGN feedback quenched star formation, one would naively expect the brightest AGN to be housed in the least star forming galaxies, but the observed correlation between SFR and AGN luminosity is positive. The impact of AGN on star formation and outflows of their host galaxies is therefore complex and an ongoing area of research (Smethurst et al., 2016; Zubovas & Bourne, 2017; Pontzen et al., 2017).

2.2.5 Modelling feedback from active galactic nuclei in simulations

One way to isolate the impact of a particular process is via simulations; all current large-scale cosmological simulations include a model of AGN feedback. Some simulations, such as HORIZON (Dubois et al., 2014b) and Illustris (Genel et al., 2014) implement a two-mode feedback model, returning either energy or momentum to the gas surrounding the SMBH depending on whether accretion rates are high or low. Other simulations, such as the EAGLE simulation (Schaye et al., 2014) or the semi-analytic model used with MassiveBlackII (Khandai et al., 2015; Di Matteo et al., 2005) use only a single, energy based feedback prescription at all accretion rates. Each model has a set of free parameters that need to be tuned to an observational result, such as the GSMF at $z = 0$ (Illustris, Genel et al., 2014; Sijacki et al., 2015), or the $M_{\mathrm{SMBH}} - \sigma$ also at $z = 0$ (HORIZON, Dubois et al., 2014b; Volonteri et al., 2016). Despite the different implementations, all four simulations produce good agreement with a range of observations and report that AGN feedback is necessary to reproduce the number density of galaxies at the knee of the GSMF and above (see Figure 2.1). While the large number of galaxies included in each of these simulations make them ideal tools to study statistical measures such as the black hole correlation relations and mass functions, their necessarily limited resolution of ~ 1 kpc means their predictive power on small scales is limited.

Higher resolution simulations of individual objects frequently report AGN driven outflows (Naab & Ostriker, 2017), but find varying impact on the star formation. In some simulations, star formation is significantly reduced in the presence of AGN (Di Matteo et al., 2005; Sijacki et al., 2007; Dubois et al., 2013a), while others find it unaffected despite strong outflows (Cattaneo & Teyssier, 2007; Gabor & Bournaud, 2014; Roos et al., 2015). Studying the impact of AGN in more detail by comparing simulations with and without AGN feedback, Dubois et al. (2013a) found strong evidence that AGN feedback heats the gas at galaxy and halo scales. Not only does this reduce star formation by diminishing the supply of cold, dense gas (Dubois et al., 2013b), it also powers hot superwinds that blow away cold filamentary inflows, preventing the gas supply from being replenished. This combination of reduced inflows and increased outflows lowers the baryon fraction of the galaxy by up to 30% within the halo. Pontzen et al. (2017) also report that boosted outflows and reduced inflows quench star formation and stress that the latter is particularly important to prevent star formation in the long term.

In conclusion, simulations and observations agree that feedback from AGN can have a profound effect on the evolution of the host galaxy. While the picture is far from complete, there are strong hints that the energy and momentum deposited by an AGN produces a combination of heating, inflows and outflows. In combination with other processes that influence galaxy evolution, such as cosmic accretion and stellar feedback, it can significantly alter the gas content, and as a consequence the star formation, of its host galaxy. This question is investigated further in Chapter 7.

2.2.6 The origin of supermassive black holes

If SMBHs do significantly influence the evolution of massive galaxies, as both observations and simulations suggest, then there must be a robust sample of SMBHs in place by redshift $z = 6$, scattered throughout the most massive galaxies. Their genesis occurs during a time for which little observational information is currently available. Instead, we can probe their origin using simulations to build an evolution history for black holes, bridging the gap from the overdensities at the time of last scattering imprinted in the CMB to the wealth of observational evidence at low redshift. The difficulty in simulating the origin and evolution of SMBHs is the large range of scales involved, both spatially and temporally, which cannot all be captured in a single simulation. Different parts of their evolution must therefore be investigated in different contexts.

There is by now a large body of numerical work concerned with the formation of black hole seeds, from which three models for black hole seed formation have emerged (described in detail in Section 2.2.6). Many of these simulations use extremely high resolutions, down to the au scales probed in Latif et al. (2013a) but in return are run only until the compact object collapses. Following the evolution of black holes once they are born requires much longer timescales to be covered, which comes at a cost. There are two approaches, both of which rely on sub-grid models to capture the physics of black hole accretion and feedback as the physical scales of the black hole remain unresolvable with current technology. Firstly, one can use simulations of isolated galaxies at high resolution, sacrificing the cosmological context and evolution of the host galaxy, to investigate the early growth phase of newly born black holes (Gan et al., 2014; Emsellem et al., 2014; Lupi et al., 2016; Curtis & Sijacki, 2016b). Alternatively, zoom simulations at pc to kpc resolution are able to cover the cosmological context and coevolution with the host galaxy to redshift $z = 6$ and beyond, into the observational regime (Di Matteo et al., 2008; DeGraf et al., 2012; Dubois et al., 2012a; Curtis & Sijacki, 2016a). Recent advances in this type of simulation allow for a more sophisticated treatment of physics, including radiative transfer and magnetic fields (Pacucci & Ferrara, 2015). Levine et al. (2008, 2010) attempted to combine cosmological context and high resolution by running low resolution simulations until a point of interest and then adding extra resolution to study a particular episode in detail. Finally, large box hydrodynamical simulations at kpc resolution provide large populations over the entire history of the universe (Dubois et al., 2014b; Genel et al., 2014; Schaye et al., 2014), able to probe the statistical quantities associated with black hole and galaxy coevolution, but at a cost of resolution and complexity of algorithms.

From this work, SMBHs form at redshift $z = 30 - 10$ via one of three possible seed formation mechanisms and then grow via a combination of gas accretion and mergers with other black holes, as explained in more detail in the next two sections.

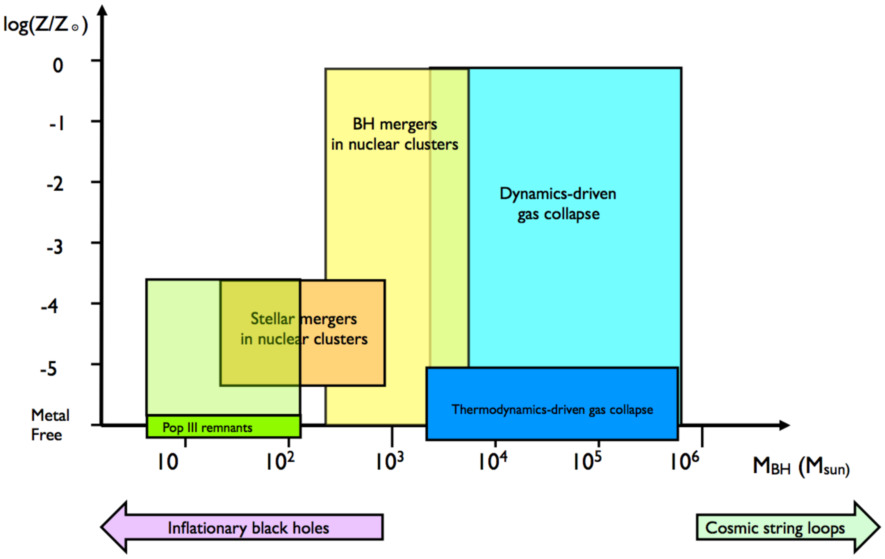

Figure 2.5: Expected black hole seed masses for a range of seed formation models, at a range of gas metallicities (Reprint of Figure 1 from Volonteri et al., 2013).

2.2.6.1 Black hole seed formation in the early Universe

Three main formation channels have been suggested for the progenitors of SMBHs, based on a combination of analytic models and high resolution simulations (Turk et al., 2009; Schleicher et al., 2010; Regan et al., 2014b; Taylor & Kobayashi, 2014; Ballesteros-Paredes et al., 2015; Choi et al., 2015). Each process requires different initial conditions, and results in a different distribution of seed masses.

The first category is stellar mass seeds, which are the remnants of the first massive stars, so called Pop III stars. As the first stars, by definition, formed from pristine gas, a lack of metal line cooling suppressed fragmentation and results in a population of massive stars (Bromm et al., 1999; Nakamura & Umemura, 2001). If their stellar masses are outside the mass range for pair instability supernovae ($140-260$ $M_{\odot}$) which leave no remnant (Fraley, 1968), they will collapse to a remnant black hole after a short ($<$ Myr) lifetime (Volonteri, 2010). The first stellar mass seeds are expected to form at redshift $z = 30 - 20$ along with the first stars, in so called “mini-halos” with halo masses of approximately 10^6 $M_{\odot}$, and have black hole seed masses in the range $10 -$ a few $\times$ 10^2 $M_{\odot}$ (Johnson & Bromm, 2007; Turk et al., 2009).

The disadvantage of this model is that the seed black holes are quite small in comparison to the other formation channels, which means that their initial accretion rates will be limited by their low masses. On the other hand, stellar mass seeds are expected to form early and therefore have more time to accrete until the first observations. In contrast to the other formation channels, black hole formation as stellar remnants is well understood, and recent observations by the LIGO consortium prove the existence of black holes in the mass range $10 - 10^2$ (Abbott et al., 2016a). Our poor understanding of the initial mass function (IMF) of Pop III stars makes it difficult to estimate their number density, which could be very high in comparison to the observed number density of SMBHs. While there are possible scenarios in which only a subset of the stellar mass seeds grow to be supermassive, one would then expect the remainder to form a population of IMBHs (Tanaka & Haiman, 2009), which is not (yet?) observed.

Heavier black hole seeds could be formed by runaway stellar cluster collapse of a second generation of low mass stars, formed from gas pre-enriched by the Pop III stars, which could form compact nuclear clusters (Regan et al., 2014b). During runaway collapse, these clusters could produce a black hole remnant with a mass of $10^2 - 10^4$ $M_\odot$ (Devecchi & Volonteri, 2009). The limitations of this model are mainly its late formation time, with the black hole seed expected to form around $z \approx 10$, as it requires two generations of stars and the dynamical evolution of a star cluster (Davies et al., 2011). There is also little evidence for this type of runaway cluster collapse at low redshift.

The third formation process is again unique to high redshift as it relies on the presence of metal-free gas. So called "direct collapse" black holes are postulated to form when gas in halos with virial temperatures above 10^4 K cools as far as possible via Hydrogen cooling and then continues to contract adiabatically. To prevent the object from fragmenting when cooling too efficiently, UV radiation by a close neighbour is needed to dissociate molecular hydrogen in the collapsing halo (Shang et al., 2010; Schleicher et al., 2010; Glover, 2015). Simulations show that rotation halts contraction at a radius of ~ 20 pc. If these objects are able to redistribute angular momentum, for example through a bars-within-bars mechanism (Begelman et al., 2006), each object will form a central supermassive star with a mass of $10^4 - 10^6$ $M_\odot$, which then collapses into a black hole in the mass range $10^4 - 10^5$ $M_\odot$ (Latif et al., 2013b; Lupi et al., 2014; Choi et al., 2015) with the remaining mass ejected. For direct collapse black holes to form, several conditions need to be met, so their number density varies over cosmic time (Dijkstra et al., 2014; Habouzit et al., 2016; Regan et al., 2017). The IMF of these objects is also very dependent on the (still poorly understood) state of mini-halos at high redshift (Ferrara et al., 2014).

It is challenging to observationally distinguish between different seed formation models, as the long evolution history of the Universe means few of the original objects remain. Efforts have been made to differentiate models by comparing their predicted number densities to the number density of SMBHs today, but efficient accretion has washed out any trace of the initial population (Volonteri, 2010; Volonteri & Bellovary, 2012). Nevertheless, there is hope that probing the population of black holes at higher redshift, using upcoming observational missions such as the James-Webb Space Telescope (JWST) and eLisa, will allow different models to be differentiated. Another possible way to probe the seed formation mechanism is via the occupation fraction of local low-mass and dwarf galaxies, which have evolved less strongly than more massive galaxies, and might preserve a distribution of black holes similar to the original one. For models with lower seed formation efficiencies, a larger fraction of halos will be without a massive black hole (Van Wassenhove et al., 2010). However, this strategy requires a complete sample of black holes in dwarf galaxies, which is a very challenging observational task (Greene, 2012).

2.2.6.2 Supermassive black hole growth and feedback

Once the seed black hole has formed, it needs to grow rapidly as even in the best case, it has to gain three orders of magnitude in mass in a limited timeframe to match observations (Haiman & Loeb, 2001). Accounting for long merger timescales, low observed accretion rates and high observed radiative efficiencies, achieving this level of growth is challenging (Shapiro, 2005; Pelupessy et al., 2007; Johnson et al., 2013).

The gas accretion onto the black hole is limited by the available gas supply in its vicinity. Simulations have shown that the gravitational collapse of mini-halos produces dense central gas concentrations in the proto-galaxies where black hole seeds are born (Pelupessy et al., 2007; Greif et al., 2008). These dense central regions are sufficiently concentrated to enable a black hole's rapid early growth (Bromm & Loeb, 2003; Greif et al., 2008). Unfortunately, both the formation of the black hole itself and its growth phase involve energetic feedback processes which have been found to blow away the remaining gas, starving the newborn black hole (Johnson & Bromm, 2007; Alvarez et al., 2009; Whalen et al., 2010; Shlosman et al., 2016). Timescales for the gas to recover and flow back are on the order of a local Hubble time, which could severely limit the black hole's ability to grow. As a consequence, simulations have been struggling to produce black holes of the requisite masses (Di Matteo et al., 2008; Costa et al., 2014b; Dubois et al., 2015).

Recent simulations have shown that the interplay between local gas supply, the black hole and its feedback are very subtle and difficult to capture in an analytic model. For example, Pacucci & Ferrara (2015) find that black holes with seed masses above 10^5 $\mathrm{M}_\odot$ enter an accretion dominated regime and grow rapidly, but lower mass seeds end up feedback dominated if radiative feedback is treated self-consistently. Lupi et al. (2016) report an early rapid accretion phases when the black hole is born in a dense, massive clump, a phenomenon only seen when structures on the scale of giant molecular clouds are captured in simulations.

It has been argued that a short Super-Eddington accretion boost, particularly early on, significantly increases the chance of producing SMBHs by $z = 6$ (Volonteri & Rees, 2005). Alexander & Natarajan (2014) suggest that stellar dynamics at redshift $z > 15$ could prevent low mass black holes from being embedded in slowly draining accretion discs, and instead launch them into a Super-Eddington phase, allowing them to grow to 10^4 $\mathrm{M}_\odot$ in a few Myr.

Evidence varies for the importance of early black hole - black hole mergers. Some argue for the importance of early mergers to quickly build up the mass of a seed black hole (Arun et al., 2009; Volonteri, 2010), which then allows for more efficient gas accretion. In line with this argument, current black hole models often assume mergers to be efficient and merge black holes when they are in close proximity and at low relative velocity (Dubois et al., 2010; Costa et al., 2014a). However, others argue that dynamical times for mergers are long (Sesana et al., 2007; Chapon et al., 2013; Goicovic et al., 2017) and while binaries might commonly form, mergers contribute little to early mass growth (Dubois et al., 2014a).

Not only does feedback from the black hole itself impact the black hole's own accretion history, current simulations show that by the time we observe the first quasars at $z = 6$, black holes have already significantly altered the evolution history of their host galaxy (Di Matteo et al., 2012; Dubois et al., 2013a). The first billion years of the Universe are therefore crucial not only for the population of SMBHs themselves, but also the population of massive galaxies in the local universe. While a lot of work has been done on the origin of black holes, we still struggle to reconcile observed masses with our limited knowledge about their early evolution. The current challenge is to dramatically increase the amount of information available on the black hole and its immediate environment, as black hole growth is highly non-linear and very dependent on local conditions. We also need to develop more complete feedback models, including radiative effects, magnetic fields and cosmic rays, as the further evolution of black holes at high redshift seems to sensitively depend on their own feedback. Both problems are inherently numerically expensive, and will require novel algorithms to make simulations affordable. The question of increased resolution is one of the challenges addressed in this thesis, which will study the coevolution of black holes and galaxies using a range of hydrodynamical simulations produced using the code RAMSES, which is introduced in the next chapter.

Chapter 3

Numerical methods: RAMSES

The origin and evolution of galaxies, and all their components, is a highly non-linear problem whose complexity does not lend itself well to analytic analysis. It does however break down into a set of distinct physical laws and processes, such as gravity, cooling and star formation, which can be combined and integrated to simulate the evolution of the Universe. RAMSES (Teyssier, 2002) is one code capable of such simulations and the one employed for the work presented here. Its relevant elements are described in this chapter. The work presented in Chapters 4, 5 and 6 are based on the publicly available version from October 2013 (Teyssier, 2013).

One common problem in astrophysics is that the range of scales involved in a given problem can be huge, from atomic to galactic, and as such, it is extremely difficult to fully capture a physical process in a single simulation. This is especially true for processes that suffer from an 'inverse cascade' problem, where (unresolved) small scale behaviour influences the outcome on (resolved) large scales. For these reasons, one has to rely on sub-grid models, which aim to replicate the impact of unresolved, small scale behaviour on scales relevant to the simulation at hand, using only information available in the simulation. This chapter therefore also includes a description of the sub-grid algorithms used in Chapter 4 to 7. Algorithms and parameters only applicable to a particular section of this thesis are explained in the relevant chapter.

3.1 Nomenclature

For the remainder of this thesis, vector quantities are denoted in bold, $\boldsymbol{x}$, scalar quantities in italics, x and volume average quantities with a bar $\bar{x}$. Time averaged values are denoted by triangular brackets, $\langle x \rangle$. Quantities denoted with a ∞ subscript, X_∞, are measured "far from the gravitational influence of the black hole" or analytic values, and quantities as measured on the fly by the cloud particles of the sink are denoted as $x_\bullet$. The details of what the last two quantities mean are explained in Section 4.2.3.

The term “local” refers to the immediate vicinity of the black hole, whereas the term “global” describes conditions far from the black hole.

3.2 Structure

RAMSES is an adaptive mesh refinement (AMR) code that also includes collisionless particles. To integrate the equations of fluid dynamics and gravity, the distribution of matter in the universe has to be discretised, for which RAMSES uses one of these two complementary approaches, depending on whether the distribution to be modelled is subject to pressure forces or not.

Dark matter (DM) only interacts via gravity and so can be approximated as a collisionless fluid, which in turn can be discretised as a set of particles that carry the information of the flow. The same holds for stars, as their collisional cross sections are so small as to be negligible. The evolution of these distributions is then simply determined by the equations of motion for each individual particle in a gravitational potential, namely

$$\begin{aligned} \frac{d\boldsymbol{x}_\mathrm{p}}{dt} &= \boldsymbol{v}_\mathrm{p} \\ \frac{d\boldsymbol{v}_\mathrm{p}}{dt} &= -\boldsymbol{\nabla}\phi_\mathrm{g} \end{aligned} \tag{3.1}$$

where $\boldsymbol{x}_\mathrm{p}$,$\boldsymbol{v}_\mathrm{p}$ and ϕ_g are the position, velocity and gravitational potential respectively. This could be solved by direct summation of the interaction between each particle, but this approach is extremely numerically expensive, scaling as N^2 with the number of particles, N.

There is also the added complication that not all matter in the Universe is collisionless. Gas contributes significantly to the gravitational potential and is both compressible and subject to pressure forces. Its evolution is described by the Navier-Stokes equations

$$\begin{aligned} \frac{\partial\rho}{\partial t} + \boldsymbol{\nabla}\cdot(\rho\boldsymbol{u}) &= 0 \\ \frac{\partial(\rho\boldsymbol{u})}{\partial t} + \boldsymbol{\nabla}\cdot(\rho\boldsymbol{u}\otimes\boldsymbol{u}) &= -\boldsymbol{\nabla}P + \boldsymbol{\nabla}\cdot\boldsymbol{\sigma} - \rho\boldsymbol{\nabla}\phi_g \\ \frac{\partial E}{\partial t} + \boldsymbol{\nabla}\cdot(\boldsymbol{u}(E+P)) &= \boldsymbol{\sigma}\cdot\boldsymbol{\nabla}u \end{aligned} \tag{3.2}$$

where, $E, \rho, \boldsymbol{u}, P, \boldsymbol{\sigma}$ and ϕ_g are the energy, density, velocity, pressure, viscosity and gravitational potential respectively. This reduces to the Euler equations in the absence of viscosity, as $\sigma \to 0$.

To solve this set of equations, the distribution of gas is discretised on a Cartesian grid. Evolving it requires the fluid to be advected across cell boundaries, which in RAMSES is done by solving the local Riemann problem: each boundary is treated as a discontinuity, and the propagation of states is used to solve for fluxes across cell boundaries. To do this numerically, RAMSES uses the HLLC Riemann solver (Harten, 1983; Toro et al., 1994), which evaluates half timestep fluxes in a so-called Godunov scheme (Godunov, 1962). From the fluxes, the cell states are reconstructed using piecewise linear approximations at half timesteps, using a Monotone Upstream-centred Schemes for Conservation Law Limiter (MUSCL), in combination with a MinMod slope limiter to minimise undesirable oscillatory features in regions of steep gradients. A range of functions could in theory be used to reconstruct the states, where higher order schemes are less diffusive, but also less numerically stable.

To calculate the overall distribution of matter, and the resulting gravitational potential, the information contained in particles and the information stored on the grid have to be combined. In RAMSES, this is done by assigning the mass stored in particles to the grid, using a so-called cloud-in-cell (CIC) scheme, where the mass of each individual particle is uniformly distributed across the nearest cells using the assignment function

$$W(x - x_\mathrm{p}) = \begin{cases} 1 - |x - x_\mathrm{p}|/\Delta x & |x - x_\mathrm{p}| \leq \Delta x \\ 0 & \text{otherwise.} \end{cases} \tag{3.3}$$

Here, x is the position of a cell, x_p is the position of a particle and Δx is the cell spacing. Once the total (dark matter, stars, gas and black holes) distribution of mass, ρ, has been computed by this method, one can solve the Poisson equation

$$\nabla^2 \phi_\mathrm{g} = 4\pi G \rho \tag{3.4}$$

for the gravitational potential ϕ_g, which in turn can be used to calculate the acceleration on particles. To solve this partial differential equation, RAMSES uses a multi-grid approach (Guillet & Teyssier, 2011).

A timestep has to be set to integrate the positions of particles and the hydrodynamical properties of the gas on the grid. In RAMSES, this timestep is chosen such that each cell can be updated using only information contained in its direct neighbours. In practice this means the timestep is set so that no particle can cross more than a single cell per step, no sound wave can cross more than a single cell per step (the Courant condition), and the timestep does not not exceed the free-fall time of the gas in any cell. For cosmological simulations, the scale factor should not grow more than 10% during a single timestep. The timestep is set such that all conditions are fulfilled for the smallest cells in the simulation. When working with a non-uniform grid, RAMSES provides the option of sub-cycling, where larger cells are updated on a timestep scaled by their cell size, to reduce numerical costs.

One common problem in astrophysical simulations is the range of length scales covered in a single simulation, which can be vast, making it prohibitively expensive to fill the volume with a uniform grid. Instead, one can reduce numerical costs for a given scientific question by concentrating the smallest cells in regions of interest, and filling the remaining volume with larger cells. In RAMSES, this can be done by pre-defining a set of grids at varying resolutions, which are then kept fixed in position throughout the simulation. This approach lends itself well to isolated simulations with a fixed geometry.

More complex setups, such as galaxy evolution simulations, see the regions of interest move and evolve, which requires an adaptive criteria to update refinement as needed. First, the entire box is refined uniformly to a minimum level l_{min}. RAMSES then takes a Cartesian, cell-based refinement approach to AMR (Khokhlov, 1998). Cells flagged for refinement are each split into 8 cubic "child cells", i.e. each new level of refinement reduces the edge length of a cell by a factor 2. The new cells are then appended to its parent's OctTree, a data structure used to handle refinement information within the simulation. Each cell retains information on its parent's cell, 6 neighbouring parent cells, 8 child cells, previous and next grid cells. Cells are refined until all refinement criteria are met, up to a maximum refinement level, l_{max}. Cells without further refinement are referred to as "leaf cells" and make up the grid covering the box.

In order to determine whether a cell should be refined, RAMSES commonly uses a number of different criteria, with the option to define custom ones for the specific context of a given simulation. First and foremost is a quasi-Lagrangian criterion, where a cell is refined if it's mass exceeds a threshold mass. For this purpose, baryons and dark matter are handled separately. A cell is refined when $M_{\rm DM} > n_{\rm refine} M_{\rm DM,init}$ or when $M_{\rm bar} = M_* + M_{\rm gas} + M_{\rm BH} > n_{\rm refine} M_{\rm bar,init}$ where $n_{\rm refine}$ is commonly set to 8. $M_{\rm DM,init}$ and $M_{\rm bar,init}$ are the average initial dark matter and average baryonic cell masses respectively. In order to resolve the collapse length scale at all times, RAMSES also refines such that the local Jeans length[1], $l_{\rm J}$, satisfies

$$l_{\rm J} = \sqrt{\frac{15 k_B T}{4\pi G \rho m_p}} > 4\Delta x \tag{3.5}$$

where k_B is the Boltzmann constant, T is the temperature and m_p the mass of a proton. The criteria can be combined as the user sees fit. In order to avoid sudden spatial transitions in refinement level, cells with neighbours that are 2 or more levels higher refined are also flagged for refinement, until the level difference between neighbours is only a single level everywhere.

RAMSES is parallelised using the message passing interface (MPI) libraries. To distribute the information across processors, each processor handles a connected sub-volume of the simulation, identified using a section of the volume filling Peano-Hilbert Curve (Peano, 1980; Hilbert, 1891) threaded through the grid. Each processor also handles the appropriate particles for the sub-volume, as well as a passive copy of immediately adjacent cells (so called “ghost cells”) for use in neighbour calculations. The information is redistributed across processors with user-defined frequency.

3.3 Cooling and heating

Radiative gas cooling occurs via a range of different atomic and molecular processes. When a free electron collides with a bound electron, it transfers energy, which either knocks the bound electron to an excited state (collisional excitation) or knocks it free (collisional ionisation). The opposite happens during recombination when a free electron recombines with an ion and radiates away the kinetic energy of the electron as well as the recombination energy. Finally, during free-free emission, a free electron is accelerated by an ion, emitting a photon. The dominant cooling channel as well as the abundance of ions responsible for cooling are a function of temperature and the overall cooling efficiency therefore depends on the composition of the cooling gas. This information can be summarised in so called “cooling functions”, which act as look-up tables to determine the volumetric radiative cooling rate, Λ, for a given composition and temperature. The cooling functions assume local thermal equilibrium (Sutherland & Dopita, 1993).

Ideally, one would track the full chemical evolution of the cosmic gas but this is currently prohibitively computationally expensive. A common approach is to assume that the distribution of species is fixed, for example to the relative abundances of

---

[1] See section 3.5.1 for a derivation

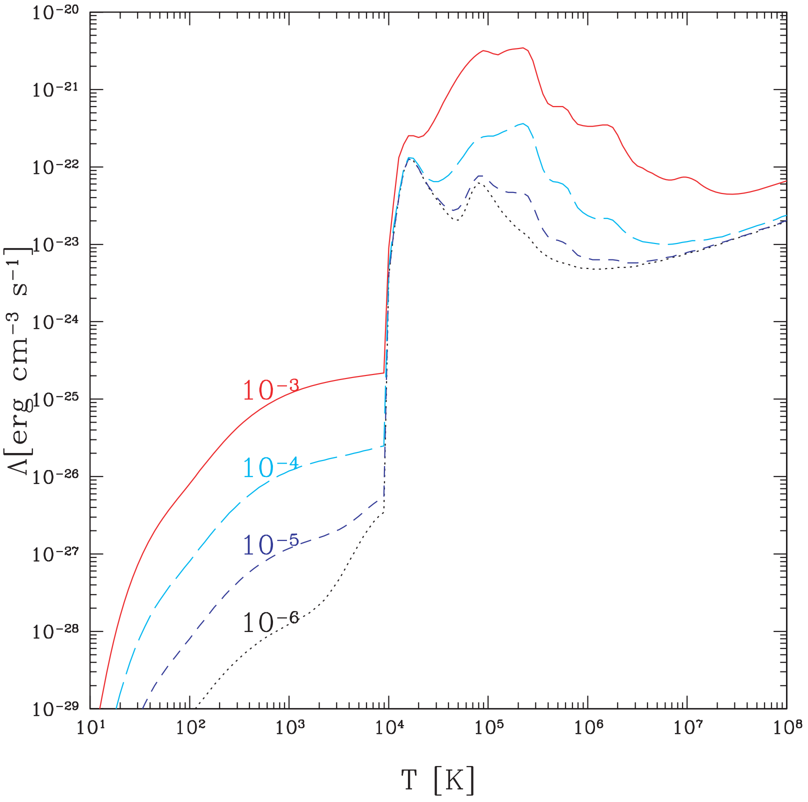

Figure 3.1: Cooling function including H, He, H_2, HD and metal line cooling for different metallicity values, with abundances appropriate for the early Universe (Reprint of Figure 4 from Maio et al., 2007).

elements found in the sun. This reduces the number of free parameters to two, the temperature and the overall metallicity[2]. The relative abundances of different metals change with redshift (Galli & Palla, 2013). One example of such a cooling function can be seen in Figure 3.1, which includes hydrogen, helium, metal lines and H_2, and shows how even a small difference in metallicity, here measured in units of solar metallicity, significantly affects the cooling rate. The specifications of each cooling model used in the simulations presented here are described in the relevant chapter.

The gas is assumed to be ideal and evolves according to an equation of state,

$$P = \rho(\gamma - 1)u \tag{3.6}$$

where P is the pressure, ρ is the density, γ is the adiabatic index and $u = U/M$ is the specific internal energy of the gas (see Equation 3.12). For an adiabatic processes, $\gamma = 5/3$ for a monoatomic gas or $\gamma = 7/5$ for a diatomic gas.

At a redshift around $z = 10$, the Universe went through a period of reionisation. As the propagation of photons is not followed in hydrodynamical simulations, the impact of reionisation is captured using photon heating from a redshift dependent uniform background, following Haardt & Madau (1996).

3.4 Star formation and stellar feedback

Stars form in dense molecular clouds, which leads to a correlation between a galaxy's gas surface density, Σ_{gas}, and the star formation surface density, Σ_{SFR}, the so called Kennicutt–Schmidt law (Schmidt, 1959; Kennicutt, Jr., 1998):

$$\Sigma_{\mathrm{SFR}} = (2.5 \pm 0.7) \times 10^{-4} \left(\frac{\Sigma_{\mathrm{gas}}}{\mathrm{M_{\odot} pc^{-2}}}\right)^{1.4\pm0.5} \mathrm{M_{\odot} yr^{-1} kpc^{-2}}. \tag{3.7}$$

Extrapolating to three dimensions (Krumholz & Tan, 2007), one can link the star formation density to the density of the gas:

$$\dot{\rho}_* = \epsilon_* \rho / t_{\mathrm{ff}}, \tag{3.8}$$

where $\dot{\rho}_*$ is the star formation rate density, $\epsilon_* = 0.02$ a (constant) star formation efficiency, ρ the gas density and t_{ff} the local free fall time of the gas. When unable to resolve the formation of individual stars in a simulation, the stellar population of a galaxy can be approximated using larger star particles that each represent a

---

[2] "Metal" is here taken to have its common meaning in astrophysics: any element except Hydrogen and Helium

cluster of stars with an analytic or empirical initial mass function (IMF), such as those suggested by Chabrier (2003) and Salpeter (1955).

In the model used in this thesis, stars are formed when the gas number density exceeds a gas density threshold, ρ_0 (see individual chapters for specific values), with a star formation rate calculated according to Equation 3.8. Star particles are generated using a Poisson random process (Rasera & Teyssier, 2006; Dubois & Teyssier, 2008) with a stellar mass resolution set by the density threshold and the minimum spatial resolution of the simulation. As the star particles are treated as point masses, two particles with identical initial properties (mass, location and velocity) will follow identical trajectories throughout the simulation. To reduce computational costs, star particles are therefore spawned with integer multiples of the minimum stellar resolution. Each individual star particle represents a population of stars with a Salpeter (1955) IMF, using a low and high mass cutoffs of $0.1\mathrm{M}_\odot$ and $100\mathrm{M}_\odot$ respectively.

In the work presented in this thesis, only the HORIZON simulation (Chapter 7) includes a prescription for stellar feedback. Stellar feedback is a collective term that is here meant to include all channels through which stars return energy, mass, metals and momentum to their surroundings, both during and at the end of their lifetime. In an effort to account for stellar feedback as comprehensively as possible, the (subgrid) model implemented in HORIZON includes stellar winds, Type II and Type Ia supernovae. Mechanical feedback energies from Type II supernovae and stellar winds are computed using STARBURST99 (Leitherer et al., 2010, 1999). Specifically, it uses a Padova model (Girardi et al., 2000) with thermally pulsating asymptotic branch stars (Vassiliadis & Wood, 1993), and stellar winds are calculated as in Leitherer et al. (1992). The frequency of Type I supernovae (SNe) is estimated from Matteucci & Greggio (1986), assuming a binary fraction of 5%. More detailed discussions of the stellar feedback model can be found in Dubois & Teyssier (2008); Kimm et al. (2015) and Rosdahl et al. (2017).

In the HORIZON simulations in Chapter 7, gas with a density above the star formation threshold ρ_0 evolves according to a polytropic equation of state, $T = T_0(\rho/\rho_0)^{\gamma-1}$ where $\gamma = 4/3$. This prevents numerical fragmentation by adding non-thermal pressure support to ensure that the Jeans length, l_J, remains resolved (see Equation 3.5 in Section 3.2 for details). It has been argued that this also mimics the effect of stellar heating by young stars (Springel & Hernquist, 2003).

3.5 Supermassive black holes

Black holes in RAMSES are modelled as Lagrangian sink particles (Krumholz et al., 2004; Dubois et al., 2010), i.e. each black hole is a massive particle that moves over the grid, removing gas from a small accretion region centred on its current location (see Bate et al., 1995; Springel et al., 2005, for the equivalent in Smooth Particle Hydrodynamics codes). Each particle has a mass, position, velocity and birth epoch. Its dynamical evolution is handled by the CIC scheme and multigrid approach described in Section 3.2, but a series of sub-grid algorithms are employed to handle different aspects of black hole physics, namely black hole formation, growth and feedback. Formation sites are identified using density criteria or structure finding algorithms. Accretion proceeds via the Bondi-Hoyle-Lyttleton (BHL) formula (see Section 3.5.2), using local mass-weighted quantities computed from cell values in the accretion region. Two different feedback models are available to return energy to the surrounding gas.

3.5.1 Seed formation

Before a black hole can be created, suitable formation sites need to be identified. For simple simulations, like the ones in Chapter 4, the initial black hole mass, location and velocity are specified as part of the input parameters and a sink particle with the appropriate properties is inserted before time integration begins. For more complex galaxy evolution simulations, physically motivated black hole formation sites need to be identified on the fly. There are a range of ways to accomplish this. Physically, black holes are formed as the end-product of gravitational collapse, possibly via the intermediate step of a star (see Section 2.2.6). Therefore identifying suitable regions for black hole formation in simulations requires the identification of connected regions of gas whose gravitational attraction is able to overcome its internal support.

A self-gravitating system will adhere to the virial theorem, which states that:

$$\frac{1}{2}\left\langle\frac{d^2I}{dt^2}\right\rangle = 2 < U > +W, \tag{3.9}$$

where I is the scalar moment of inertia of the system, U is the internal energy of the system and W is the total potential energy. Triangular brackets denote the time average. For a system in equilibrium, in which the internal energy provides support against gravitational collapse, $\left\langle\frac{d^2I}{dt^2}\right\rangle = 0$, so the total potential energy of the system can be related to the total internal energy via:

$$W = -2U. \tag{3.10}$$

If the internal energy is smaller than $-\frac{1}{2}W$, then internal forces are unable to offset gravitational attraction and the system collapses. If it is larger, then the system expands.

Following Caroll & Ostlie (2008) the total potential energy of a self-gravitating spherical cloud of gas of uniform density ρ, mass $M = \frac{4\pi}{3}\rho r^3$ and radius r is

$$W = -\frac{3}{5}\frac{GM^2}{r}. \tag{3.11}$$

Assuming the internal energy is entirely thermal, and the gas is ideal and monoatomic,

$$U = \frac{3Nk_bT}{2}, \tag{3.12}$$

where $N = \frac{M}{m_p}$ is the number of particles and T is the temperature of the gas.

Substituting the expression for total internal energy (Equation 3.12) and total potential energy (Equation 3.11) into the virial theorem yields a maximum diameter for a cloud of a given density and temperature, the so called Jeans length, l_J:

$$l_\mathrm{J} = \sqrt{\frac{15k_BT}{4\pi G\rho m_p}}, \tag{3.13}$$

which can also be recast as a maximum mass, the Jeans mass, M_J

$$M_\mathrm{J} = \frac{4\pi}{3}\rho\Big(\frac{l_\mathrm{J}}{2}\Big)^3 = \frac{1}{6\sqrt{\pi\rho}}\Big(\frac{15k_BT}{Gm_p}\Big)^{3/2}. \tag{3.14}$$

Any cloud of uniform density, ρ, and temperature, T, either more massive or more extended, will collapse under its own gravity.

A computationally cheap way to identify black hole formation sites, that lends itself well to large scale cosmological simulations like the HORIZON simulations in Chapter 7, is a simple density criterion similar to the one for star formation. Here, a black hole with a predefined seed mass M_seed is inserted when a cell exceeds a user-defined threshold density, ρ_BH, measured in co-moving coordinates so it is unaffected by cosmological expansion (Krumholz et al., 2004; Dubois et al., 2012a). Additional checks can be added to further narrow down the choice of location. For example, the HORIZON simulation also requires that the gas is Jeans unstable and located a minimum user-defined distance from an existing black hole (Dubois et al., 2014b). The latter criteria avoids spawning a series of black holes in quick succession from the same cell.

To test the gravitational stability of the gas in a cell, one can evaluate Equation 3.14. Assuming that the mass in the cell can be approximated as a cloud of uniform density, if more resolution was available the cell would collapse if $M_{\rm cell} > M_{\rm J}(\rho_{\rm cell}, T_{\rm cell})$. Therefore, given the information available, this is a way to identify suitable black hole formation sites. Once a cell has been flagged for black hole formation, its gas mass is reduced by the seed mass $M_{\rm seed}$, and a Lagrangian particle of equal mass is inserted in the centre of this cell, inheriting the velocity vector of the parent cell. Mass and momentum are conserved.

A more physically motivated method that is computationally more expensive involves running a structure finding algorithm on the fly to identify gravitationally bound, collapsing clouds. In RAMSES this is done using the parallel hierarchical watershed (PHEW) algorithm (Bleuler & Teyssier, 2014; Bleuler et al., 2015). All cells above a user-defined density threshold (see Figure 3.2, panel *i*) are sorted in order of density and then assigned to their densest neighbour, with local density maxima flagged with a peak identifier. This creates a watershed segmentation, where saddle points act as separators of geometrically connected density peaks (see Figure 3.2, panel *ii*). To identify only the most prominent features, clumps whose peak to saddle point ratio is smaller than 1.5 are merged to their neighbour, through their highest saddle point (see Figure 3.2, panel *iii*). Isolated clumps failing this relevance check are discarded. If requested, the remaining peaks can be merged into composite clumps using a user-defined saddle threshold (see Figure 3.2, panel *iv*).

The densest cells of the remaining peaks are then considered potential black hole formation locations. For each location, a sphere of radius $4 \times \Delta x_{min}$ around the densest cell of the peak, the size of the eventual black holes accretion region, is investigated further for its suitability as a black hole formation site. If it is found to be collapsing, able to overcome any sources of internal pressure of the gas and is a minimum distance from existing sinks, it is flagged for sink formation. Collapse under self gravity is evaluated using Equation 3.9, requiring the centre of mass scalar moment of inertia of the sphere under investigation to fulfil $\left\langle \frac{d^2 I_{\rm cm}}{dt^2} \right\rangle < 0$. Collapse to a point-like object occurs when all three timescales associated with the eigenvectors of the tensor moment of inertia, $I_{\rm cm}$, are negative. Once a peak has passed all three requirements, additional user-defined checks can be introduced to further narrow down the sink formation location. Otherwise, a black hole sink particle is inserted in the densest cell, inheriting its properties as described above.

3.5.2 Accretion

One problem for hydrodynamical simulations involving black holes, particularly ones covering a large range of scales, is to determine how much of the gas, present on

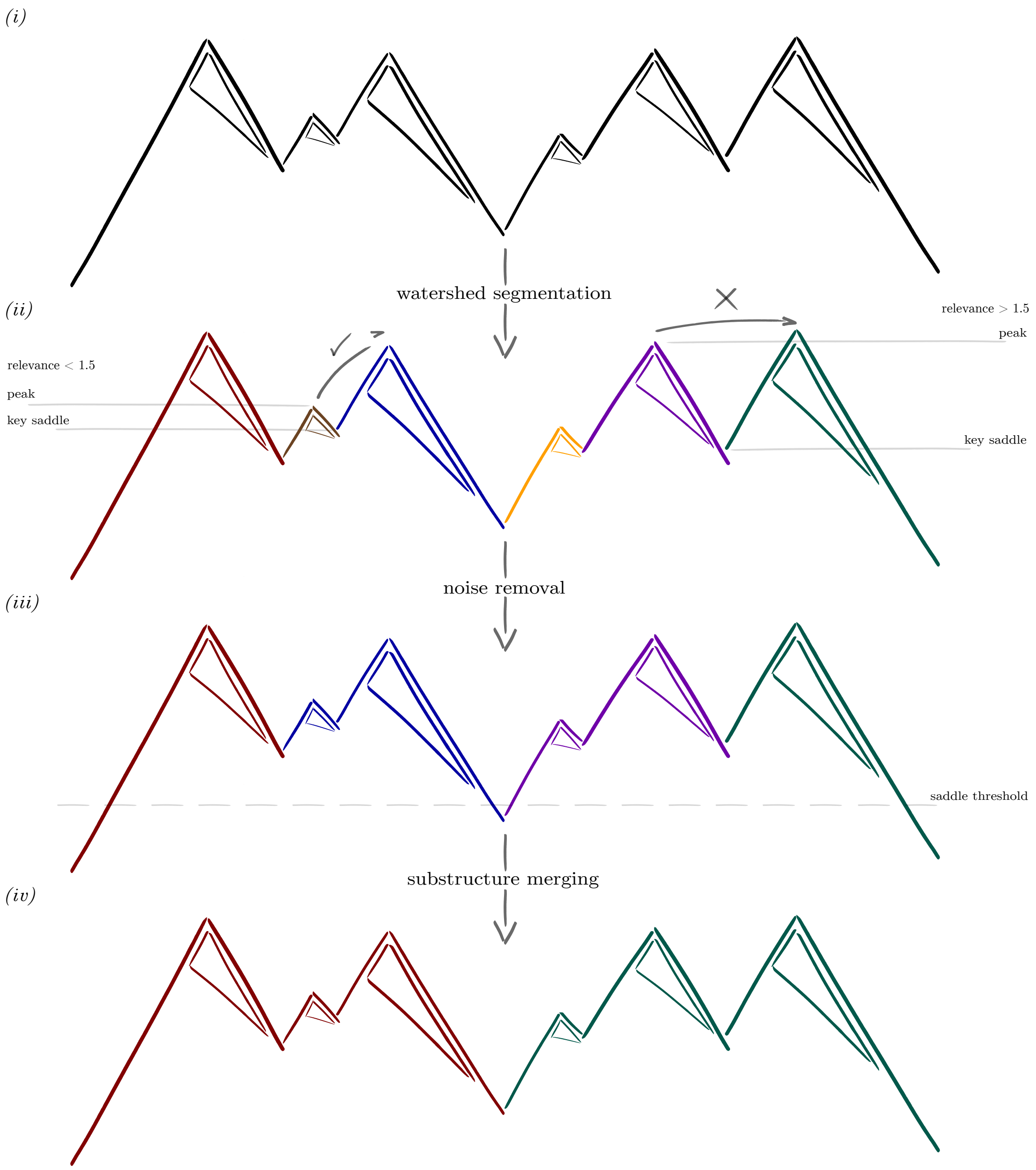

Figure 3.2: The structure finding algorithm PHEW uses a watershed segmentation, followed by a series of hierarchical mergers, to identify connected clumps of gas, which can then be investigated for their potential suitability as black hole formation locations (Reprint of Figure 1 from Bleuler et al., 2015).

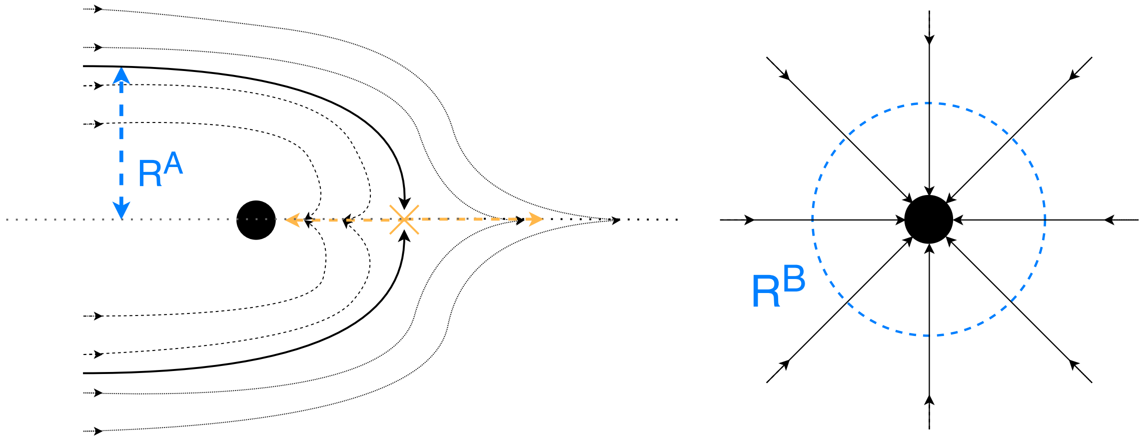

Figure 3.3: Accretion onto a point mass from a homogeneous medium moving at constant velocity relative to the sink (left) or at rest (right), based on an analysis by Hoyle & Lyttleton (1939) and Bondi & Hoyle (1944). The black hole is marked by the filled black circle and the relevant scale radius of each problem is annotated in blue. The stagnation point is marked in yellow.

scales far beyond the influence of the black hole itself, gets accreted by the black hole. As information on gas properties in the vicinity of the black hole is usually limited by resolution, one has to turn to analytic or empirical models. One such analysis was performed by Hoyle & Lyttleton (1939) and Bondi & Hoyle (1944), who studied the accretion onto a point mass from a uniform background medium under the gravitational influence of the point mass. Their combined work forms the basis of many modern sub-grid accretion models for black holes, including the one used in this work, and it is therefore explored here in some detail.

3.5.2.1 Bondi-Hoyle-Lyttleton accretion

Hoyle & Lyttleton considered the case of a homogeneous distribution of gas, of constant density, ρ_∞, and sound speed, $c_{s,\infty}$, at infinity, being perturbed by a point mass of mass $M_{\rm sink}$, moving at a constant velocity, v_∞, relative to the gas at infinity. Infinity is a loosely defined term, here taken to mean far from the sink, where the properties of the gas are approximately unperturbed by the gravitational potential of the sink. Thinking about the fate of streamlines of gas bending in the sink's gravitational potential, Hoyle & Lyttleton discovered that there is a point downstream of the sink at which a test particle of gas will have exactly zero velocity along the axis of symmetry of the problem, the so called stagnation point. The problem is illustrated in the left panel of Figure 3.3, with the stagnation point marked in yellow, and the arrested streamline in bold.

This streamline has an impact parameter, i.e. a maximum radial separation from the axis of symmetry, of

$$R^A_\infty = \frac{2GM_{\rm sink}}{v^2_\infty}, \tag{3.15}$$

which is also called the accretion radius. Any gas on a streamline intersecting the axis of symmetry between the sink and the stagnation point will be accreted, while everything else will flow downstream (see left panel of Figure 3.3). One can therefore calculate the mass flow through a circle with radius R^A, far upstream of the sink, to determine the accretion rate onto the sink. This analysis returns the Hoyle-Lyttleton accretion rate:

$$\dot{M}^{\rm HL} = \frac{4\pi G^2 M^2_{\rm sink}\rho_\infty}{v^3_\infty} \tag{3.16}$$

Thinking instead about a point mass at rest relative to the gas at infinity (right hand panel in Figure 3.3), Bondi & Hoyle (1944) discovered that this problem also has a characteristic scale radius, the Bondi radius,

$$R^{\rm B}_\infty = \frac{GM_{\rm sink}}{c^2_{s,\infty}}, \tag{3.17}$$

defining the distance out to which the pressure in the gas approximately balances the gravitational attraction from the sink. Flow outside R^B_∞ is subsonic and the density is almost uniform. Inside R^B_∞, the flow becomes supersonic and the gas is compressed, approaching a free-fall solution. Neither analysis included the effect of cooling or self-gravity of the gas and the analysis by Hoyle & Lyttleton (1939) also disregarded pressure effects.

Based on the similarities between the two solutions, Bondi (1952) proposed an interpolation formula for the accretion rate onto the sink, intended to cover the full range of Mach numbers, $\mathcal{M}$, known as the Bondi-Hoyle-Lyttleton (BHL) accretion rate:

$$\dot{M}^{\rm BHL} = \frac{4\pi G^2 M^2_{\rm sink}\rho_\infty}{(c^2_{s,\infty} + v^2_\infty)^{3/2}}. \tag{3.18}$$

Which of the two scale radii is relevant in a given situation depends on the Mach number at infinity, $\mathcal{M}_\infty = \frac{v_\infty}{c_{s,\infty}}$, as follows:

$$R^{\rm S}_\infty = \begin{cases} R^{\rm B}_\infty = \frac{GM_{\rm sink}}{c^2_{s,\infty}} & \text{if } \mathcal{M}_\infty \leq 1 \\ R^{\rm A}_\infty = \frac{2GM_{\rm sink}}{v^2_\infty} & \text{if } \mathcal{M}_\infty > 1. \end{cases} \tag{3.19}$$

It can be approximated using an interpolated scale radius,

$$R^{\rm BHL}_\infty = \frac{GM_{\rm sink}}{v^2_\infty + c^2_\infty}, \tag{3.20}$$

which can be used to assess whether a quantity, α^i, measured at a distance r^i from the sink, is "at infinity", where $r^i >> R_\infty^{\mathrm{BHL}}$, or close to the sink, where $r^i \sim R_\infty^{\mathrm{BHL}}$.

While based on comparatively simple assumptions and neglecting much of the gas physics involved in the problem, numerical experiments have shown that the BHL accretion rate provides a good fit to numerical experiments (Shima et al., 1985; Ruffert, 1995a,b; Edgar, 2004). It has become the standard sub-grid model for accretion onto black holes for simulations in which the smallest resolution element is much larger than the scale radius of the problem, i.e. where $\Delta x_{min} >> R^{\mathrm{BHL}}$, as the information available in such simulations provide a good approximation to the assumptions of the BHL analysis. It does however have some notable shortcomings, including its complete disregard for the angular momentum of the infalling gas and the lack of both cooling and pressure effects.

3.5.2.2 Numerical accretion models

When employing a BHL based model in a galaxy evolution simulation, one major difficulty lies in correctly estimating the input parameters: ρ_∞, v_∞ and $c_{s,\infty}$. Limited resolution will artificially depress ρ and artificially increase c_s by employing a cooling floor. Booth & Schaye (2009) therefore argue for a boost factor in the accretion rate, to compensate for unresolved cold, dense gas phases. Other attempts have been made to include some of the missing physics in a boost factor. Pelupessy et al. (2007) separately account for the cold and hot phases of the gas, and Vogelsberger et al. (2013) vary the boost factor according to a pressure criteria to avoid overestimating the accretion rate when no star-forming gas is present in the vicinity of the black hole. DeGraf et al. (2017) use a stochastic boost factor to account for cloud formation on scales below those resolved in the simulation. To compensate for the lack of angular momentum in the BHL model, some schemes scale down the BHL accretion rate according to the vorticity of the accreted gas (Curtis & Sijacki, 2016b, based on work by Krumholz et al. (2004)), or scale the BHL accretion rate by a factor involving the viscous timescale of the accreted gas to account for angular momentum (Rosas-Guevara et al., 2015; Schaye et al., 2014). Other schemes remove gas from the grid according to the BHL rate, but store it temporarily in an "accretion disc", using the viscous timescale of the material to calculate the true accretion rate from this reservoir (Power et al., 2011; Dubois et al., 2012a).

Several efforts have been made to develop accretion algorithms that use large-scale properties of the host galaxy to calculate the accretion rate onto the black hole. Hopkins & Quataert (2011) developed a model in which the accretion rate onto the

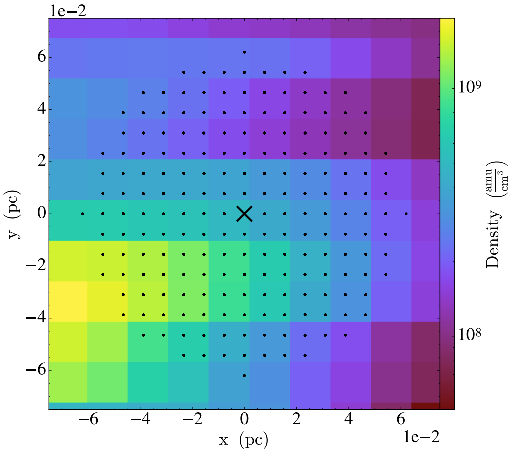

Figure 3.4: Simulation snapshot showing the distribution of cloud particles (black dots) around the sink (black cross), plotted over a density distribution from Chapter 5.

sink is calculated from gravitational torques within the galaxy, and DeBuhr et al. (2010) calculate the accretion rate onto the sink from the mean surface density of the galaxy, the angular rotation frequency and a free viscosity parameter. Hobbs et al. (2011) use a "ballistic" accretion model based on the velocity dispersion of the gas in the bulge.

All of these models have in common that they estimate the mass accretion onto the black hole from information on scales far above the gravitational influence of the black hole itself. While this does represent the current best guess, and has led to many insights into the evolution of black holes, it can only ever be a highly simplistic approximation, particularly for such a non-linear problem as the evolution of gas in galaxies. Once the length scales of the black hole itself are included in a simulation, one will have to think carefully about the approximations made in sub-grid algorithms and their viability in the new regime. Chapter 4 presents work on this subject for both BHL accretion and the analytic drag force algorithm described in Section 3.5.4.

3.5.2.3 Bondi-Hoyle-Lyttleton accretion in RAMSES

The standard sink particle accretion algorithm in RAMSES uses the BHL formula for accretion. To calculate ρ_∞, v_∞ and $c_{s,\infty}$, RAMSES uses a cloud-particle system (Dubois et al., 2010; Teyssier et al., 2011), where a total of 2109 cloud particles are scattered around the sink with a spacing of $\Delta_{cloud} = \Delta x_{loc}/2$, filling a sphere with radius $r_{cloud} = 4 \times \Delta x_{loc}$, where Δx_{loc} is the local cell width (see Figure 3.4 for an example of a 2D slice through the cloud of particles). The cloud particles allow the sink to sample its surrounding volume, and smooth the sudden discontinuity in properties when passing from one cell into the next.

Each cloud particle samples the properties of its host cell. The mass weighted average gas properties for the BHL formula are then calculated by summing over the cloud particles according to

$$x_\bullet = \frac{\sum x_i m_i \alpha_i}{\sum \alpha_i \sum m_i}, \tag{3.21}$$

where the contribution of each cloud particle i is weighted by a Gaussian kernel,

$$\alpha_i = \exp\Big(- \Big(\frac{r_i}{r_{kernel}} \Big)^2 \Big). \tag{3.22}$$

r_i is the cloud particle's distance from the sink and

$$r_{\text{kernel}} = \begin{cases} \Delta x_{loc}/4 & \text{if } r^{\text{BHL}}_{\text{host}} < \Delta x_{loc}/4 \\ 2\Delta x_{loc} & \text{if } r^{\text{BHL}}_{\text{host}} > 2\Delta x_{loc} \\ r^{\text{BHL}}_{\text{host}} & \text{otherwise,} \end{cases} \tag{3.23}$$

and

$$r^{\text{BHL}}_{\text{host}} = \frac{GM_{\text{sink}}}{v^2_{\text{host}} + c^2_{\text{host}}}, \tag{3.24}$$

is the interpolated scale radius in Equation 3.20. v_{host} and c_{host} are the relative velocity and sound speed of the sink's host cell respectively.

Gas mass is removed from local cells, and added to the sink particle by looping over the cloud particles and removing mass from each host cell according to

$$\Delta m_i^{\text{cell}} = \frac{\dot{M}^{\text{BHL}}_\bullet dt}{\sum \alpha_i} \times \alpha_i \tag{3.25}$$

where dt is the local timestep. $\dot{M}^{\text{BHL}}_\bullet = \dot{M}^{\text{BHL}}(\rho_\bullet, v_\bullet, c_\bullet)$ is the accretion rate calculated on the fly according to the BHL formula in Equation3.18, evaluated using local average mass weighted quantities.

To avoid creating numerical instabilities by removing too much gas mass from a single cell, the total mass accreted per timestep, per cloud particle, is capped at 75% of the cell's gas mass. This criteria is commonly employed in RAMSES and ensures that no single cell is emptied when the sink particle moves across the grid. In simulations where the sink is forced to remain in a particular host cell (Chapter 4), a density floor is introduced that ensures $\rho > \rho_{min} = 10^{-5}\ \mathrm{H\,cm^{-3}}$ to avoid numerical problems. The results are insensitive to ρ_{min}, as long as $\rho_{min} << \rho_\infty$.

3.5.3 Feedback modes of active galactic nuclei

As described in detail in Section 2.2.3, active galactic nuclei (AGN) interact with their environment by emitting radiation, accelerating winds and driving jets. While these interactions are thought to be the source of their coevolution with galaxies, they are challenging to include in the type of large scale cosmological simulations used to study this coevolution, because they originate on scales far below the resolution limit.

Assuming that a fraction, ϵ_{f}, of the accreted energy is liberated as feedback, the total energy produced by an AGN is

$$\dot{E}_{\mathrm{AGN}} = \epsilon_{\mathrm{f}}\epsilon_{\mathrm{r}}\dot{M}_{\mathrm{BH}}c^2, \tag{3.26}$$

where ϵ_{r} is the radiative efficiency, c is the speed of light and $\dot{M}_{BH}$ is the accretion rate onto the sink. The HORIZON simulation in Chapter 7 use two feedback modes, modelled on the observed radiative and jet mode respectively (Dubois et al., 2012b).

At high accretion rates, i.e. for Eddington ratios $\chi = \ 0.01$, the quasar mode (Booth & Schaye, 2009; Teyssier et al., 2011) deposits thermal energy isotropically into a sphere of radius Δx, centred on the black hole. This energy is deposited with an efficiency of $\epsilon_{\mathrm{f}} = 0.15$, a free parameter derived by tuning the results to the $M_{BH} - M_{\sigma}$ relation at $z = 0$.

At low accretion rates, $\chi \leq 0.01$, the AGN transitions to a radio mode (Dubois et al., 2010) and deposits kinetic energy into bipolar outflows with jet/wind velocities of 10^4 kms^{-1}, along an axis aligned with the angular momentum of the accreted material following the model of Omma et al. (2004). The total rate of energy deposited is given by Equation 3.26, albeit using a higher efficiency of $\epsilon_{\mathrm{f}} = 1$.

3.5.4 Dynamical friction due to a gaseous background

One way in which the black hole interacts with gas in its vicinity is through dynamical friction. The relative velocity between the two creates an overdensity downstream of the black hole, and the enhanced gravitational pull that ensues acts as a drag force. This process transfers momentum from the sink to the gas — contrary to accretion which transfers momentum from the gas to the black hole — thus boosting the tendency of the sink to stay attached to overdense local gaseous structures. When this drag is unresolved, black holes often dynamically decouple from the gas, leading in the worst cases to their ejection from the galaxy disc. This has been a recurring problem in galaxy simulations including supermassive black holes (Sijacki et al., 2007; Volonteri et al., 2016; Biernacki et al., 2017).

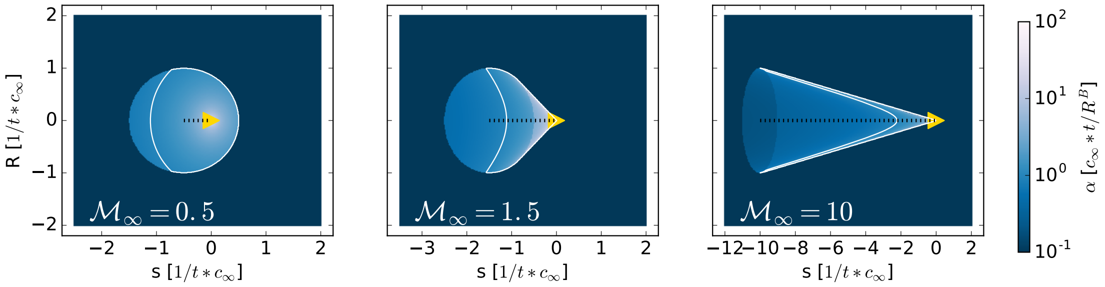

Figure 3.5: Gravitational wake of point mass moving through a uniform density medium, as calculated analytically by Ostriker (1999). The sink is moving to the right and is marked by the yellow triangle. The black dotted line shows the path of the sink during the buildup of the wake. The contour in white shows an overdensity of $\alpha = 1$.

Ostriker (1999) calculated the over-density distribution in the wake of a massive point source, $M_{\rm sink}$, using time-dependent linear perturbation theory based on work by Just & Kegel (1990). They found that the finite-time perturbation takes the form of “a loaded ice-cream cone” downstream of the sink (see Figure 3.5 for examples at a variety of mach numbers). The resulting drag force depends on the same gas properties at “infinity” as the BHL accretion rate, including the Mach number, $\mathcal{M}_\infty = \frac{v_\infty}{c_{s,\infty}}$. The force takes the form

$$\mathbf{F}^D_\infty = -I_{drag} \times \frac{4\pi (G M_{\rm sink})^2 \rho_\infty}{v^2_\infty} \hat{\mathbf{v}}_\infty = F^D_\infty \hat{\mathbf{v}}_\infty, \tag{3.27}$$

using

$$I_{drag} = \begin{cases} \frac{1}{2}\ln\left(\frac{1+\mathcal{M}_\infty}{1-\mathcal{M}_\infty}\right) - \mathcal{M}_\infty & \text{if} \quad \mathcal{M}_\infty < 1 \\ \frac{1}{2}\ln\left(\mathcal{M}^2_\infty - 1\right) + \ln(\Lambda) & \text{if} \quad \mathcal{M}_\infty > 1, \end{cases} \tag{3.28}$$

where $\ln(\Lambda) = \ln\left(\frac{r_{max}}{r_{min}}\right)$ is the Coulomb logarithm and $\hat{\mathbf{v}}_\infty$ is the unit vector pointing in the direction of the relative velocity between the sink and the gas. $r_{max} = (c_{s,\infty} + v_\infty) \times t$ is the maximum distance information has travelled downstream of the sink and r_{min} is a smoothing length used to curtail the diverging density profile near the sink. I_{drag} diverges as $\mathcal{M} \to 1$, so the subsonic and the supersonic regime is joined using a linear interpolation for $0.98 < \mathcal{M} < 1.02$. The resulting magnitude of the force depends strongly on Mach number, as can be seen in Figure 3.6. It peaks around $\mathcal{M}_\infty \approx 1.1$, depending on $\ln(\Lambda)$, and tails off for both $\mathcal{M} \to 0$, where the density perturbation becomes symmetric and $\mathcal{M} \to \infty$, where the opening angle of the cone goes to zero.

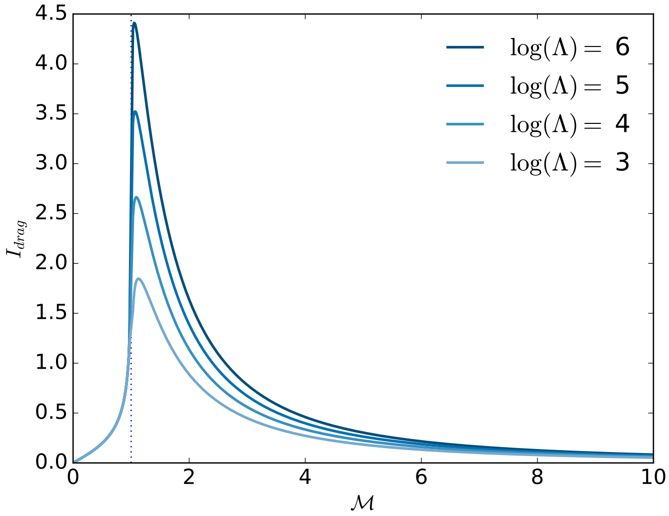

Figure 3.6: Analytic drag force on the sink due to dynamical friction according to the model by Ostriker (1999), for different coulomb logarithms.

RAMSES includes a sub-grid model for this force where the analytic formula in Equation 3.27 is evaluated using the same mass-weighted, kernel-weighted quantities used in the BHL accretion, $\rho_\bullet$, $c_{s,\bullet}$ and $v_\bullet$, and where $\ln(\Lambda) = 3.2$ based on work by Chapon et al. (2013). The choice for $\ln(\Lambda)$ is re-examined in the next chapter, as the value from Chapon et al. (2013) is extracted from black hole merger simulations that do not include accretion.

3.6 Structure finding

To compare results to observations, individual galaxies and halos have to be identified in the continuous distribution of matter produced by the simulation. As defined in Section 2.1.1, a halo is a bound structure of DM, and a galaxy is a bound structure of gas, stars and stellar remnants. DM and stars are handled as collisionless particles in RAMSES, and halos and galaxies can therefore be identified by finding bound clumps of particles. This requires a structure finding algorithm. One such algorithm frequently used with RAMSES is ADAPTAHOP (Aubert et al., 2004; Tweed et al., 2009), which applies a similar methodology to the PHEW algorithm described in Section 3.5.1, but operates on particles instead of grid cells. Its functionality is here described for halos, identified in the distribution of DM particles, but it can also be applied to star particles to identify the location of galaxies.

ADAPTAHOP computes a density for each particle, using its n_{hop} closest neighbours, to calculate a density field. Halos are then defined to be connected structures above a global density threshold $\rho_t = n \times \bar{\rho}_{box}$, where $\bar{\rho}_{box}$ is the global average in the box. $n = 80$ is a choice based on the density of the outer regions of virialised halos as calculated in the spherical top hat collapse model. All halos containing less than a minimum number of particles, a free parameter, are discarded immediately. To identify substructure, the density field within each halo is decomposed into local maxima, connected by saddle points. The statistical relevance of each substructure is then assessed by evaluating $(\bar{\rho} - \rho_t)/\rho_t > f/\sqrt{N}$, where $\bar{\rho}$ is the average density within the substructure, N the number of particles it contains and f a free parameter. To be considered relevant, a (sub)halo must also fulfil $\rho_{max} > \alpha\bar{\rho}$ and $r_{halo} > f_e\bar{d}_{part}$ where α and f_e are free parameters, and ρ_{max}, r_{halo} and $\bar{d}_{part}$ are the local density maximum, the radius of the (sub)halo and the mean separation between particles respectively. Typical parameter choices are $\alpha = 1.0$, $f_e = 0.05$ and $f = 4.0$. The exact choice of parameters depends on the resolution of the simulation and whether one is analysing dark matter or star particles. ADAPTAHOP outputs a location, based on the densest particle in the halo, total mass, virial mass, largest radius, virial radius and centre of mass velocity for each (sub)halo, as well as information on which subhalo belongs to which parent halo.

Chapter 4

Bondi or not Bondi: the impact of resolution on accretion and drag force modelling for supermassive black holes

4.1 Introduction

Supermassive black holes (SMBHs) in cosmological or idealised galaxy simulations are a problem that is challenging to tackle numerically as it involves a vast range of length scales. It is currently impossible to adequately track the gas from where it originated all the way to the black hole in a single simulation. As discussed in detail in Section 3.5, in adaptive mesh refinement (AMR) particle-mesh codes such as RAMSES, SMBHs are typically modelled as "sink" particles (Krumholz et al., 2004; Dubois et al., 2010), and accretion usually proceeds at the Bondi-Hoyle-Lyttleton (BHL) rate using the analytic formula in Equation 3.18. In this approach, quantities measured "at infinity" should in principle be used to infer the mass accretion rate onto the sink.

Whilst a good starting point in situations with limited information, this approach has several notable shortcomings, particularly the question of where "infinity" is to be defined in a busy galaxy simulation. The analytic solution also does not account for the background gravitational potential of the host galaxy or the self-gravity of the gas, nor does it consider density or velocity gradients, angular momentum or gas effects such as pressure or instabilities (see Edgar, 2004, for a detailed review of BHL accretion). The BHL analytic solution assumes a uniform density and velocity of the gas at infinity, a situation that most closely resembles simulations where all characteristic lengthscales associated with the SMBH are much smaller than the local resolution and the gas reservoir per cell is large compared to the accreted mass. Indeed in this situation, "infinity" can reasonably be understood to mean the gas cells surrounding the sink, as the black hole's gravitational potential is unresolved and therefore has little impact on the local gas flow. This is particularly true in the absence of feedback, as the cold, dense phase of the interstellar medium (ISM) that is expected to feed the black hole is also under-resolved in low resolution simulations. However, while this has led some authors to introduce a simple boost factor to the Bondi rate to compensate for the lack of resolution (Booth & Schaye, 2009), Negri & Volonteri (2017) recently showed that the interplay between local gas density, accretion rate and resolution is much more subtle and becomes even more so in the presence of black hole feedback.

Computational advances, including the zoom-within-zoom algorithm presented in Chapter 5, now allows a much wider range of scales to be resolved in a single simulation, fast approaching the physical scale (Schwarschild or Kerr radius) of the black hole itself. In this chapter, the behaviour of the popular BHL accretion algorithm is investigated as the gravitational influence of the black hole becomes better resolved and the local cell mass becomes comparable to the mass accreted per time resolution element. In this respect, the work presented here is similar to Ruffert (1995a,b), who investigated the validity and convergence of the analytic BHL formula using a range of fixed spherical, inflow-only regions to represent the sink. However, the work in this chapter focuses less on the validity of the BHL analytic solution itself — even though it is discussed how the observed instabilities cause systematic deviations from the analytic solution — than on testing the impact of representing the black hole as a sink particle, as well as estimating local gas averages from partially or fully resolved density features within the accretion region of the sink.

The work in this chapter also builds on Krumholz et al. (2004), who first introduced the use of Lagrangian sink particles in grid codes, and Dubois et al. (2010) who implemented it in RAMSES. More specifically, presented here is the most thorough exploration of the BHL parameter space performed to date using such a model, both in Mach number and resolution, and for two gas adiabatic indices. This allows the sub-grid accretion algorithm behaviour to be tested in a variety of regimes commonly encountered in galaxy simulations, including two features only briefly mentioned in Krumholz et al. (2004): a change in accretion mode at sufficiently high resolution, and accretion becoming inefficient when the resolution approaches the scale radius of the black hole.

As the total mass accreted onto a SMBH in galaxy simulations depends not only on the amount of gas removed, but also on the ability of the black hole to stay attached to local high density structures, this chapter includes a detailed look at the sub-grid algorithm described in Section 3.5.4 that is used to account for the dynamical friction exerted by the gaseous gravitational wake behind the sink particle. The particular simulations presented here were performed with RAMSES, but the conclusions reached are widely applicable to sub-grid algorithms where the size of the accretion region, over which local gas properties are measured and from which gas is removed, decreases with increasing resolution.

The structure of this chapter is as follows. Section 4.2 presents the setup of the simulations. Section 4.3 investigates accretion for a range of Mach numbers and resolutions, including a detailed study of the Bondi problem (Section 4.3.1) and the Hoyle-Lyttleton problem (Section 4.3.2). Section 4.4 presents results for the impact of Mach number and resolution on dynamical friction, both resolved on the grid and calculated analytically. Conclusions can be found in Section 4.6. Appendix A briefly discusses a numerical issue with small accretors and uniform initial conditions encountered for the simulations presented in this chapter.

4.2 The simulations

The gas dynamics in the vicinity of the sink particle were studied using a series of 3D isolated boxes where the sink particle is embedded in a uniform gas flow. Each simulation is parametrised by a Mach number, $\mathcal{M}_\infty = v_\infty / c_{s,\infty}$, where $c_{s,\infty}$ and v_∞ are the sound speed and flow velocity respectively. As the sink is held fixed at the centre of the box, v_∞ is both the absolute flow velocity and the flow velocity relative to the sink. Resolution is measured as the number of cells, N, within the sink particle's scale radius, R^S_∞, so that $N = R^S_\infty / \Delta x_{\rm min}$, where $\Delta x_{\rm min}$ is the size of the smallest grid cell in the simulation.

This characteristic scale radius of each simulation is defined according to Equation 3.19 and depends on $\mathcal{M}_\infty$. Setting $G = c_{s,\infty} = \rho_\infty = 1$ for all simulations reduces the number of parameters of the problem. Under this assumption, $M_{\rm sink} = R^S_\infty / L_{\rm box}$, so the mass of the sink determines the relative size of the local scale radius to the size of the box. In order to reduce edge effects, $L_{\rm box} = 1000 \times R^S_\infty$, which leads to $M_{\rm sink} = 0.001$. The final parameter is the gas pressure, which is set to $P = 1/\gamma$ for all simulations. Simulations were performed for two values of γ: an adiabatic one, where $\gamma = 1.3334$, and a quasi-isothermal one, where $\gamma = 1.0001$.

Most work on BHL accretion in the literature parametrises resolution by r^*/R^S, where r^* is the size of the accretor. However, with the sink particle algorithm used here, r^* cannot be unambiguously defined, as gas is removed from a region spanned by a Gaussian kernel, whose width varies from $\Delta x_{\text{min}}/4 < r_{\text{kernel}} < 2\Delta x_{\text{min}}$, depending on local conditions (see Section 4.2.3 for details). N is therefore used as the resolution parameter instead, but $r^* \approx 2\Delta x_{\text{min}}$, i.e. $r^*/R^S_\infty \approx 2/N$, is employed when comparing to previous work.

4.2.1 Nomenclature

For the remainder of this chapter, quantities denoted with a ∞ subscript, i.e. X_∞, are analytical values which parametrise the simulations presented here, defining both initial and boundary conditions. Quantities with a $\diamond$ subscript, i.e. $X_\diamond$, are calculated from cell values across the entire simulation box, and quantities denoted with a $\bullet$ subscript, i.e. $X_\bullet$, are calculated numerically on the fly and are based on mass-averaged quantities within the accretion region of the sink, as laid out in Section 3.1.

The input parameters for each simulation are summarised in its name as follows: a simulation labelled m$\mathcal{M}_\infty$nNx is adiabatic ($\gamma = 1.3334$) when $x = a$, and quasi-isothermal ($\gamma = 1.0001$) when $x = i$. For example, a simulation called m10.0n4.8a has $\gamma = 1.3334$, $\mathcal{M}_\infty = 10$ and $N = 4.8$.

4.2.2 Simulation setup

All simulations presented here are performed with the AMR code RAMSES (Teyssier, 2002) and consist of a simple three-dimensional isolated cubic box, with a sink particle kept fixed at the centre. Gravity is prescribed by an analytic gravitational field for a point mass, which inherits the mass and location of the sink and a gravitational softening length equal to the smallest cell size, Δx_{min}. There is no radiative cooling and the gas is not self-gravitating.

Gas enters the simulation box diagonally, from the bottom left corner, to avoid issues associated with grid aligned flows[1]. All ghost regions (cells outside of the simulation domain) are set using a zero-gradient scheme, except in the $\mathcal{M} = 0$ case where constant inflow boundary conditions are used instead to avoid edge effect propagation. Accretion proceeds via the usual RAMSES sink particle algorithm (Krumholz

---

[1] Such as odd-even decoupling (Quirk, 1994) or the Carbuncle phenomenon at shock fronts (Peery & Imlay, 1988; Elling, 2009)

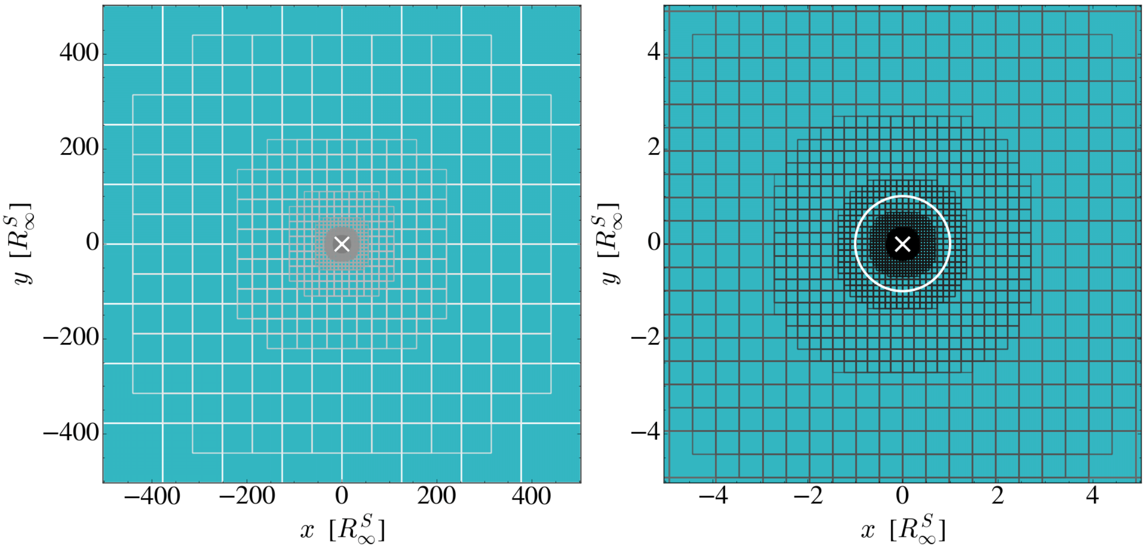

Figure 4.1: Grid structure showing the size of the box (left panel) and the nested grids in the vicinity of the sink (right panel) for a representative simulation (mach0.0n100a). The black hole position is indicated by a white cross, and the white circle in the right plot denotes R^S_∞.

et al., 2004; Dubois et al., 2010). A linear reconstruction scheme and a Courant factor of 0.36 are used as higher order reconstructions and/or larger Courant factors lead to numerical artefacts.

To reach the required number of levels of refinement within a sufficiently large box, a fixed nested grid strategy is employed, as seen in Figure 4.1. In all cases, the black hole is surrounded by concentric shells of progressively lower refinement down to a root grid of 16^3 cells, with the highest resolution shell determined by the chosen resolution value of N. Test of various nested grid configurations show that the results presented here are insensitive to the exact grid layout, so the radius of grids was doubled for each level to balance the size of refined regions with computational cost. For all simulations with $N \leq 100$, all levels of refinement are present from the beginning of the simulation. Simulations with $N > 100$ developed a shock feature during the initial settling phase, which led to a permanent breaking of the symmetry of the flow (see Appendix A). This issue can be prevented by running simulations with $N > 100$ with a lower initial resolution of $N \approx 60$ until $t \approx 3$, and refine to the desired level after this pre-conditioning period. To further avoid numerical effects when refining, only a single additional level of refinement is added per timestep until the correct N is reached. In order to give all simulations enough time to damp initial condition transients, all time-averaged values in this chapter are measured for $t > 10$,

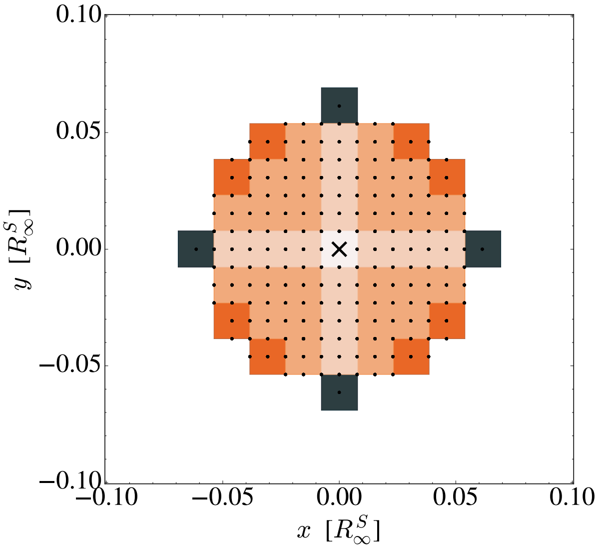

Figure 4.2: Cloud particles are scattered at regular distances from the sink, to calculate mass-averaged gas quantities of the local black hole environment for use in the Bondi-Hoyle accretion formula. The sink is denoted by a black cross and the cloud particles by black dots. Cells probed by the cloud particles are highlighted in colour, while cells outside the accretion region are shown in white. Colours reflect the number of cloud particles present in that cell.

where the time, t, is given in units of scale radius crossing time:

$$[t] = \frac{R^S_\infty}{(c_{s,\infty} + v_\infty)}. \tag{4.1}$$

4.2.3 The accretion algorithm

The standard sink particle accretion algorithm in RAMSES is based on the BHL formula (Equation 3.18) which is described in detail in Section 3.5.2. For the remainder of this chapter, we distinguish between three accretion rates:

1. $\dot{M}^{BHL}_\infty$ is the analytic BHL accretion rate in Equation 3.18, evaluated using the boundary conditions ρ_∞, v_∞ and $c_{s,\infty}$.

2. $\dot{M}^{\mathrm{BHL}}_\bullet = \dot{M}^{\mathrm{BHL}}(\rho_\bullet, v_\bullet, c_{s,\bullet})$ is the accretion rate calculated on the fly according to the BHL formula in Equation 3.18, evaluated using local mass-weighted average quantities.

3. $\dot{M}_\bullet$ is the accretion rate onto the sink, based on $\dot{M}^{\mathrm{BHL}}_\bullet$ but limited by the mass supply within the accretion region.

To calculate the non-local gas properties for use in $\dot{M}_\bullet^{\rm BHL}$, RAMSES uses a cloud-particle system (Dubois et al., 2010; Teyssier et al., 2011), where a total of 2109 cloud particles are scattered around the sink with a spacing of $\Delta_{\rm cloud} = \Delta x_{\rm loc}/2$, filling a sphere with radius $r_{\rm cloud} = 4 \times \Delta x_{\rm loc}$, where $\Delta x_{\rm loc}$ is the local cell width. For simulations more complex than the isolated case presented here, the sink is free to move across the grid, making the cloud particles invaluable in sampling local cells, whether the sink is positioned at the cell centre or not. However, fixing the sink at the centre of a cell in combination with the spacing of $\Delta x_{\rm loc}/2 = \Delta x_{\rm min}/2$ creates ambiguity when assigning cloud particle to host cells, as some of these particles are located exactly on cell boundaries. RAMSES assigns ambiguous particles preferentially to the cell located downwards and to the left in the coordinate system used here, creating a preferential accretion direction. To avoid this issue, all cloud particles are moved radially inward by distributing them with a separation of $\Delta x_{\rm loc}/2.001$. This restores spherical symmetry as each cloud particle is unambiguously assigned to a host cell (See Figure 4.2 for a 2D slice through the cloud of particles). As the sink particle is forced to remain in a specific host cell, a density floor of $\rho_{\rm min} = 10^{-5}$ is introduced, below which accretion is not permitted. The results presented here are insensitive to the choice of $\rho_{\rm min}$ as long as $\rho_{\rm min} \ll \rho_\infty$.

4.2.4 The drag force algorithm

In the simulations presented here, the sink particle is held at a fixed position, so the drag force sub-grid algorithm described in Section 3.5.4 is not employed. Instead, the analytic formula given by Equation 3.27 is compared to values measured directly from the gravitational wake of the sink particle when the latter is resolved. As with the BHL accretion rate, the analytic model (Equation 3.27) calculates the drag force from gas flow properties at infinity. However, the sub-grid model implemented in RAMSES evaluates the drag force on the sink based on the same local mass-weighted quantities used for BHL accretion, $\rho_\bullet, \mathbf{v}_\bullet$ and $\mathcal{M}_\bullet$.

In this chapter, the drag force is calculated using four different methods:

1. F_∞^D is the drag force given by Equation 3.27, using the quantities at infinity.
2. $F_\bullet^D = F^D(\rho_\bullet, v_\bullet, \mathcal{M}_\bullet)$ is the drag force given by Equation 3.27, using mass-weighted quantities from the accretion region, i.e. the drag force as estimated by the sub-grid model.

3. $F^D_\diamond = \sum \frac{GM_{\rm sink} m_i}{r_i^2}\hat{\mathbf{r}}_i$ is the net gravitational force exerted on the sink by gas on the grid. It is found by summing over all cells, i, in the box, where m_i and r_i are the cell mass and cell position vector relative to the sink respectively. As the setup is symmetric, this is equal to the total gravitational attraction of the wake downstream of the sink.

4. $F^D_{\rm tot} = F^D_\bullet + F^D_\diamond$ is the total drag force acting on the sink when the sub-grid algorithm is active.

4.3 Accretion

The BHL formula in Equation 3.18 is based on two limiting cases. The pure Bondi problem, where the sink is at rest relative to the uniform density background, i.e. $\mathcal{M}_\infty = 0$, and the Hoyle-Lyttleton problem, where $\mathcal{M}_\infty \gg 1$. Both cases are investigated in detail, in this section and the next, before exploring the parameter space further.

4.3.1 The Bondi problem

As expected from the analytic work by Bondi & Hoyle (1944), Figure 4.3 shows that the global flow pattern is symmetric and directed radially towards the sink even at very low resolutions, with the sink located at the density peak. Only the immediate sink environment is influenced by the resolution. There is an initial period during which the simulation is dominated by initial condition transients, which last until $t \approx 5$ (see Figure 4.4). After the simulation has settled into a steady flow pattern, the accretion rate onto the sink, $\dot{M}_\bullet$, converges. Note that the actual accretion rate onto the sink, $\dot{M}_\bullet$, is not necessarily the same as the BHL accretion rate evaluated from the local gas properties, $\dot{M}^{\rm BHL}_\bullet$, as $\dot{M}_\bullet$ is limited by the total gas mass present in the accretion region, $\sum(\rho_i \Delta x_i^3 \alpha_i)$.

The analytic solution for the BHL accretion rate assumes that gas properties are measured far from the influence of the sink, a situation that is best captured in the simulations presented here when the cell size is significantly larger than the sink's gravitational scale radius, i.e. $\Delta x \gg R^S_\infty$. Figure 4.4 shows that for the unresolved case (m0.0n0.05a), the accretion rate onto the black hole, as well as the local gas properties as calculated using the cloud particles, reflect the analytic solution closely. The local overdensity before accretion is very shallow, at $\bar{\rho} \sim 1.035\rho_0$ (see m0.0n0.05a

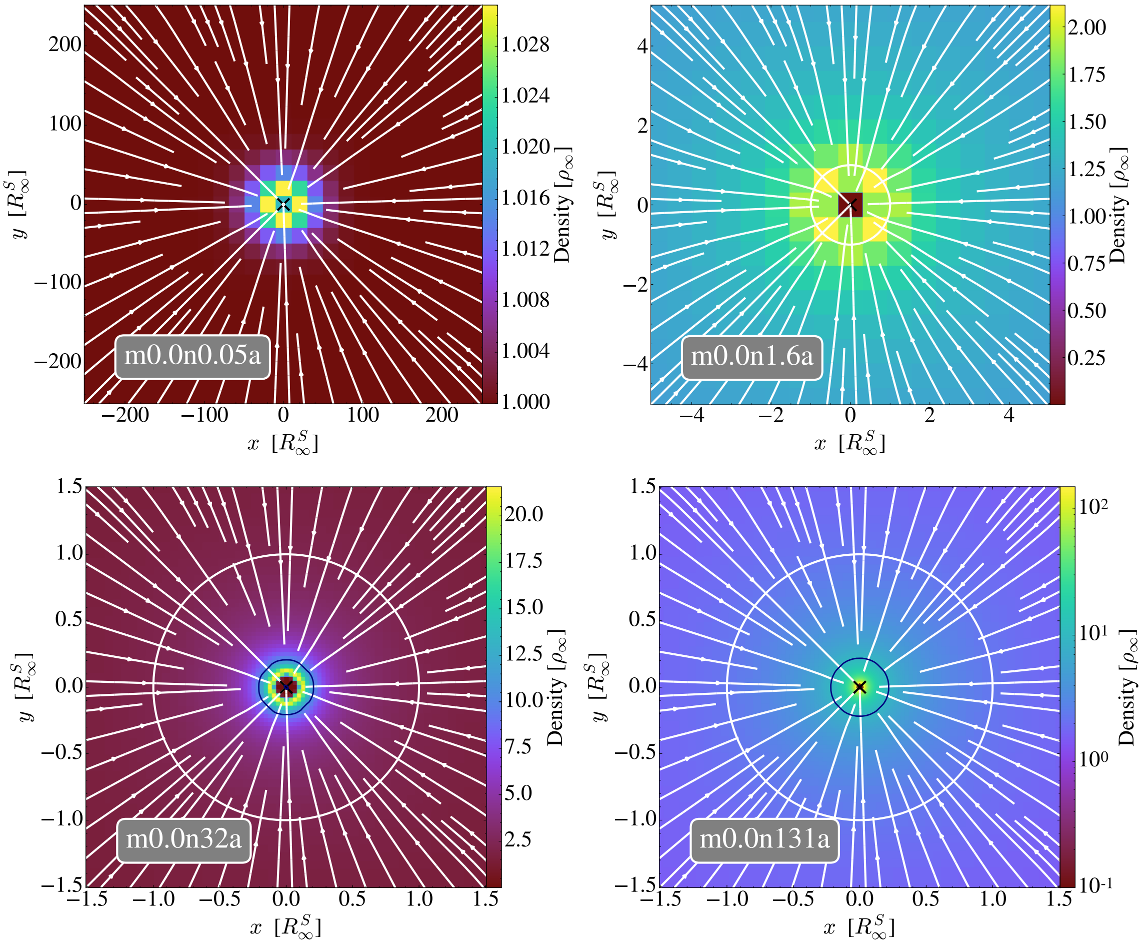

Figure 4.3: Slices through the central density peak surrounding the sink for resolutions of $N = 0.05, 1.6, 32$ and 131 respectively. R_∞^S is denoted by a white circle. The sink is located in the centre of the box and denoted by a black cross. The circular dark blue contour indicates the sonic surfaces in each slice, when they are resolved (only visible in the bottom two plots). Note that the top two panels are zoomed out in comparison to the bottom ones to show the relevant features at each resolution, hence the circles marking R_∞^S are much smaller in these panels, particularly in the top left

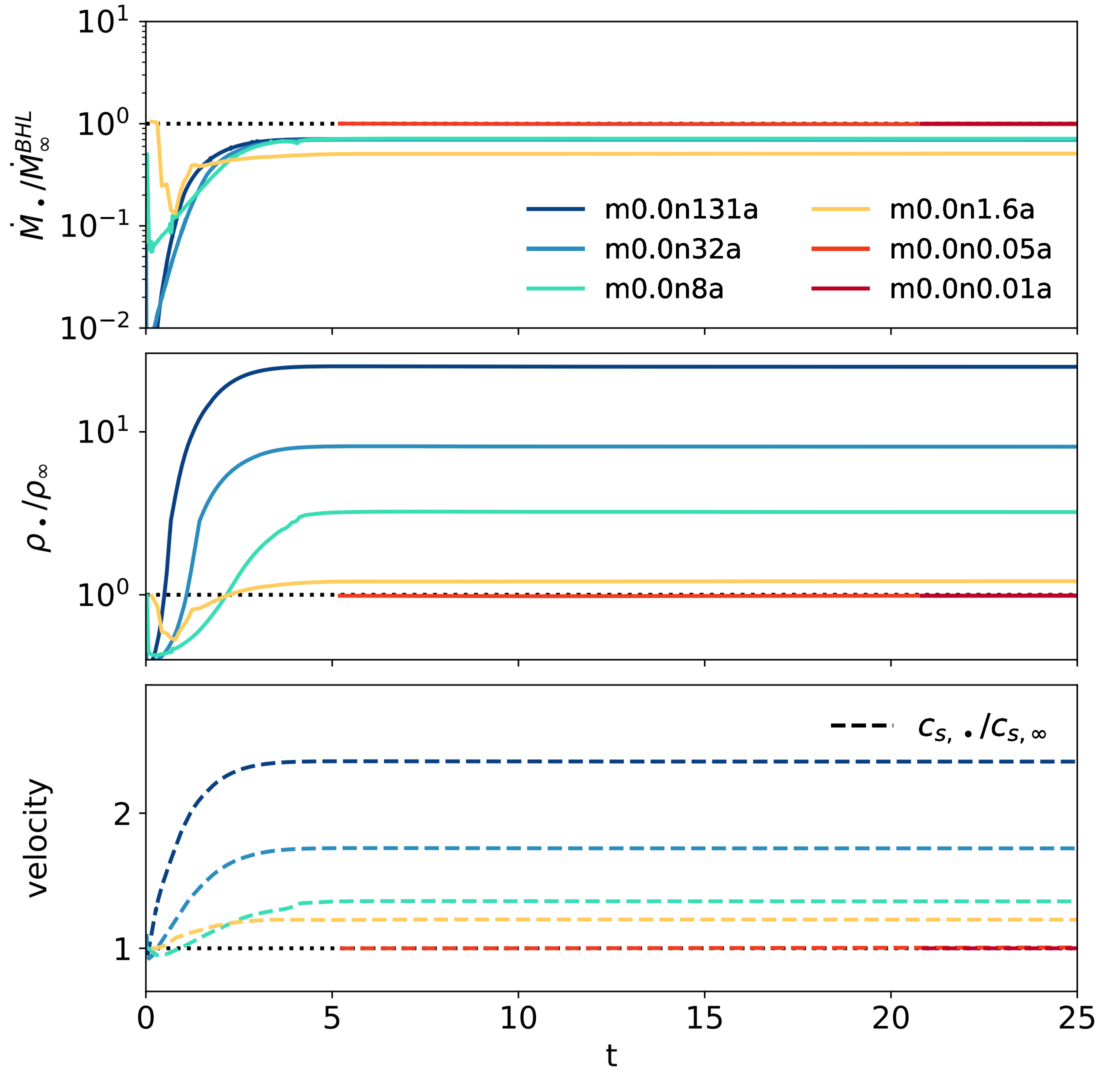

Figure 4.4: Accretion rate and gas properties as sampled by the cloud particles for a variety of resolutions in the pure Bondi case, where $\mathcal{M}_{\infty} = 0.0$.

in the top left panel of Figure 4.3), and is suppressed within the cell containing the sink once accretion starts.

The other extreme is the most highly resolved case probed here (m0.0n131a, bottom right panel of Figure 4.3), where an overdense peak develops within the Bondi radius. When the gas crosses the sonic radius, it transitions from subsonic to supersonic flow as it evolves towards a free fall solution before being accreted. The rising local density increases the BHL rate computed on the fly and the maximum amount of gas permitted is removed from the central cells at each time step. This means that the accretion algorithm effectively transitions from the BHL algorithm to a flux limited accretion scheme, where the accretion rate onto the sink is set by the gas inflow rate into the spherical accretion region of the sink, which has a radius of $\approx 2\Delta x_{\rm min}$. This is referred to as flux limited accretion (FLA) throughout this thesis.

Note that at this resolution, the accretion rate onto the sink settles well below the analytic BHL rate, to $\dot{M}_\bullet/\dot{M}^{\rm BHL}_\infty \sim 0.78$, in agreement with results in Edgar (2004). Simulations with higher force resolution probe the density profile on smaller scales, and therefore measure higher densities $\rho_\bullet$. As the contraction is adiabatic, higher densities have correspondingly higher sounds speed, $c_{s,\bullet}$, as can be seen in Figure 4.4. For this reason, although the analytic model postulates that the gas should transition to supersonic near the Bondi radius, comparing the white circle to the dark blue contours in Figure 4.3 shows that the transition occurs at a smaller radius. This is again in agreement with previous numerical simulations. Finally, the accretion rate has already converged to its $0.78\dot{M}^{\rm BHL}_\infty$ value even at the comparatively modest resolution of $N = 32$, in simulation m0.0n32a.

The partially resolved case, m0.0n1.6a, where $R^S_\infty \sim \Delta x_{\rm min}$ shows intermediate behaviour, with a shallower central density feature and a smaller evacuated region. This simulation also shows more noticeable grid effects, both in the streamlines and in the central density peak, as spherical symmetry is poorly described by the small number of Cartesian resolution elements when $0.1 < N < 4$. The resulting steady state accretion rate is lower than the converged value, with $\dot{M}_\bullet/\dot{M}^{\rm BHL}_\infty \sim 0.6$, as the local density feature feeding the black hole is not replenished efficiently.

Arguably the most worrisome numerical aspect of transitioning to FLA is that the force due to the pressure gradient, artificially created by the low density region developing in the immediate vicinity of the sink, might dominate the gravitational force on the gas. Figure 4.5 shows that while such a pressure gradient does reinforce the gravitational pull on the gas at the edge of the accretion region, it is not the dominant force. Moreover, the contribution of this pressure force decreases for simulations with higher

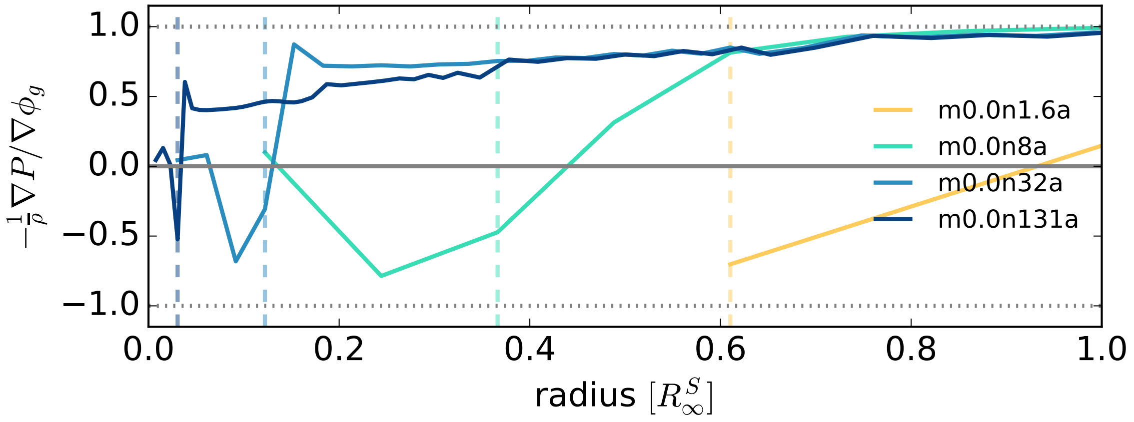

Figure 4.5: Acceleration due to the pressure gradient within the Bondi radius compared to the gravitational acceleration. The dashed vertical lines mark the point of highest pressure, i.e. the edge of the accretion region.

resolution as the gravitational acceleration $-\nabla\phi_g$ increases faster than the pressure gradient, ∇P, within the Bondi radius. As previously mentioned, this effect is also insensitive to the choice of density floor, as long as $\rho_{\text{min}} \ll \rho_\infty$, as for sufficiently small pressure inside the accretion region, $\nabla P \simeq \Delta P/\Delta x = (P_{\text{edge}} - P_{\text{in}})/\Delta x \to P_{\text{edge}}/\Delta x$. ∇P therefore only depends on the cell size Δx and the pressure at the edge of the accretion region, P_{edge}, not the pressure inside the accretion region P_{in}.

From the pure Bondi problem presented here, I conclude that the accretion algorithm is well behaved at all resolutions. Indeed, at low resolution when $N < 0.05$, local gas properties measured in the vicinity of the sink, using cloud particles, produce an accretion rate in excellent agreement with the analytic BHL formula. At high resolution, where $N > 8$, accretion is driven by the local gas supply into the accretion region and the algorithm transitions to FLA. The accretion rate onto the sink converges to the correct value of $\dot{M}_\bullet/\dot{M}_\infty^{\text{BHL}} \sim 0.78$ in that case, fed by supersonically free-falling gas well within the Bondi radius. For intermediate resolutions of $0.05 < N < 8$, grid effects lead to poorer spherical symmetry and lower accretion rates, as gas neither reflects the values at infinity nor forms a central gas profile able to efficiently feed the sink, a conclusion also reached by Krumholz et al. (2004). However, independently of resolution, the accretion rate onto the black hole estimated using the sub-grid algorithm remains within 20% of the correct $0.78 \times \dot{M}_\infty^{\text{BHL}}$ value.

4.3.2 The Hoyle-Lyttleton problem

The other analytic solution was developed by Hoyle & Lyttleton (1939) for an accretor moving supersonically through a uniform medium, where the bulk velocity dominates over the local sound speed, such that $\dot{M}_{\infty}^{\rm BHL}$ approaches $\dot{M}_{\infty}^{\rm HL} = GM_{\rm sink}/v_{\infty}^{2}$. In this section accretion onto the sink in the highly supersonic case is investigated, where $\mathcal{M}_{\infty} = 10$.

4.3.2.1 The adiabatic case

For the adiabatic case, when $\gamma = 1.3334$, in agreement with expectations from the analytic solutions (Hoyle & Lyttleton, 1939; Ostriker, 1999), a conical wake develops downstream of the sink, as is evident in Figure 4.6. In the unresolved case, m10n0.08a, the overdensity around the sink is small, and the streamlines and properties of the gas are only mildly perturbed by the presence of the sink. Therefore, $\dot{M}_{\infty}^{\rm BHL} \approx \dot{M}_{\bullet}^{\rm BHL} \approx \dot{M}_{\bullet}$, and the sink accretes according to the analytic solution (top panel in Figure 4.7). The gravitational wake is especially prominent because the simulations presented here are isolated. If the sink was embedded in a non-uniform medium, such as is typically found for black holes in galaxy simulations, local inhomogeneities would quickly wash out the gravitational focusing effect of the black hole.

With increasing resolution, such as m10n13a in Figure 4.6, the flow patterns resembles the analytic solution by Hoyle & Lyttleton (1939), with bent streamlines, a stagnation point, and an accretion column clearly visible in the density slices. The bow-shock with the characteristic increase in density towards the edge of the shock, predicted by Dokuchaev (1964); Ruderman & Spiegel (1971); Rephaeli & Salpeter (1980), also becomes apparent. At these intermediate resolutions, the shock is attached to the accretor and the solution is stable (see Figure 4.7). As in the pure Bondi case, $\dot{M}_{\bullet} < \dot{M}_{\bullet}^{\rm BHL}$, and the accretion algorithm transitions to FLA, with the under-dense accretion region visible in the density slices (Figure 4.6).

With even more resolution and an even smaller accretor, such as in m10n26a, the bowshock detaches from the accretor and the solution becomes unstable, in agreement with results by Ruffert (1995b). While the accretion rate begins to fluctuate due to local instabilities, the time-averaged value remains similar to the case before the bowshock detached (compare m10n13a and m10n26a in Figure 4.7) as an overdense peak feeding the black hole continues to exist. In this regime, the wake alternates between episodes when instabilities die down and the flow returns to a more symmetric configuration (top row, Figure 4.8), and episodes when instabilities dominate flow

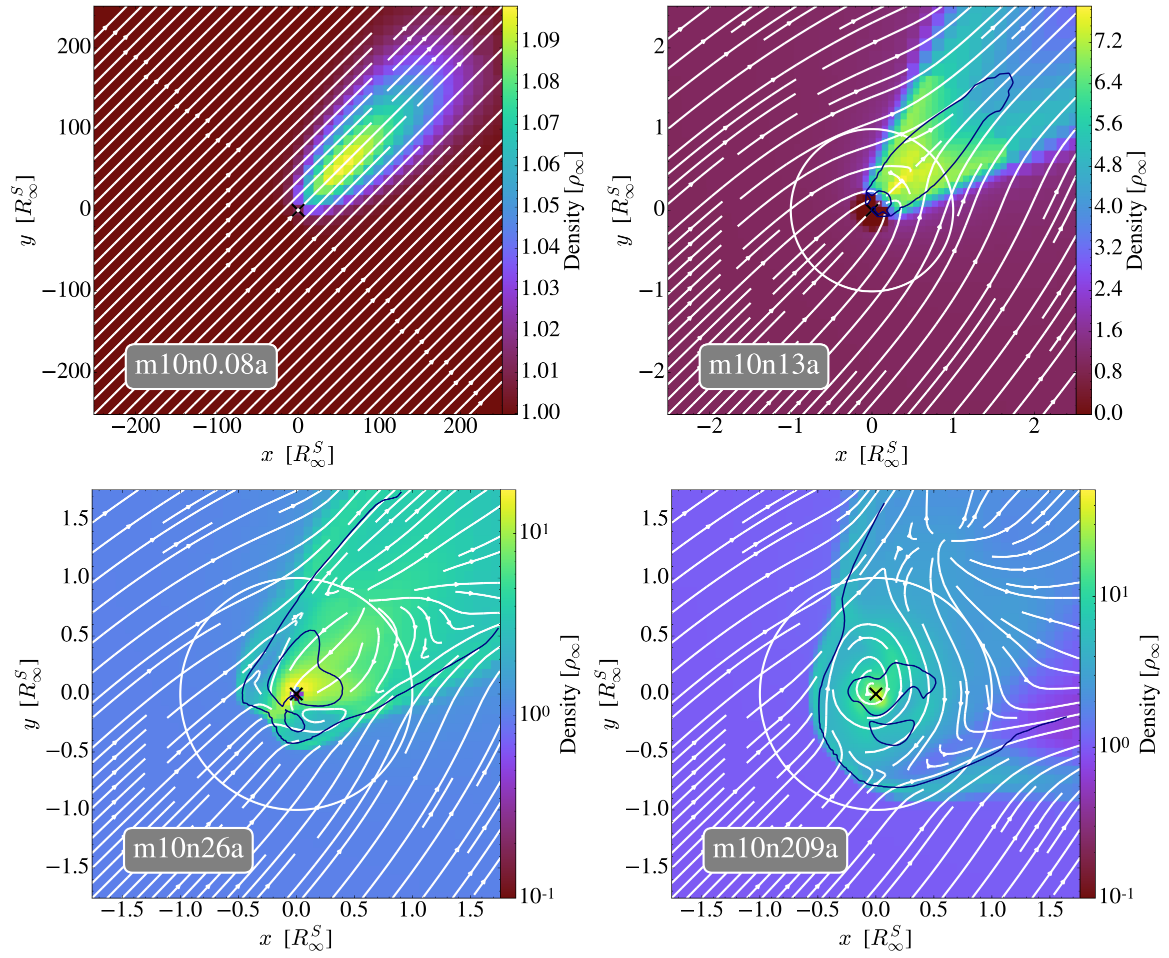

Figure 4.6: Slices of the central density feature for resolutions of $N = 0.08, 13, 26$ and 209. The white circle denotes the size of the accretion radius R^S_∞ and the black hole location is marked by a black cross. Note that the top left panel is zoomed out considerably in comparison to the other three, making R^S_∞ difficult to see. The sonic surface for each slice is delineated by a dark blue contour when resolved and streamlines are shown in white.

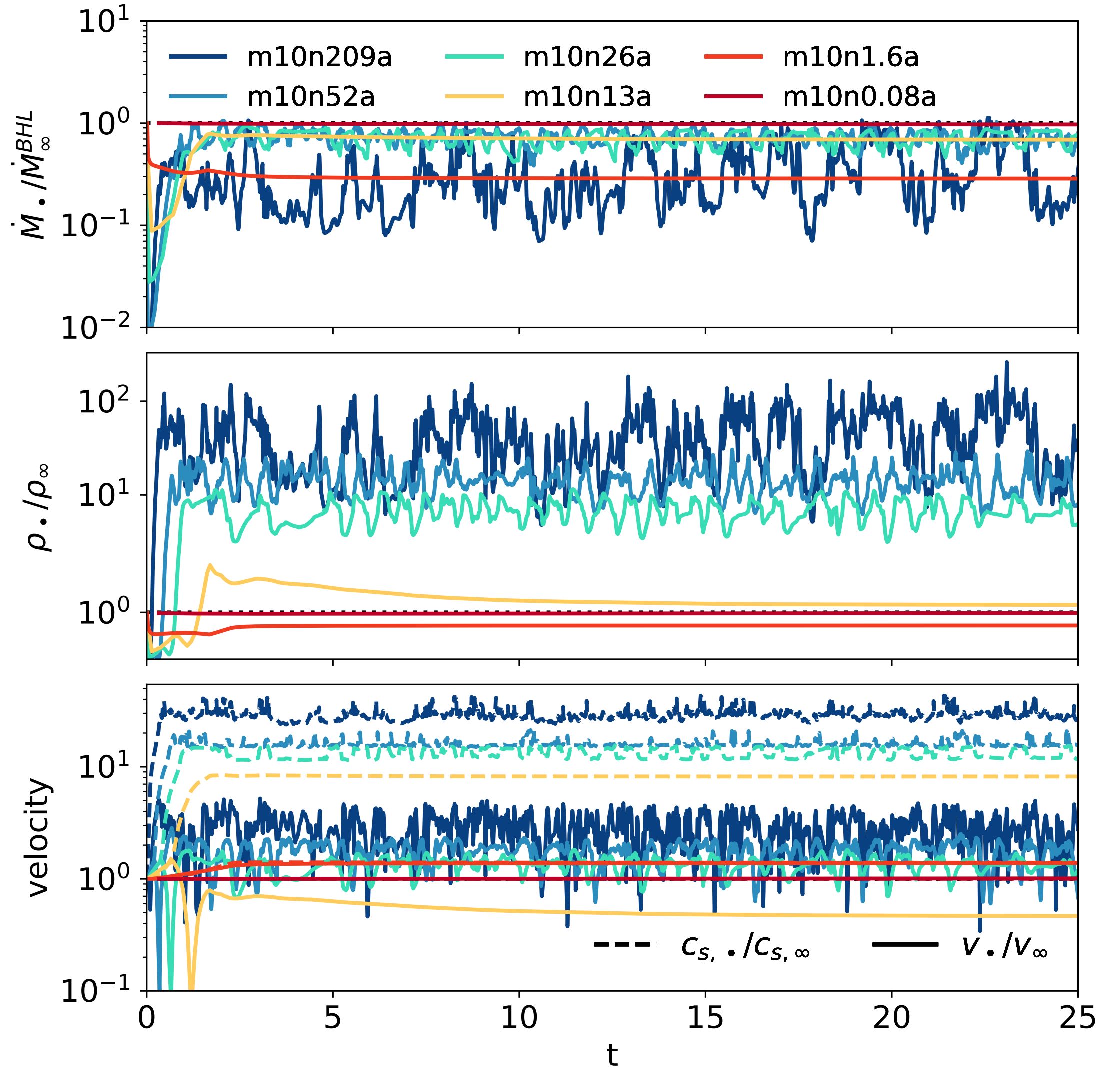

Figure 4.7: Accretion rates (top panel) and mass weighted gas properties, including density (middle panel) and velocity and sound speed (bottom panel) at a variety of resolutions for the supersonic Hoyle-Lyttleton case, where $\mathcal{M}_\infty = 10.0$. All gas properties are measured by the kernel-weighted cloud particles within the accretion region of the sink.

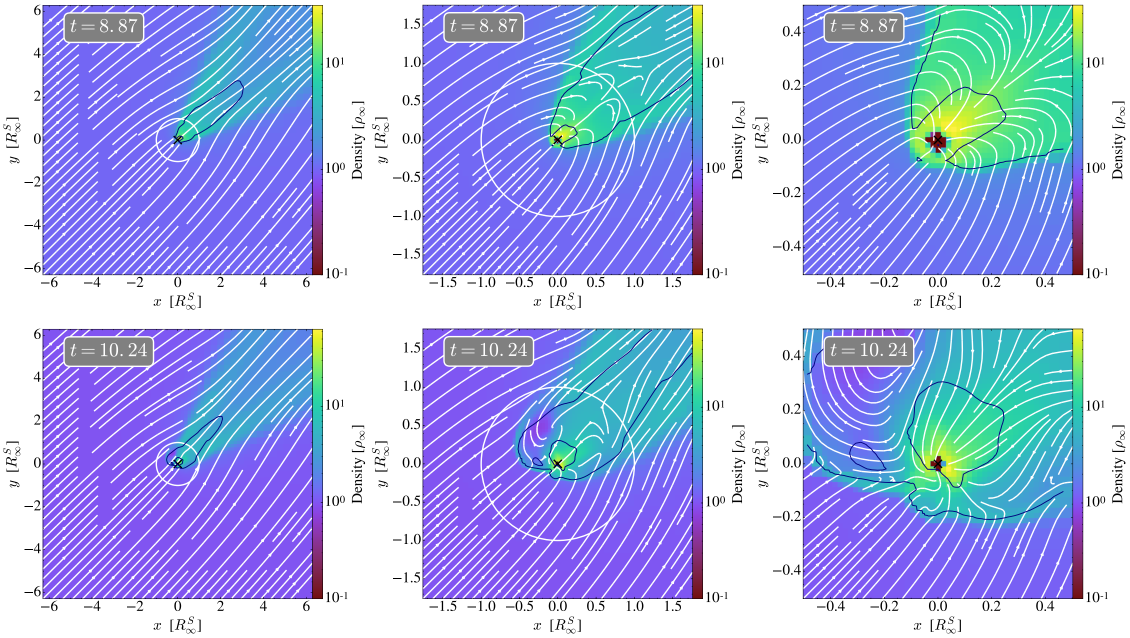

Figure 4.8: Zoom in on flow patterns in and around the wake at two different times for m10n50a. R^S_∞ is shown as a white circle. Streamlines are indicated in white, and the sonic surface as dark blue contours. The global wake is stable, but the front of the bowshock alternates between periods of symmetric flow (top row, $t = 8.87$) feeding the black hole efficiently, and instability dominated flow (bottom row, $t = 10.24$).

patterns, which are severely disruptive (bottom row, Figure 4.8). When not disrupted by instabilities, the gravitational attraction of the sink creates a peaked density profile around the accretion region, which like in the Bondi case in Section 4.3.1 leads to a resolution dependence of the sound speed at the edge of the accretion region.

While the global wake remains stable, eddies develop behind the shock, suggesting that the instability is caused by the physical acoustic-advection instability reported in Foglizzo et al. (2005) for $\gamma > 4/3$, $\mathcal{M}_\infty > 3$ and "sufficiently small accretors". According to these authors, the instability is caused by entropy perturbations that advect from the shock to the sonic surface around the accretor, where they excite acoustic waves due to the inhomogeneity of the flow, which in turn propagate outwards back towards the shock, where they excite new entropy perturbations. The instabilities are therefore expected to occur in the subsonic region, where the gas has been decelerated by the shock and not yet sufficiently reaccelerated by the gravitational potential of the sink (see the bottom rows of Figure 4.6 and Figure 4.8). The lack of subsonic region between shock and accretor in the temporarily settled configuration in Figure 4.8 (top row) supports this conclusion.

The accretion algorithm creates a low density region around the sink, surrounded by a high density shell replenished by inflowing material. While most of the mass that enters the accretion region is removed, the sink particle algorithm does not implement a strict inflow criteria, with any surplus gas free to leave the accretion region during the next timestep. Accretion onto the sink varies on short timescales, as the turbulent flow feeds the black hole intermittently (see m10n26a in Figure 4.7), with an average value of $\dot{M}_\bullet/\dot{M}_\infty^{\mathrm{BHL}} \simeq 0.8$, in agreement with simulations discussed in Edgar (2004). This is also the value recovered for the steady state solution of extremely small accretors ($N = 1000$) in the 2D axisymmetric simulations of El Mellah & Casse (2015).

The picture changes drastically for the highest resolutions and smallest accretors (m10n209a in Figure 4.6), as the instabilities become more severe. The bow shock opening angle increases, supported by a strong rotational flow around the sink. Global flow patterns, which feed the sink at lower resolutions, break down and the solution is entirely dominated by instabilities, as evident by the oscillations in Figure 4.7. Temporary periods of ordered flow that replenish the gas supply in the accretion region of the sink, seen for resolutions $14 \leq N \leq 50$, disappear. The time averaged accretion rate drops to $\dot{M}_\bullet/\dot{M}_\infty^{\mathrm{BHL}} \simeq 0.11$, with order of magnitude fluctuations around this value, so that extrapolating results from lower resolution runs provides a poor estimate of the accretion rate onto the sink. These results are in agreement with work by MacLeod & Ramirez-Ruiz (2015), who investigate Hoyle-Lyttleton type accretion in the presence of a density gradient at infinity and find that the resulting circularisation of gas behind the (warped) shock reduces the accretion rate onto the sink by over an order of magnitude. In the simulations presented here, the circularisation is driven by the advective-acoustic instability and not a global density gradient. However, both cases showed that the presence of significant angular momentum in the gas behind the shock reduces the accretion onto the sink by more than an order of magnitude.

Overall, in contrast with the Bondi case, the accretion onto the sink for an adiabatic Hoyle-Lyttleton flow converges to a value well below the analytic BHL solution, as it becomes entirely dominated by instabilities for very small accretors, i.e. $r^*/R^S < 0.01$ (equivalent to $N \geq 200$).

4.3.2.2 The quasi-isothermal case

Foglizzo et al. (2005) argue that the advective-acoustic instability should disappear for $\gamma < 4/3$, whereas Ruffert (1995a) report unstable flow for $\mathcal{M}_\infty = 10$ at any γ. To further investigate the possible presence of instabilities, a set of quasi-isothermal simulations with $\gamma = 1.0001$ was added to the suit of simulations presented in this

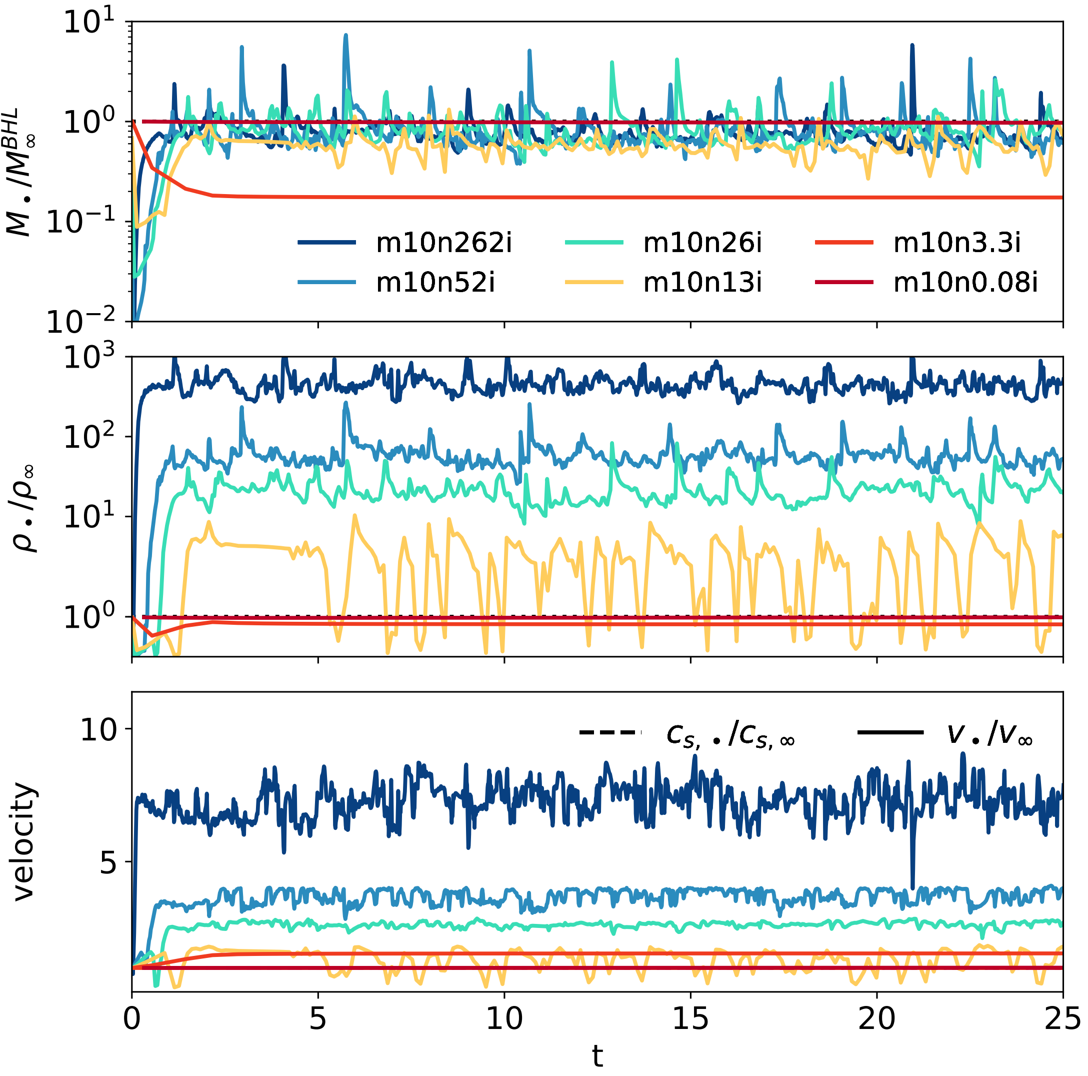

Figure 4.9: Accretion rates and gas properties within the accretion region in the supersonic Hoyle-Lyttleton case, where $\mathcal{M}_\infty = 10.0$. Gas is quasi-isothermal i.e. $\gamma = 1.0001$, and each line represents a different resolution, as indicated in the top panel of the figure.

chapter. They show a gas flow pattern broadly similar to the adiabatic case presented above, albeit with very different consequences for the accretion onto the sink. As can be seen in Figure 4.10, at very low resolution, for m10n0.08i, only a small overdensity forms downstream of the sink, which accretes according to the analytic BHL formula (see the top panel in Figure 4.9). For higher resolutions, i.e. $N \geq 3$, the local flow pattern again shows curved streamlines, a stagnation point, a shock and FLA. However, the shock opening angle is significantly smaller. Figure 4.9 also shows that the accretion onto the sink is unstable for $N > 3$, a significantly lower resolution threshold than for the $\gamma > 4/3$ runs. Instead of an under-dense region forming along the axis of symmetry downstream of the stagnation point, the quasi-isothermal simulations feature an overdensity. By definition, the sound speed (bottom panel of Figure 4.9) remains constant in the quasi-isothermal case, so as resolution increases the relative velocity of the gas with respect to the sink increases much more than in the adiabatic case (bottom panel of Figures 4.7 and 4.9).

As a result, and contrary to the adiabatic case, the shock remains attached to the accretion region regardless of resolution and the flow stays supersonic everywhere except in a small narrow region around the stagnation point (see Figure 4.10). Despite this, strong instabilities develop in the wake, particularly for small accretors, which also lead to accretion rate variations on short timescales. Instead of originating at the bowshock, these instabilities occur downstream of the sink and affect the wake globally, disrupting the accretion column onto the sink. No eddies are visible in the streamlines (see Figure 4.10). Rather, the narrow wake clumps and is distorted in the direction perpendicular to the axis of symmetry of the wake, in agreement with the seminal work of Ruffert (1995b). Taken at face value, the persistent instabilities for $\gamma \simeq 1$ seem to contradict predictions by Foglizzo et al. (2005), who argue that the acoustic-advective instability should disappear for $\gamma < 4/3$. However these authors do point out that a different type of instability could occur at low γ, based on vorticity (rather than entropy) perturbations between shock front and accretor. The instabilities seen here are also reminiscent of the ones discussed in Cowie (1977) and Soker (1990), who study the fate of small overdensities when the wake is modelled as an accretion line. Beyond the exact nature of the instabilities, the main difference between the adiabatic and the quasi-isothermal simulations is that the strong drop in average accretion rate for the smallest accretors, which have $N > 200$, is absent for quasi-isothermal simulations. At the highest resolution probed here, the $\gamma = 1.0001$ accretion rate converges to $\dot{M}_\bullet / \dot{M}_\infty^{\rm BHL} \simeq 0.74$: while instabilities exist, they do not efficiently prevent accretion onto the sink.

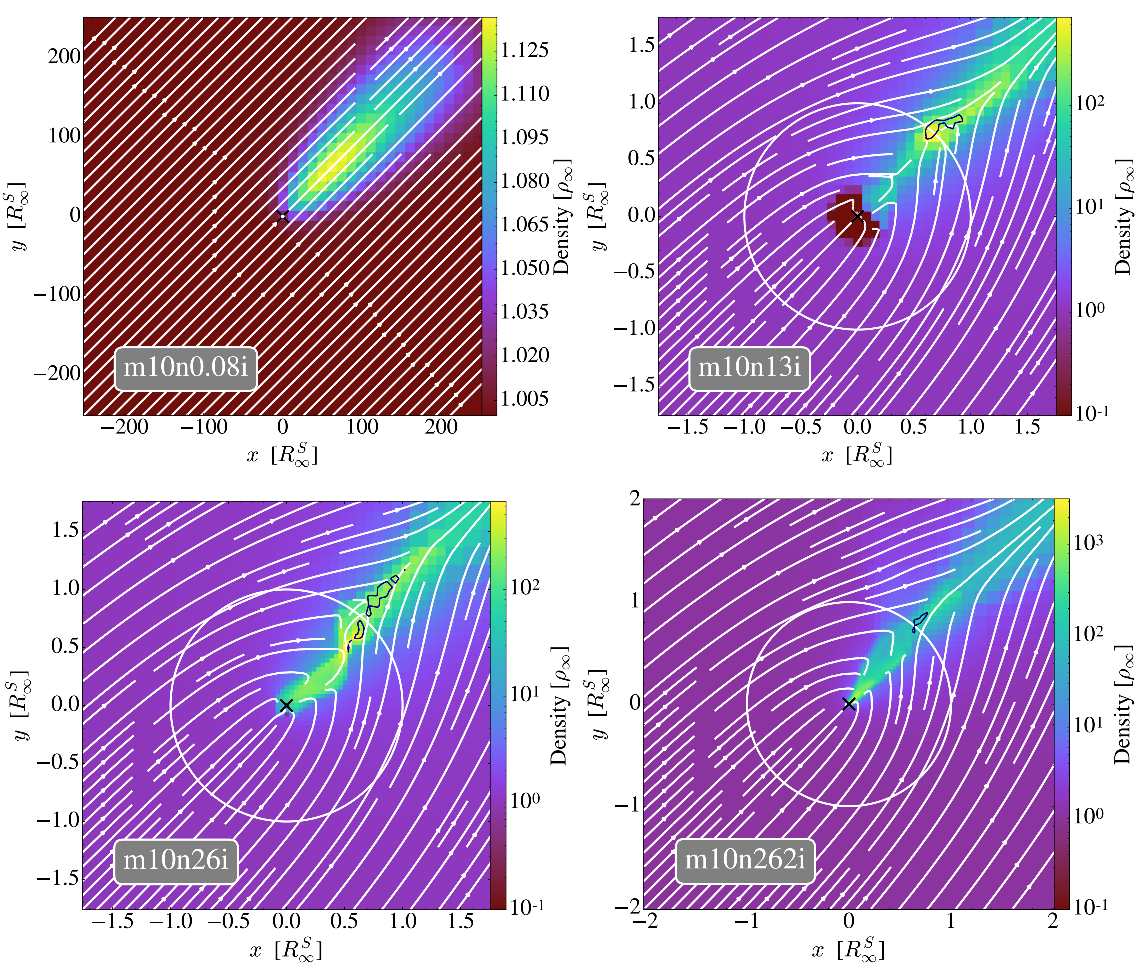

Figure 4.10: Density slices for the quasi-isothermal runs, where $\gamma = 1.0001$, for a variety of resolutions. The accretion radius, R^S_∞, is shown as a white circle and is invisible in the top left panel due to the large edge-length. The sink is marked by a black cross. Sonic surfaces are represented as dark blue contours and are located near the stagnation points but are difficult to notice as they are small and narrow: the flow remains supersonic almost everywhere. Streamlines are shown in white.

In summary, for a black hole moving supersonically, with $\mathcal{M}_\infty = 10$, accretion proceeds via the BHL algorithm at low resolution, $N < 8$, and transitions to FLA at higher resolution. The wake is unstable for both values of γ, with the instabilities in the adiabatic case originating in the subsonic region between the sink and the detached bowshock and near the stagnation point in the quasi-isothermal case. With increasing resolution and decreasing size of the accretor, the flow becomes unstable on progressively shorter timescales. In the adiabatic case, instabilities dominate for $N > 100$ which leads to an order of magnitude reduction in the time-averaged accretion rate onto the sink, whilst in the quasi-isothermal case this averaged accretion rate converges to a value close to that of the BHL case.

4.3.3 Exploring the full range of Mach numbers

In this section, the evolution of flow patterns and accretion rates in the adiabatic case are explored for a wider range of Mach numbers and different resolutions. Note that in the case of intermediate Mach numbers ($0.3 \lessapprox \mathcal{M}_\infty \lessapprox 3.0$), the analytic BHL formula (Equation 3.18) is not expected to yield as accurate an estimate of the accretion onto the sink as in the low and high Mach number cases previously studied as it is merely an educated interpolation between these two extremes cases.

However, Figure 4.12 presents the average accretion rate, in units of the analytic BHL accretion rate $\dot{M}_\infty^{\mathrm{BHL}}$, for a variety of Mach numbers as a function of resolution. As was already be seen for the Bondi and the Hoyle-Lyttleton problem (Section 4.3.1 and 4.3.2 respectively), the unresolved case with $N < 0.01$ closely follows the analytic formula. As resolution increases, the behaviour becomes more complicated. In the intermediate regime, where $0.01 < N \leq 50$, the simulations diverge from the analytic formula in a way that depends non-monotonically on the Mach number. Sub- and supersonic simulations ($\mathcal{M}_\infty = 0.5$ and $\mathcal{M}_\infty = 3$ respectively) show average accretion rates systematically lower than the BHL formula by up to a factor 5. Trans-sonic simulations, on the other hand, with $0.9 \leq \mathcal{M} \leq 1.5$, feature accretion rates which are larger by up to a factor of 2.3. The BHL formula, used here to normalise results, is most uncertain for the trans-sonic regime, where both the bulk velocity, v_∞, and the sound speed, $c_{s,\infty}$, have a significant influence on the flow. The high resolution results ($N > 50$) support this conclusion as the accretion rates indeed converge to higher values in the trans-sonic regime. I caution that at intermediate resolutions, the pressure force into the low resolution region around the sink can dominate over the local gravitational force, possibly funnelling extra gas into the accretion region and thus leading to an overestimate of the accretion rate. However, this effect is very

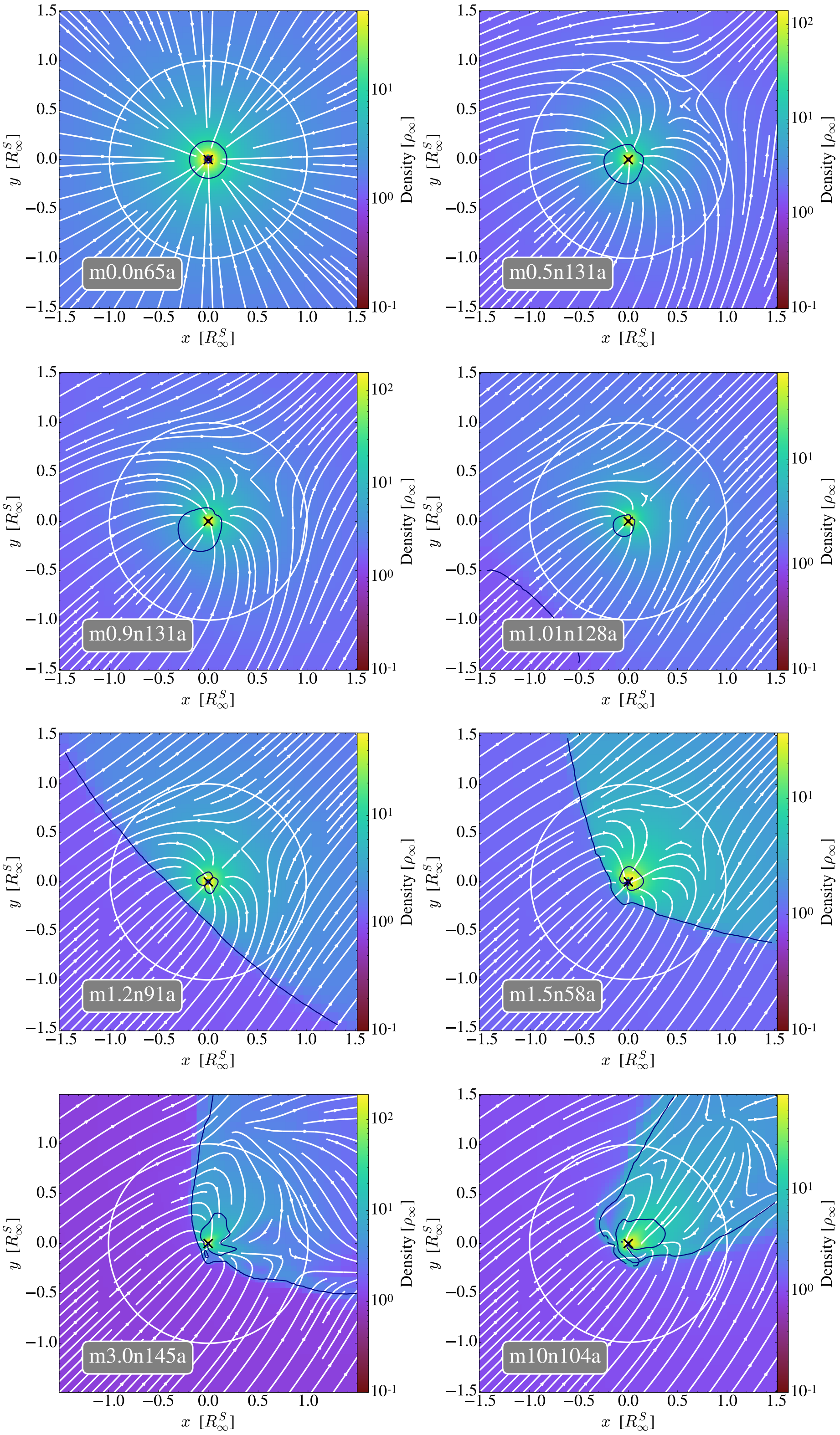

Figure 4.11: Density slices for a range of Mach numbers, ordered from the lowest $\mathcal{M}_\infty$ in the top left to the highest in the bottom right. The sink is denoted by the black cross at the centre of each panel. Flow lines are represented by white lines, R^S_∞ is indicated by a white circle and dark blue contours mark sonic lines.

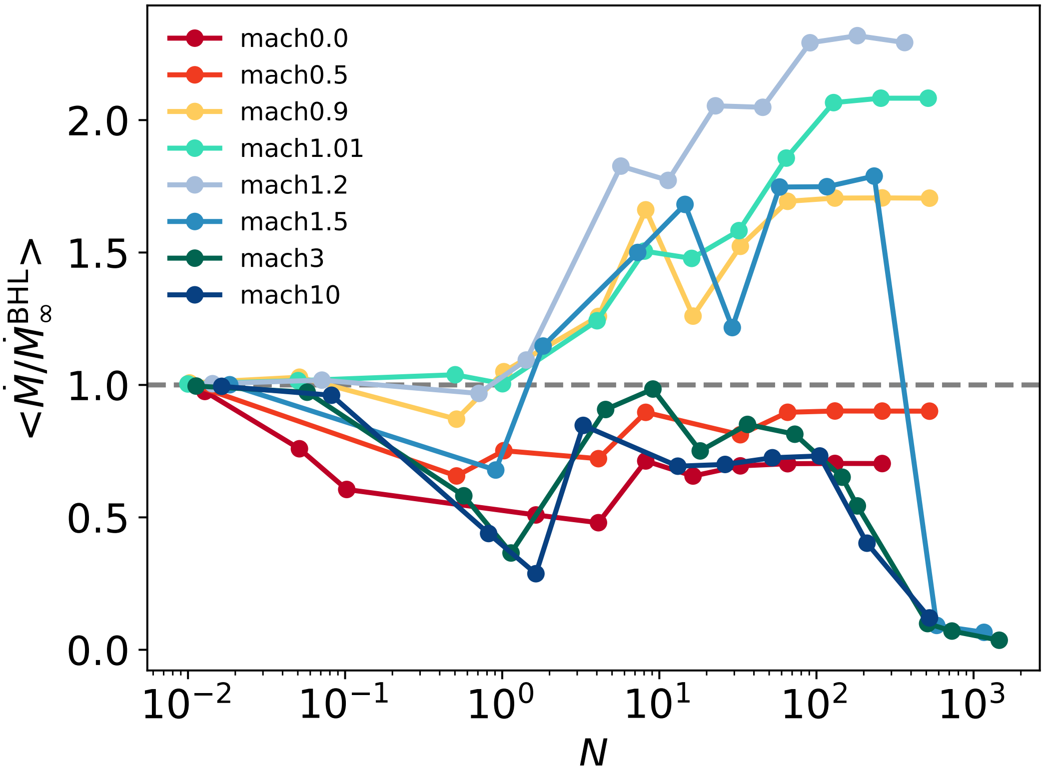

Figure 4.12: Dependence of the accretion rate on resolution for a variety of Mach numbers in adiabatic simulations with $\gamma = 1.3334$. Each data point represents the time averaged value for $t > 10$ of a specific run, as indicated on the panel.

localised as it only occurs at the edge of the accretion region and is alleviated by the kernel function which creates a gradual transition of density within the region covered by the cloud particles. However at high resolution ($N > 100$), the gravitational force comfortably dominates over the pressure force for all cases studied here (see Figure 4.5 for a measure in the Bondi case). Moreover, all resolved simulations with Mach numbers $\mathcal{M}_\infty \leq 1.2$ show steady state solutions (Figure 4.11).

From $\mathcal{M} > 1.5$, eddies begin to form behind the shock. When the accretor becomes small enough, instabilities become stronger and begin to influence accretion onto the sink more significantly. The density slices for both $\mathcal{M}_\infty = 3$ and $\mathcal{M}_\infty = 10$ show strong instabilities that disrupt the flow patterns and decrease the time-averaged accretion rate onto the sink by up to an order of magnitude below the analytic value. While a lot of care has been taken to minimise the impact of initial conditions (see Appendix A), the seeding of the instabilities could still be due to the way the simulations are initialised. This is likely to affect the exact resolution and/or Mach number at which the instability dominated regime appears but unlikely to make it vanish altogether.

In summary, the sink particle algorithm, using a locally evaluated BHL accretion rate as described in Section 3.5.2.3, is a versatile sub-grid model that smoothly adapts to a variety of resolutions. For highly resolved simulations ($N > 100$), the kernel function used to remove mass ensures that the maximum accreted mass per timestep, dominated by the dense cells at the edge of the accretion region, always exceeds the local gas supply and the sub-grid model automatically transitions to FLA. Intermediate and low resolutions ($N \leq 50$) lead to mixed results and appear to be a difficult regime when approximating the accretion rate onto the black hole from local gas properties. However, they still manage to capture the accretion rate within a factor ≈ 2, at least in the moderate Mach number regime ($\dot{M}_\infty^{\mathrm{BHL}} < 1.5$). It is only at higher Mach numbers that they deviate from resolved time averages by more than an order of magnitude, an effect that can significantly impact the final mass of the sink. This is potentially important when simulating the cosmological growth of supermassive black holes, where an early accretion boost is crucial to enable the black holes to reach observed masses within the limited timeframe available (see Volonteri, 2010, for a review).

4.4 Drag force

The gravitational force of the wake formed downstream of a sink particle in the presence of a significant bulk flow exerts dynamical friction on the sink particle, opposite to the direction of motion. This causes it to reduce its relative velocity with respect to the gas over time. However, the total drag force of the wake depends sensitively on the mass contained in the wake, particularly close to the sink.

Returning to the $\mathcal{M}_\infty = 10$ adiabatic simulations discussed in detail in Section 4.3.2.1, one can see in the density slices in Figure 4.6 that the wake develops even for low resolution simulations. Figure 4.13 shows the mass distribution and drag force profile of each wake at $t = 25$, plotted against the distance, s, to the sink measured along the axis of symmetry of the wake. The mass distribution close to the sink, where $|s| < 0.5R^S_\infty$, shows some variation with increasing resolution due to the locally unstable flow. However, the global structure of the wake at larger radii converges quickly for $N \geq 13$ and remains stable. The highest contributions to the drag force are found in the immediate vicinity of the accretor but as symmetric contributions upstream and downstream of the sink cancel out, the larger scale structure of the wake (within $s \approx R^S_\infty$ in this case) contributes the bulk of the net drag force onto the sink. As a result, the overall drag is adequately captured even with moderate resolution: integrating $\mathrm{d}F^D_\diamond/\mathrm{d}s$ over s for simulations m1013a, m10n26a, m10n52a, m10n209a yields values for $F^D_\diamond$ of 0.057, 0.063, 0.071, 0.059 respectively in the dimensionless units used in this work, so that even in the most violently unstable case (m10n209a), the drag force fluctuates by less than 20%.

Figure 4.14 shows the spatial contribution of density slices along the wake for adiabatic simulations at $t = 25$ for a range of Mach numbers, the density slices for which can be found in Figure 4.11. As expected, the wakes contain a significant amount of mass on relatively large scales (up to $s \approx 25 \times R^S_\infty$), especially for trans-sonic ($\mathcal{M}_\infty \simeq 1$) configurations. However, the inverse square dependence on the distance to the sink means that for all Mach numbers investigated here, most of the gravitational drag force comes from a region within $r_{\max} \simeq 10 \times R^S_\infty$. Note that the intensity of the force also depends on the opening angle of the wake, with a similar mass profile exerting a stronger pull in the direction of motion if confined to a narrower cone. Finally, material in front of the sink particle, pooling behind the detached shock, exerts a non-negligible gravitational force in the opposite direction and reduces the overall drag, especially in the sub- and trans-sonic regimes. While this feature was also observed in Chapon et al. (2013), it appears more prominently

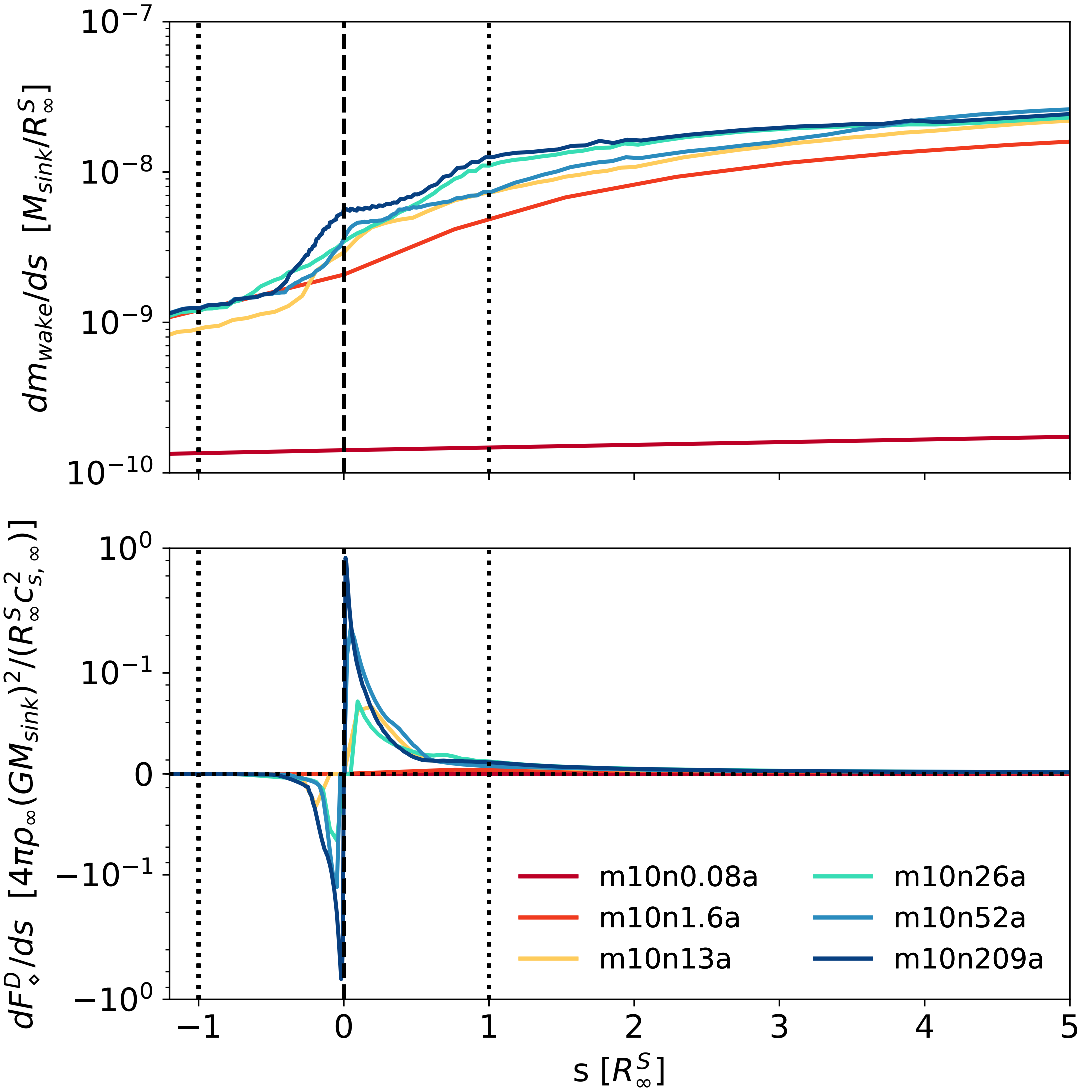

Figure 4.13: Differential mass (top panel) and drag force (bottom panel) profiles at time $t = 25$ for a highly supersonic adiabatic flow ($\mathcal{M}_\infty = 10$) at a variety of resolutions, as labelled on the bottom panel. s is the distance of a given mass slice to the sink particle measured along the axis of symmetry of the wake. Negative values indicate density slices upstream of the sink and the sink location is denoted by the vertical dashed line. Dotted vertical lines stand for $s = R^S_\infty$.

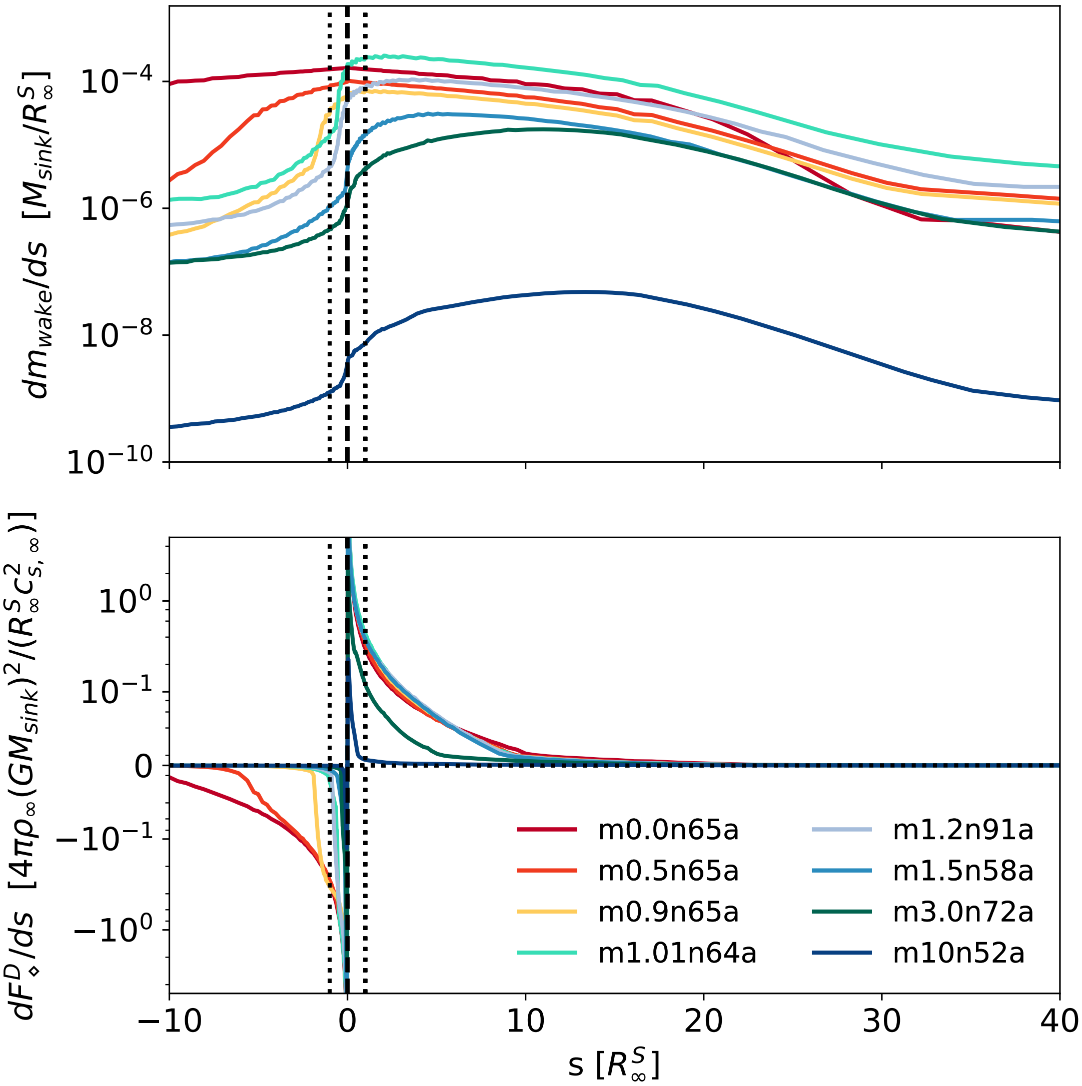

Figure 4.14: Same as Figure 4.13, but for different Mach numbers, as labelled on the bottom panel. Note the increased upstream contribution to the drag force as the Mach number decreases.

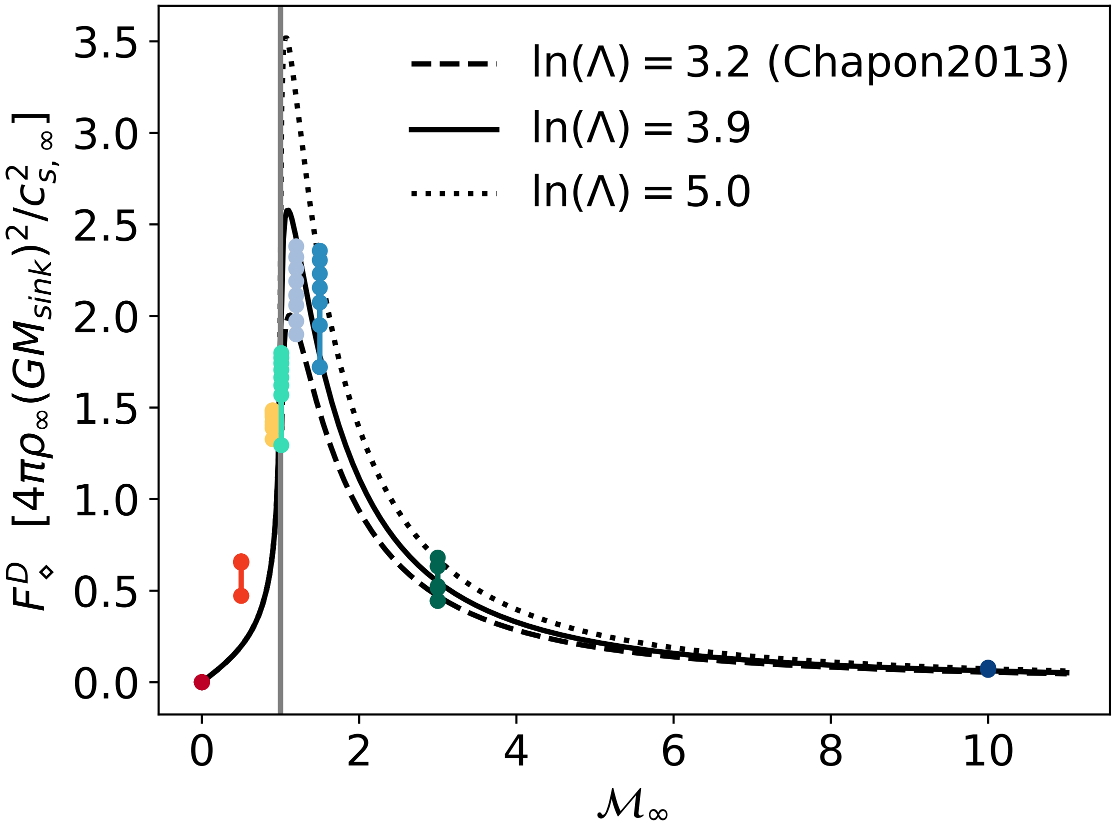

Figure 4.15: Drag force due to the gravitational wake as a function of the Mach number of the flow, for simulations with a resolution $N \approx 65$ (see bottom panel of Figure 4.14 for the exact resolution). The different curves represent the analytic formula of Ostriker (1999) in Equation 3.27 for different values of the Coulomb logarithm, as indicated on the panel. Each simulation was sampled at times $t = [10, 12, 14, 16, 18, 20, 22, 24]$ to give an idea of the dispersion in the drag force measurements, hence the multiple data points for any given Mach number.

in the simulations presented here and is completely absent in the analytic solutions for supersonic black holes by Ostriker (1999), who state that the sink particle only generates a density wake within the rear Mach cone.

Comparing the total net gravitational drag for the set of resolved simulations in Figure 4.14 to analytic estimates in Figure 4.15, and considering that the size of the accretor sets the smallest scale $r_{\rm min} \simeq r^* \simeq 2\Delta x_{\rm min} \simeq \frac{1}{32} R^S_\infty$, the Coulomb logarithm evaluates to $\ln(\Lambda) = \ln(r_{\rm max}/r_{\rm min}) \simeq 5.8$, larger than the value of 3.2 reported by Chapon et al. (2013) in the trans-sonic regime. It is clear from the bottom panel of Figure 4.14 that $r_{\rm max}$, the characteristic size of the medium which the accreting object traverses, is quite a sensitive function of Mach number. It drops

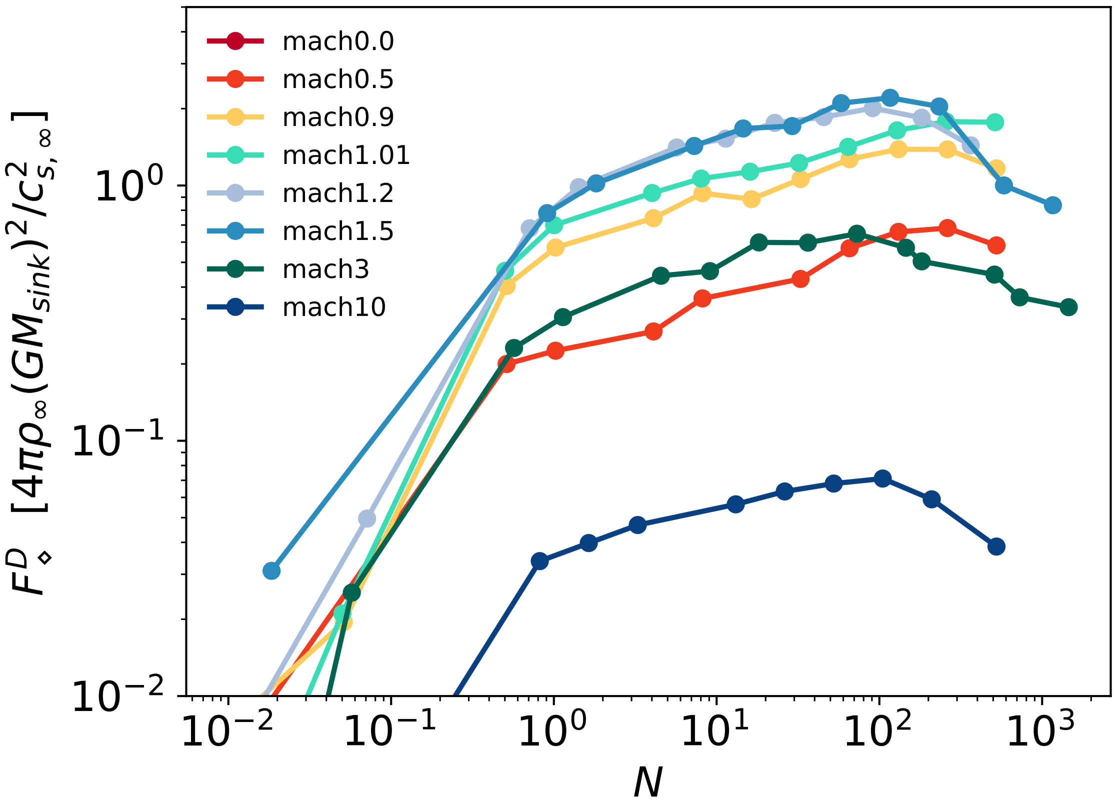

Figure 4.16: Time averaged gravitational drag force as a function of resolution for a variety of Mach numbers, as indicated on the panel. The drag force for $\mathcal{M}_\infty = 0$ does not appear on the plot as it is negligible.

from $\approx 10 \times R^S_\infty$ in the trans-sonic regime to approximately R^S_∞ at Mach $\mathcal{M}_\infty = 10$, corresponding to $\ln(\Lambda) \simeq 3.5$. Fitting over the whole range of Mach numbers covered by the simulations, a time averaged best fit Coulomb logarithm of $\ln(\Lambda) = 3.9^{+1.1}_{-0.7}$ is found, where errors describe an lower (upper) bound to individual time measurements. The discrepancy with the Chapon et al. (2013) results is potentially partly due to the different value of γ used in these authors' simulations, but more likely caused by the absence of accretion onto their black holes. This is somewhat corroborated by the fact that an excess in gravitational drag for subsonic sinks is found here, which show a more prominently asymmetric density profile than in the non-accreting analytic solutions.

Contrary to the accretion rates in Figure 4.12, the drag force due to the wake forming behind the sink shows rapid convergence at a surprisingly low resolution of $N \geq 1$, as shown in Figure 4.16. It is only vastly underestimated at resolutions as low as $N < 1$, i.e. when the scale length corresponding to the gravitational influence of the black hole, R^S_∞, is small in comparison to the minimal cell size, so that gravitational focusing is inefficient. The drag force due to the wake is only moderately influenced (up to 50%) by the instabilities developing behind the bow shock, as it is dominated by larger scale contributions.

On top of the direct computation of the gravitational drag exerted on the sink by the overdensity in its wake, RAMSES includes a sub-grid algorithm that calculates the drag force based on the same local mass weighted average quantities used to estimate the accretion rate, $F^D_\bullet(M_{\rm sink}, \rho_\bullet, v_\bullet, c_{s,\bullet})\hat{\mathbf{v}}_\bullet$, with $\ln(\Lambda) = 3.2$. It is designed to compensate for the lack of drag force in low resolution simulations, clearly visible in Figure 4.16 for $N < 1$. As the drag force on the sink is a vector quantity, it inherits its direction from the local relative velocity, $\mathbf{v}_\bullet$.

For the supersonic case, $\mathcal{M}_\infty = 10$, the top panel of Figure 4.17 shows that the magnitude of the cell based force, $F^D_\diamond$, converges for $t > 10$ and $N > 10$. The instabilities in the flow for $N > 20$ cause small variations of about 10% in magnitude (filled symbols). To investigate the impact of resolution both on the magnitude and direction of the drag force, a fiducial drag $\mathbf{F}^D_{\rm fid} = -F^D_\diamond(N = 209, t = 25)\hat{\mathbf{v}}_\infty$ is defined, based on the flow velocity at infinity and the converged drag force intensity $F^D_\diamond(N = 209, t = 25)$. Figure 4.17 shows that the drag force due to the wake is always parallel to the fiducial force (filled markers, middle panel). As expected from the direction of the wake, the force acts in the opposite direction to the global flow velocity and slows the sink down. For low resolution, $N < 2$, the drag force due to the wake can be underestimated by a large factor, but remains steady. For high resolutions, $N > 10$, the magnitude of the force rapidly converges to within 10% of the fiducial value. At the highest resolutions, $N > 20$, short term variations in the magnitude of the force are visible (filled markers, top panel) but no significant deviation from the axis of symmetry of the wake (middle panel).

The sub-grid based drag force, $\mathbf{F}^D_\bullet$ (see top two panels of Figure 4.17), by contrast shows a very erratic behaviour for $N > 20$. Its magnitude, $F^D_\bullet$ (top panel, solid lines), varies considerably on the shortest timescale probed here, the finest timestep of the simulations, and can be both significantly larger or significantly smaller than the $F^D_\diamond$ value. Moreover, higher resolutions show larger fluctuations. Not only does the magnitude of the force fluctuate rapidly but it is also directed in the opposite direction to the fiducial force most of the time (actually at all times for $10 < N < 200$). This can be easily understood because the bulk of the mass enters the accretor through the accretion column, which has a flow direction directly *opposite* to the global flow (see Figure 4.6 for some examples). As a result of mass weighting, the velocity of the

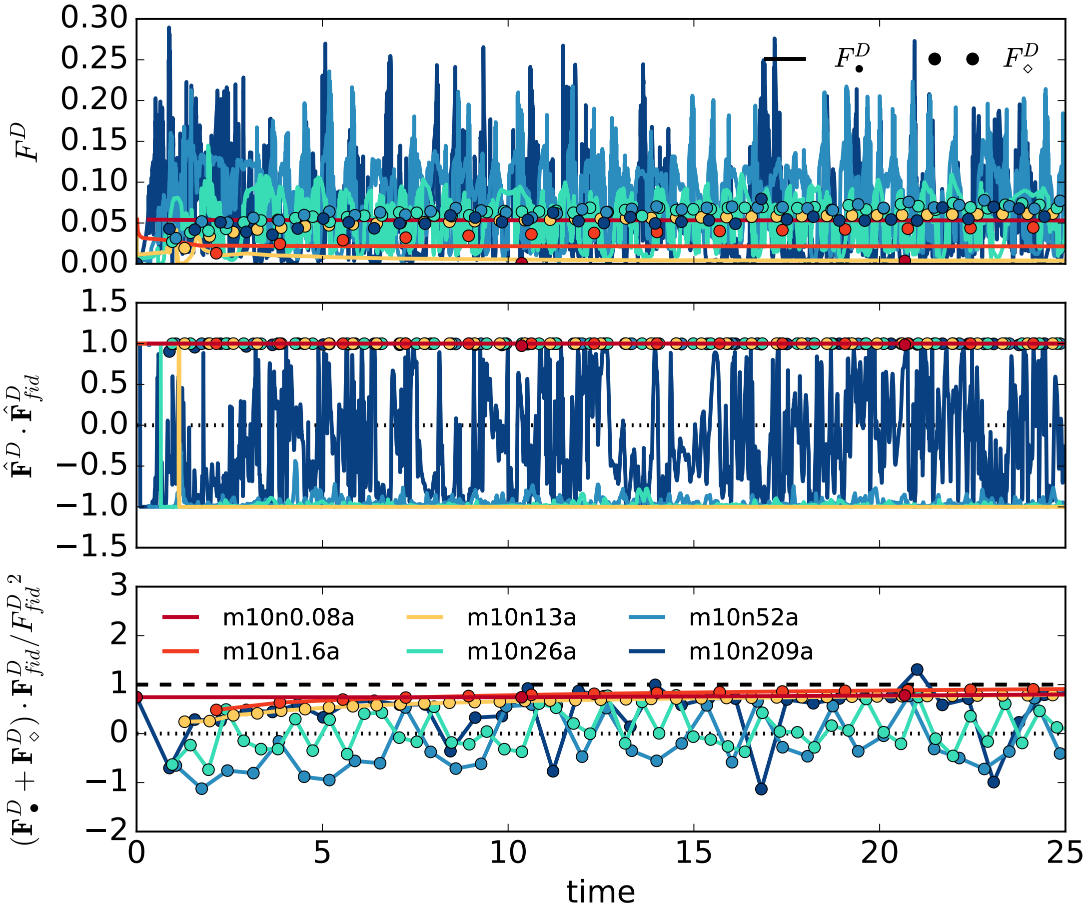

Figure 4.17: Comparison of the drag force calculated in two different ways: i) directly from the contribution of each gas cell of the simulation, $F^D_{\diamond}$, shown as filled markers and ii) using the sub-grid algorithm based on local mass weighted quantities and $\ln(\Lambda) = 3.2$ (Equation 3.28), $F^D_{\bullet}$, shown as solid lines. The top panel shows the magnitude of the two forces, the middle panel quantifies the direction of the forces relative to a fiducial force $\mathbf{F}^D_{\rm fid}$ and the bottom panel compares the total force to the fiducial force. $F^D_{\diamond}$ is sampled at coarse timesteps of the simulation, whereas $F^D_{\bullet}$ is sampled at fine timesteps, hence the different number of datapoints. The horizontal dashed line denotes the point when the instantaneous force is equal to the fiducial force, $F^D_{\rm fid}$, taken to be the force measured in m10n209a simulation at $t = 25$ (see text for details). The horizontal dotted line shows where the instantaneous force is either zero or perpendicular to the fiducial value.

accretion column thus dominates the local flow velocity, $\mathbf{v}_\bullet$, and since $\mathbf{F}^D_\bullet \propto \mathbf{v}_\bullet$ the sub-grid drag force flips direction as soon as the accretion column forms. The full extent of the problem becomes apparent when calculating the total drag force in the presence of the sub-grid algorithm, $\mathbf{F}^D_{\rm tot} = \mathbf{F}^D_\diamond + \mathbf{F}^D_\bullet$ (bottom panel of Figure 4.17). For the resolved cases, $N > 20$, $\mathbf{F}^D_\bullet$ is frequently the dominant term. This causes a net force that *accelerates* the sink relative to the global gas flow which is clearly unphysical. It is entirely caused by the fact that the local mass weighted relative velocity, $\mathbf{v}_\bullet$, reflects neither the direction nor the magnitude of the value at infinity, as soon as the accretion column forms and the bow shock detaches from the accretor. As expected, the total drag force for the unresolved cases, $N < 2$, is unaffected, as the local velocity reflects the value at infinity because the accretion column has not (fully) developed. The sub-grid based drag force therefore significantly and accurately contributes to the overall drag force in that case. This contribution naturally drops as the accretion column builds up, leading to a decrease in relative velocity and an increase in sound speed as the flow in the vicinity of the accretor becomes sub-sonic (see bottom panel of Figure 4.7). From Figure 4.17 (top and middle panel), the flip in drag force direction occurs around at $N \simeq 10$, which does not create a significant problem at this resolution as the resolved drag force term dominates. As a result, $\mathbf{F}^D_{\rm fid} > \mathbf{F}^D_{\rm tot}(N \leq 10) \geq 0.8 \times \mathbf{F}^D_{\rm fid}$ (bottom panel of Figure 4.17) for the standard drag force sub-grid model used in RAMSES with $\ln(\Lambda) = 3.2$. The best fit value of $\ln(\Lambda) = 3.9$ (not shown) shows near perfect agreement at $N < 2$.

Based on these results, and taking a conservative approach, sub-grid algorithms for the drag force onto the sink should be avoided as soon as the characteristic scale radius, R^S_∞, becomes larger than the size of the accretor, $2\Delta x_{\rm min}$. Note that R^S_∞ itself depends on the relative velocity of the black hole and the ISM, and can therefore be difficult to determine in more complex simulations than the idealised Bondi-Hoyle-Lyttleton flows investigated here, where the values at infinity are known. The criterion at which the sub-grid drag force on black holes becomes unphysical might therefore have to be revised for galaxy evolution simulations, the analysis of which is left to the next chapter.

4.5 Are the instabilities at high Mach number physical?

To the best of my knowledge, no complete analytic analysis yet exists to explain the instabilities regularly found for accretion wakes in the Hoyle-Lyttleton problem.

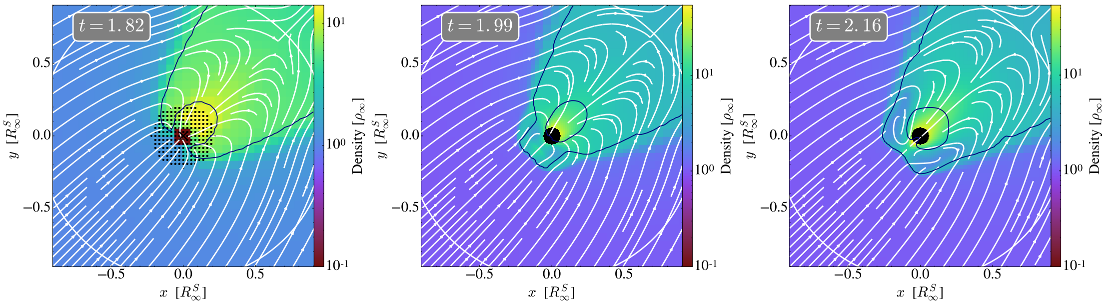

Figure 4.18: Density slices across the characteristic radius for a shrinking accretor, here from $N = 36$ to $N = 72$, equivalent approximately to $r^*/R^S_\infty = 0.06 \rightarrow 0.028$ embedded in an adiabatic supersonic ($\mathcal{M}_\infty = 3$) flow. While the larger accretor shows a stable flow pattern, short lived eddies appear behind the bow shock for the smaller accretor. The size of the accretor is annotated by the cloud particles (black dots) and R^S_∞ by the solid white circle. Sonic surfaces are denoted as black contours.

So the question remains whether the observed instabilities are physical or numerical in nature. The work on the danger of using small accretors with uniform initial conditions (see Appendix A) shows that a numerical origin is extremely difficult to rule out. However, a vast amount of efforts to stabilise simulations have failed, as instabilities have been reported using a wide variety of codes, coordinate systems and models for the accretor (see Foglizzo et al., 2005, for a review). In agreement with the results presented here, a majority of authors report a clear link of the appearance of instabilities with the size of the accretor.

Claims of the existence of steady state solutions for small accretors and high Mach numbers, such as the work by El Mellah & Casse (2015) ($N = 1000$ and $\mathcal{M} = 16$) and Pogorelov et al. (2000) ($N = 20$ and $\mathcal{M} = 20$) have also been made. However, all such results are based on 2D axisymmetric simulations and it is well known that for Hoyle-Lyttleton simulations, certain instabilities only appear for particular configurations. For example the flip-flop instability, frequently observed in 2D, is entirely absent in 3D simulations (Blondin & Pope, 2009).

To investigate how a shrinking accretor impacts the stability of the flow, an established, stable simulation with an intermediate accretor size (m3n32a) has refinement levels added so the accretor shrinks, reproducing the conditions in m3n72a where instabilities unsettle the flow. Both simulations are supersonic, with $\mathcal{M}_\infty = 3$, and m3n32a has reached a steady state before the accretor is shrunk.

The density slices in Figure 4.18 confirm observations from the $\mathcal{M}_\infty = 10$ case presented in Section 4.3.2. Shrinking the accretor creates a subsonic region upstream of the accretor, forming between the supersonic region surrounding the accretor and the shock front, which is absent for the steady state solution at lower resolution. Instabilities in the form of small eddies develop as soon as the subsonic region appears, and move downstream along the wake, disturbing the symmetry of the flow and eventually unsettling the accretion column feeding the sink. During the next 10 dynamical times, the simulation does not resettle into a steady state.

As previously discussed, Foglizzo (2009) suggest that the advective-acoustic instability should unsettle the flow in exactly this manner. It is characterised by small entropy perturbations forming behind the shock, which are advected towards the accretor, where the local rise in density causes them to reflect back to the shock front, exciting further perturbations. According to the authors, for this phenomenon to occur, the accretor needs to be small enough that the shock detaches and the subsonic region opens, in good agreement with results presented here. As was shown in section 4.3.3, all of the simulations which have $\gamma > 4/3$, a sufficiently strong shock (which occurs for $\mathcal{M}_\infty \geq 1.5$), and enough resolution for the shock to detach, show these types of instabilities, providing strong support for the physical origin advocated in the analysis of Foglizzo (2009).

For the isothermal case, the shock remains attached and the flow is supersonic everywhere, but the wake again becomes globally unstable for sufficiently small accretors. This is particularly important for simulations of SMBHs, as their galactic environment is strongly affected by radiative cooling. While the instabilities for the quasi-isothermal case appear to have a different — and arguably less well analytically understood — origin than the adiabatic ones, they lead me to conclude that Hoyle–Lyttleton accretion onto galactic black holes is very unlikely to reach a steady state, whether radiative cooling plays an important role or not.

Indeed, repeating the experiment to explore the origin of the instability in the quasi-isothermal case, by shrinking the accretor for a $\mathcal{M}_\infty = 3$ simulation from $N = 3.5$ (m3n3.5i), which leads to a steady state solution, to $N = 7$ (m3n7i), which is unstable (see Figure 4.19) reveals that the instability originates behind the sink in the region of the stagnation point of the flow. This could be the instability described in Cowie (1977), where the authors consider the impact of a small density perturbation in the accretion column and find evidence for an oscillatory instability. While the flow lines bend for the steady state at low resolution, the spatial extent of the accretor still encompasses the stagnation point. The reverse flow, up the accretion column and towards the sink, does not exist. After shrinking the accretor, the stagnation point appears in the flow pattern and oscillates perpendicularly to the axis of symmetry of the wake. This conclusion is supported by Figure 4.10, where the largest

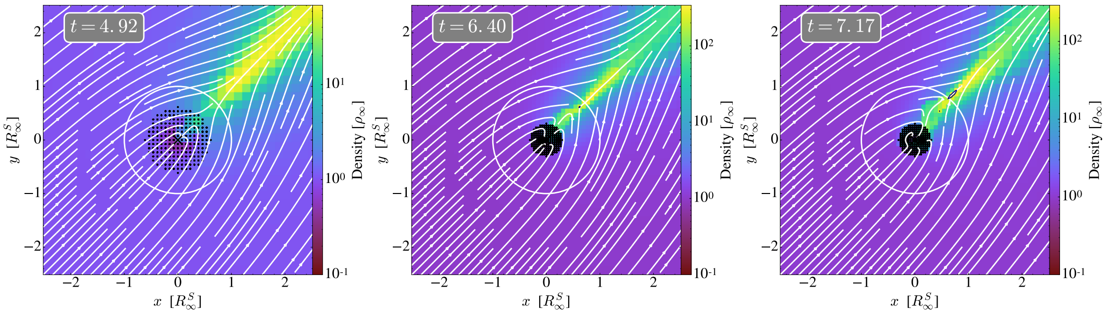

Figure 4.19: Density slices through the characteristic scale radius for a shrinking accretor, here from $N = 3.5$ to $N = 7$, embedded in a quasi isothermal supersonic flow with $\mathcal{M}_\infty = 3$. The size of the accretor is annotated by the cloud particles (black dots) and R^S_∞ by the solid white circle. Sonic surfaces are denoted as black contours: the flow remains supersonic everywhere except in a very narrow region around the stagnation point. Flowlines are shown in white.

perpendicular displacement of the wake occurs at or near the stagnation point for all three simulations with $N \geq 7$.

A variety of other instabilities have been suggested to potentially arise from Hoyle-Lyttleton 'like' accretion flows, such as individual vortex rings in Kim & Kim (2009) for a non-accreting massive perturber embedded in an adiabatic 2D supersonic flow, or the shock cone vibrations reported in Lora-Clavijo & Guzman (2013) who investigate 2D adiabatic BHL accretion using general relativity instead of Newtonian gravity. Neither of these phenomena are observed here. Finally, the main numerical limitation of the simulations presented here is that the size of the accretor is intrinsically tied to the resolution, with $r^* \approx 2 \times \Delta x_{\rm min}$. To truly investigate the nature of the instabilities would require exploring the impact of resolution and accretor size independently, which is beyond the scope of this work.

4.6 Conclusions

As numerical capabilities increase, a larger range of length scales can be covered in simulations. In this chapter, the impact of resolution on the accretion and drag force for a sink particle in grid simulations was investigated, using sub-grid algorithms based on the Bondi-Hoyle-Lyttleton (BHL) interpolation formula for accretion and the linear analytic drag force estimate of Ostriker (1999). In all simulations presented here, the size of the accretion region shrunk with increasing resolution and both analytic formulas were evaluated using mass-averaged, kernel-weighted quantities in the immediate vicinity of the accretor. Adopting an idealised BHL setup, an originally homogeneously distributed gas with a uniform Mach number was allowed to settle in the gravitational potential of the accreting particle, with the relevant scale radius, $R^S_\infty(\mathcal{M}_\infty)$, resolved by a fixed number of resolution elements, $N = R^S_\infty(\mathcal{M}_\infty)/\Delta x_{\rm min}$, throughout the simulation. This setup was used to investigate the impact of a wide range of resolutions, $0.01 < N < 500$, Mach numbers $0 \leq \mathcal{M}_\infty \leq 10$ and two values of the adiabatic index $\gamma = 1.3334, 1.0001$, on the accretion rate and drag force sub-grid algorithms onto the accretors, modelled as sink particles with radius of $r*/R^S_\infty \simeq 2/N$.

As expected, for very low resolutions, i.e. $N < 1$, where the local mass weighted quantities reflect the ones at infinity because the gravitational influence of the sink is small, the accretion sub-grid algorithm closely follows the analytic value. Accretion rates converged for $N > 100$ and Mach numbers $\mathcal{M}_\infty < 1.5$, although not to the analytic value set by the accretion algorithm as accretion onto the black hole transitioned from BHL to flux limited accretion. However, in this situation, the accretion rates measured onto the accretor all converged to values within a factor $\simeq 2$ of the BHL analytic interpolation formula, with the largest discrepancies found for trans-sonic flows.

However, at higher Mach numbers, $\mathcal{M}_\infty \geq 1.5$, and higher resolution, instabilities appeared and started to dominate the flow to such an extent that the accretion rate dropped to $\dot{M}/\dot{M}^{\rm BHL}_\infty \approx 0.1$ in the adiabatic runs with $\gamma = 1.3334$. In the quasi-isothermal case ($\gamma = 1.0001$), accretion rates remained within a factor $\simeq 2$ of the BHL rate over the whole range of Mach numbers probed here, despite the presence of instabilities in the accretion column at high Mach numbers ($\mathcal{M}_\infty \geq 1.5$).

These findings are in agreement with previous work by Ruffert (1995b) and Foglizzo et al. (2005) as to the origin of these instabilities, at least in the adiabatic case when $\gamma \simeq 4/3$. They are caused by an acoustic-advective feedback loop which develops in the sub-sonic region between the shock front and the accretor, as described analytically in Foglizzo et al. (2005) and Foglizzo (2009). In the quasi-isothermal case, where $\gamma = 1.0001$, an instability was observed that originates near the stagnation point in the accretion column downstream of the accretor and locally displaces this column perpendicularly to the axis of symmetry of the problem, reminiscent of effects discussed in Cowie (1977) and Soker (1990).

The drag force due to the wake, calculated from the cell densities in the box, converged quickly with resolution, again in agreement with results by Ruffert (1995a,b). However, as the bow shock and accretion column began to be resolved by the simulation, the local mass weighted quantities used in the sub-grid algorithm to evaluate the analytic drag force according to the formula of Ostriker (1999), started to poorly reflect the values of these quantities “at infinity”. Indeed, not only did the magnitude of the drag force estimated in this way fluctuate on extremely short time scales but its direction actually flipped, so that the net force on the accretor remained an *acceleration* with respect to the gas flow, rather than a drag, even after adding the contribution from the resolved wake. This clearly unphysical behaviour solely occurs because the local relative velocity becomes dominated by gas flowing down the accretion column. Therefore, sub-grid algorithms for the drag force should be avoided as soon as the characteristic radius becomes larger than the accretor size.

The main limitation of the conclusions about the sub-grid algorithms presented in this work is that quantities “at infinity” are poorly defined in more realistic galaxy evolution or cosmological simulations. For example, the scale radius for quantities “at infinity”, taken to be the boundary of the box, is a clearly defined quantity in the simulations presented here. However, embedded in a galaxy, the flow far from the accretor is not uniform or isothermal, and the gravitational field due to the accreting black hole not necessarily dominant. I also caution that it is important to resolve R^S_∞, which can be much smaller than the Bondi radius for highly supersonic flows which are quite frequent in galaxy simulations.

Moreover, while fairly similar trends are measured in both the adiabatic and quasi-isothermal cases probed here, the simulations do not include radiative cooling explicitly, which could significantly influence the sound speed in the vicinity of the accretor, and through it both the accretion rate and drag force. In addition, the drag force is calculated but the sink is not allowed to move under its influence. This is particularly important for the instability dominated runs, where one would not expect the accretor to remain located on the axis of symmetry of the bow shock at all times. In less idealised simulations than those presented in this chapter, the sink is also expected to encounter non-homogeneous gas moving non-uniformly, other compact objects and gradients in the local gravitational potential, all of which could dominate the local flow. The impact of additional physics and an evolving environment, whether in a disc galaxy in Chapter 5 or in a full cosmological context in Chapter 6, is investigated in the following chapters.

However, it is clear that the sub-grid models described in this work go beyond a basic implementation of an analytic BHL prescription. They are more versatile and reliable, capable of handling a wide range of resolutions and complex flow configurations. In particular, the smooth transition from an algorithm based on the analytic BHL accretion formula to FLA, when local density and velocity features are resolved, ensures that the accretion rate onto the black hole makes the best use of the local information available in the simulation and should quite naturally converge to the correct Newtonian solution once the size of the accretion region approaches the physical size of the black hole. By contrast, the drag force sub-grid algorithm becomes unphysical as resolution increases and extra caution regarding its implementation in less idealised simulations is needed: a question which is investigated in the next chapter.

Chapter 5

A zoom-in on the art of planting supermassive black holes in galaxies

5.1 Introduction

First observed to power quasars at redshift $z \geq 6$ (Fan et al., 2006), supermassive black holes (SMBHs) formed during the first billion years of the Universe. This epoch saw not only the formation of SMBHs but also of many other key components of the Universe: the first halos collapsed, the first gas streamed in and the first stars ignited. Given the current paucity of observations from this era, we have to rely almost entirely on simulations to understand the early evolution of SMBH, but numerical efforts are hampered by the vast range of scales involved. Including all relevant length scales to monitor the co-evolution of a single SMBH and its host galaxy, from the Mpc scales of cosmic filaments through the kpc and pc scales of the galaxy, down to the sub-pc scales on which the black hole accretes gas, in a single simulation constitutes a tremendous computational challenge. The problem does not lend itself well to idealised simulations because the evolution of a black hole depends crucially on the evolution of its host galaxy, and the host galaxy itself is evolving rapidly. Cosmological zoom simulations, where a high resolution region containing all the progenitors of a particular object is embedded in a lower resolution background, is one way to maximise the range of scales resolved in a single simulation. Zoom simulations have given crucial insight into galaxy evolution in the early Universe (Di Matteo et al., 2008; Dubois et al., 2012b; Ceverino et al., 2017), but even in the most resolved zoom simulations available to date, the black hole environment remains under-resolved.

Physically, three viable channels have been proposed to form the progenitors of SMBHs, which produce black hole seed masses in the range $10^1 - 10^5$ $\mathrm{M}_\odot$ (see Section 2.2.6). Whatever seed mass the black hole forms with, it needs to gain orders of magnitude in mass to reach the observational estimates of $\sim 10^9$ $\mathrm{M}_\odot$ at $z = 6$ (Fan et al., 2006), so gas accretion remains a key process of its early evolution. Even in the highest resolution simulations currently available, the smallest resolution extends orders of magnitude beyond the physical size of the black hole, whose event horizon is best measured in au. In the absence of sufficient resolution, one has to rely on sub-grid algorithms that include the impact of unresolved scales using analytic or empirical models. Arguably the most common way to describe unresolved accretion onto black holes is through the Bondi-Hoyle-Lyttleton (BHL) model (Hoyle & Lyttleton, 1939; Bondi & Hoyle, 1944; Bondi, 1952), introduced in Section 3.5.2 and discussed in detail in Chapter 4.

In galaxy evolution simulations, the main difficulty with such BHL-like models lies in correctly estimating the input density, sound speed and relative velocity of the gas and the black hole. To compensate for unresolved cold, dense gas, some authors advocate using a boost factor in the accretion rate (Booth & Schaye, 2009; Vogelsberger et al., 2013; DeGraf et al., 2017). Conversely, other authors prescribe scaling the accretion rate down, as the BHL model does not account for angular momentum, which significantly influences gas dynamics (Krumholz et al., 2004; Power et al., 2011; Dubois et al., 2012a; Rosas-Guevara et al., 2015; Curtis & Sijacki, 2016b). Negri & Volonteri (2017) study the accuracy of BHL accretion for a black hole with feedback at various resolutions, and find that it can lead to both over- and under-estimates of the black hole mass, depending on the resolution and the way variables are calculated. Gaspari et al. (2013) systematically break the assumptions of Bondi accretion by cooling, heating or stirring the gas, and find that the accretion rate is boosted several orders of magnitude above the Bondi value. MacLeod & Ramirez-Ruiz (2015) study the accretion onto a point mass in the presence of a pressure gradient at infinity, and find the accretion rate onto the black hole to be reduced by up to two orders of magnitude.

In light of these discrepancies, efforts have been made to develop alternative accretion algorithms, which use large-scale properties of the host galaxy to calculate the accretion rate onto the black hole, and naturally account for the angular momentum of accreted gas. Anglés-Alcázar et al. (2013), building on a model by Hopkins & Quataert (2011), compute the accretion rate onto the black hole based on a measure of large scale gravitational torques in the galaxy. DeBuhr et al. (2010) calculate the accretion rate onto the black hole from the mean surface density of the galaxy, the angular rotation frequency and a free viscosity parameter, and Hobbs et al. (2011) use a "ballistic" accretion model based on the velocity dispersion of the gas in the bulge.

All these models estimate the mass accreted by the black hole from information on scales much larger than the gravitational zone of influence of the black hole itself. While this does represent an educated best estimate, and has led to many insights into the evolution of black holes, it can, on its own, only ever constitute an overly simplistic treatment of the problem. To underpin such an approach, it is essential to perform large sets of higher resolution simulations that improve our knowledge of the behaviour of the gas flow in the vicinity of the black hole in a variety of situations (e.g. embedded in gas rich/poor, interacting/quiescent galaxies) by (at least) reaching scales where the black hole gravity dominates.

With such a goal in mind, this chapter introduces a new refinement algorithm for grid codes, dubbed "zoom-within-zoom" that allows a black hole's environment to be resolved in the full context of its host galaxy at affordable computational cost. It is used to study the impact of up to two extra orders of magnitude in spatial resolution on the local gas properties around a black hole in an highly unsettled, gas rich galaxy intended to mimic a high-redshift mini-halo. Building on work in Chapter 4, a particular focus is placed on testing the behaviour of BHL accretion in such a highly non BHL-like context.

Section 5.2 presents the simulation setup, Section 5.3 gives a detailed description of the novel refinement algorithm and Section 5.4 establishes its numerical convergence. Section 5.5 discusses the choice of black hole seed mass and its impact on the evolution of a black hole's accretion history at a given resolution. Section 5.6 then performs a comparative study of the impact of resolution on the accretion history. A discussion can be found in Section 5.7, and conclusions in Section 5.8.

5.2 The simulations

All simulations presented in this chapter use the adaptive mesh refinement code RAMSES (Teyssier, 2002), using a Courant factor of 0.6 (see Chapter 3 for details). The simulation box is refined up to a root grid of 64^3, and then adaptively refined up to level l_{glob} outside the zoom-within-zoom region, and to l_{zoom} within this region (see Sections 5.3 for details). The size of an individual cell at level l is equal to $\Delta x_l = \frac{30}{2^l}$, as $L_{\mathrm{box}} = 30$. Refinement is determined using a quasi-Lagrangian criterion: a cell is split into 8 when its total baryonic mass exceeds 8 times the average gas mass it initially contained. To avoid numerical fragmentation (Truelove et al., 1997), cells are also refined so the local Jeans length exceeds the cell size at all times.

5.2.1 Initial conditions

In order to test the refinement algorithm in a controlled environment, it is first implemented in an isolated cooling halo with a Navarro-Frenk-White profile (NFW profile) (Navarro & White, 1994) for the gas and dark matter components, similar to the one presented in Dubois & Teyssier (2008), before being used in a cosmological context in Chapter 6. The dark matter potential of the halo is modelled as a fixed analytic potential, where:

$$\rho_{\mathrm{NFW}}(r) = \frac{\rho_s}{(r/r_s)(1+r/r_s)^2}, \tag{5.1}$$

so the halo has a total integrated mass

$$M(<r) = 4\pi\rho_s r_s^3 \bigg(\ln(1+r/r_s) - \frac{r/r_s}{1+r/r_s} \bigg), \tag{5.2}$$

where r_s and ρ_s are the halo's characteristic radius and density respectively. $\rho_s = 200\rho_{\mathrm{c}}$, where $\rho_{\mathrm{c}} = \frac{3H_0^2}{8\pi G}$ is the current critical density, assuming a Hubble parameter of $H_0 = 70\ \mathrm{km\,s^{-1}Mpc^{-1}}$. The characteristic radius, r_s, is set using the concentration parameter $c = \frac{r_{200}}{r_{\mathrm{s}}} = 3.5$ in combination with $v_{200} = 85\ \mathrm{km\,s^{-1}}$. The halo is truncated at two virial radii, with densities outside set to a uniform value of $\rho_{\mathrm{IGM}} = 3.27 \times 10^{-9}\ \mathrm{H\,cm^{-3}}$, and embedded in a box of $L_{\mathrm{box}} = 1.0$ Mpc. Gas is initiated as $\rho_{\mathrm{gas}} = 0.15\rho_{\mathrm{NFW}}$ within the halo, with the initial pressure profile set according to hydrostatic equilibrium. The total halo mass is therefore $M_{\mathrm{tot}} = 5.1 \times 10^{11}\ \mathrm{M_\odot}$.

When adding rotation (Section 5.5 and Section 5.6 only), angular momentum is distributed according to $j(r) = j_{\mathrm{max}}\frac{M(<r)}{M_{\mathrm{vir}}}$, following Bullock et al. (2001), normalised to a spin parameter $\lambda = J|E|^{1/2}/(GM_{\mathrm{vir}}^{5/2}) = 0.04$, where J and E are the total angular momentum and total energy respectively.

5.2.2 Cooling, heating and star formation

Above 10^4 K, the gaseous mix of H, He and electrons radiatively cools through collisions following the pristine cooling function tabulated by Sutherland & Dopita (1993). Below this temperature radiative cooling occurs via H_2 emission following Grassi et al. (2014). The gas follows an ideal equation of state and is assumed to be mono-atomic with adiabatic index $\gamma = 5/3$.

Star formation proceeds as described in Section 3.4, using a constant star formation efficiency of $\epsilon_* = 0.02$ and a star formation number density threshold of $\rho_0/\mu m_\mathrm{H} = 4.73\times10^3\ \mathrm{H\,cm^{-3}}$, where m_H is the atomic weight of a hydrogen atom and μ the mean molecular weight. Star particles are spawned with masses of $M_* = 149.2\ \mathrm{M_\odot}$, or integer multiples thereof. No stellar feedback is included.

5.2.3 Black hole formation and accretion

Black hole formation sites are identified on the fly using the structure-finding PHEW algorithm (Bleuler et al., 2015), which locates gravitationally bound, collapsing gas clumps in cells above the threshold density of $\rho_\mathrm{t} = 10^3\ \mathrm{H\,cm^{-3}}$ (see Section 3.5.1). After a potential formation site has been identified, a sink particle (Krumholz et al., 2004; Dubois et al., 2010) is inserted in the clump's densest cell, with a user-defined initial sink mass, M_seed. To conserve mass and momentum, an equivalent amount of mass is removed from the host cell, and the new sink particle inherits the host cell's velocity vector. Only one black hole is formed per simulation.

Accretion proceeds according to the BHL accretion rate in Equation 3.18 (as described in detail in Section 3.5.2) and the characteristic scale radii, R^B and R^A, are calculated according to Equation 3.15 and Equation 3.17 respectively. The accretion rate onto the sink therefore depends on the properties of the gas from which it accretes. To sample the local environment, RAMSES uses so-called "cloud particles", which are distributed with constant spacing $\Delta x_\mathrm{min}/2$ around the sink particle, filling a sphere with radius $r_\mathrm{cloud} = 4\Delta x_\mathrm{min}$ (see Section 3.5.2.3). Each cloud particle samples the gas properties of the cell that contains it. Unlike in Chapter 4, the simulations in this chapter do not include a density floor. Accretion is also not limited to the Eddington rate (Equation 2.5), which is plotted for comparison only.

5.2.4 Dynamical friction

Section 5.5 explores the option of adding a sub-grid drag force to the black hole to compensate for unresolved dynamical friction due to the gaseous background. It is calculated according to an analytic formula by Ostriker (1999), following the model described in Section 3.5.4 and using $\ln(\Lambda) = 4.0$ based on Chapter 4.

5.3 The zoom-within-zoom refinement algorithm

In order to reach high resolution in the vicinity of the black hole, but keep the simulations numerically affordable, I developed a refinement algorithm called "zoom-

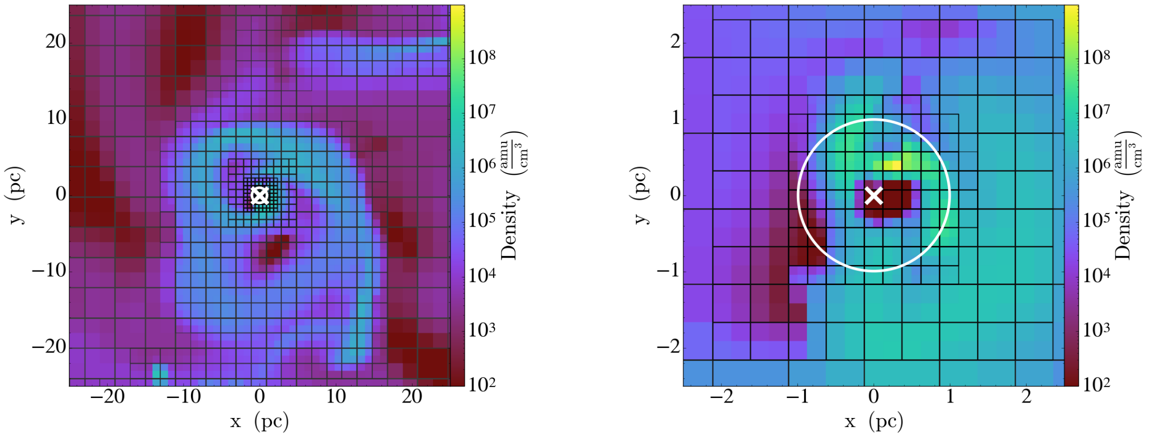

Figure 5.1: Grid structure due to the zoom-within-zoom refinement algorithm, at different length scales, plotted over a gas density projection. The example shown here has three extra levels, so $l_{\mathrm{zoom}} = l_{\mathrm{glob}} + 3$. The black hole is marked by the white cross, and the radius of the zoom-within-zoom region, r_{zoom}, by the white circle. Outside the zoom-within-zoom region, the galaxy is adaptively refined based on a quasi-Lagrangian criterion.

within-zoom". This super-Lagrangian refinement scheme surrounds the sink particle with a spherical region at fixed, user-defined resolution $\Delta x_{\mathrm{zoom}} = L_{\mathrm{box}}/2^{l_{\mathrm{zoom}}}$, embedded in an adaptively refined galaxy at lower resolution, $\Delta x_{\mathrm{glob}} = L_{\mathrm{box}}/2^{l_{\mathrm{glob}}}$, where l describes a refinement level. The highly refined spherical region around the black hole has a radius of $r_{\mathrm{zoom}} = n_{\mathrm{zoom}}\Delta x_{\mathrm{zoom}}$, where n_{zoom} is a free parameter kept fixed throughout the simulation (see Section 5.4.1 for a convergence study). In order to minimise edge effects at grid boundaries, the high refinement region is surrounded by concentric shells of progressively lower refinement until Δx_{glob} is reached, as can be seen in Figure 5.1. This ensures that neighbouring cells differ at most by a single refinement level. The refinement region is centred on the sink particle position at all times, tracking its movement through the box, with cells (de-)refining as appropriate.

Once a black hole particle has been created, the zoom-within-zoom algorithm is activated. To allow newly created cells to be distributed over several processors, and to minimise numerical artefacts due to the number of levels being added, cells are only allowed to be refined by two extra levels per coarse timestep of the simulation. The simulation is load-balanced after each coarse timestep during the level triggering phase. Accretion onto the black hole is only activated once the target level, l_{zoom}, has been reached, to allow the structure of the gas flow to emerge and avoid contaminating the early accretion history of the black hole. As explored in detail in Section 5.5, a maximum drag force is applied to the sink during the initial level trigger phase which updates the relative velocity, $v_\bullet$, such that $v_\bullet = 0$ at each timestep.

The mass of newly created star particles is dependent on the mass of the cells in which they form, with smaller cells forming lighter star particles. To avoid overwhelming the simulation with a large number of star particles that carry only a small amount of mass, star formation is prevented in any cell with $\Delta x < \Delta x_{\rm glob}$. This has negligible influence on the total stellar mass of the galaxy (see Section 5.6).

RAMSES uses a cloud-in-cell solver to calculate the total density distribution to solve the Poisson equation, where the mass of each particle is distributed over local grid cells according to

$$W(x - x_p) = \begin{cases} 1 - |x - x_{\rm p}|/\Delta x & |x - x_{\rm p}| \leq \Delta x \\ 0 & \text{otherwise,} \end{cases} \tag{5.3}$$

where x and $x_{\rm p}$ are the positions of the cell centre and particle, and Δx is the cell size. In simulations with a large range of refinement levels, this can lead to spurious local maxima in the density field if a massive particle is deposited into an extremely small cell. To avoid this issue in zoom-within-zoom, star particles are deposited at a maximum level $l_{\rm glob}$. The OctTree structure is used to distribute mass into child cells at higher refinement levels. This is a commonly used technique for massive dark matter particles in cosmological zoom simulations (e.g. Powell et al., 2011).

5.4 The behaviour of black holes in spherically collapsing halos

To test the convergence behaviour of the algorithm and thus set the values of the parameters, $n_{\rm zoom}$ and $l_{\rm zoom}$, a black hole of mass $M_{\rm BH} = 260\ {\rm M}_\odot$ is placed at the centre of a non-rotating version of the halo described in Section 5.2.1 and the gas is allowed to cool. The accretion rate onto the sink, as well as the mass weighted density, sound speed and relative velocity as measured by the cloud particles around the sink, are compared for a range of parameters. The initial density distribution is refined up to $l_{\rm glob}$, so a higher $l_{\rm glob}$ leads to a more peaked initial density profile. The black hole is inserted at the centre of the simulation at $t = 0 \times\ t_{\rm ff}/10^3$, where $t_{\rm ff}$ is the free fall time of the halo. The zoom-within-zoom levels are added during the first coarse timesteps of the simulation until $\Delta x_{\rm min} = \Delta x_{\rm zoom}$, where $\Delta x_{\rm min}$ is the size of the smallest cell in the simulation.

Convergence test simulations

name	$l_{\rm glob}$	$l_{\rm zoom}$	$\Delta x_{\rm glob}$ [pc]	$\Delta x_{\rm zoom}$ [pc]	$n_{\rm zoom}$
C_l20n2	16	20	15.8 pc	0.99 pc	2
C_l20n4	16	20	15.8 pc	0.99 pc	4
C_l20n8	16	20	15.8 pc	0.99 pc	8
C_l20n16	16	20	15.8 pc	0.99 pc	16
C_halo16*	16	-	15.8 pc	–	-
C_halo20*	20	-	0.99 pc	–	-
C_l18n8	16	18	15.8 pc	3.9 pc	8
C_l20n8	16	20	15.8 pc	0.99 pc	8
C_l22n8	16	22	15.8 pc	0.25 pc	8
C_l24n8	16	24	15.8 pc	0.06 pc	8

* equivalent to setting $n_{\rm zoom} = 0$

Table 5.1: Parameters for simulations used for the convergence tests in Section 5.4.

5.4.1 The impact of the size of the zoom-within-zoom region

To quantify the the impact of varying $n_{\rm zoom}$, zoom-within-zoom simulations with $2 \leq n_{\rm zoom} \leq 16$ are compared to two fiducial simulations using only the standard quasi-Lagrangian refinement, called C_halo16 and C_halo20 respectively (see Table 5.1 for details). When not using zoom-within-zoom, the black hole host cell is forced to remain at $l_{\rm glob}$ to avoid spurious local de-refinement. For the zoom-within-zoom simulations, $l_{\rm glob} = 16$ and $l_{\rm zoom} = 20$, leading to a resolution of $\Delta x_{\rm zoom} = 0.99$ pc within the zoom-within-zoom region, and a maximum of $\Delta x_{\rm glob} = 15.8$ pc outside.

Figure 5.2 shows that zoom-within-zoom refinement allows the gas in the vicinity of the black hole to evolve from the initial conditions of a low resolution simulation, C_halo16, to the solution of the high resolution simulation, C_halo20. Convergence behaviour of mass weighted density, $\rho_\bullet$, and sound speed, $c_\bullet$, is very good for $n_{\rm zoom} \geq 4$. When $n_{\rm zoom} < 4$, gas properties fail to converge as the maximum extent of the cloud particles, set at $r_{\rm cloud} = 4\Delta x_{\rm zoom}$, is larger than the size of smallest zoom-within-zoom grid, $r_{\rm zoom} = n_{\rm zoom}\Delta x_{\rm zoom}$. When this happens, the outermost cloud particles probe regions outside $r_{\rm zoom}$ where gas properties reflect the density profile a factor of two further away than if the cell were refined an extra level.

One notable feature of Figure 5.2 is the discontinuity in the evolution of the gas near $t = 1.8\ t_{\rm ff}/10^3$. This change in local quantities occurs when the black hole transitions to flux limited accretion (FLA). While the sub-grid algorithm remains unchanged, the amount of mass it schedules for accretion (calculated using Equation 3.18), $\dot{M}_{\rm BHL} \times \Delta t = G M_{\rm BH}^2 \rho_\bullet \Delta t/(v_\bullet^2 + c_\bullet^2)^{3/2}$, exceeds the total gas mass available within the accretion region, $\rho_\bullet V_\bullet$, where $V_\bullet$ is the constant volume of the

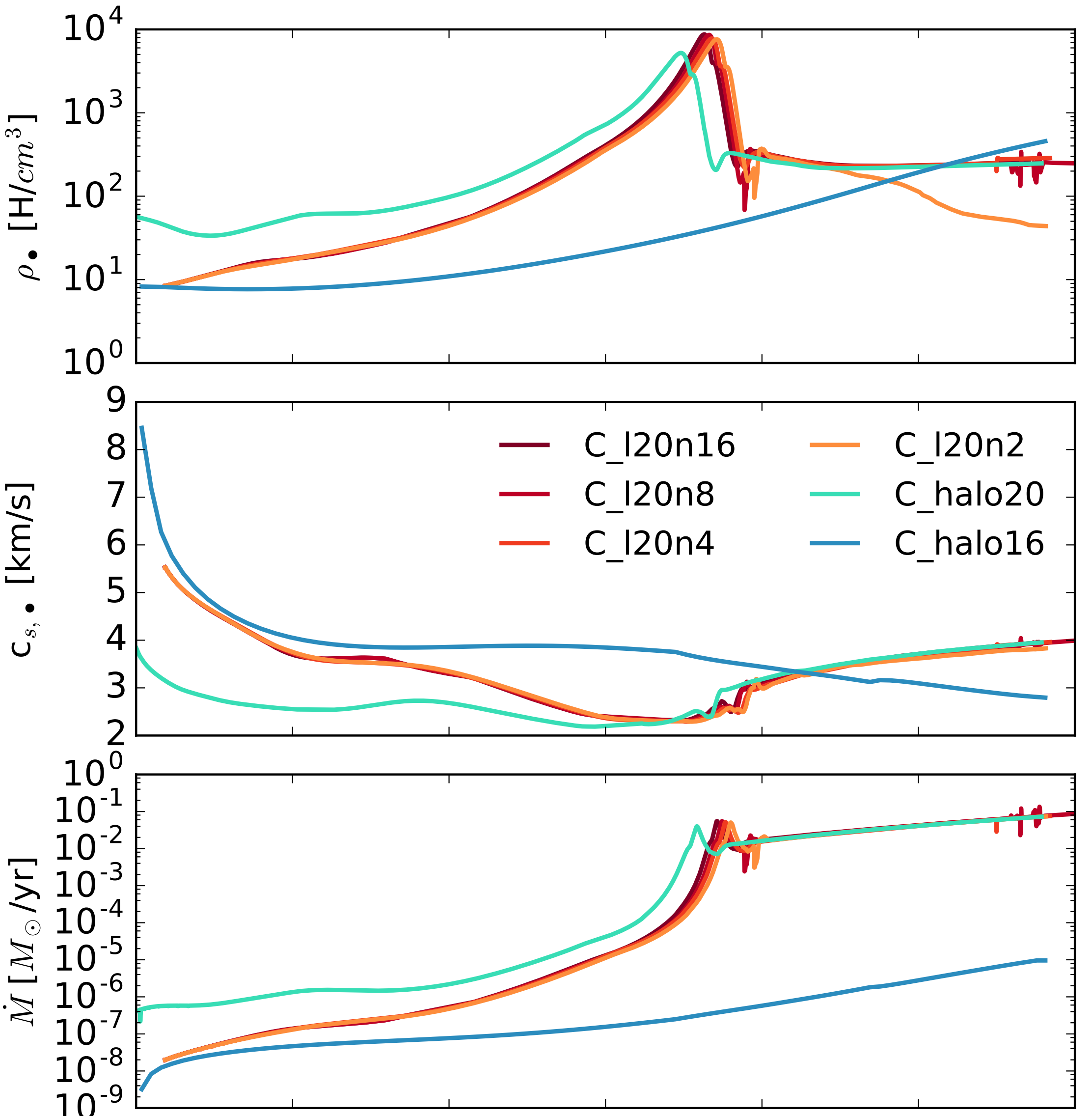

Figure 5.2: Density (top panel), sound speed (middle panel) and accretion rate (bottom panel), as measured by the black hole, for a range of different radii of the zoom-within-zoom region, $r_{\rm zoom} = n_{\rm zoom}\Delta x_{\rm zoom}$ (see Table 5.1). Convergence is achieved for $n_{\rm zoom} \geq 4$.

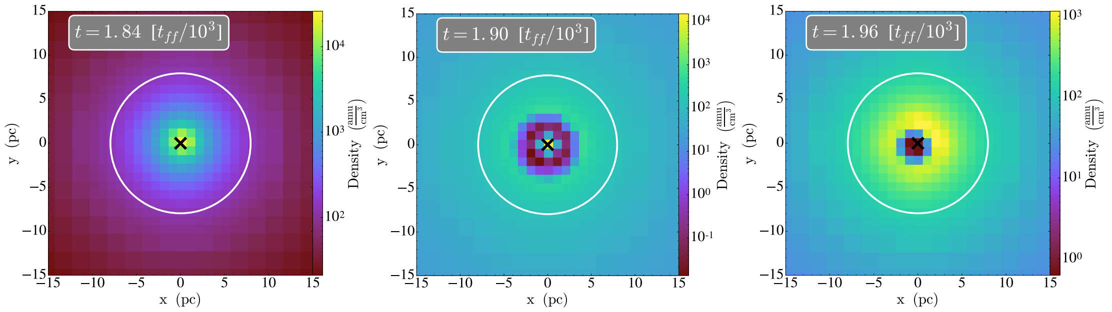

Figure 5.3: Density slices of the gas in the vicinity of the black hole in C_l20n8 at various epochs during the transition from BHL accretion to FLA. The black cross denotes the position of the black hole, and r_{zoom} is marked by the white circle.

accretion region and Δt is the timestep. This occurs for a minimum black hole mass of:

$$M_{\text{FLA}} = \sqrt{\frac{V_\bullet}{G^2 \Delta t}} (v_\bullet^2 + c_{s,\bullet}^2)^{3/2}. \tag{5.4}$$

As mass is conserved, $M_{\text{BH}} > M_{\text{FLA}}$ results in the maximum available mass being removed at each timestep, transitioning the accretion rate from being based on the BHL formula to being set by the mass flux into the accretion region. This occurs approximately when the scale radius is equal to the resolution (see Section 4.3).

The black hole remains in FLA as long as $M_{\text{BH}} > M_{\text{FLA}}$. At fixed l_{zoom}, $V_\bullet$ and Δt are approximately constant. In the spherically symmetric case studied here, $v_\bullet \approx 0$, and Figure 5.2 shows that $c_{s,\bullet}$ varies by less than a factor of two over the course of these simulations. For all intents and purposes, the right-hand side of Equation 5.4 can therefore be considered constant, and the black hole remains in FLA as $M_{\text{BH}} > M_{\text{FLA}}$ continues to hold. If there exists a non-negligible relative velocity $v_\bullet > 0$, conservation of momentum during accretion decreases the relative velocity, lowering M_{FLA} while increasing M_{BH} and thus driving the system towards the FLA regime. In less idealised simulations, where $c_{s,\bullet}$ and $v_\bullet$ can vary strongly, it is possible for the black hole to transition back to BHL accretion, but again, gas cooling and accretion will tend to reduce sound speed and relative velocity respectively, thus lowering M_{FLA} while increasing M_{BH} and steering towards FLA.

During the transition to FLA, the central density structure is accreted by the black hole and replaced with a low density accretion region (see Figure 5.3), decreasing the kernel weighted average $\rho_\bullet$. For reasons of numerical stability, only 75% of a cell mass can be removed in a single accretion step, so densest cells are emptied more slowly (middle panel of Figure 5.3). As evident in Figure 5.2, the transition to FLA is therefore not an instantaneous process.

In conclusion, the zoom-within-zoom scheme is not sensitive to the choice of n_{zoom} as long as $n_{\mathrm{zoom}} \geq 4$, so that the highest refinement region is larger than the accretion region, i.e. $r_{\mathrm{zoom}} \geq r_{\mathrm{cloud}}$.

5.4.2 The impact of resolution in the high refinement region

In order to investigate the impact of resolution within r_{zoom} on the accretion history of the black hole, this section presents a comparison of simulations which have $l_{\mathrm{glob}} = 16$, $n_{\mathrm{zoom}} = 8$, and $16 \leq l_{\mathrm{zoom}} \leq 20$ (see Table 5.1). $l_{\mathrm{zoom}} = 16$ means that $l_{\mathrm{glob}} = l_{\mathrm{zoom}}$, which is equivalent to C_halo16.

Figure 5.4 shows that black hole accretion rates and masses converge after the transition to FLA for all resolutions probed here. Despite starting out with identical initial conditions, simulations with higher l_{zoom} transition earlier because gas collapses more effectively in the black hole vicinity. The dense gas increases the accretion rate during the BHL phase, allowing the black hole to build up mass faster. Simulations with higher l_{zoom} also have smaller accretion regions, which lead to smaller transition masses as $M_{\mathrm{FLA}} \propto \sqrt{V_\bullet(l_{\mathrm{zoom}})}$.

Figure 5.4 shows that the time-averaged value of the density, $\rho_\bullet$, is resolution dependent. In FLA, exactly 75% of the mass contained in the accretion region is removed. The residual mass, M_{res}, left in the accretion region at the end of an accretion step, adjusts until the inflowing mass between accretion steps, M_{inflow}, becomes exactly equal to 75 % of the total mass from which the black hole accretes, i.e. until $M_{\mathrm{inflow}} = 3M_{\mathrm{res}}$. The process is self-balancing, as surplus gas is deposited in, or removed from, M_{res} until it holds. For the free-falling gas in this test case, M_{inflow} is constant and independent of the radius through which it is measured, and therefore approximately resolution-independent, but M_{res} depends on the size of the accretion region. To keep M_{res} constant, simulations with smaller accretion regions have higher density, as seen in Figure 5.4.

Considering the halo collapse is spherically symmetric, the relative velocity as measured by the cloud particles, $v_\bullet$, is surprisingly high, with more resolved simulations exhibiting a higher relative velocity. All simulations with $l_{\mathrm{zoom}} \geq 20$ develop an oscillatory behaviour that appears in the density, $\rho_\bullet$, the relative velocity, $v_\bullet$, and to a lesser extent in the sound speed, $c_\bullet$. Higher resolution simulations get unsettled earlier and oscillate with shorter periods.

To investigate the origin of these oscillations and the high $v_\bullet$, the original simulation, C_l20n8, where momentum is conserved during accretion and the black hole moves under gravity, is compared to two simulations where the black hole position is

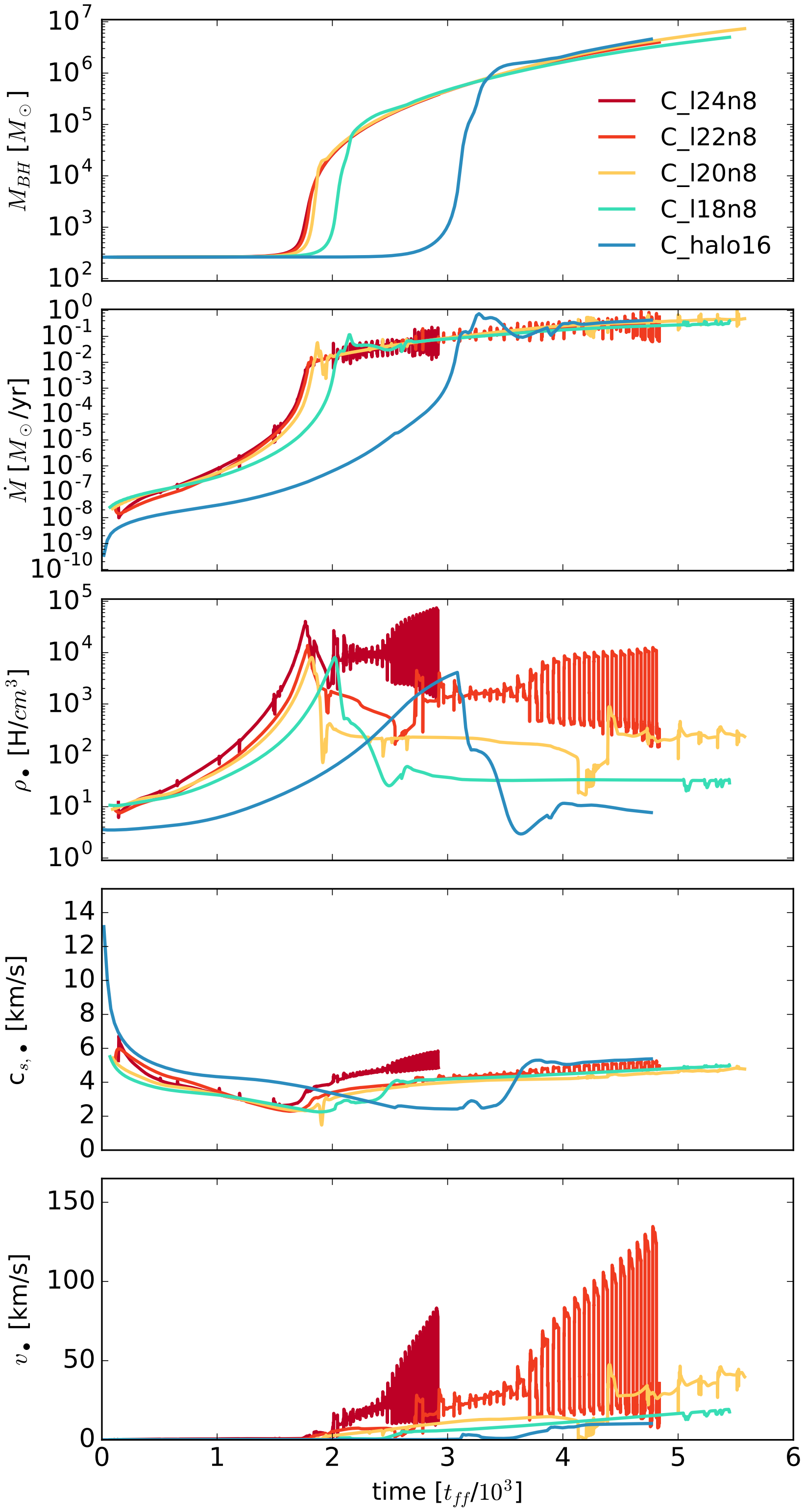

Figure 5.4: Time evolution of gas and black hole properties, for simulations with different $\Delta x_{\rm zoom}$ but otherwise identical in all aspects (see Table 5.1 for details).

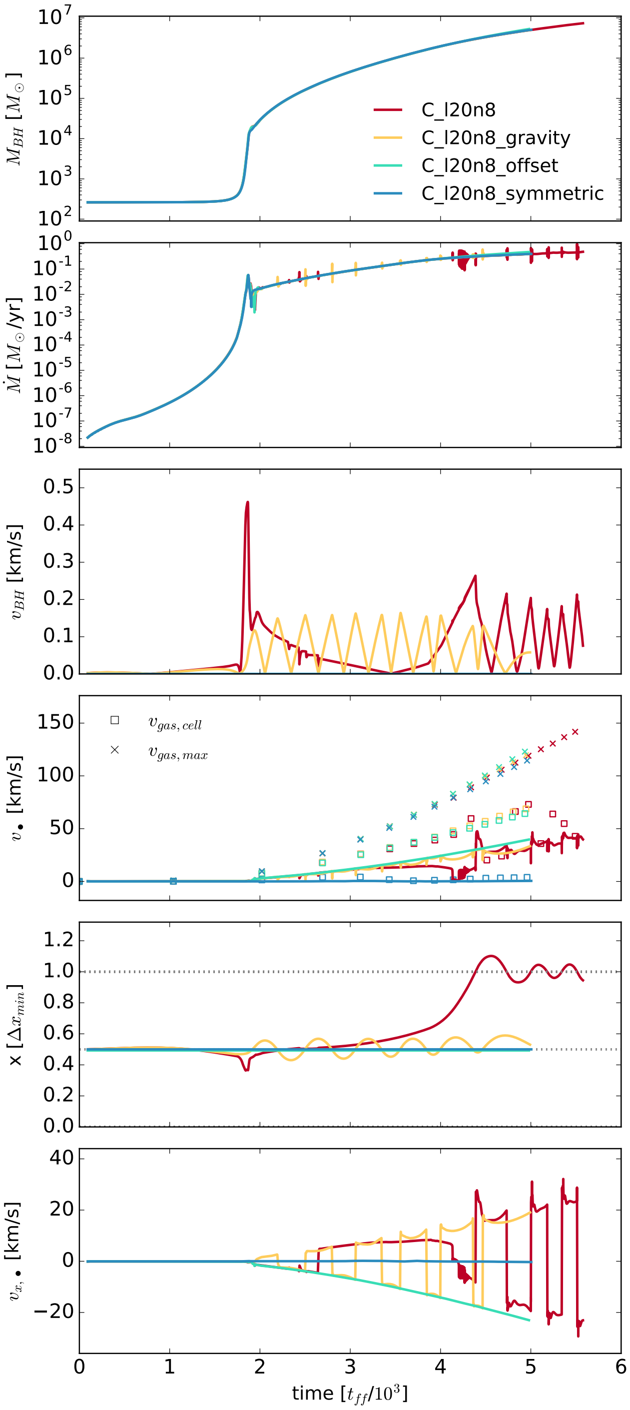

Figure 5.5: Time evolution of various quantities associated with the black hole including mass, M_{BH}, and accretion rate, $\dot{M}_{\mathrm{BH}}$. v_{BH} is the absolute velocity of the black hole in the frame of the box, $v_\bullet$ is the magnitude of relative velocity between gas and black hole as measured by the cloud particles and $v_{x,\bullet}$ is its x-component. Square markers denote the relative velocity between black hole and its host cell, $v_{\mathrm{gas,cell}}$, and crosses show the maximum relative velocity within the region probed by the cloud particles $v_{\mathrm{gas,max}}$. x is the x-coordinate of the black hole's position and dotted grey lines on the fifth panel denote the cell boundaries. Details of the simulation setups can be found in the text.

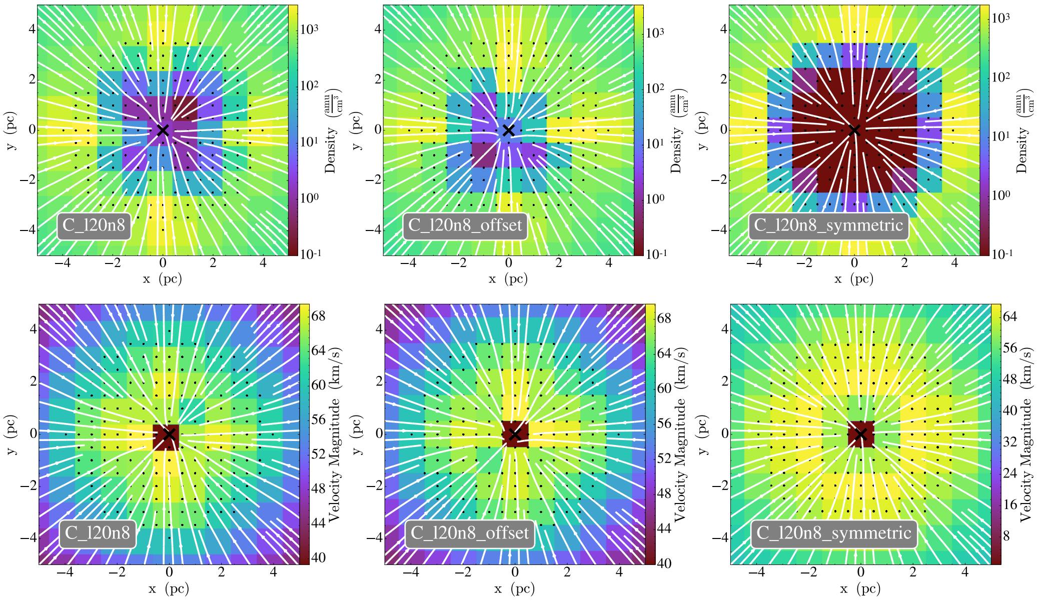

Figure 5.6: Gas density and velocity magnitude slices in the accretion region of the black hole for C_l20n8 (left), C_l20n8_offset (center) and C_l20n8_symmetric (right) show that only cells from which mass is removed during accretion are affected by the lack of symmetry. Streamlines are annotated in white, and the location of the black hole is marked by a black cross. Cloud particle positions are shown as black dots.

held fixed. In C_l20n8_symmetric, the black hole is located exactly at the cell centre, and the cloud particles are spaced using $\Delta x_{\mathrm{cloud}} = \Delta x_{\mathrm{zoom}}/2.02$ to preserve symmetry and avoid ambiguity in deciding a cloud particle's host cell. In C_l20n8_offset, the black hole is also fixed but its position is shifted in all three dimensions by $-0.01 \times \Delta x_{\mathrm{zoom}}$ relative to the cell centre. In this simulation the cloud particles remain spaced at $\Delta x_{\mathrm{zoom}}/2.0$ and now preferentially probe cells to the bottom left of the black hole in the frame of the box. For comparison one final simulation is added, C_l20n8_gravity, in which the black hole is only accelerated by gravity, i.e. where momentum is not conserved during accretion. For all simulations, the local minimum of the analytic potential is set to the initial position of the black hole.

Figure 5.5 shows that, as expected, accretion rates as well as black hole masses converge for all four simulations. As well as $v_\bullet$, the fourth panel from the top of Figure 5.5 shows the maximum inflow velocities at the bottom of the potential well (crosses), and the velocities of the cell hosting the black hole (squares), which coincide with the local velocity minimum. In all simulations, $v_\bullet$, is significantly lower than the corresponding inflow velocity and equal to or lower than the corresponding velocity of the black hole host cell. The velocity of the black hole with respect to the box, v_{BH}, is very small in comparison to $v_\bullet$, but does show the same oscillations.

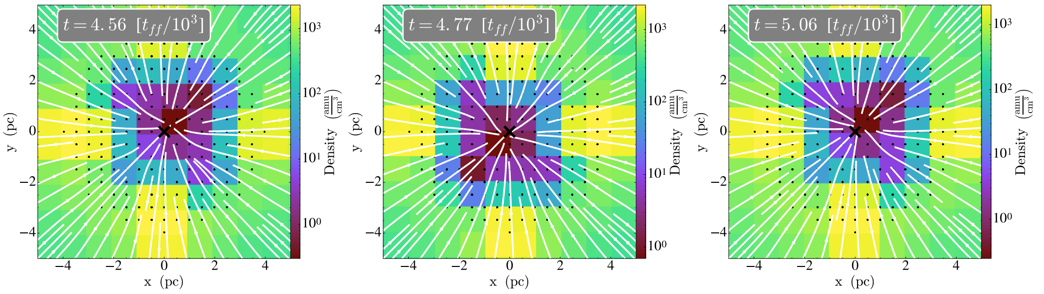

Figure 5.7: Gas density and velocity magnitude slices in the accretion region of the black hole at three points in time show the flipping symmetry as the black hole passes across a cell boundary. Streamlines are annotated in white, and the location of the black hole is marked by a black cross. Cloud particles are shown as black dots.

Figure 5.6 shows that both for C_l20n8 and the deliberately asymmetric simulation, C_l20n8_offset, the density distribution at the edge of the accretion region is not symmetric (see top panel), and the central gas cell has a non-negligible velocity relative to the box (see bottom panel). By contrast, the perfectly symmetric case, C_l20n8_symmetric, has both a spherically symmetric density distribution and a noticeably lower velocity in the host cell (0.01 $\mathrm{km\,s^{-1}}$ compared to $\sim$ 40$\mathrm{km\,s^{-1}}$). The relative velocity as measured by the black hole, $v_\bullet$, therefore simply reflects the fact that fluxes at the bottom of the potential well cancel imperfectly. The vector addition over the cloud particles successfully smoothes out some of the local discretisation in the gas, but a remnant velocity remains. Higher resolution simulations have higher relative velocities because the gas is more strongly accelerated by the deeper potential well.

When the sink is moving only under gravity, in C_l20n8_gravity, $v_\bullet$ has the same magnitude as in the fixed C_l20n8_offset, but its direction flips as the black hole traces out a small circle within the host cell (see bottom two panels of Figure 5.5). As there is no momentum transferred during accretion and the black hole's initial position coincides with the bottom of the analytic potential, the black hole must be accelerated by the gravitational force exerted by the local gas distribution. Figure 5.7 shows that after the transition to FLA at $t = 2\ t_{\mathrm{ff}}/10^3$, the black hole is surrounded by an irregular shell of leftover gas, which indeed imparts a small net gravitational pull on the black hole.

In C_l20n8, the black hole is accelerated by gravity and by momentum conservation during accretion. Even though the relative velocity at $t = 1.9\ t_{\mathrm{ff}}/10^3$ is comparatively small at $v_\bullet = 2.3\ \mathrm{km\,s^{-1}}$, the large amount of mass accreted ($M_{\mathrm{acc,tot}} > 10^5\ \mathrm{M_\odot}$) compared to the initial black hole mass ($M_{\mathrm{init}} = 260\ \mathrm{M_\odot}$) means accreted momentum dominates. The black hole therefore receives a velocity kick during the transition to FLA that dislodges it from its position at the cell centre (see third and fifth panel of Figure 5.5). When entering a new cell, gas is removed preferentially from a different set of cells, changing the symmetry of the local density distribution (see Figure 5.7). The local symmetries in the density distribution flip, the black hole reverses and passes back into the original cell, the symmetries flip again, and the oscillations continue. Note that this issue is purely numerical and only involves cells within the accretion region, outside of which the gas properties remain as spherically symmetric as possible on a Cartesian grid.

This movement of the black hole causes all local mass weighted quantities to vary by several orders of magnitude on short timescales, including the accretion rate, but the cumulative effect on the mass evolution of the black hole is negligible as the accretion algorithm is self-balancing. Any mass not accreted at a given time-step remains on the grid until accreted. As can be seen in the top panel of Figure 5.4, the mass of C_l24n8 is indistinguishable from the mass of stable simulations such as C_l18n8 after more than 2 $t_{\rm ff}/10^3$. These types of numerical oscillatory features are expected to be less prominent for massive black holes in a full astrophysical context. Simulations involving turbulence, other compact objects and angular momentum are unlikely to produce this type of perfectly symmetric collapse for extended periods of time, and local asymmetries in the density distribution around the black hole are thus more likely to be caused by the gas flow itself than by black hole accretion.

I therefore conclude that the accretion algorithm is robust, despite local mass weighted quantities $\rho_\bullet$, $v_\bullet$ and $c_{s,\bullet}$ becoming increasingly less representative of global values as $\Delta x_{\rm zoom}$ decreases, due to small sink movements and asymmetric sampling. However, the impact of FLA on local gas quantities could have important consequences for other sub-grid algorithms, such as the analytic drag force introduced in Section 5.2.4, whose efficiency peaks at Mach number $\mathcal{M} = 1$. When overestimating $v_\bullet$, the black hole will appear supersonic, despite negligible movement with respect to the bulk of the gas, and any attempt to correct for this apparent supersonic motion could un-physically accelerate the black hole.

Note that the simplicity of the setup studied in this section might contribute to the black hole mass convergence. As both the gas and the black hole are gravitationally bound at the bottom of the potential well, the point at which mass is transferred from the grid to the sink particle has no effect on the long term evolution. In a more complex environment, the lifetime of the cloud feeding the black hole could become a concern. From Figure 5.4, the transition to FLA happens around $t \simeq 1.7\ t_{\rm ff}/10^3$,

Disc galaxy simulations			
name	l_{zoom}	M_{seed} [$\mathrm{M}_{\odot}$]	Drag force
D_l26_tiny	26	260	None
D_l26_small	26	5×10^3	None
D_l26_medium	26	2.6×10^4	None
D_l26_big	26	1×10^5	None
D_l26_huge	26	2.6×10^5	None
O_l26	26	260	Ostriker1999
F_l26_a	26	10^{-2}	Max
F_l26_b	26	10^{-1}	Max
F_l26_c	26	1	Max
F_l26_d	26	260	Max
F_l26_e	26	10^4	Max
F_l26_f	26	10^5	Max

Table 5.2: Parameters for simulations in Section 5.5. All simulations have $l_{\mathrm{glob}} = 20$ and $n_{\mathrm{zoom}} = 8$.

but it is significantly delayed for C_l18n8 and C_halo16. In other words, should the cloud feeding the black hole be disrupted at $t \simeq 1.9\ t_{\mathrm{ff}}/10^3$, before the transition to FLA for C_l18n8 and C_halo16, the difference in black hole mass at that point would have led to very different evolution histories for these two simulations. These issues are addressed in the following section.

5.5 Seeding black holes in collapsing clouds

An important parameter in simulations of early black hole evolution is the choice of seed mass for the black hole. On the one hand, there are physical considerations. As laid out in Section 2.2.6, different formation channels lead to black holes with vastly different masses, from the 100 $\mathrm{M}_{\odot}$ remnants of early stars, to the 10^5 $\mathrm{M}_{\odot}$ black holes formed by direct collapse. On the other hand, there are numerical considerations as well, which are particularly connected to seeding location. In this section, the impact of different seed masses on the evolution of the black hole formed in a collapsing cloud is explored in detail, before presenting an algorithm to automatically determine the black hole seed mass based on the structure of the cloud at the final resolution, Δx_{zoom}.

5.5.1 The role of dynamical friction

To study the formation of black holes in more realistic clouds, rotation is added to the halo described in Section 5.2.1 and it is evolved for 100 Myr at a resolution of $\Delta x_{\mathrm{glob}} = 0.99$ pc (using $l_{\mathrm{glob}} = 20$) to allow the gas to cool and fragment. A collapsing cloud is then identified using the PHEW algorithm, and a sink particle of a given mass M_{seed}, which inherits the position and velocity of its parent cell, is inserted. Once the black hole has been seeded, the zoom-within-zoom algorithm adds levels up to $l_{\mathrm{zoom}} = 26$, using $n_{\mathrm{zoom}} = 8$. The parameters for simulations in this section are summarised in Table 5.2.

As can be seen in Figure 5.8, the evolution history of the black hole depends crucially on the choice of M_{seed}. The two simulations with the most massive seeds, D_l26_big and D_l26_huge, using $M_{\mathrm{seed}} = 1 \times 10^5$ $\mathrm{M}_\odot$ and $M_{\mathrm{seed}} = 2.6 \times 10^5$ $\mathrm{M}_\odot$ respectively, rapidly converge. Both simulations accrete the central dense feature of the cloud and transition to FLA as soon as accretion commences (see Figure 5.9). For the remainder of the simulations, while not identical due to the chaotic nature of the problem, their accretion rates, density $\rho_\bullet$ and relative velocity, $v_\bullet$, show the same trends, including intermittent periods of high frequency fluctuations (see central three panels of Figure 5.8). Their gas supply is replenished by small-scale spiral and bar-like features (Figure 5.9) so their accretion rates remain orders of magnitude above their respective Eddington limits (Equation 2.5, dotted line in Figure 5.8). After a delay of 0.6 Myr, and prolonged growth phase, the mass of D_l26_medium also converges with the two biggest seed masses but the two smallest black holes investigated here, D_l26_tiny and D_l26_small, fail to transition to FLA and do not gain appreciably in mass.

D_l26_tiny and D_l26_small, as well as the early ($t < 100.8$ Myr) evolution of D_l26_medium, show regular oscillations in $\rho_\bullet$, $v_\bullet$ and the accretion rate, with periods much longer than the corresponding oscillations in D_l26_big and D_l26_huge (see Figure 5.8). Contrary to the features seen in the collapsing halo in Section 5.4.2, these oscillations do indicate a global movement of the black hole relative to the cloud. As can be seen in Figure 5.10, the cloud studied here has sufficient angular momentum to collapse to a disc. During this collapse, black holes with $M_{\mathrm{BH}} \geq 10^5$ $\mathrm{M}_\odot$ remain within the plane of the disc, (right hand columns of Figure 5.10), but lighter black holes (left hand columns of Figure 5.10) oscillate perpendicularly to the disc, with an amplitude of vertical motion significantly higher than the disc scale height. The black holes alternate between passing with high velocity through a dense medium (high $\rho_\bullet$, high $v_\bullet$ in the disc plane) and slowing down in a low density environment

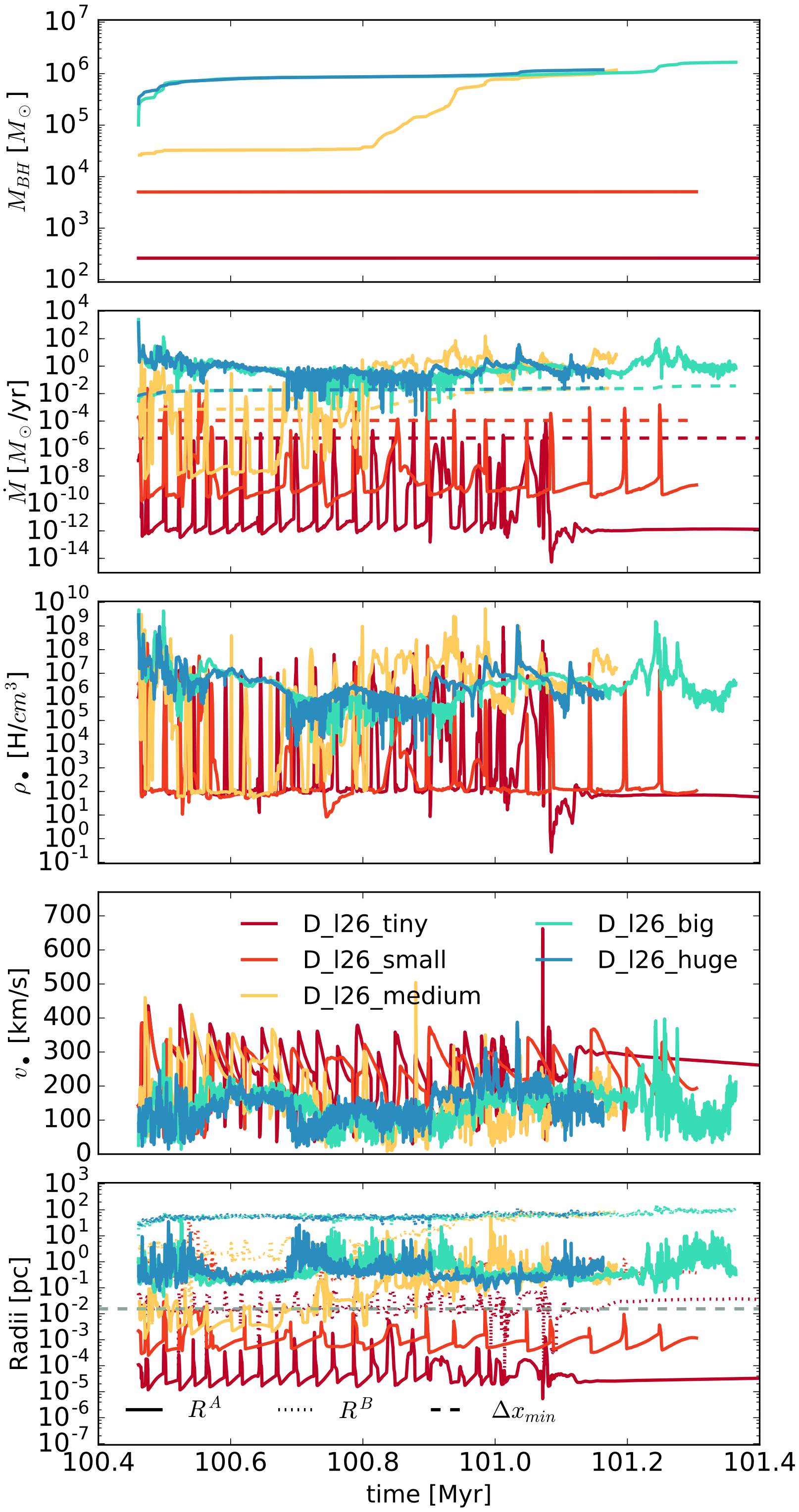

Figure 5.8: Time evolution of mass averaged quantities in the accretion region of black holes seeded with various black hole seed masses in identical collapsing clouds (see Table 5.2 for details). The bottom panel shows the scale radii for each simulation, R^{A} and R^{B} (see Equations 3.15 and 3.17), in comparison to the minimum resolution of the simulation, $\Delta x_{\mathrm{min}} = \Delta x_{\mathrm{zoom}}$, which is plotted as a grey dashed line.

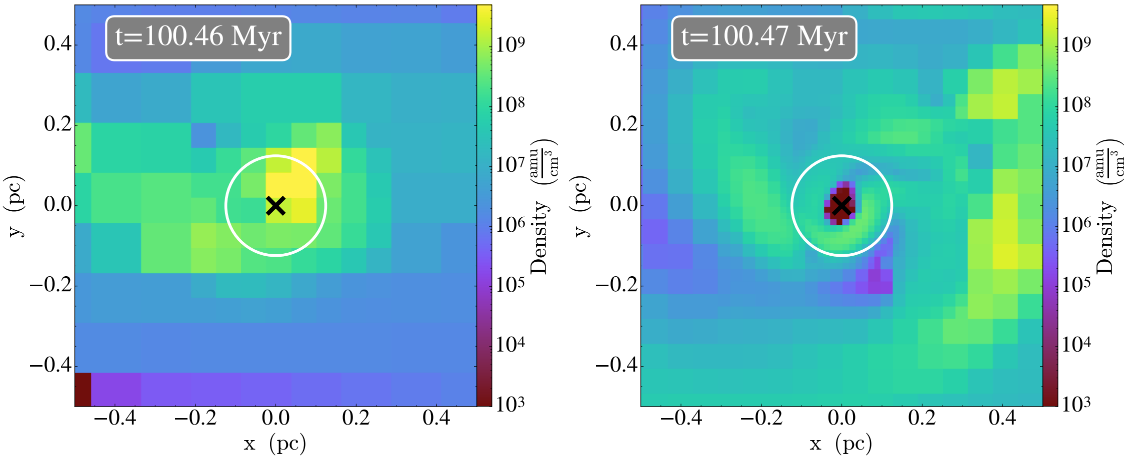

Figure 5.9: The black hole in D_l26_big accretes the central density feature of the collapsed cloud as soon as the highest resolution has been reached, and transitions to FLA. The black hole location is marked by a black cross and the white circle shows the size of the zoom-within-zoom region, r_{refine}. Note that the highest resolution had not yet been reached in the left panel. The free fall time of the gas contained within r_{refine} in the left panel, $t_{\mathrm{ff}} = 1.2 \times 10^{-3}$ Myr, is an order of magnitude smaller than the time between snapshots.

(low $v_\bullet$, low $\rho_\bullet$ above and below the disc). As $\dot{M}_{\mathrm{BHL}} \propto \rho_\bullet / v_\bullet^{2/3}$, neither situation allows the black hole to grow efficiently in the BHL formalism. While it is possible that the choice of sub-grid accretion algorithm has some influence on the results here, the bigger limitation is clearly the inability of the black hole to remain attached to the cloud during collapse.

Keeping sink particles attached to gas features is a common challenge in hydrodynamical simulations (Sijacki et al., 2007; Volonteri et al., 2016; Biernacki et al., 2017). Physically, a black hole has two mechanism to exchange momentum with the gas: (i) accretion, which transfers momentum from the gas to the black hole and (ii) gravitational focusing of the gas into an overdense wake downstream of the black hole, whose gravitational attraction acts as a drag force (see Just & Kegel, 1990, for an analytic derivation). Resolving this wake on the grid requires the gravitational scale radius of the black hole to be similar to the local resolution, i.e. $R^{\mathrm{A}} = \frac{2GM_{\mathrm{BH}}}{v_\infty^2} \sim \Delta x_{\mathrm{zoom}}$ for a supersonic black hole (as explored in detail in Section 4.4), where v_∞ is the relative velocity between black hole and the bulk of the gas. Resolving the drag force is therefore dependent on the local cell size (identical for all D_l26 simulations) and the mass of the black hole. Unfortunately, the drag force is not numerically self-correcting in the sense that accretion is. Indeed, if $R^{\mathrm{A}} > \Delta x$, the black hole reduces its relative velocity, v_{BH}, by transferring momentum to the gas, increasing R^{A} further. By contrast,

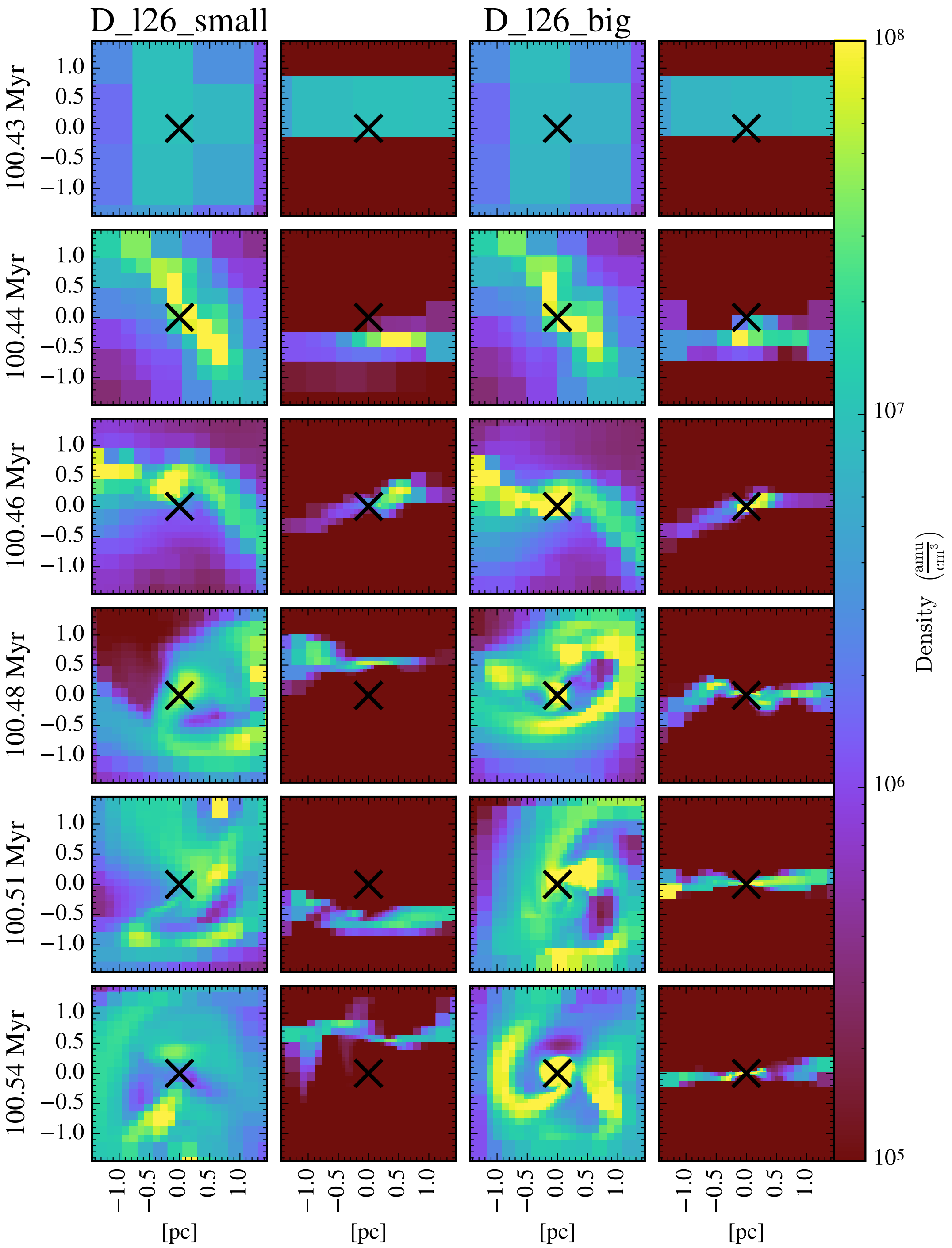

Figure 5.10: Density projections face-on (1st and 3rd column) and density slices edge-on (2nd and 4th column) of the collapsing cloud in D_l26_small (left columns) and D_l26_huge (right columns) at 5 consecutive points in time. The black hole location is marked by a black cross in each panel.

if $R^{\mathrm{A}} < \Delta x$ originally, the relative velocity remains high and R^{A} remains unresolved. Of course, the analytic work on which R^{A} is based assumes that gas on scales much larger than R^{A} is homogeneous, has a uniform velocity, is not self-gravitating, and not subject to an external gravitational potential other than that of the black hole, none of which are reasonable assumptions in the context presented here. However, the final panel of Figure 5.8 shows that, despite these limitations, the accretion radius $R^{\mathrm{A}}_{\bullet}$, evaluated using local mass-weighted quantities, does remain a reliable test of drag force resolution. For the two heaviest seeds, D_l26_big and D_l26_huge, $R^{\mathrm{A}}_{\bullet} > 10\Delta x_{\mathrm{zoom}}$ at all times. The two smallest simulations, D_l26_small and D_l26_tiny, by contrast have $R^{\mathrm{A}}_{\bullet} << \Delta x_{\mathrm{zoom}}$. D_l26_medium has an accretion radius, $R^{\mathrm{A}}_{\bullet}$, that is marginally resolved. Early on, the black hole oscillates but the brief intervals when $R^{\mathrm{A}}_{\bullet} > \Delta x_{\mathrm{zoom}}$ produce sufficient drag on the black hole for the oscillations to decay, allowing the black hole to settle into the disc around $t = 100.7$ Myr. It then accretes the disc's dense core during a period of sustained accretion, transitions to FLA, and converges with D_l26_big and D_l26_huge.

The relative velocity, $v_{\bullet}$, remains high even for black holes settled in the disc, so $R^{\mathrm{A}}_{\bullet} \propto 1/v^2_{\bullet}$ is potentially underestimated. For D_l26_tiny and D_l26_small, $v_{\bullet}$ reflects the motion of the black hole relative to the background medium, apparent in the slice plots shown in Figure 5.10. For the simulations where the black hole has settled at the bottom of the disc's potential well, $v_{\bullet}$ reflects the inflow (or possibly angular) velocity of the gas, as explored in Section 5.4.2. Fortunately, resolving the local radius based on $v_{\bullet}$ ensures the global radius, based on the relative velocity between the black hole and the bulk of the cloud, is also resolved, as $v_{\bullet}$ tends to always be larger than v_{∞}.

Choosing an appropriate black hole seed mass is particularly challenging in the zoom-within-zoom algorithm because of the larger range of scales involved. Too light, and the black hole never grows. Too massive, and too much mass is removed on pre-collapse scales, smoothing out the internal structure of the cloud and polluting the early evolution of the black hole, the very process to be investigated. Just right, and the black hole mass after the FLA transitions reflects the mass of the cloud core at full resolution (see mass convergence in the top panel of Figure 5.8). For the cloud studied in this section, at a resolution of $\Delta x_{\mathrm{glob}} = 0.99$ pc and $\Delta x_{\mathrm{zoom}} = 0.01$ pc, this yields a seed mass range 5×10^4 $\mathrm{M}_{\odot} < \mathrm{M}_{\mathrm{seed}} < 5 \times 10^5$ $\mathrm{M}_{\odot}$, which is quite narrow. Therefore choosing M_{seed} requires detailed prior knowledge about the cloud in question at the chosen resolution. Furthermore, this seed mass will inevitably vary from cloud to cloud.

5.5.2 Choosing a seed mass

In order to make the simulations less reliant on the choice of M_{seed}, which should only reflect the structure of the cloud in which the black hole is formed, the sink particle needs to remain attached to the cloud during the initial phase of the collapse when the extra zoom-within-zoom levels are added. This can be achieved by updating the black hole velocity according to the evolving velocity of the gas, either by simply subtracting the relative velocity (which is equivalent to applying the maximum possible drag force), or by using a more physically motivated algorithm that compensates for the unresolved dynamical friction, such as the analytic model by Ostriker (1999) briefly introduced in Section 5.2.4. Parameters for all simulations used in this section can be found in Table 5.2.

Figure 5.11 shows a comparison of the mass evolution for both methods, including data during the initial phase where the zoom-within-zoom levels are added and accretion is prevented, $t < 104.8$ Myr. The converged solution for massive seeds is represented by D_l26_big. O_l26 denotes the simulation that uses a sub-grid drag force calculated according to the analytic formula by Ostriker (1999). The simulations in blue (M_l26_a - M_l26_f) employ a maximum drag force, which sets $v_{\bullet} = 0$ at each fine timestep. All simulations with a sub-grid drag force only apply the force while $R_{\bullet}^{\mathrm{A}} < 0.2\Delta x_{\mathrm{loc}}$, as work presented in Section 4.4 shows that applying an analytic force based on local mass weighted quantities when the accretion radius is resolved can unphysically accelerate the black hole. Δx_{loc} is the size of the black hole's host cell at a given point in time. Once all refinement levels have been triggered, $\Delta x_{\mathrm{loc}} = \Delta x_{\mathrm{zoom}}$.

As can be seen in the top panel of Figure 5.11, the model based on the linear analytic dynamical friction estimate is unable to solve the problem: the black hole in O_l26 does not grow. Comparing the Mach number for O_l26 to the fiducial D_l26_big shows that the drag force does not significantly influence the motion of the black hole. Physically, the magnitude of the drag force due to the wake is very dependent on the Mach number of the black hole (see Figure 3.6) as increasing the Mach number of the perturber decreases the opening angle of the gravitational wake, which reduces the wake's mass and decreases its gravitational attraction. For a highly supersonic black hole, dynamical friction is inefficient, so this is a case where overestimating $v_{\bullet}$ has significant consequences as it is used to calculate the Mach number $\mathcal{M}_{\bullet}$, which is in turn used to calculate the magnitude of the drag force.

Figure 5.11 shows that the problem occurs as soon as the black hole is formed. Even though the black hole is initiated with zero velocity relative to its host cell, the cloud particles span several cells and $v_{\bullet}$ is not necessarily equal to zero due to the

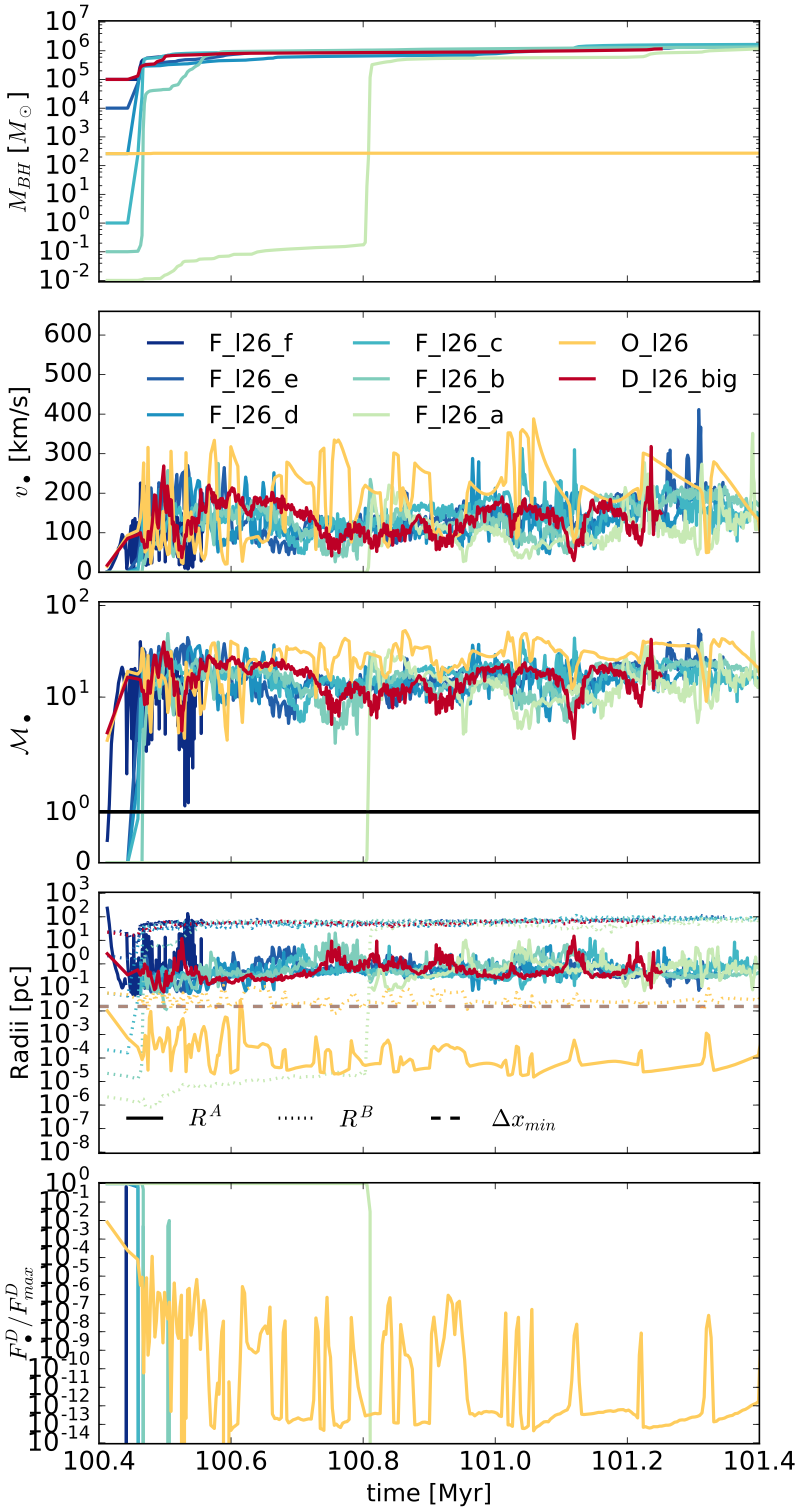

Figure 5.11: Time evolution of black hole related quantities for a range of different seed masses and drag force algorithms (see Table 5.2 for details), including the magnitude of the drag force, F^{D}, and the Mach number, $\mathcal{M}_\bullet$. All curves are time averaged over 200 data-points for clarity. The second to last panel also includes the minimum cell size of the simulation as a dashed grey line, $\Delta x_{min} = \Delta x_{zoom}$, for comparison.

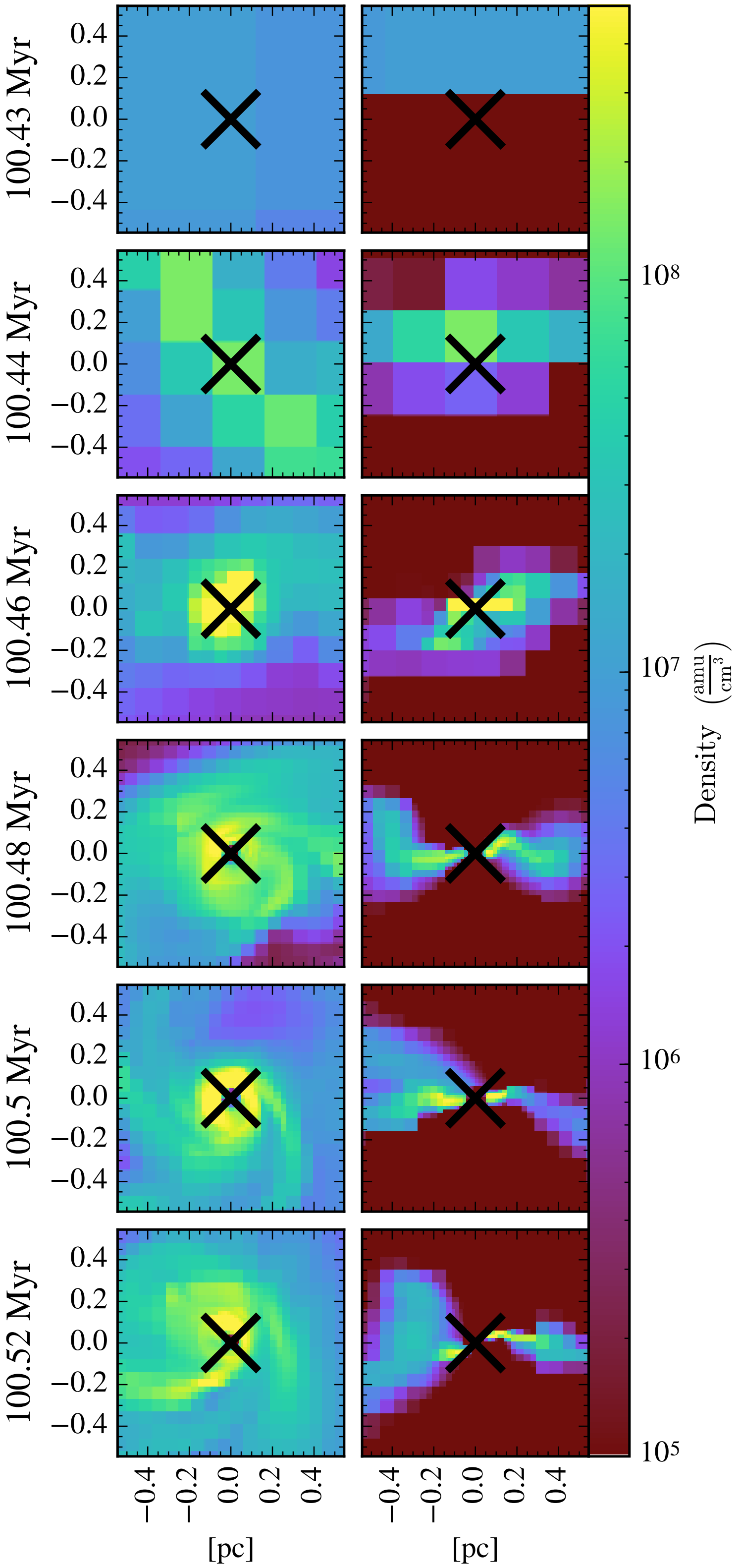

Figure 5.12: Edge-on and face-on density slices for F_l26_a ($M_{\mathrm{seed}} = 1\ \mathrm{M}_{\odot}$) at various epochs, when applying the maximum drag force while $R_{\bullet}^{\mathrm{A}} < 0.2\Delta x_{\mathrm{loc}}$. The black hole is marked by the black cross.

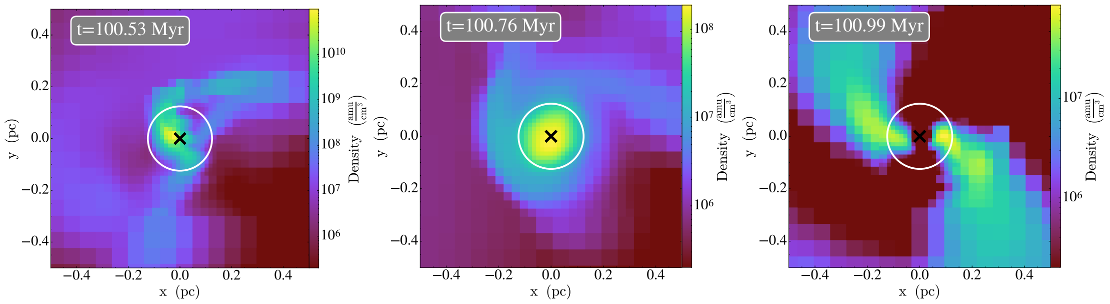

Figure 5.13: Density slices of F_l26_a at three points in time. The location of the black hole is marked by a black cross, and the size of r_{zoom} by a white circle.

local velocity dispersion. For example, in the simulations shown here, $\mathcal{M}_{\bullet,init} \approx 7$ (see O_l26 in the third panel of Figure 5.11). The magnitude of dynamical friction falls off rapidly for $\mathcal{M} > 5$ (see Figure 3.6), and is therefore unable to keep the black hole attached to the collapsing cloud in this simulation. Figure 5.11 shows that this magnitude remains very low throughout the simulation (see bottom panel) and the black hole fails to settle in the potential, as evident by the low-frequency oscillations in the relative velocity (see second panel).

On the other hand, when applying the maximum drag force to the black hole by updating its velocity so that $v_\bullet = 0$ at each fine timestep, it remains attached to the cloud independently of seed mass (see F_l26_a - F_l26_f in Figures 5.11 and 5.12). Only black holes with the lightest seed (F_l26_a, and to some extent F_l26_b) take a significant amount of time to converge. All other black holes accrete the cloud's core once the maximum resolution Δx_{zoom} has been reached, and transition to FLA in the process. As can be seen in the density slices of F_l26_c in Figure 5.12, even black holes with very small seed masses (F_l26_c has $M_{\text{seed}} = 1\ \mathrm{M}_\odot$) settle smoothly into the emerging gas disc. Once in FLA, $R_\bullet^{\mathrm{A}} > 0.2\Delta x_{\text{zoom}}$ for all simulations presented here, and the sub-grid algorithm becomes inactive.

For F_l26_a, the black hole initial mass is below the transition mass (Equation 5.4) and so it grows via BHL until it is sufficiently massive to transition to FLA. While in the BHL regime, the drag force remains active (bottom panel in Figure 5.11), and the black hole stays attached to the cloud core (see density slices in Figure 5.13). Like in the spherical collapse case presented in Section 5.4.2, both gas and black hole are gravitationally bound so the delay in accretion has no influence on the long-term evolution. This could change in the presence of feedback which could disrupt the cloud before the black hole is sufficiently massive.

One difficulty in applying an analytic drag force, whether based on dynamical

Disc galaxy simulations				
name	l_{zoom}	Δx_{min} [pc]	M_{seed} [$\mathrm{M}_{\odot}$]	CPU hours*
R_l20	20	0.99	5×10^{3}	10
R_l23	23	0.12	700	62
R_l26	26	0.01	40	404

*per Myr of evolution on 36 cores

Table 5.3: Parameters for simulations in Section 5.6. All simulations have $l_{\mathrm{glob}} = 20$ and the two zoom simulations have $n_{\mathrm{zoom}} = 8$.

friction or not, is that it requires a measure of the relative velocity, which can be defined in several ways. As shown in Section 5.4.2, $v_{\bullet}$ is the better choice for symmetric collapse, where the relative velocity of the black hole host cell, v_{cell}, is even more overestimated compared to the bulk relative velocity. In non-spherically symmetric cases, like the ones presented in this section where the scale height of the disc is smaller than the radius of the accretion region and cloud particles probe both sides of the disc simultaneously, the local velocity dispersion will increase $v_{\bullet}$ relative to v_{cell}. Setting $v_{\bullet} = 0$ was chosen because results show that this choice allows the black hole to remain sufficiently attached to the cloud to continue accreting from its core until it transitions to FLA, as it does naturally when self-consistently resolving dynamical friction (compare right two columns of Figure 5.10 to Figure 5.12). The reader is cautioned that this does not necessarily mean that the black hole is perfectly attached to the core at all times.

In summary, a small initial black hole mass, in combination with forcing $v_{\bullet} = 0$ until the dynamical friction of the black hole is resolved, is a robust way of producing a seed black hole that reflects the emerging structure of the collapsing cloud from which it is formed. To avoid delaying the transition to FLA, the seed mass should be chosen such that $R^{\mathrm{B}} = \frac{GM}{c_{s,\mathrm{cell}}^{2}} \approx \Delta x_{\mathrm{zoom}}$, where $c_{s,\mathrm{cell}}$ is the sound speed of the host cell at formation so that the accretion radius is marginally resolved at the end of cloud collapse. However, it is noted that the outcome of the simulation does not sensitively depend on this choice.

5.6 The impact of resolution on black hole accretion

To investigate the impact of resolving small-scale features on the evolution history of the black hole, two simulations with different l_{zoom}, R_l23 and R_l26, are compared to one simulation without zoom-within-zoom, R_l20 (see Table 5.3 for details). All

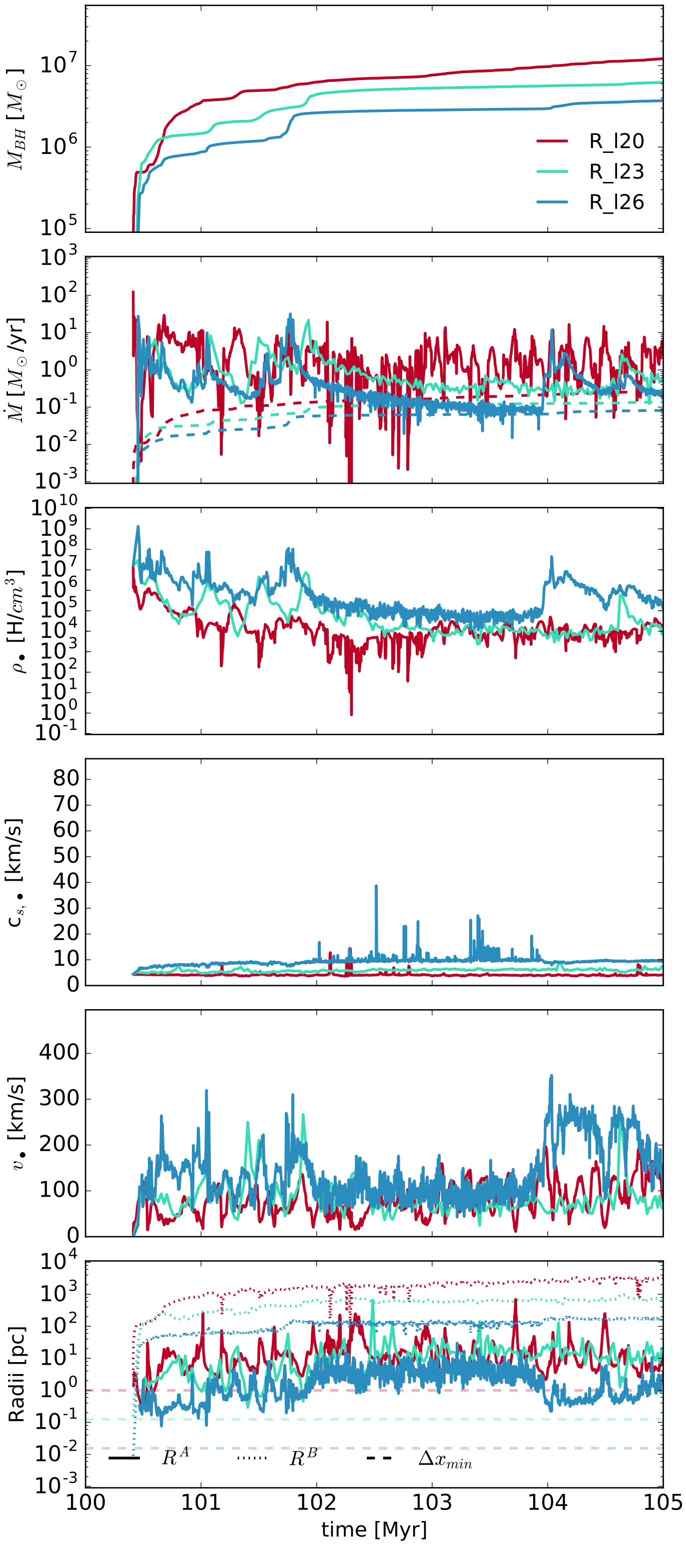

Figure 5.14: Time evolution of black hole and local gas properties at different resolutions (see Table 5.3 for details). All data is time averaged over 300 data-points for clarity.

three simulations have $l_{\rm glob} = 20$ and $n_{\rm refine} = 8$. The cooling halo and seed formation location are identical to the simulations presented in Section 5.5.

Improving resolution in the vicinity of the black hole has an immediate and significant impact on its evolution, as can be seen in Figure 5.14. While it may seem to only mildly affect the black hole's rapid initial mass growth (all three simulations manage to grow a black hole of $3-6\times10^5$ $\rm M_\odot$ in less than 0.1 Myr despite starting from order of magnitude different seed masses), the later accretion patterns and the gas properties in the immediate vicinity of the black hole differ so notably that the simulations never converge. This analysis is corroborated by the density slices presented in Figure 5.15: they show that the large scale structure of the cloud is similar in all three simulations (top row) but that the central structure varies significantly.

The black hole in R_20 accretes the entire dense core of the cloud (Figure 5.15, second row), emptying an irregularly shaped accretion region whose size is of the same order as in-falling clumps. Any cloud that subsequently falls into the centre is disrupted and accreted immediately by the black hole, without re-forming a dense core. By contrast, the accretion region of black holes in the zoom-within-zoom simulations R_l23 and R_l26 is significantly smaller than the physical extent of the core. Instead of being accreted, the core collapses into a nuclear gas disc whose rotationally supported structure is captured by the simulation. The difference between accreting the entire core and embedding a black hole in the centre of the resolved core results in a factor 5 difference in black hole mass after just 0.5 Myr of evolution and has a lasting influence on black hole accretion.

The mass profiles in Figure 5.16 confirm that the zoom-within-zoom algorithm leaves large-scale structure unaffected and smoothly extends the gas profile to smaller scales. $M_{\rm tot}$ and M_* converge for all three simulations on scales above 2 pc, the size of the accretion region in R_l20. At scales below 10 pc, the gas profile is curtailed prematurely for R_l20. $M_{\rm gas}$ converges in the region resolved by both R_l23 and R_l26, 0.5 pc $< r <$ 10 pc, showing that zoom-within-zoom brings out the existing internal structure of the cloud. This is also confirmed by the density projections in Figure 5.15 which show that the physical extent of the nuclear disc is the same in R_l26 and R_l23. With a factor 8 higher resolution, R_l26 captures more of the internal structure of the core, in the form of spiral features, than R_l23, in which the core appears smoother (compare R_l23 and R_l26 in the third row of Figure 5.15). The change in stellar profile M_* at small radii is due to the different gravitational smoothing lengths of the simulation.

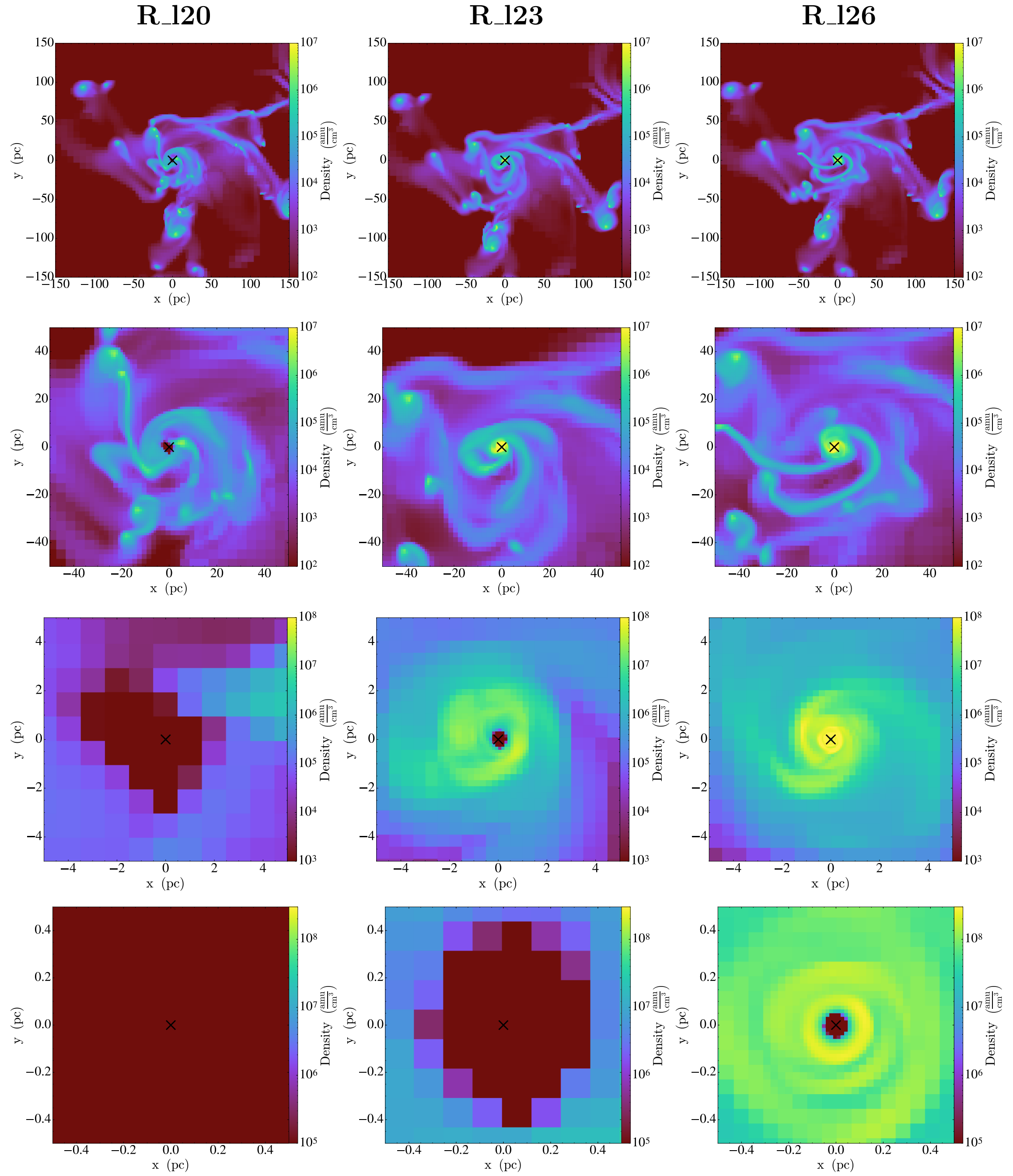

Figure 5.15: Gas density projections for R_l20 (left), R_l23 (middle) and R_l26 (right) at $t = 102$ Myr, at a range of different magnifications. The black hole location is marked by a black cross.

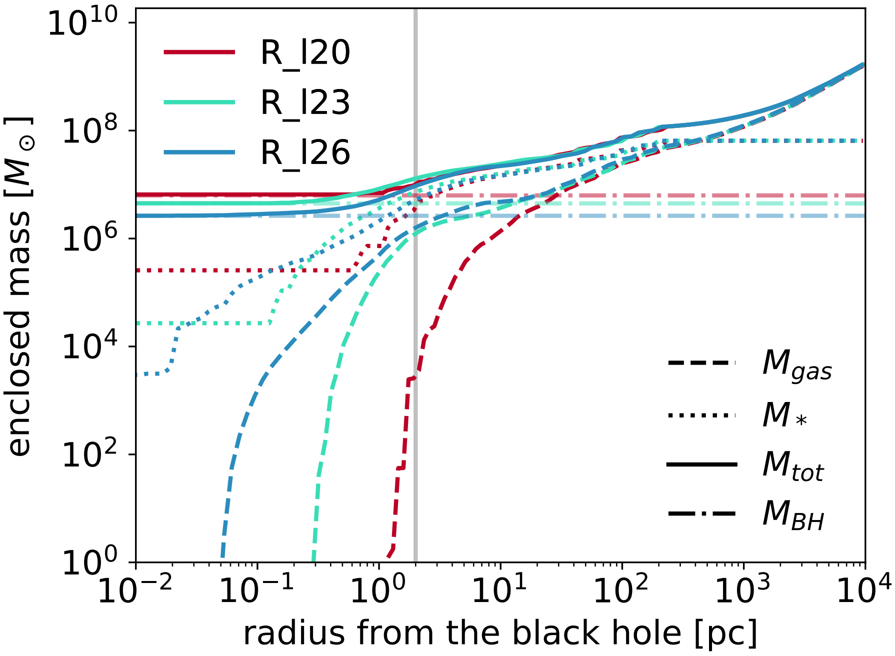

Figure 5.16: Cumulative spherical mass profiles for three different l_{glob}, including gas mass, stellar mass, black hole mass and total mass. The vertical grey line marks the approximate extent of the nuclear disc.

Comparing the duty cycle of R_l20 to R_l23 and R_l26 in the second panel of Figure 5.14 shows that resolving the nuclear disc around the black hole markedly changes the accretion pattern. The black hole in R_l20 grows continuously, with an accretion rate that fluctuates rapidly by orders of magnitude with a period of less than 0.1 Myr. The other two simulations, by contrast, grow in an episodic fashion, where long periods of smooth accretion are interspersed with accretion bursts that last up to 0.5 Myr.

While R_l26 and R_l23 show similar accretion patterns, the factor 8 difference in resolution affects the gas properties measured in the vicinity of the black hole. The initial black hole mass is smaller in R_l26 as the volume covered by the accretion region, and therefore the mass it contains when transitioning to FLA, is resolution dependent. The black hole is fed by gas that loses sufficient angular momentum to enter the accretion region at the centre of the nuclear disc. Black holes in higher resolution simulations have smaller accretion regions, so the gas needs to lose more angular momentum before being accreted, which results in a lower accretion rate during smooth accretion (compare R_l23 to R_l26 around $t = 103$ Myr in Figure 5.14). R_l26 shows more high-frequency fluctuations due to the movement of the sink particle at the bottom of the gravitational potential well, as discussed in Section 5.4.2.

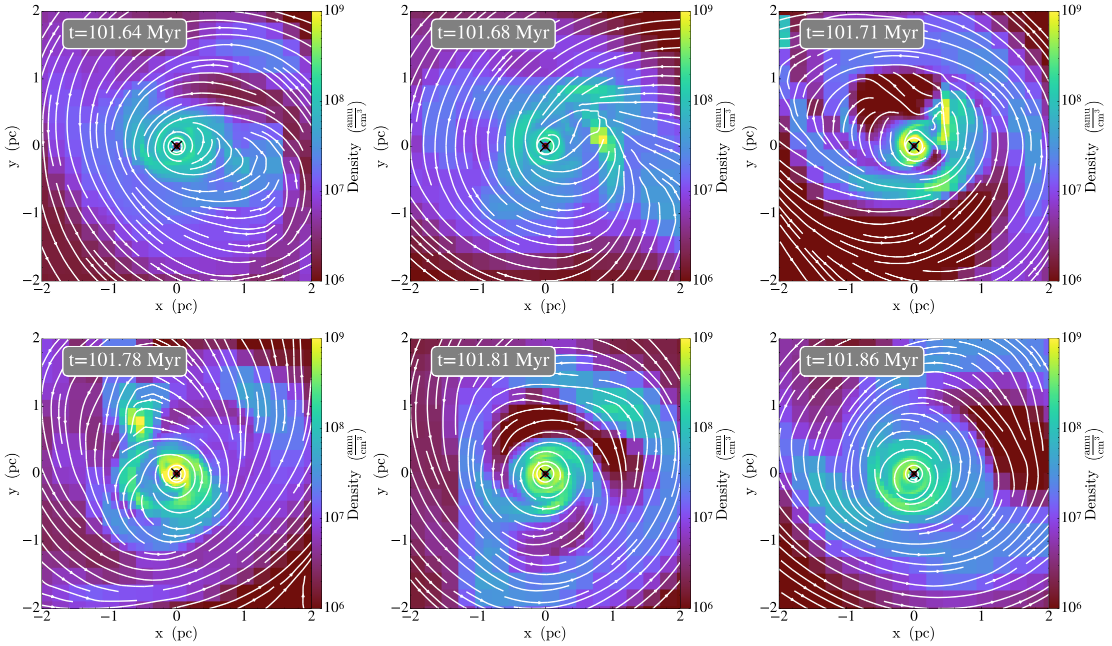

Figure 5.17: Gas density projections of the black hole environment during the accretion burst around $t = 101.8$ Myr in R_l26. The accretion spike is caused by the disruption of a gas clump. The black hole is marked by the black cross and streamlines are shown in white.

As it removes the gas deeper in the black hole potential well, it measures higher local densities and somewhat higher local sound speeds than R_l23.

The density projections for the accretion burst in R_l26 at $t = 101.6 - 102$ Myr, shown in Figure 5.17, show that accretion episodes occur when the nuclear disc is disrupted by an in-falling clump. The resulting redistribution of angular momentum allows gas to fall into the accretion region, boosting the accretion rate, before the disc resettles and the black hole returns to smooth accretion. These accretion bursts are the dominant accretion mechanism in both R_l23 and R_l26, as in both simulations the nuclear gas disc is resolved. The growth by clump-capture for sufficiently high resolution is also reported in Lupi et al. (2016), who perform numerical simulations of seed black holes in a galactic context at resolutions comparable to R_l23.

The differences in local gas properties and accretion duty cycles at higher resolution will have consequences for black hole feedback. R_l20 would produce a "rapid fire" feedback, consisting of fast bursts carrying over an order of magnitude more feedback energy than feedback due to smooth accretion in R_l26, assuming $\dot{E}_{\mathrm{BH}} \propto \dot{M}_{\mathrm{BH}}$. By contrast R_l23, and particularly R_l26, would produce much weaker sustained feedback, interspersed with strong feedback bursts that would encounter the dense gas structures of the nuclear disc. These different duty-cycles are expected to significantly impact the effect of feedback on the black hole host galaxy, particularly in combination with the effect reported in Negri & Volonteri (2017). They reported that resolution impacts the ability of black holes to self-regulate via feedback as it affects the size of the gas reservoir into which the feedback energy is deposited. Like the accretion and dynamical friction algorithms, feedback sub-grid algorithms will have to be carefully re-examined for use in the high-resolution environment presented here. An investigation into the effect of feedback on black hole accretion is therefore postponed to future work.

One advantage of the zoom-within-zoom algorithm is its comparatively low computational cost, as can be seen in Table 5.3. The biggest limitation on increasing l_{zoom} is the shrinking timestep, set (among other criteria) by the sound-crossing time or the flow crossing time of the smallest cell in the simulation. Adding an extra level of resolution at identical gas conditions would halve the timestep. Figure 5.14 shows that simulations with higher l_{zoom} also have a higher sound speed in the smallest cells near the black hole, as well as a higher flow velocity into the black hole's potential, further decreasing the timestep. Some of the extra cost can be mitigated by subcycling and by distributing the smallest cells over a larger number of processors, but ultimately there will be a limit of the number of extra resolution levels that can be added using zoom-within-zoom.

5.7 Discussion

5.7.1 Sub-grid algorithms for accretion

Even at the improved resolutions available in zoom-within-zoom, the size of the accretion region remains orders of magnitude above the event horizon of the black hole and the nuclear disc resolved here is much larger than the expected accretion disc around the black hole itself. Using Equation 4 in Morgan et al. (2010), the accretion disc around a black hole with $M_{\mathrm{BH}} = 10^6$ $\mathrm{M}_{\odot}$ has a maximum radius of about 8×10^{-6} pc, far below the 0.01 pc resolution of even the most highly resolved black hole studied here, R_l26. As discussed in Section 3.5.2.2, several sub-grid schemes have been proposed to include some of the unresolved black hole physics in large-scale simulations.

Comparing densities in the vicinity of the black hole (Figure 5.14) shows that lower resolution simulations have lower density in the black hole vicinity. Despite this, I do not advocate compensating with a boost factor, as suggested in Booth & Schaye (2009), for two reasons. Firstly, despite the lower density, R_l20 *over-accretes* in comparison to R_l26. Secondly, adding a boost factor to the BHL formula is inconsequential once accretion has transitioned to FLA, as the mass accreted by the black hole is no longer determined by the value calculated by the BHL accretion rate in Equation 3.18, boosted or otherwise.

FLA makes two assumptions: all gas entering the accretion region is accreted by the black hole, and the timescale between gas entering the accretion region and entering the event horizon of the black hole is short. The former holds for the simulations presented here, particularly R_l23 and R_l26 where the black hole dominates the gravitational potential at the edge of the accretion region. The question of accretion timescale is more difficult, as comparing R_l20 to R_l23 and R_l26 shows that in the presence of rotation, accretion can be significantly delayed even on the scales resolved here. Therefore, even if all gas will eventually be accreted, one would expect a significant time delay, which could also have important consequences for black hole feedback.

As suggested in Krumholz et al. (2004) and Curtis & Sijacki (2016b), one solution to account for this delay could be to scale down the accretion rate onto the black hole by a factor reflecting the vorticity of the gas in, or just around, the accretion region. This concept would be difficult to implement in the current accretion algorithm, as FLA is an emergent phenomenon, not an explicitly calculated accretion rate. For the same reason that scaling up the BHL accretion rate to compensate for under-resolved density has no effect, scaling it down to compensate for vorticity is not necessarily going to have the desired effect either. One of the main justifications for zoom-within-zoom is precisely that it is able to account for the vorticity of the gas self-consistently, reducing the need for further sub-grid schemes.

An explicit flux-based scheme for sink particles in RAMSES is introduced in Bleuler & Teyssier (2014), based on Gong & Ostriker (2013), which calculates the accretion rate onto the sink using the Riemann fluxes at the cell boundaries. The authors compare this scheme to a BHL based scheme using local quantities extrapolated to infinity and note that the explicitly flux based scheme performs better at high resolution, whereas a BHL based scheme is more successful at low resolution. The flux scheme of Bleuler & Teyssier (2014) has the advantage that it produces a smooth transition of the flow into the accretion region and thereby avoids low densities within the accretion region and associated pressure forces at the accretion edge boundary. The advantage of the scheme presented here is that it automatically transitions from BHL to FLA, without the need for an explicit transition criterion. A resolution dependent transition criteria will depend on R^{A} and can therefore be difficult to define in a galactic context, as discussed in Section 5.5.2.

The highly coherent gas accreted by the black hole in R_l23 and R_l26 would lend itself to being combined with a sub-grid accretion disc scheme, like the ones presented in Power et al. (2011) and Dubois et al. (2012b), particularly as the size of the accretion region for R_l26 approaches the boundary conditions for a supermassive black hole's accretion disc[1]. In these schemes, gas is removed from the grid and added to a sub-grid "accretion disc", from which the accretion rate onto the black hole itself is calculated, taking the viscous timescale into account. This would also give an opportunity to include models for physical processes not included in hydrodynamical simulations, such as magnetic fields. How one would accurately capture the accretion duty cycle with such a scheme, and what the accretion duty cycle of a black hole even looks like on the relevant length scales, remains an open question.

5.7.2 Other refinement schemes

The work presented in this chapter is similar to Curtis & Sijacki (2015), who present a super-Lagrangian refinement algorithm for the black hole vicinity, implemented in the moving-mesh code AREPO. One significant difference between the two works is that Curtis & Sijacki (2015) define a maximum radius, set to the gravitational smoothing length of the black hole, within which resolution is increased as a function of distance to the black hole. In zoom-within-zoom, a minimum radius is defined instead, within which the maximum resolution is maintained. It is surrounded by regions of progressively lower resolution until the resolution of adaptively refined cells is reached. Maintaining a fixed grid around the black hole guarantees that accretion only proceeds from cells at full resolution. This is important as results in Section 5.4.1 have shown that accreting from mixed resolution leads to non-convergence of the mass evolution (see C_l20n2, which has $r_{\rm refine} < 2\Delta x_{\rm min}$, in Section 5.4.1).

Another important difference between these two works is that in Curtis & Sijacki (2015), the smallest cell is of the order of the black hole's Bondi radius, $R^{\rm B}$. In the simulations presented here, the Bondi radius is consistently well resolved for black holes of comparable seed mass. Accretion in Curtis & Sijacki (2015) therefore continues to rely on the BHL accretion rate, and uses a resolution at which the BHL accretion rate based on local quantities can differ as much as a factor 5 from the analytic value (see Section 4.3.3). The authors combine BHL accretion with a non-isotropic scheme where gas is only accreted from cells with a temperature below a

---

[1] The accretion disc around a black hole with $M_{\rm BH} = 10^9$ $\rm M_\odot$ has a maximum radius of 10^{-3} pc, using Equation 4 in Morgan et al. (2010). This is only an order of magnitude below the resolution in R_l26, which has $\Delta x_{min} = 0.01$ pc.

threshold temperature, $T_{\rm cold}$, when the fraction of cold gas within the softening length of the black hole exceeds 25%. This is designed to mimic accretion via a cold, dense disc that is expected to dominate black hole accretion in a galactic environment. No such sub-grid model is employed here as results in Section 5.6 show that the disc feeding the black hole can be resolved self-consistently using zoom-within-zoom.

The issue of keeping the black hole attached to local gas flows is addressed in Curtis & Sijacki (2015) by calculating separate dynamical and sub-grid masses for each black hole. The former is used for calculations of the gravitational potential whereas the latter is the 'physical' mass of the black hole, equivalent to $M_{\rm BH}$ in the work presented here. The advantage of a dynamical mass model in comparison to the drag force model used here is that it does not rely on measuring local relative velocities and therefore avoids the associated difficulties. However, as stated in Curtis & Sijacki (2015), using a dynamical mass for local gravity calculations will introduce significant changes to the central potential, and have a lasting impact on local gas dynamics, when the black hole mass exceeds the local cell masses. At the resolutions studied in this chapter, the black hole mass frequently exceeds the local cell mass even for simulations with unresolved dynamical friction. The ratio $M_{\rm BH}/M_{\rm cell} \propto M_{\rm BH}/(\rho_{\rm cell}\Delta x^3_{\rm zoom})$ (important for the dynamical mass model) depends on the local density $\rho_{\rm cell}$ whereas the ratio $R^{\rm A}/\Delta x_{\rm zoom} \propto M_{\rm BH}/(\Delta x_{\rm zoom} v^2)$ (important for the drag force model) depends on the relative velocity, v, instead. A supersonic black hole in a low density environment can therefore dominate the local potential but still have under-resolved dynamical friction. One such example is the black hole in D_l26_small in Section 5.5.1 that fails to settle in the disc despite its black hole mass exceeding local cell masses by over an order of magnitude.

Another super-Lagrangian refinement scheme for black holes in RAMSES is proposed in Lupi et al. (2015), in which the black hole host cell is forced to the highest resolution available in the simulation. Their scheme differs from the work presented here in that the black hole environment is only resolved as well as the densest gas, not better, as in zoom-within-zoom. The refinement scheme of Lupi et al. (2015) was independently implemented in the simulations presented here, including the fiducial runs without zoom-within-zoom to avoid local cells de-refining when the density in the accretion regions drops during FLA.

5.8 Conclusions

The origin of supermassive black holes is a complex problem that involves gas flows and gravitational torques over many orders of magnitude in length scale, from the Mpc scales of the cosmic web to the au scales of the event horizon. To extend the range of scales included in a single simulation, this chapter introduced the "zoom-within-zoom" refinement algorithm, a super-Lagrangian scheme for the AMR code RAMSES in which the black hole is surrounded by a spherical region of high resolution. Simulations presented in this chapter use zoom-within-zoom to improve the resolution in the vicinity of the black hole by two orders of magnitude. This refinement algorithm reproduces the same gas properties in the vicinity of the black hole as measured in adaptively refined simulations and is not sensitive to the radius of the highest resolution sphere, as long as it is larger than the size of the black holes accretion region, i.e. $r_{\rm zoom} > 4\Delta x_{\rm zoom}$ (see Section 5.4).

Building on the idealised and isolated case in Chapter 4, this chapter also tested two black hole sub-grid algorithms in a highly resolved galactic environment: BHL accretion onto a sink particle and an analytic drag force. Work in this chapter confirmed that the BHL algorithm remains robust even when the gravitational potential of the black hole is well resolved, as first reported in Section 4.3. When sufficient resolution in the vicinity of the black hole makes the problem locally strongly non-Bondi, the code automatically transitions to FLA, where the accretion rate onto the black hole is determined by the mass flux into the accretion region, not by the BHL accretion rate. The advantage of continuing to use the BHL sub-grid algorithm, instead of an explicit flux-based scheme such as the one employed in Bleuler & Teyssier (2014) and Lupi et al. (2016), is that the BHL algorithm is very versatile. When the local gas properties are such that the gravitational influence of the black hole is unresolved, BHL accretion provides a reliable sub-grid model. When resolution becomes sufficiently high to resolve the black hole's potential, the black hole transitions to FLA automatically, without the need for an explicit transition criterion. It will also automatically transition back if local gas properties change again.

To evaluate local gas properties in the vicinity of the black hole, the sink-particle algorithm uses so-called cloud particles, filling a sphere with radius $r_{\rm cloud} = 4\Delta x_{\rm zoom}$, spaced at $\Delta x_{\rm zoom}/2$. Measuring a discretised distribution of gas using this discretised set of particles leads to two numerical features in the local gas properties, particularly in spherically symmetric collapse: small movements of the black hole cause the measured gas properties to fluctuate as the cloud particles preferentially probe different cells, and the relative velocity between the black hole and the gas is over-estimated. Neither the long-term mass evolution, nor the time-averaged gas properties (except for the relative velocity) are affected.

In this chapter it has been demonstrated that to obtain a black hole initial mass that reflects the internal structure of its parent cloud, revealed by the extra zoom-within-zoom resolution, an initial seed mass for which the Bondi radius, R^{B}, is approximately equal to the smallest cell at full resolution — i.e. $M_{\mathrm{seed}} = \Delta x_{\mathrm{zoom}} c_{s,\mathrm{cell}}^2 / G$, where $c_{s,\mathrm{cell}}$ is the sound speed of the cell in which the sink particle is formed — is adequate. This result holds provided that unresolved dynamical friction is compensated by applying the maximum drag force while $R_{\bullet}^{\mathrm{A}} < 0.2\Delta x_{\mathrm{zoom}}$. Once full resolution has been reached, the black hole accretes the mass of the cloud collapsing into its accretion region, and transitions to FLA. The outcome of the simulation does not sensitively depend on the choice of M_{seed}, as overly light seeds merely experience a small time-delay before converging to the same solution than more massive seeds. The simulations presented here therefore do not distinguish between the three formation models, described in Section 2.2.6, as the rapid gas accretion dilutes any sign of the seed mass in less than a Myr. However this is expected to change in the presence of black hole feedback, where the initial rapid accretion phase is expected to produce strong feedback that could halt cloud collapse at different stages, depending on the initial seed mass of the black hole.

Work presented in this chapter showed that the evolution history of a black hole is driven by the internal structure of the clouds from which it accretes. Resolving the gravitational potential of the black hole is therefore a necessary but not a sufficient condition to capture its accretion history in a particular cloud. When comparing the evolution of black holes with identical initial conditions, whose local environment was resolved at 0.99 pc, 0.16 pc and 0.01 pc respectively, the gas properties in the immediate vicinity of the black hole differed so notably that the simulations never converged. This is despite the fact that all three simulations had a well resolved Bondi radius, with $R^{\mathrm{B}}/\Delta x_{\mathrm{min}} > 10^3$ at all times. In the simulation resolved at 0.99 pc, the black hole accretes continuously but chaotically, as gas clumps are disrupted and accreted as soon as they fall into the accretion region of the black hole. At higher resolution, by contrast, the simulation resolves a nuclear gas disc that smoothly feeds the black hole. Both high resolution simulations experience accretion bursts when in-falling clumps cause disc instabilities in the nuclear disc, which occur much more rarely and last much longer than the oscillations in the accretion rate at 0.99 pc.

This simple example has shown that the accretion behaviour of the black hole is highly non-linear and dominated by small scale features in the immediate vicinity of the black hole. It has also shown that resolving the scale length of the black hole's potential provides an upper limit to gas accretion, as internal structure of the gas cloud feeding the black hole, such as rotational support on sub-pc scales, has long-term consequences for black hole accretion. The work presented here therefore highlights the need for algorithms like zoom-within-zoom that allow the gas evolution in the black hole vicinity to be captured self-consistently, particularly once feedback will be included. As the evolution history of the black hole is determined by that of its host cloud, simulations on longer timescales than the 5 Myr studied here have to be embedded in a environment able to adequately capture the evolution of the galactic gas supply. Work presented in the next chapter will therefore employ the zoom-within-zoom algorithm in a cosmological context to investigate the origin of the first quasars.

Chapter 6

The role of gas accretion in the early evolution of supermassive black holes

6.1 Introduction

The earliest supermassive black holes (SMBHs) were observed powering bright quasars between redshifts $z = 7.1$ and $z = 6$ (Fan et al., 2006; Mortlock et al., 2011; De Rosa et al., 2014) by the time the Universe was barely a billion years old. Their origin and early evolution must have occurred in the context of a rapidly evolving host galaxy that was itself only just beginning to develop, in conditions very different to what is observed in the local Universe.

The first stars formed are thought to have formed between redshift $z = 50$ and $z = 20$ in mini-halos with dark matter masses of 10^{5-6} $\mathrm{M}_{\odot}$ (Bromm et al., 1999). As these stars collapsed from massive clumps of pristine gas, which could not fragment further in the absence of metal line cooling, these so-called first population III stars (Pop IIIs) probably have had a comparatively top-heavy initial mass function (IMF) (Glover, 2005; Latif & Ferrara, 2016). Supernovae from this first generation of stars were sufficiently energetic to eject gas from the shallow potential well of their host halo and shut of further star formation (Bromm et al., 2003; Kitayama & Yoshida, 2005; Whalen et al., 2008) and had a complex and lasting influence on gas cooling. On the one hand, the high energy radiation emitted by the first stars dissociated molecules and ionised the primordial gas. This reduced the availability of crucial cooling agents, such as H_2, and therefore limited the ability of gas to cool below 10^4 K, the bottom of the cooling curve for Hydrogen (Maio et al., 2007; Johnson et al., 2013). On the other hand, the first supernovae also created the first metals, whose line transitions allow gas to cool much more effectively than by H_2 and H cooling alone. The ejected, pre-enriched gas from Pop III stars mixed with primordial gas on kpc scales before falling back into halos (Cen & Riquelme, 2008; Greif et al., 2010; Johnson et al., 2013) with dark matter masses in the range 10^{7-8} $\mathrm{M}_\odot$. Here the pre-enriched gas formed the second generation of stars (Omukai et al., 2005), called population II stars (Pop IIs), which were housed in halos with sufficiently deep potential wells to retain gas even in the presence of stellar feedback: the first galaxies had formed.

As described in detail in Section 2.2.6, black hole progenitors formed via one of three channels, which were active at different redshifts and produced black holes with different seed masses M_{seed}: remnants of the first or second generation of stars ($M_{\mathrm{seed}} = 100$ $\mathrm{M}_\odot$, $z = 30-20$), run-away cluster collapse ($M_{\mathrm{seed}} = 10^{3-4}$ $\mathrm{M}_\odot$, $z = 12-10$) and direct collapse ($M_{\mathrm{seed}} = 10^5$ $\mathrm{M}_\odot$, $z = 15-12$). Once formed, black holes grow via a combination of mergers and gas accretion, with simulations showing varying results for their relative importance. Some authors argue that mergers, particularly early mergers, were an important channel for early black hole mass growth (Arun et al., 2009; Volonteri, 2010). Others report that merger timescales are long and stress the importance of gas accretion for black hole growth (Sesana et al., 2007; Chapon et al., 2013; Goicovic et al., 2017).

Even with a combination of gas accretion and mergers, the limited timeframe between the first stars and the first observations of SMBHs, about 700 Myr, remains a challenge. Simulations have shown that there is ample gas within the galaxy, at pc to kpc scales from the black hole (Levine et al., 2008; Emsellem et al., 2014; Dubois et al., 2015; Smidt et al., 2013). However, it is uncertain how much of this gas is accreted, particularly in the presence of feedback which can blow away the local gas supply and starve the black hole (Gan et al., 2014; Pacucci & Ferrara, 2015; Lupi et al., 2016). At high redshift, the timescales for the gas to recover and flow back are on the order of the local Hubble time, introducing a serious time delay to early black hole accretion (Johnson & Bromm, 2007; Whalen & Fryer, 2012). Simulations have also shown that rotational support of the gas is a significant limiting factor to black hole growth, as viscous timescales are long (Choi et al., 2013; Emsellem et al., 2014; Shlosman et al., 2016). For the black hole to accrete efficiently, it has to either avoid settling into a disc (Alexander & Natarajan, 2014), or momentum has to be efficiently transferred via a mechanism like bars-within-bars (Begelman et al., 2006). Recent work at a range of scales has shown that the problem is highly non-linear (Latif et al., 2013a; Lupi et al., 2016; Curtis & Sijacki, 2016a; Negri & Volonteri, 2017), so it is difficult to accurately extrapolate the accretion rate onto the black hole from large scales, as the mass accreted, as well as the accretion duty cycle, depend on the local structure. As the host galaxy is rapidly evolving on similar timescales as the black hole, the early evolution of black holes has to be studied in a cosmological context.

In this chapter, the role of gas accretion for early SMBH growth is investigated. Using a hydrodynamical simulation with sub-pc resolution in the black hole vicinity allows the gravitational potential of the black hole to be resolved in a full cosmological context. To establish an upper limit for gas accretion, feedback from both stars and black holes is omitted. This chapter is structured as follows: Section 6.2 introduces the simulation. A detailed description of the choice of host halo is given in Section 6.3. Section 6.4 presents the coevolution of the black hole and its host galaxy, with a particular focus on the black holes dynamical evolution in Section 6.5. The future evolution of the black hole is discussed in Section 6.6. Section 6.7 presents a discussion as to whether the results presented here provide an upper limit on the contribution of gas accretion to the early mass evolution of potential SMBHs. Conclusions can be found in Section 6.8.

6.2 The simulation

The simulation presented in this chapter is a cosmological zoom simulation run using RAMSES (Teyssier, 2002), using a Courant factor of 0.8 (see Chapter 3 for details). Initial conditions at $z = 500$ were produced with MUSIC (Hahn & Abel, 2011), following a standard ΛCDM cosmology, consistent with WMAP-9 data (Bennett et al., 2013), where $\Omega_\mathrm{m} = 0.279$, $\Omega_\Lambda = 0.721$, $\Omega_\mathrm{b} = 0.0463$, $H_0 = 70.0\,\mathrm{kms^{-1}\,Mpc^{-1}}$, $\sigma_8 = 0.821$ and $n_s = 0.972$. The 1 Gpc^3 box is refined up to a root grid of 128^3. Within it lies a zoom region that encompasses all progenitors to a halo with mass 6.7×10^{11} $\mathrm{M_\odot}$ at redshift $z = 6$, which is the most massive progenitor to a halo with mass 4×10^{15} $\mathrm{M_\odot}$ at $z = 0$ (see Section 6.3 for details). The dark matter mass resolution in the zoom region is 1.48×10^5 $\mathrm{M_\odot}$. The zoom region is adaptively resolved at 1 pc in physical units, kept fixed during cosmological expansion by the addition of new levels. Refinement is triggered according to a quasi-Lagrangian refinement criteria which adds a level when the mass in a cell exceeds 8 times the initial dark matter or baryon mass.

The gas cools to 10^4 K using atomic cooling following Sutherland & Dopita (1993), and to 130 K using $\mathrm{H_2}$ cooling following Grassi et al. (2014). Gas follows a monoatomic equation of state with $\gamma = 5/3$. Star formation takes place in cells above a density threshold of $\rho_* = 4.73 \times 10^3$ $\mathrm{H\,cm^{-3}}$ and follows a Kennicutt-Schmidt law

(Schmidt, 1959; Kennicutt, Jr., 1998; Krumholz & Tan, 2007) with a fixed star formation efficiency of $\epsilon_* = 0.02$. Star particles are generated using to a Poisson random process (Rasera & Teyssier, 2006; Dubois & Teyssier, 2008) and have a minimum mass of 129 $M_\odot$ or integer multiples thereof. There is no stellar feedback.

The simulation contains a single black hole, modelled as a sink particle (Krumholz et al., 2004; Dubois et al., 2010). It is seeded with a mass of $M_{\rm seed} = 260$ $M_\odot$ in a collapsing clump identified using the PHEW structure-finding algorithm (see Bleuler & Teyssier, 2014, and Section 3.5.1), using a density threshold of $\rho_0 = 10^3$ $\rm H\,cm^{-3}$. The environment of the black hole is resolved using the zoom-within-zoom algorithm (see Section 5.3), using $n_{sink} = 4$ and a maximum resolution $\Delta x_{sink} = 0.016$ pc. The highest resolution region therefore has a radius of approximately 0.064 pc. The total zoom-within-zoom region, including all intermediate levels, has a radius of approximately 5 pc. The exact values depend on the grid configuration at a particular point in time and can vary somewhat (see Figure 5.1 for a visual representation).

The black hole formally accretes using the Bondi-Hoyle-Lyttleton (BHL) formula (Equation 3.18, see Section 3.5.2 for details). At the high resolution available with the zoom-within-zoom refinement strategy, accretion onto the black hole is limited by the availability of gas supply in the accretion region and the black hole accretes via flux limited accretion (FLA) (see Section 5.4). This simulation does not include a prescription for black hole feedback.

6.3 Choosing the host halo

While the population of quasars observed at $z = 6$ and above is steadily growing, only the brightest objects active at the time are visible with modern telescopes. Investigating the origin of these quasars is therefore a quest for statistical outliers which are likely to be found in exceptional environments. Quasars at high redshift have a number density of approximately 1 $\rm Gpc^{-3}$ (McGreer et al., 2013). On the assumption that the most overdense environments produce the biggest black holes, the halo hosting the black hole investigated in this chapter was chosen to have the same number density as the observed quasars, i.e. to be the most massive $z = 6$ progenitor of the most massive halo contained in a $\rm Gpc^3$ box.

To make the required resolution affordable in such a big box requires using a zoom simulation (Katz et al., 1994; Navarro & White, 1994), in which a high resolution region containing the target halo is embedded in a large box at much lower resolution. The low resolution region provides evolving gravitational boundary conditions for the

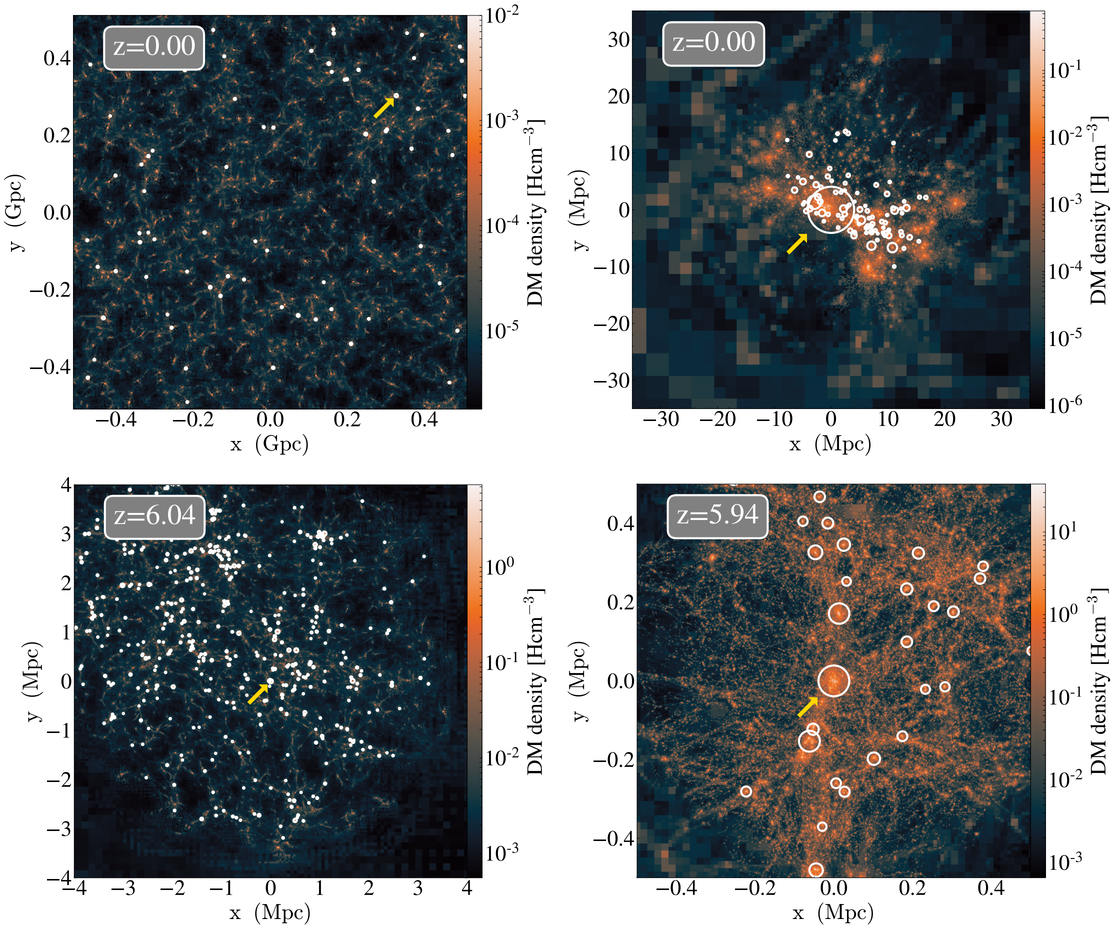

Figure 6.1: Dark matter density projections of the N-body simulations used to identify the most massive progenitor of the most massive halo in the box. Top left: whole box at $z = 0$. Top right: zoom of the most massive halo at $z = 0$. Bottom left: progenitors of the most massive halo at $z = 6.04$. Bottom right: zoom of most massive progenitor at $z = 5.94$. White circles denote the virial radii of halos, and the relevant halo in each projection is marked by a yellow arrow. For clarity, only the most massive halos are annotated. From top left to bottom right, the minimum halo masses marked on the plot are 1.98×10^{14} M$_\odot$, 6.05×10^{12} M$_\odot$, 6.05×10^{11} M$_\odot$ and 1.48×10^{10} M$_\odot$.

high resolution region. To ensure that the region of interest is not contaminated by information from low resolution regions, the initial conditions for the zoom region have to encompass all components contained within the target halo at the end of the simulation. In RAMSES, each dark matter particle is assigned a unique identifier. Regions collapsing into a particular halo can therefore be identified by tracing the dark matter particles, which make up the halo at the target redshift, back to the initial redshift. The region containing all particles at the initial redshift determines the size of the zoom region.

The initial conditions for the simulations presented here were produced using MUSIC (Hahn & Abel, 2011), which combines a convolution of Gaussian white noise with a real-space transfer function kernel to produce a three-dimensional multi-grid density and velocity distribution. This distribution has the same power spectrum as the density perturbations in the early Universe, as observed today in the cosmic microwave background (CMB).

Producing the initial conditions for a zoom simulation of the most massive progenitor to the most massive halo was a multi-step process. Simulating the whole Gpc^3 box at full mass resolution down to redshift $z = 0$ would have been prohibitively computationally expensive, but staggering several sets of zoom simulations makes the computation affordable. First, a suitably massive halo was found at redshift $z = 0$ using a N-body (dark matter only) simulation on a uniform grid of 128^3, with mass resolution of 1.98×10^{13} M$_\odot$. The random seed produced a halo with $M_{\rm DM} = 4.06 \times 10^{15}$ M$_\odot$ at $z = 0$ (see top left panel in Figure 6.1). Its mass significantly exceeded that of any other halo in the box, with the next most massive object containing $M_{\rm DM} = 2.96 \times 10^{15}$ M$_\odot$. To ensure that its progenitors at $z = 6$ are sufficiently well resolved and unambiguously identifiable, the most massive halo was resimulated to $z = 0$ using a dark matter only zoom-simulation with a mass resolution of 6.07×10^8 M$_\odot$ (see top right panel, Figure 6.1).

Tracing the dark matter particles contained in the halo at $z = 0$ back to redshift $z = 6$ revealed that the most massive progenitor has a mass of 6.7×10^{11} M$_\odot$ (see bottom left panel in Figure 6.1). To reduce overall numerical costs and improve the mass resolution, the zoom region was recalculated to only cover this progenitor halo and the region from which it collapsed. The final set of initial conditions embeds a zoom region with a physical radius of 500 kpc at $z = 6$ within a 140 Mpc^3 box, equivalent to 1 Gpc^3 at redshift $z = 0$ (see bottom right panel in Figure 6.1 for a projection of the final configuration). It encompasses all information relevant to the formation of a 6.7×10^{11} M$_\odot$ halo at $z = 6$, at a dark matter mass resolution

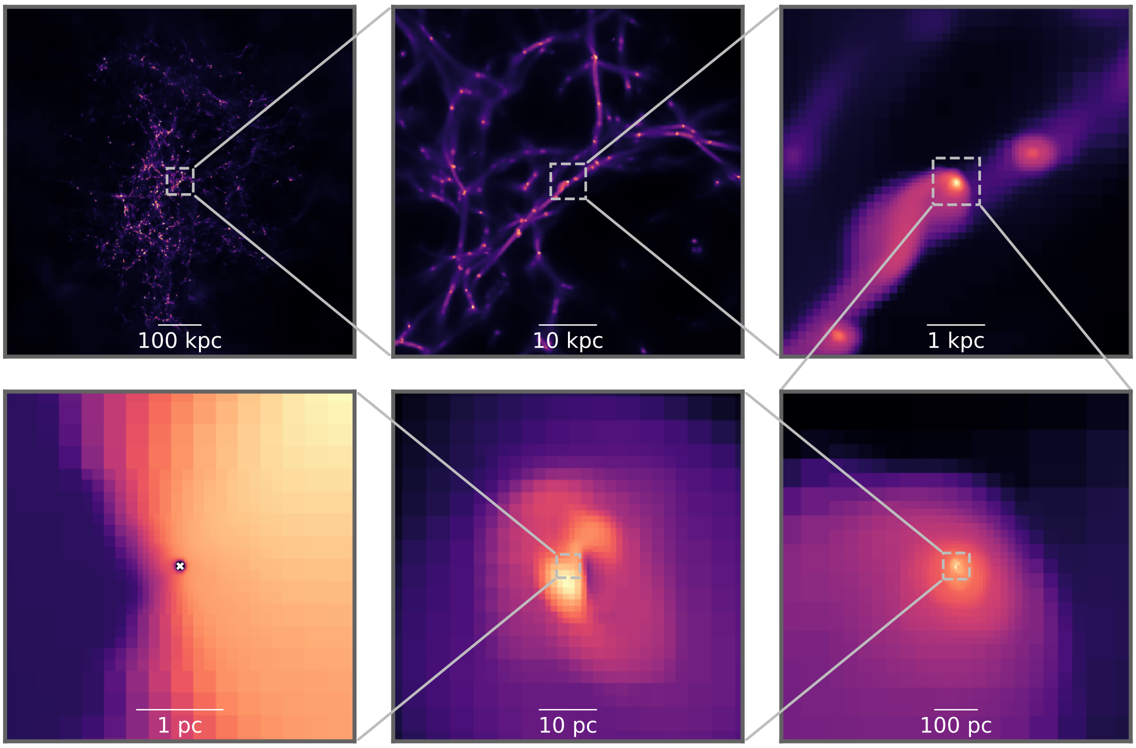

Figure 6.2: Gas density projections showing the range of scales covered in the simulation, from the cosmic web to sub-pc resolution near the black hole. This snapshot was taken at $z = 16.4$. Each frame is centred on the black hole, whose location is marked by a white cross in the bottom left panel. To cover the large range of densities, the colour scale is rescaled for the minimum and maximum of each panel. The bottom left panel shows a density slice instead of a density projection.

of 1.48×10^5 M$_\odot$. These initial conditions, in combination with zoom-within-zoom refinement around the black hole, produced a cosmological zoom simulation able to follow gas flows from Mpc to sub-pc scales, as can be seen in the density projections in Figure 6.2.

The zoom region at redshift $z = 6$ has a radius of 500 kpc, significantly bigger than the virial radius of the target halo at 39.1 kpc. The high resolution region therefore includes a large number of the target halo's neighbours (see bottom right panel in Figure 6.1). To avoid the black hole forming within a halo that will not merge into the target halo, a maximum distance criteria to one of the target halo's dark matter particles was added to the checks performed when identifying a black hole formation location. Using the PHEW structure finding algorithm, the black hole forms at redshift $z = 19.3$ in a halo with a mass of 3.02×10^8 M$_\odot$. According to Lukić et al. (2007), this halo is at the massive end of the halo mass function (HMF) at $z = 20$ and has a number density of 0.33 Mpc^{-3}. The halo chosen here has similar properties to the halo studied in Smidt et al. (2013), who successfully produced a SMBH from a 10^5 $\mathrm{M}_\odot$ seed via gas accretion, at 35 pc resolution. It is also similar to the halo studied in Dubois et al. (2013a), whose final black hole mass of 5×10^7 $\mathrm{M}_\odot$ falls short of observations but does include the impact of active galactic nuclei (AGN) feedback. Their work confirms that cold, dense accretion flows (Di Matteo et al., 2012; Dubois et al., 2013a; Smidt et al., 2013) into the halo should provide an ample gas supply for black hole accretion.

6.4 Black hole and galaxy coevolution

After being formed at redshift $z = 19.3$ with a seed mass of 260 $\mathrm{M}_\odot$, the black hole grows rapidly early on. During the first extended accretion burst, which lasts approximately 10 Myr, it reaches 10^6 $\mathrm{M}_\odot$ (see Figure 6.3). The black hole then enters a sustained high accretion phase, during which the accretion rate can exceed the Eddington rate by three orders of magnitude for brief periods of time. At $z = 16$, accretion drops to an average Eddington ratio of $\chi = 5.8 \times 10^{-6}$. For the next 200 Myr, the accretion rate onto the black hole remains low, with an average Eddington ratio of $\chi = 1.28 \times 10^{-5}$ and becomes episodic, where extended periods of smooth accretion are interspersed with brief accretion bursts (see $z = 16.5$ and $z = 15.3$ in Figure 6.3). Overall, the black hole assembles 90% of its mass via super-Eddington accretion, despite only spending 10% of the time accreting in this regime. The black hole has reached 5.05×10^6 $\mathrm{M}_\odot$ by $z = 11.1$, the end of the simulation

Like the black holes of comparable mass and resolution in an isolated galaxy studied in Chapter 5, the black hole presented here transitions to FLA immediately. However, the mass gained during this transition (166.8 $\mathrm{M}_\odot$, which has a free-fall time of 0.75 Myr) is far less than the total mass accreted in the first accretion burst at $z = 19.2$, during which the black hole gains 10^4 $\mathrm{M}_\odot$ in 2.38 Myr. Accretion onto the black hole is therefore driven by the continuing collapse of the cloud hosting the black hole and not by the shift in accretion algorithm (See Section 5.4 for details). For the remainder of the simulation, the accretion rate onto the sink is determined by the mass flux into the accretion region of the black hole, which has a radius of $4 \times \Delta x_{min} = 0.06$ pc (see low density region surrounding the black hole in the bottom left panel of Figure 6.2). A similar early accretion burst is reported in Shlosman et al. (2016), who find that their central black hole seed gains over 10^6 $\mathrm{M}_\odot$ in 2 Myr.

The 200 Myr investigated in this work not only cover the early evolution of the black hole but also the early evolution of the black hole's host galaxy, whose stellar

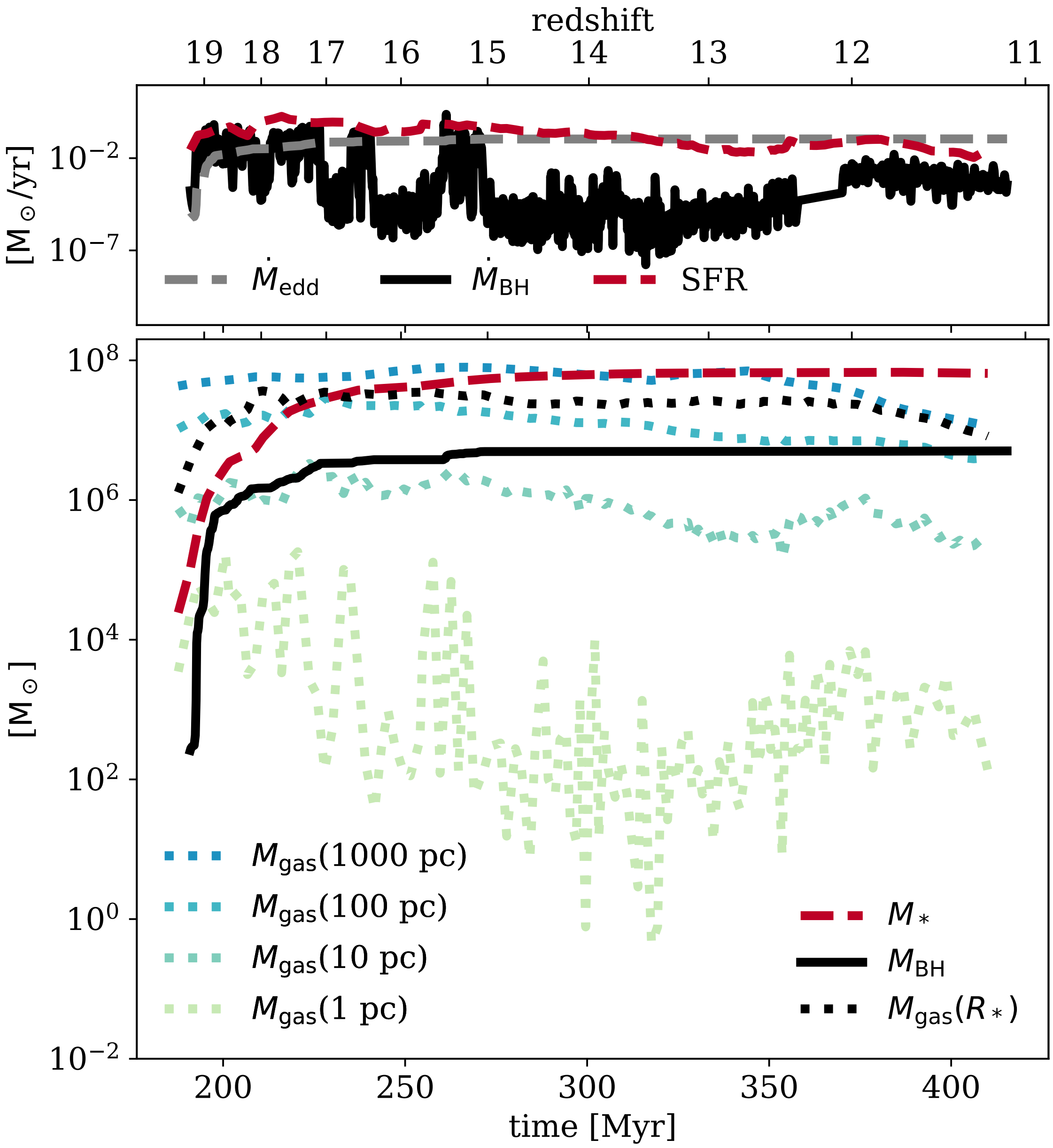

Figure 6.3: Time evolution of black hole and host galaxy mass properties. The top panel shows the black hole accretion rate $\dot{M}_{\rm BH}$ and the star formation rate. The bottom panel shows mass evolutions, including black hole mass, $M_{\rm BH}$, stellar mass, M_*, and gas masses measured within a sphere of a given radius, r_i, centred on the black hole, $M_{\rm gas}(r_i)$. All quantities are sampled at 5 coarse timesteps except for the black hole accretion rate, which is sampled at 200 fine timesteps. The smooth black hole accretion rate, $\dot{M}_{\rm BH}$, at $z = 12.3$ is due to missing data.

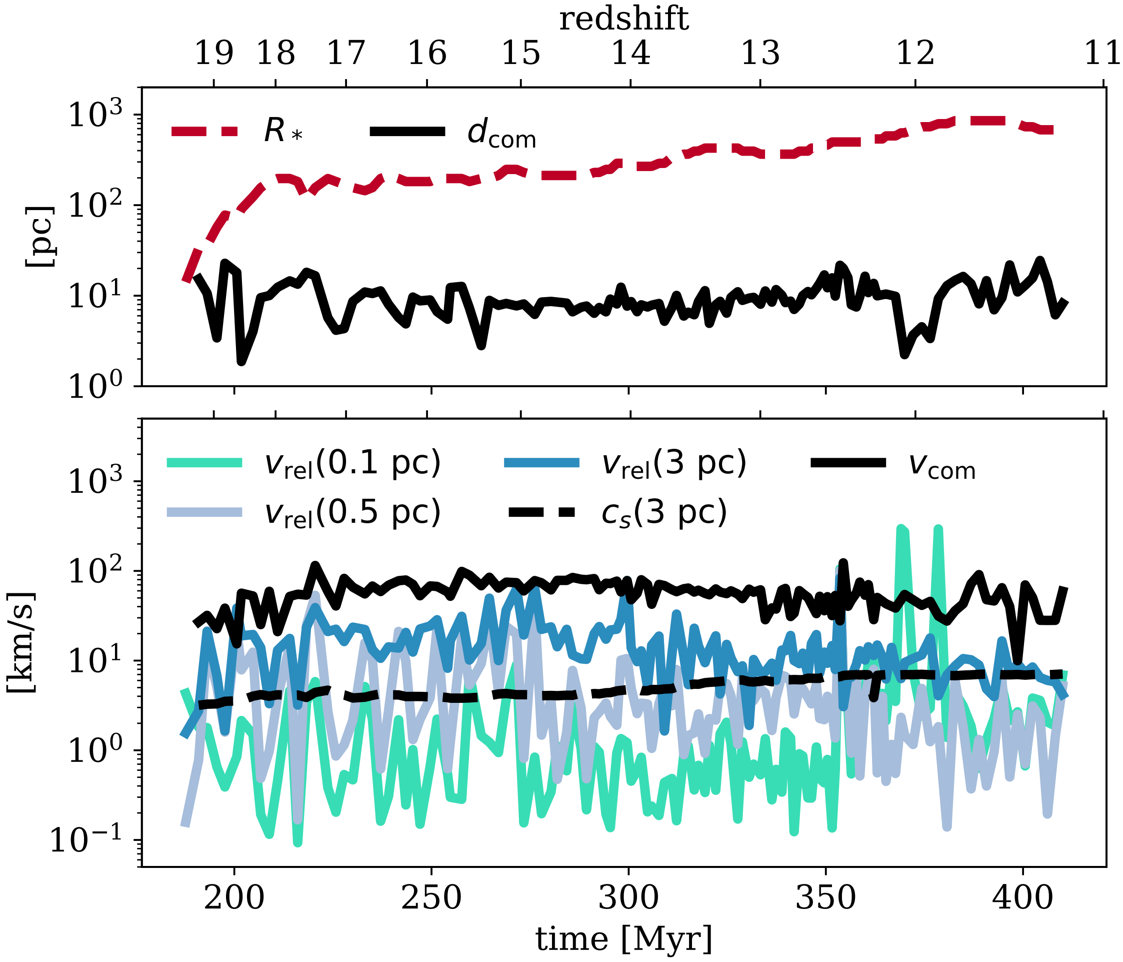

Figure 6.4: Time evolution of radii and velocities associated with the black hole and host galaxy. The top panel shows the stellar radius of the galaxy, R_*, and the distance from the black hole to the stellar centre of mass, $d_{\rm com}$. The bottom panel shows the sound speed, c_s, and the relative velocity of the black hole to the gas centre of mass within a given radius, $v_{\rm rel}(r_i)$. $v_{\rm com}$ is the relative velocity between the black hole and the stellar centre of mass. All quantities are sampled at 5 coarse timesteps.

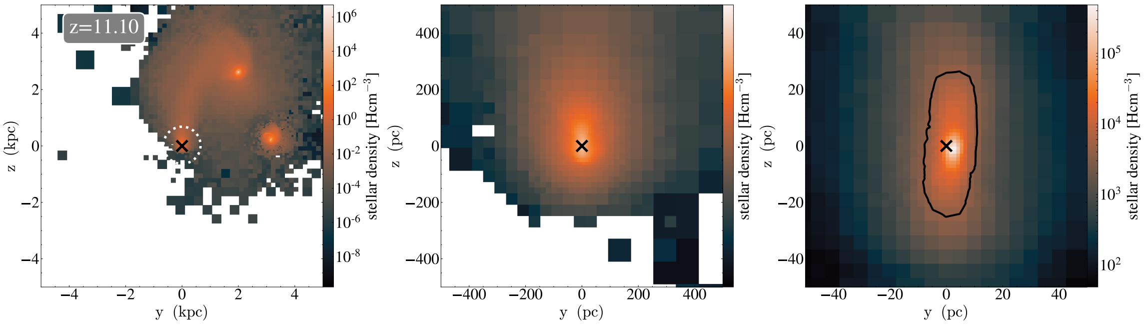

Figure 6.5: Projection of the stellar density when deposited on the grid for three different zooms. White cells contain no stars. The black hole location is marked by a black cross, and black contours in the right hand panel show the star forming region. A white dotted line denotes R_*. Note that these projections portray the galaxy edge on, as opposed to the face-on projections in Figure 6.6

mass evolves from 2.44×10^4 M$_\odot$ at $z = 19.3$ to 6.51×10^7 M$_\odot$ at $z = 11.1$. Throughout this evolution, the galaxy remains very compact. R_*, which is defined to be the radius that contains 99% of the galaxy's stellar mass at a given point in time (see top panel, Figure 6.4) increases from 14.2 pc at $z = 19.3$ to 628 pc at $z = 11.1$. The value increases as the galaxy interacts with a nearby object, which forms an extended tidal tail (see Figure 6.5).

Until $z = 16.78$, the galaxy is extremely gas rich, with the gas mass contained within R_* exceeding both the stellar mass and the black hole mass, both of which grow rapidly as a result (see the second panel in Figure 6.3). Gas is accreted and converted into stars faster than it is replenished and the gas mass within the galaxy drops, first only at the centre of the galaxy (compare $M_{\mathrm{gas}}(10\ \mathrm{pc})$ in Figure 6.3 at $z = 17$ to $z = 11$), but later also at large scales (see $M_{\mathrm{gas}}(1000\ \mathrm{pc})$ in Figure 6.3 between $z = 13$ and $z = 11$). Black contours in the central column of Figure 6.6 show that the region containing star forming gas shrinks with time, from an extended distribution with a radius of more than 20 pc at $z = 15.06$ to a compact clump with a radius of less than 5 pc at $z = 11.1$. Even when the distribution of star-forming gas is more extended, the bulk of star formation occurs within the central 10 pc of the galaxy, as can be seen by the correlation between the star formation rate (SFR) and the evolution of the gas mass within the central 10 pc of the galaxy, $M_{\mathrm{gas}}(10\ \mathrm{pc})$. By $z = 11.1$, the gas fraction within the galaxy, $M_{\mathrm{gas}}(R_*)/M_*$, has fallen to 0.126. Star formation therefore clearly provides strong competition for early black hole accretion.

Even this early in the history of the galaxy, rotation plays a significant role and the gas in the host galaxy shows large-scale spiral features at all times (see Figure 6.6).

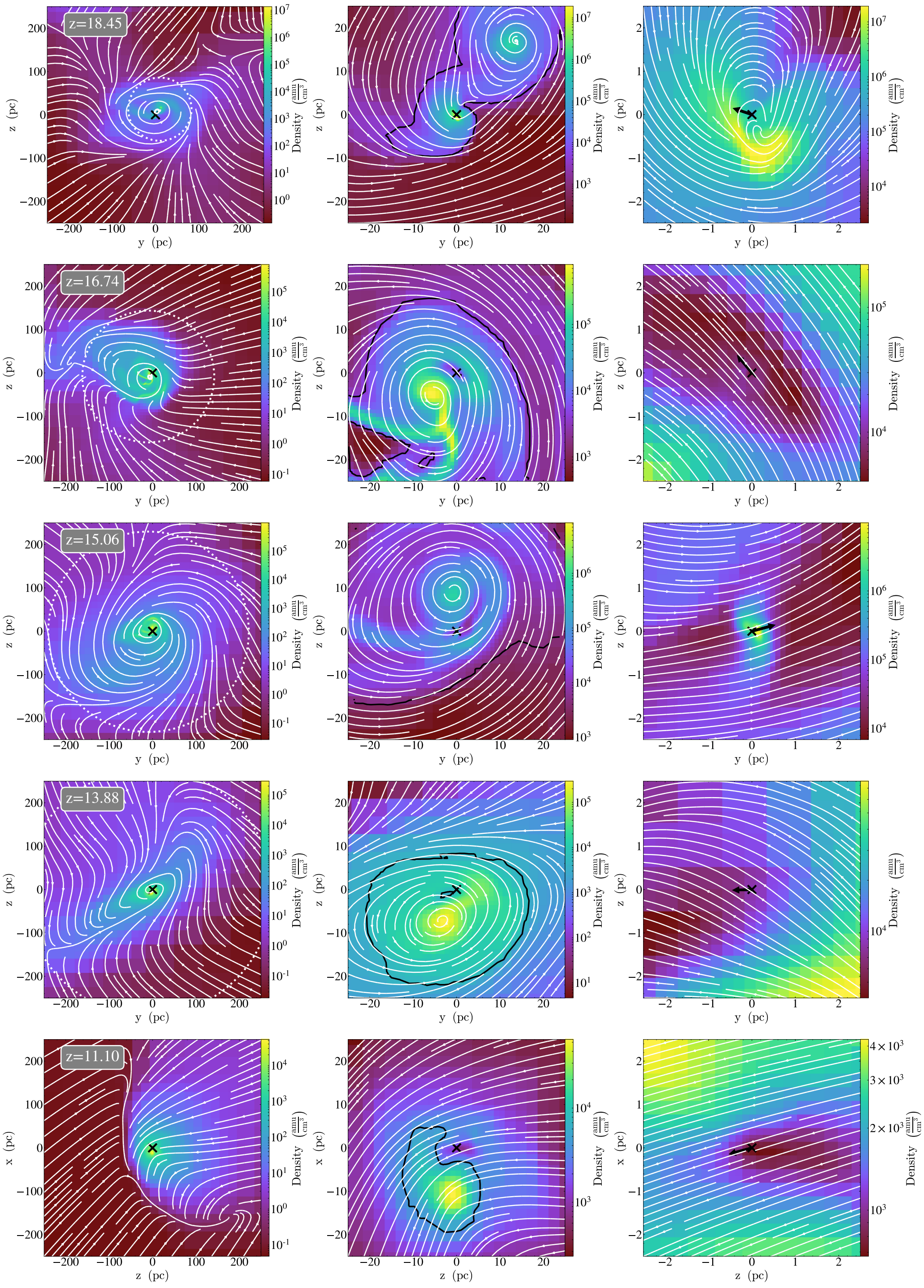

Figure 6.6: Projected gas density at three different zooms for 5 different epochs. The line of sight at each time is chosen to be the closest grid axis to the angular momentum vector of the gas contained within R_*. Streamlines are annotated in white and calculated in the centre of mass frame of gas contained within R_*, which is marked as a dotted white circle in the left panel when smaller than the edge length of the projection. The black hole location is marked by a black cross, and the black arrow in the right column denotes the direction of the instantaneous, normalised sink velocity. Black contours in the central column mark star forming gas.

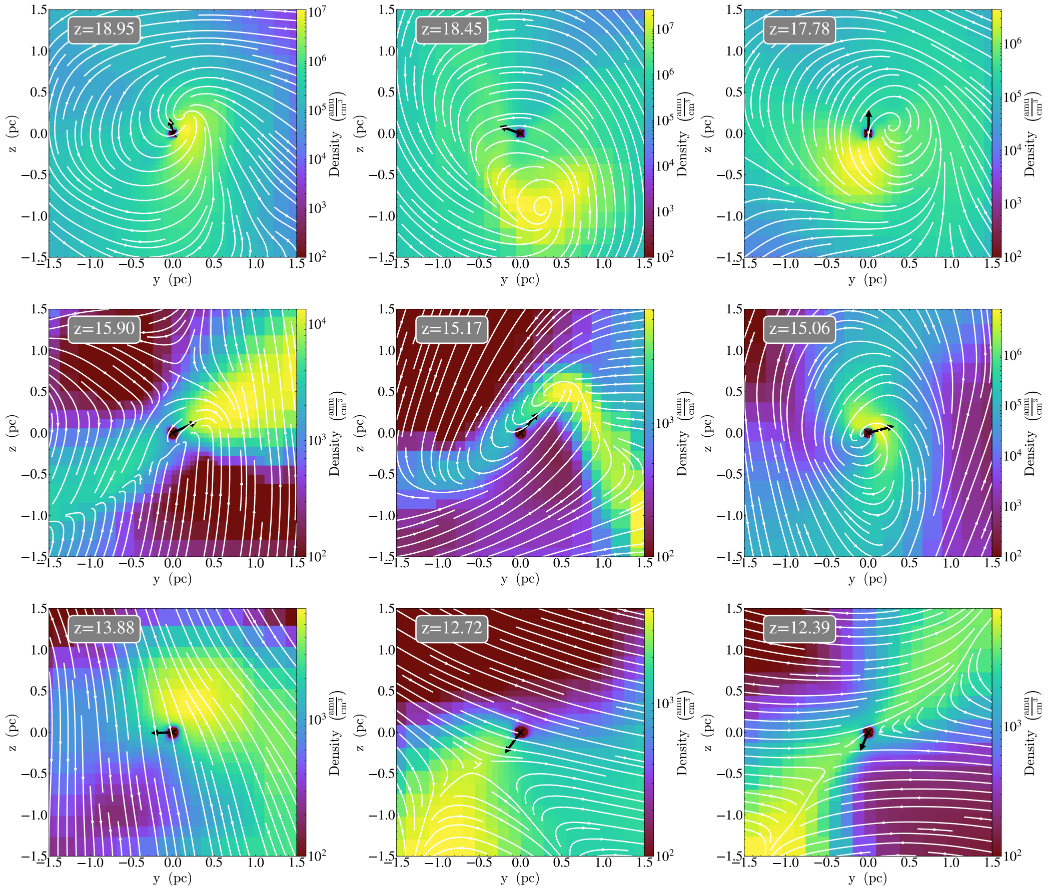

Figure 6.7: Gas density slices in the black hole vicinity at 9 different epochs. Streamlines are annotated in white and calculated in the frame of the centre of mass within a 1.5 pc sphere centred on the black hole. The black hole location is marked by a black cross, and the black arrow denotes the direction of the instantaneous, normalised sink velocity.

A diameter of ~ 30 pc for the galactic disc is in agreement with work by Shlosman et al. (2016). The gas breaks into a few large clumps with diameters on the order of 10 pc or more, which have dense cores with diameters of 1 pc or less. The formation of dense clumps at high redshift is in agreement with comparable numerical work on early black hole evolution (Regan et al., 2014a; Lupi et al., 2016), who highlight the importance of these dense clumps on early black hole accretion. However, the clumps formed here are large in comparison to simulations that resolve turbulence (Yoshida et al., 2008; Latif et al., 2013a), which report clump diameters in the range $0.01 - 0.03$ pc.

Two factors determine the accretion rate onto the black hole: the gas supply within the galactic centre (see $M_{\text{gas}}(10 \text{ pc})$ in Figure 6.3) and the black hole's proximity to the central density peak. Figure 6.6 shows that accretion bursts occur when the black hole interacts with the cores of dense clumps, such as at $z = 18.45$ (top row) and $z = 15.06$ (third row) and slows down when the black hole is located further out in the gas disc (second, fourth and fifth row). During periods of smooth accretion, under-densities form within the region gravitationally dominated by the black hole (second, fourth and fifth rows of Figure 6.6) which suggests that the black hole remains attached to one section of the disc on sufficiently long timescales to gravitationally drain the local gas supply. The immediate environment of the black hole (see Figure 6.7) is dominated by sub-pc size features that contain between 1 and 10^5 $\mathrm{M}_\odot$ (see $M_{\rm gas}$(1 pc) in Figure 6.3) and show coherent rotation aligned with the large-scale disc early on (top row, Figure 6.7). At later times, the local gas structures have a more filamentary nature and contain significantly less mass, with a peak density of $\sim 10^4$ $\mathrm{H\,cm^{-3}}$ rather than $\sim 10^7$ $\mathrm{H\,cm^{-3}}$. The black hole appears to struggle to remain attached to these sub-pc clumps, despite the fact that it dominates the local gravitational potential. The velocity arrow frequently points in the opposite direction to local gas flow, particularly when the accretion flow is highly directional (Figure 6.7). The dynamics of the black hole are discussed further in the next section.

6.5 Black hole dynamics

The density projections in Figure 6.6 show that for most of its evolution (bottom four rows), the black hole appears to be on an orbit with a 10 pc radius around the centre of the galaxy. This is confirmed by the distance between the black hole and the stellar centre of mass (see Figure 6.4), $d_{\rm com}$, which is of the order of 10 pc for most of the simulation. It does however show significant variation on < 10 Myr timescales, suggesting that the black holes does not remain attached to a fixed orbit.

The density projections in the right hand column of Figure 6.6 show that the black hole velocity vector is in general well aligned with the gas flow on pc scales. However, studying the black hole's relative velocity to the local gas in more detail is challenging, as there is no unambiguous way to define a velocity vector between the black hole and the local gas, particularly in rapidly evolving, clumpy environments. Chaotic sub-pc scale structures in the black hole vicinity, such as the ones in Figure 6.7, indicate the presence of significant velocity dispersion on pc or sub-pc scales. On larger scales, the relative velocity measures the orbital velocity of the black hole, and on smaller scales, the velocity of cells just at the edge of the accretion region can be dominated by the inflow velocity into the black holes gravitational potential (see

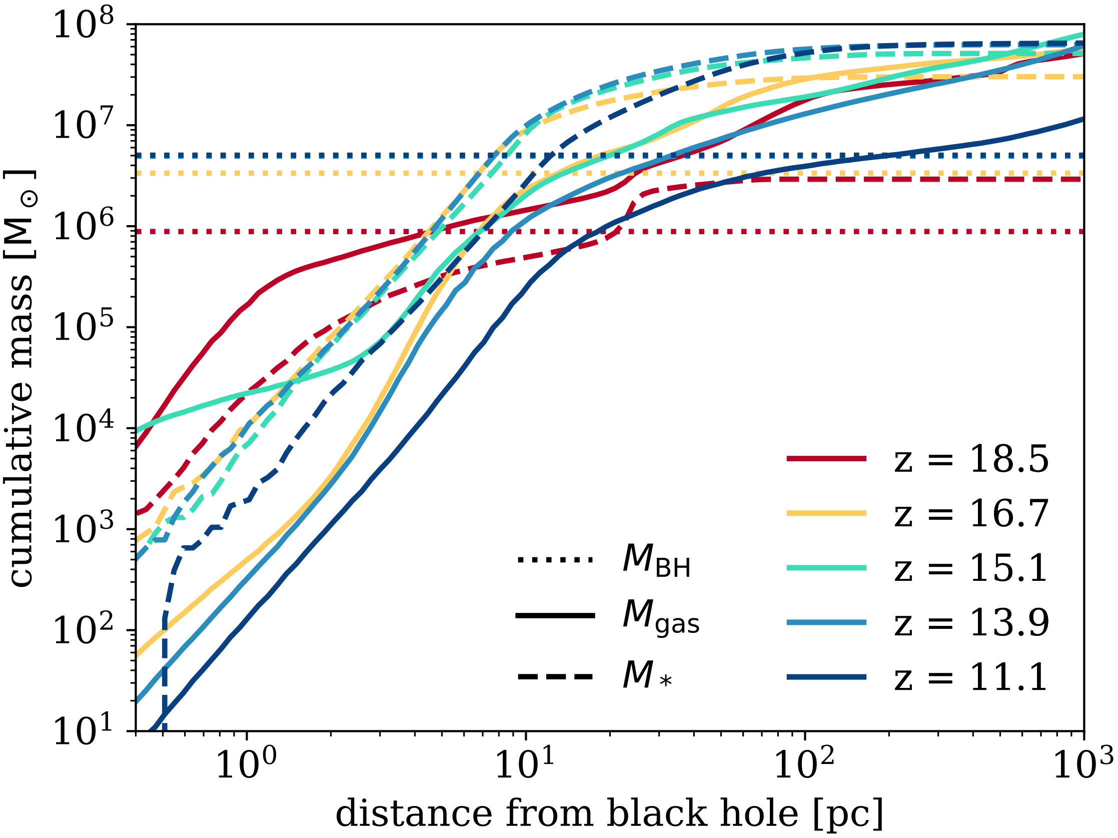

Figure 6.8: Cumulative spherical mass profiles for gas, stars and the black hole at the same epochs as the density projections in Figure 6.6. Radii are measured from the position of the black hole.

Section 5.4). Within the accretion region itself, the gas velocity can be significantly affected by the choice of sub-grid algorithm (Bleuler & Teyssier, 2014).

To alleviate some of these difficulties, the black hole's velocity is compared to the centre of mass velocity of the gas contained within a sphere centred on the black hole for a range of radii. As can be seen in Figure 6.8, the black hole mass dominates over both the stellar and gas mass contained within a sphere with a minimum of 5 pc, i.e. on scales larger than the density projections in the right hand panels of Figure 6.6. The bottom panel of Figure 6.4 shows that the relative velocity to gas at the edge of the accretion region, $v_{rel}(0.1\text{ pc})$, is subsonic. Compared to gas on larger scales, e.g. within 3 pc of the sink, the black hole is moving supersonically, suggesting that it is not dynamically attached to a particular part of the galactic gas disc. The velocity in the frame of the stellar centre of mass reflects the black hole's orbital velocity. For example, at $z = 11.1$ the Keplerian velocity of a 10 pc orbit is 73.32 kms^{-1}, in good agreement with $v_{\rm com}$. The rapid variations in $v_{rel}(0.5\text{ pc})$ and $v_{rel}(0.1\text{ pc})$ are driven by rapidly evolving clumps within the black hole's potential.

On the one hand, the dynamics of the black hole shows that a common issue in hydrodynamical simulation, in which the black hole is ejected from its host galaxy, is avoided here (Sijacki et al., 2007; Volonteri et al., 2016; Biernacki et al., 2017). On the other hand, the black hole is sufficiently massive that dynamical friction due to a gravitational wake downstream of the black hole should transfer momentum from the black hole to the gas (see Ostriker, 1999, and Section 3.5.4), and allow the black hole's orbit to decay. Chapon et al. (2013) report a decay time of 10 Myr for a black hole binary of comparable black hole masses and local gas densities to decay to 1 pc via dynamical friction, which is far shorter than the 200 Myr of evolution studied here. One would therefore expect the black hole to settle into the bottom of the galaxy's gravitational well over time, particularly in the absence of feedback, as is seen for comparable black hole's in isolated discs (see Chapter 5). Numerical issues with the orbit integrator for the black hole are ruled out by Lupi et al. (2015), who showed that the orbit of massive particles is well resolved at the resolutions probed here.

The lack of orbital decay could be due to the application of an analytic drag force on the black hole, originally designed to account for dynamical friction of the gas (see Section 3.5.4). This sub-grid algorithm remained active throughout the simulation, and the high coulomb logarithm of $\ln(\Lambda) = 250$ employed here means a strong drag force was applied to the black hole at all times, with its direction and magnitude determined by the relative velocity as measured at the edge of the accretion region. As previous work has shown (Chapters 4 and 5), the local gas properties poorly represent the gas bulk properties when the former are measured deep within the gravitational potential of the black hole. This affects both the magnitude and the direction of the drag force. This is particularly true in a turbulent environment like the one encountered here, where the immediate vicinity of the black hole is dominated by dense sub-pc gas features with a large velocity dispersion (see Figure 6.7). In the presence of this chaotic flow at the edge of the accretion region, the drag force will give the black hole unphysical velocity kicks in the frame of the galaxy, by updating the black hole's velocity according to chaotic local flows. In this scenario, $v_{\rm rel}(0.1$ pc) would be low, while relative velocities to gas on pc scales would be high, which matches the velocities of the black hole studied here (see $v_{\rm rel}(0.1$ pc) and $v_{\rm rel}(3$ pc) in Figure 6.4). In the context of the galaxy's gravitational potential, accelerating or deceleration the black hole moves it to larger or smaller orbits respectively, which leads to the variations in $d_{\rm com}$ in Figure 6.4. The overall dynamics of the black hole are therefore dominated by the galactic gravitational potential but the drag force keeps the black hole dynamically hot and prevents its orbit from decaying.

It is difficult to asses the impact of the black hole's dynamics on its long-term mass evolution. On the one hand, the black hole studied here appears to remain attached to drained regions of the disc for extended periods of time, limiting its long-term access to the galactic gas supply. Shlosman et al. (2016) studied the early evolution of a sample of seed black holes in a high-redshift galaxy and found that only those black holes located in the centre of the galactic disc underwent sustained accretion over tens of Myr. This supports the conclusions that the accretion rate in the simulation presented here might be underestimated. On the other hand, Alexander & Natarajan (2014) argue that being attached to a slowly draining gas disc decreases long-term accretion rates. Whichever approach is more relevant most likely depends on the particular gas configuration in the galaxy. It is therefore extremely important to be confident that the black holes orbit within the galaxy is captured as self-consistently as possible. To rule out any lasting impact, future work will repeat the simulation in this chapter using the seed mass prescription presented in Chapter 5.

6.6 Will this black hole become supermassive?

Is the black hole presented in this work a progenitor to the supermassive black holes with masses above 10^9 M$_\odot$ observed at redshift $z = 6$? Having reached a mass of 5.05×10^6 M$_\odot$ at $z = 11.1$, and accreting with an average accretion rate of 8.0×10^{-4} M$_\odot$yr^{-1} over the last 10 Myr, it seems unlikely. If the black hole continues growing at this rate, it would only reach 5.53×10^6 M$_\odot$ by $z = 6$. If it maintains the time averaged Eddington ratio of the last 10 Myr, $\chi = 7.3 \times 10^{-3}$, it will reach 1.01×10^7 M$_\odot$ by $z = 6$. However, as shown in Figure 6.3, the accretion rate onto the black hole can vary by 8 orders of magnitude on Myr timescales, and brief super-Eddington accretion bursts are responsible for approximately 90% of the total black hole mass gained until $z = 11.1$. Extrapolating the black holes future evolution based on its instantaneous accretion rate is therefore unreliable.

Figure 6.9 compares the coevolution of black hole mass, $M_{\rm BH}$, and stellar mass of the host galaxy, M_*, to empirical $M_{\rm BH} - M_*$ relations at redshift $z = 3, 6$ and 11 from the fit in Equation 2 of Decarli et al. (2010). The reader is advised that the model in Decarli et al. (2010) is based on observational data up to redshift $z = 3$ and the linear extrapolations at higher redshifts have to be treated with caution.

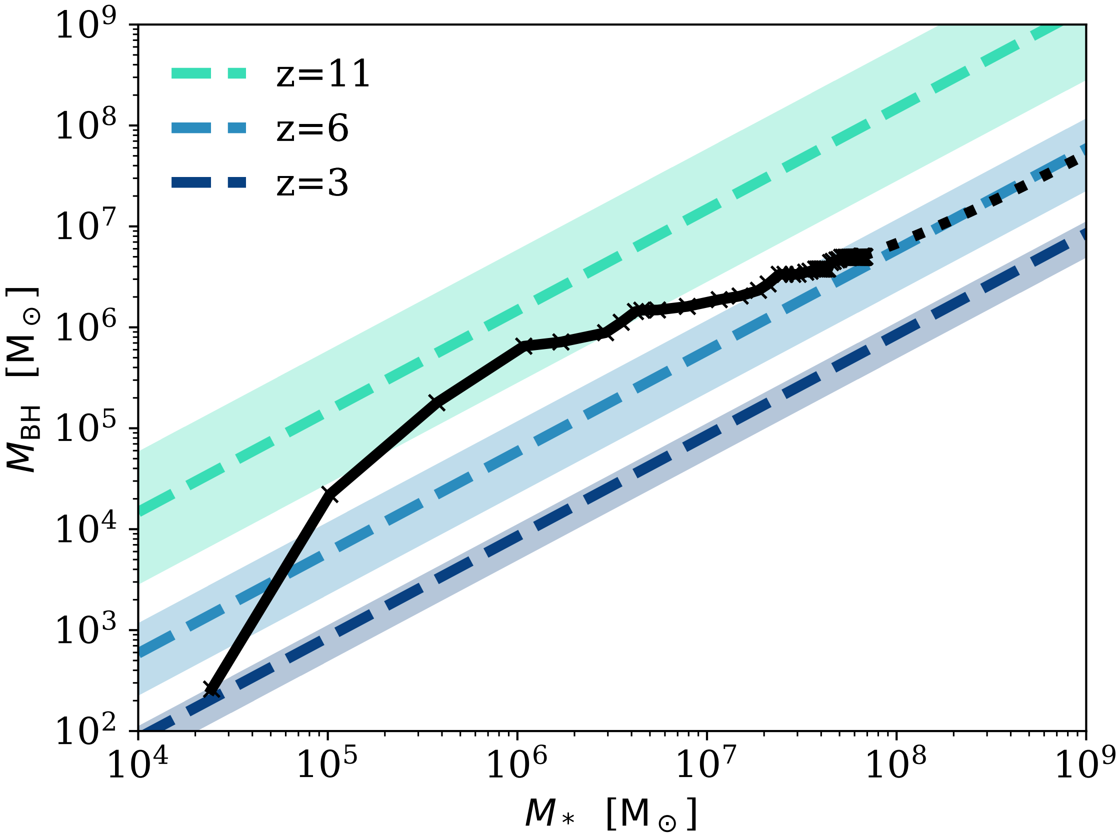

Figure 6.9: $M_{\rm BH} - M_*$ relation for the black hole and its host galaxy, including fits based on observations by Decarli et al. (2010). The stellar and black hole masses are both sampled at five coarse timesteps of the simulation. The dotted line shows the future evolution of the galaxy and black hole if $\mathrm{SFR}/\dot{M}_{\rm BH} = 0.05$ is maintained.

As suggested by Decarli et al. (2010), the black hole grows early, although not as much as a simple extrapolation of the $\mathrm{M_{BH}} - \mathrm{M_*}$ relation from $z = 3$ to $z = 11$ would suggest. The galaxy studied here then evolves to the right on the $\mathrm{M_{BH}} - \mathrm{M_*}$ diagram, as black hole accretion slows down while the SFR remains high from $z = 15$ to $z = 12$. For the last 50 Myr covered by the simulation, the gap between SFR and $\dot{M}_{\mathrm{BH}}$ narrows as star formation slows down while the black hole accretion rate increases. If the average $\mathrm{SFR}/\dot{M}_{\mathrm{BH}} = 0.05$ ratio of the last 10 Myr is maintained, the galaxy and its black hole will evolve almost perfectly along the empirical $M_{\mathrm{BH}} - M_*$ (see dotted line in Figure 6.9). However, the impending merger (see Figure 6.5) will add stellar mass to the galaxy independently of the current SFR, boosting the object to the right on Figure 6.9. Mergers will also most likely significantly affect both the SFR and $\dot{M}_{\mathrm{BH}}$ (DeBuhr et al., 2010; Gabor et al., 2015; Capelo et al., 2015), making the future evolution difficult to predict from current trends.

By redshift $z = 11.1$, M_{BH}/M_* lies within the error range of the $M_{\mathrm{BH}} - M_*$ relation at $z = 6$ from Decarli et al. (2010). This could either mean that the black hole is under-massive or that the host galaxy is over-massive. The latter is more likely in the case presented here, as the SFR will be overestimated in the absence of stellar feedback (Dekel & Silk, 1986; Benson et al., 2003; Hopkins et al., 2014; Rosdahl et al., 2017). Alternatively it could suggest that galaxies evolve onto the $M_{\mathrm{BH}} - M_*$ relation early, and then continue to grow along it. At this point, the galaxy is housed in a $M_{\mathrm{DM}} = 8.77 \times 10^8$ $\mathrm{M_\odot}$ halo, which will have grown to $M_{\mathrm{DM}} = 6.7 \times 10^{11}$ $\mathrm{M_\odot}$ by redshift $z = 6$. Assuming $M_{\mathrm{BH}}/M_{\mathrm{DM}} = 5.7 \times 10^{-3}$ is maintained for the remainder of the simulation, the black hole will grow to 3.88×10^9 $\mathrm{M_\odot}$ by $z = 6$. This value is consistent with the black hole mass reported in Smidt et al. (2013) who study a comparable halo at a significantly lower resolution of 35 pc, and exceed the mass of 8×10^7 $\mathrm{M_\odot}$ reported in Dubois et al. (2013a) for a comparable halo at 15 pc resolution under the influence of AGN feedback. There is therefore clearly scope for the black hole to become as massive as the first quasars observed at $z = 6$, but its path there remains uncertain.

6.7 Does this simulation provide an upper limit for black hole accretion?

The simulation presented in this chapter does not include feedback, neither from the black hole nor from stars. Does the work presented here therefore provide an upper limit for mass gained by gas accretion onto black holes in the early Universe?

On the one hand, the lack of black hole feedback means that the gas cannot be ejected from the galaxy. In this sense, the gas supply available in the galaxy does provide an upper limit and accretion proceeds uninterrupted from the kind of feedback episodes that cause strong outflows in other simulations, both locally and globally (Dubois et al., 2012a; Lupi et al., 2016; Curtis & Sijacki, 2016a). The halo studied here is sufficiently massive that supernova (SN) feedback will be unable to gravitationally unbind gas (Johnson et al., 2013), but it could introduce a significant time delay by destroying dense features in the galactic centre and depositing gas at larger radii. The lack of feedback also means the gas is not heated and therefore cools and collapses more effectively than in the presence of feedback. A host halo that consistently remains at the most massive end of the evolving HMF (Lukić et al., 2007) should ensure strong accretion streams that provide ample gas for black hole accretion.

On the other hand, Section 6.4 showed that the main competition for black hole accretion is star formation, which is also fed by cold, dense gas and therefore overestimated in the absence of feedback. Any mass locked up in stars is unavailable for black hole accretion for the remainder of the simulation, so it is possible that the lack of (stellar) feedback actually decreases the gas reservoir within the galaxy. This effect will be particularly relevant for galaxies without black holes. They will convert a significant share of their gas into stars prior to merging into the black hole's host, so the galactic gas reservoir will be replenished less efficiently during mergers.

One question not addressed in the work presented here is the impact of metal enrichment and ionizing radiation on the chemical composition and ionization state of the gas, which has important consequences for its ability to cool effectively. As discussed in detail in Johnson et al. (2013), the impact of feedback from the first stars on gas cooling is complex, and it can both increase and decrease cooling times at redshifts $z > 10$. The impact of metals and cooling physics on black hole accretion will be investigated in future work.

SN feedback could also shock local gas and introduce turbulence. This could aid the formation of clumps, which have been shown to play an important role particularly in early black hole accretion (Section 6.4 and Emsellem et al., 2014; Lupi et al., 2016), either by directly feeding the black hole or by disrupting local rotationally supported discs (see Chapter 5 and Levine et al., 2008). In particular the free expansion phase of the SN, which covers the first few pc of expansion and in which the SN feedback is stored in the kinetic energy of the ejecta (Johnson et al., 2013; Kimm & Cen, 2014), could play an important role in creating clumps. From this point of view, the gas accreted by black hole could be underestimated in the absence of feedback. Yoshida et al. (2008) and Latif et al. (2013a) showed that the rotation supported gas disc could be the result of under-resolved turbulence, which would require au resolution within dense clumps to capture self-consistently. Therefore, the gas studied in this simulation might be too smooth for effective black hole accretion, due to the absence of stellar feedback.

There is also the issue of dynamical friction, discussed in Section 6.5, which potentially affected the black hole's orbit and ability to accrete from the densest clumps contained within the galaxy. Solving this problem could allow the black hole to settle in the centre of the galaxy, potentially boosting accretion (Shlosman et al., 2016). On the other hand, Alexander & Natarajan (2014) argue that early dynamical disruption boosts black hole accretion, as it can prevent slowly draining accretion discs, in which case the accretion rate presented in this simulation might be an overestimate.

In conclusion, determining an upper limit on the mass gained by early Universe black holes via gas accretion is difficult, as the gas supply in the galaxy, particularly the gas supply in the black hole vicinity, is influenced by a multitude of competing processes. Therefore the results presented in this chapter most likely do not present an upper limit for gas accretion onto the black hole, as the long-term consequences of the parameter choices of the simulation presented here are difficult to ascertain.

6.8 Conclusions

This chapter investigated the first 200 Myr of a potential SMBH and its coevolution with its host galaxy. Both are housed in a halo with the same number density as quasars at redshift $z = 6$ (McGreer et al., 2013), which remains at the massive end of the local HMF throughout its evolution (Lukić et al., 2007). The black hole formed in the extremely gas rich proto-galaxy at redshift $z = 19.3$. With a seed mass of 260 $\mathrm{M}_\odot$ it is the equivalent of a stellar remnant from the first massive stars (Volonteri, 2010). During an early super-Eddington accretion burst, the black hole grows to 10^6 $\mathrm{M}_\odot$ in 10 Myr, which is matched by a similarly rapid evolution in the stellar mass of the host galaxy. As the gas supply in the galaxy becomes depleted over the next 200 Myr, star formation and black hole accretion slow down.

By the end of the simulation, at $z = 11.1$, the galaxy has a stellar mass of 6.5×10^7 $\mathrm{M}_\odot$ and houses a black hole of 5.05×10^6 $\mathrm{M}_\odot$. If the trends in $M_{\mathrm{BH}}/M_{\mathrm{DM}}$ continue, the black hole will grow to 3.88×10^9 $\mathrm{M}_\odot$ by $z = 6$, matching observations of the first quasars and low resolution work of a comparable halo reported in Dubois et al. (2013a) and Smidt et al. (2013). However, accretion onto the black hole has fallen to an average Eddington ratio of $\chi = 7.3 \times 10^{-3}$ over the last 10 Myr of evolution included in the simulation, making it unlikely that the black hole will reach 10^9 $\mathrm{M}_\odot$. Doing so would require an average accretion rate above the Eddington limit for the remainder of the black hole's evolution, or repeated, extended super-Eddington accretion bursts.

By $z = 11.1$ the black hole has fallen behind the predictions by Smidt et al. (2013), as gas within the central 35 pc (the resolution limit in Smidt et al., 2013) has settled into a rotation supported disc and a significant amount of it has been converted into stars. The black hole primarily grows via interactions with dense gas clumps with diameters of 1 pc or less, which is in agreement with other high resolution work on early black hole accretion (Emsellem et al., 2014; Lupi et al., 2016; Shlosman et al., 2016). The sub-pc gas clumps reported here can be both disc-like and filamentary and vary in mass by up to five orders of magnitude on timescales of less than a Myr. When attached to a massive clump, the black hole undergoes super-Eddington accretion bursts that contribute up to 90% of its total mass.

Unlike the work reported in Shlosman et al. (2016), where the most massive black hole settles in the centre of the galactic gas disc, the black hole presented here remains within 10 pc from the centre of the galaxy. Accretion proceeds via slow draining of the gas within the region of the disc gravitationally dominated by the black hole, which hampers its ability to grow over long periods of time. While the simulation presented here has sufficient resolution to self-consistently solve for dynamical friction due to the gravitational focusing of gas (see Chapter 5), an analytic sub-grid drag force applied throughout the simulation seems to have prevented the black hole orbit from decaying by providing small-scale velocity kicks due to directional gas flows in the black hole vicinity. The discussion on the dynamical evolution of the black hole in this chapter highlighted the importance of carefully examining sub-grid algorithms, particularly when resolving previously unattainable length scales, as analytic models can have unintended consequences that have lasting effects on the evolution of the object under investigation. This issue is the main limitation of the simulation presented in this chapter and will be addressed in future work.

With the current algorithm, star formation is prevented for all cells with a resolution below 1 pc, which span a region with a 5 pc diameter, centred on the black hole. If the black hole settled in the centre of the galaxy, this exclusion zone would cover a significant share of the central 10 pc clump responsible for the bulk of the galaxy's star formation. Developing an algorithm in which star formation is permitted within the zoom-within-zoom region, without overwhelming the simulation with light particles, will be a prerequisite for future work. As confirmed by the simulation presented in this chapter, the main numerical cost of zoom-within-zoom refinement is the integration time-step, which is halved for each additional level of refinement. The extra numerical cost of additional black holes at zoom-within-zoom resolution should therefore be limited, potentially providing an opportunity to study a sample of early black holes and investigate the importance of black hole mergers for early SMBH evolution.

In comparison to lower resolution hydrodynamical simulations of the origin of SMBH, such as Dubois et al. (2013a) and Smidt et al. (2013) who reach resolutions of 15 pc and 32 pc respectively, the work in this chapter has highlighted the importance of resolving sub-pc scales in the black hole vicinity. As demonstrated in Section 5.6 and confirmed by the work in this chapter, the evolution of the black holes studied here using sub-pc resolution is determined by the internal structure of the cloud from which it is spawned, not the choice of seed mass or accretion algorithm. Resolving the internal structure of gas clouds in the black hole vicinity has a lasting impact on the accretion history that would be difficult to translate into an analytic or empirical model. For the black hole simulation studied in this chapter, the rotational support of gas on pc scales limits how effectively the black hole's gas supply within the black hole's gravitational potential is replenished. The accretion duty cycle of the black hole is determined by pc or sub-pc size dense clumps forming within the central pc of the black hole and the relative importance of both is driven by the black hole's orbit within its host galaxy. Self-consistently tracking these processes will be even more important in the presence of feedback, which will add another non-linear effect to an already highly non-linear problem. Will the dense clumps in the black hole vicinity remain an effective way to feed the black hole even in the presence of feedback? Will the gas disc survive?

The discussion on the black hole's orbit in Section 6.5 has also highlighted the importance of self-consistently resolving forces on the black hole. As shown in Section 5.5, black holes fail to accrete efficiently when dynamical friction is under-resolved. Work in this chapter has shown that an overly strong sub-grid treatment of dynamical friction in a galactic context can give the black hole artificial velocity kicks. This issue will only become more severe when feedback accelerates gas in the black hole vicinity. With a strong drag force, the black hole can become attached to its own ejecta and be gravitationally unbound from the galaxy. This is a purely numerical phenomenon, similar to the black hole being accelerated by its own accretion column, as discussed in Section 4.4, or by local clumps, as discussed in this chapter. It will therefore become even more important to self-consistently resolve the gravitational potential of the black hole, particularly early on. For a black hole with $M_{\rm BH} = 10^4$ $\rm M_\odot$, the minimum gravitational scale length, the Bondi radius (Equation 3.17), is equal to 1.7 pc at the temperatures observed in this galaxy. If the gas is heated by feedback, or the black hole moves supersonically, it will only decrease. Maintaining sub-pc resolution in the black hole vicinity is therefore already crucial to correctly capturing its evolution. Also, the results for sub-pc structures feeding the black hole reported in this simulation are an upper limit compared to the turbulence-driven clouds reported in Latif et al. (2013a), which have a diameter of 0.01 pc and internal structure on even smaller scales.

The work presented in this chapter has proved that resolving the black hole's gravitational potential in a cosmological context is viable and even raised the possibility of including a collection of black holes within a single simulation. Tracking gas flows deep into the potential of the black hole has reduced the need for sub-grid models. However, this chapter has also highlighted several areas for improvement that will have to be addressed in future work, including the treatment of star formation within the zoom-within-zoom region and the dynamical friction on the black hole. Long term plans for future investigations include stellar feedback, a consistent treatment of metal cooling appropriate for high redshift, and eventually black hole feedback.

In conclusion, the study presented in this chapter has provided promising early insights into the origin of SMBH. Early black hole accretion relies crucially on the black hole interacting with dense, pc-size clumps. However, rotational support on scales below 100 pc cannot be neglected, as it limits long term black hole accretion. The work presented here has confirmed that the chosen host halo remains a likely candidate for SMBH formation as it provides an ample gas supply within the central 30 pc. Future work in this area will require a more holistic treatment of both the black hole and host galaxy and a wider range of physics to be included, as making sufficient gas available on sub-pc scales to grow seed black holes to the observed supermassive objects by $z = 6$ remains an unsolved problem. Once black holes have become supermassive, energy released by their feedback can have a significant impact on their host galaxy, which is a question explored in the next chapter.

Chapter 7

Cosmic evolution of stellar quenching by AGN feedback: clues from the Horizon-AGN simulation

7.1 Introduction

It has long been known that the hierarchical structure formation paradigm implied by the cold dark matter model, while very successful overall, overproduces objects at the bright and faint end of the luminosity function (White & Frenk, 1991; Kauffmann et al., 1993). Observations show much more inefficient star formation in low and high mass halos, with a peak in efficiency at the luminosity turnover (Bernardi et al., 2013; Moustakas et al., 2013; Davidzon et al., 2017). To avoid the overcooling problem and reproduce the observed luminosity function, an energetic feedback mechanism is required (Cole et al., 1994; Binney & Tabor, 1995; Cattaneo, 2001; Springel et al., 2005; Dubois et al., 2013a). Several different avenues have been suggested to provide the necessary energy input and quench star formation, including the extragalactic UV background, supernova (SN) feedback and feedback due to active galactic nuclei (AGN).

Photoionisation by the extragalactic UV background and the first generations of stars suppresses gas accretion at high redshift, causing a number of smaller dark matter (DM) halos to remain devoid of gas (Gnedin, 2000; Somerville, 2002; Okamoto et al., 2008; Geen et al., 2013). While this mechanism provides a possible solution to the overabundance of very low mass ($\leq 10^8 \mathrm{M}_\odot$) substructures in Milky-Way like halos, it has been shown to have little effect on more massive objects that collapse later (Efstathiou, 2000; Governato et al., 2007).

On the other hand, SN feedback is widely believed to play an important part in quenching star formation in halos with masses below 10^{11} $M_\odot$ (Dekel & Silk, 1986; Benson et al., 2003; Hopkins et al., 2014). Their shallow potentials allow SN driven winds with velocities comparable to the escape velocity (Mac Low & Ferrara, 1999; Christensen et al., 2016; Keller et al., 2015) to empty the host galaxy of a significant amount of gas, thereby efficiently suppressing star formation. However, even stellar feedback concentrated in intense, compact starbursts caused by major mergers or violent disc instabilities cannot quench massive galaxies (Dekel & Burkert, 2014).

It has been suggested that star formation halts in massive objects due to a slowdown in cold flows at low redshift (Feldmann & Mayer, 2014). However Mandelbaum et al. (2016) show that massive quiescent galaxies can have twice as much DM as star forming galaxies, indicating that cosmic inflows probably continue long after star formation has ceased. Furthermore, the slowdown in cold flows is expected to take place over Gyr timescales (Zolotov et al., 2015), which contradicts observational evidence that quenching of massive galaxies takes place on much shorter timescales (Barro et al., 2016; Tomczak et al., 2016). Therefore the reduction of star formation in massive galaxies is unlikely to occur solely because accretion is fading away.

Instead, AGN feedback could provide an effective route to quenching massive galaxies, as well as regulating the growth of supermassive black holes (SMBH) (Binney & Tabor, 1995; Benson et al., 2003; Di Matteo et al., 2005; Bower et al., 2006; Croton et al., 2006; Sijacki et al., 2007; Cattaneo et al., 2009; Fabian, 2012). There are two main mechanisms through which this could proceed. One possibility is that black holes and stars feed from the same cold gas supply until it is depleted by AGN feedback, at which point both processes come to a halt. Observations of galaxies that simultaneously harbour both an AGN and an active starburst provide evidence that supports this claim (Heckman & Kauffmann, 2006; Nardini et al., 2008; Lutz et al., 2008). In this scenario, AGN feedback might even accelerate star formation by further compressing the cold gas of the galaxy, in a so called positive feedback mode (Silk, 2005; Gaibler et al., 2012; Santini et al., 2012; Bieri et al., 2015). However, its main role is to prevent the gas heated/expelled by SN winds to be re-accreted at a later stage, alongside more pristine material. This is the so called maintenance mode, associated with powerful radio emission (Rafferty et al., 2006; Cattaneo & Teyssier, 2007).

Alternatively, AGN feedback could act directly on the gas content of the galaxy. It could expel the interstellar medium (ISM) out of galaxies in massive galactic winds and/or prevent star formation by directly heating the ISM gas (Miller, 1988; Pedlar et al., 1990; Springel et al., 2005; Di Matteo et al., 2005; Murray et al., 2005; Fabian et al., 2006b). This view is supported by observational evidence of frequent and fast outflows in massive galaxies (Tremonti et al., 2007) able to drive a significant gas mass (Heckman et al., 2000; Veilleux et al., 2005; Weiner et al., 2009; Sturm et al., 2011) using only $5 - 10\%$ of accretion power (Moe et al., 2009; Saez et al., 2009; Dunn et al., 2010). Whilst such outflows are common in AGN with powerful radio jets (de Kool et al., 2001; Arav et al., 2001; Reeves et al., 2003; Chartas et al., 2007; Dunn et al., 2010; Moe et al., 2009), careful analysis of higher redshift objects provides evidence that quasars can also launch powerful energy driven winds and thus cause a rapid star formation decline (Maiolino et al., 2012; Page et al., 2012; Farrah et al., 2012; Cicone et al., 2014; Costa et al., 2015; Williams et al., 2017). Both modes of AGN feedback can reproduce the observed correlations between host galaxy and black hole properties (Ferrarese & Merritt, 2000; Gebhardt et al., 2000), as shown by e.g. Di Matteo et al. (2008); Booth & Schaye (2009); Dubois et al. (2012a); Sijacki et al. (2015); Volonteri et al. (2016). However, the timescale over which quenching takes place is still a matter of debate, with evidence existing for both rapid quenching (Thomas et al., 2005) and much slower processes (Quintero et al., 2004). The timescales were thought to dependend on galaxy type (Schawinski et al., 2014) but recent research has shown that there is no correlation (Smethurst et al., 2016).

In this chapter, state-of-the-art cosmological simulations are used to investigate when and how AGN feedback affects their host galaxies. The impact of such feedback on stellar masses and large scale gas flows is isolated by comparing the evolution of a statistically representative sample of individual objects, identifying matching galaxies in two simulations, HORIZON-AGN (H-AGN) and HORIZON-noAGN (H-noAGN). As their names indicate, these simulations are identical in all aspects except one is run with, and the other without, AGN feedback. Following the evolution of twinned galaxies from redshift $z = 5$ down to $z = 0$ allows the epoch of quenching to be determined and identifies ensuing changes in the stellar masses of affected galaxies.

The chapter is structured as follows: Section 7.2 briefly introduces the HORIZON simulation suite and Section 7.3 explains the procedure used to identify pairs of corresponding objects across both simulations. Section 7.4 presents the effect of AGN feedback on galaxy stellar masses throughout cosmic time and Section 7.5 determines the causes for the measured quenching by studying the evolution of the black hole population, the gas content of halos and galaxies and gas inflow/outflow rates. Section 7.6 summarises and discusses the results.

7.2 The simulations

This chapter presents a comparative analysis of two simulations: HORIZON-AGN (H-AGN) and HORIZON-noAGN (H-noAGN). Both simulations are run from identical initial conditions and share the same technical specifications and implementations of physics. The only difference is that HORIZON-AGN also includes a sub-grid modelling of SMBHs and the associated AGN feedback (see Section 7.2.5) whereas H-noAGN does not. More details can be found in Dubois et al. (2014b).

7.2.1 Cosmology and initial conditions

Both simulations were run with RAMSES (Teyssier, 2002). Initial conditions were produced using MP-GRAFIC, which creates a three-dimensional Gaussian random field with the same power spectrum as the cosmic microwave background (CMB) (Prunet et al., 2008). The initial conditions setup is a standard ΛCDM cosmology consistent with the WMAP-7 data (Komatsu et al., 2011), with $\Omega_{\rm m} = 0.272$, dark energy density $\Omega_\Lambda = 0.728$, baryon density $\Omega_{\rm b} = 0.045$, Hubble constant $H_0 = 70.4$ kms^{-1}Mpc^{-1}, amplitude of the matter power spectrum $\sigma_8 = 0.81$, and power-law index of the primordial power spectrum $n_s = 0.967$. Both simulations were carried out until $z = 0.0$.

The simulated cube of $L_{\rm box} = 100$ Mpc h^{-1} on a side is initially refined uniformly down to physical $\Delta x = 1$ kpc, which requires a root grid with 1024^3 cells at $z \simeq 100$. Extra refinement levels are continuously triggered using a quasi-Lagrangian criterion: a grid cell is split into eight whenever its dark matter (DM) or baryonic mass exceeds eight times the initial DM or baryonic mass respectively. To keep the size of the smallest cells approximately constant in physical units, cells on the most refined level are split into eight every time the expansion factor doubles, if they fulfil the refinement criteria.

There are a total of 1024^3 DM particles in each simulation, leading to a DM mass resolution of 8×10^7 M$_\odot$. All collisionless particles, i.e. DM and stars, are evolved using a multi-grid Poisson solver with a cloud-in-cell interpolation to assign particles to grid cells (see Section 3.2).

7.2.2 Cooling and heating

The gas is allowed to cool down to 10^4 K using H, He and atomic metal cooling following Sutherland & Dopita (1993), and accounting for photon heating by a uniform UV background (Haardt & Madau, 1996) from $z_{\rm reion} = 10$ onwards. The ratio between elements is assumed to be solar in these cooling/heating calculations. The gas follows a mono-atomic equation of state, with adiabatic index $\gamma = 5/3$.

7.2.3 Stars and supernovae

Star formation is modelled using a Kennicutt–Schmidt law (Equation 3.8) with a (constant) star formation efficiency $\epsilon_* = 0.02$. Stars form when the gas number density exceeds $\rho_0/(\mu m_\mathrm{H}) = 0.1\ \mathrm{H\,cm^{-3}}$ where m_H is the mass of a hydrogen atom and μ the mean molecular weight, and star particles are generated according to a Poisson random process (Rasera & Teyssier, 2006; Dubois & Teyssier, 2008) with a stellar mass resolution of $M_* \simeq 2 \times 10^6\ \mathrm{M_\odot}$, kept constant throughout the simulation. To avoid numerical fragmentation, and to mimic the effect of stellar heating by young stars, a polytropic equation of state, $T = T_0(\rho/\rho_0)^{\kappa-1}$, is used for gas above the star formation density threshold, with $\kappa = 4/3$. Stellar feedback proceeds as described in Section 3.4 and includes stellar winds, Type II and Type Ia supernovae. On top of energy, mass and metals injected into the ISM by stellar feedback, the simulations also keep track of a variety of chemical elements (O, Fe, C, N, Mg, Si) synthesised in stars, with stellar yields estimated according to the W7 model of Nomoto et al. (2007).

7.2.4 Supermassive black hole formation and accretion

In H-AGN, black holes are seeded with an initial mass of $10^5\ \mathrm{M_\odot}$ in dense, star-forming regions, i.e. when a gas cell exceeds $\rho > \rho_0$ and is Jeans unstable, provided such regions are located more than 50 kpc away from a pre-existing black hole (Dubois et al., 2010). These black holes subsequently accrete gas at the Bondi-Hoyle-Lyttleton (BHL) rate (Equation 3.18). To somewhat mitigate resolution effects that make it difficult to capture cold, dense regions of the ISM in resolutions like HORIZON, a boost factor α is used, such that $\dot{M}_\mathrm{SMBH} = \alpha \dot{M}_\mathrm{BHL}$, following Booth & Schaye (2009) and Dubois et al. (2012a):

$$\alpha = \begin{cases} (\rho/\rho_0)^2 & \text{if } \rho > \rho_0 \\ 1 & \text{otherwise} \end{cases} \tag{7.1}$$

Accretion is capped at the Eddington rate defined in Equation 2.5. A standard radiative efficiency, typical of a Shakura-Sunyaev accretion disc around the black holes, of $\epsilon_\mathrm{r} = 0.1$ is assumed (Shakura & Sunyaev, 1973).

Black holes are also allowed to merge with one another when they are closer than 4 kpc, and their relative velocity is smaller than the escape velocity of the binary.

Number of twins per mass category			
z	Small	Medium	Large
	$M_* \leq 10^9$ M$_\odot$	10^9 M$_\odot$ < M$_*$ < 10^{11} M$_\odot$	10^{11} M$_\odot$ ≤ M$_*$
0.1	9,508	51,652	2,060
1	13,393	60,547	773
3	11,823	19,855	7
5	2,949	1,626	0

Table 7.1: Number of galaxy twins in each sub-sample, at the range of redshifts presented here.

7.2.5 AGN feedback

Two modes of AGN feedback are implemented in H-AGN as described in detail in Section 3.5.3, depending on the instantaneous accretion rate of the SMBH. At high accretion rates, the quasar mode deposits energy with an efficiency of $\epsilon_{\rm f} = 0.15$. The radio mode takes over at low accretion rates, using a higher efficiency of $\epsilon_{\rm f} = 1$.

The radiative efficiencies of the two AGN feedback modes were chosen to reproduce the scaling relations between black hole mass and galactic properties in the local Universe, $M_{\rm SMBH} - M_*$ and $M_{\rm SMBH} - \sigma_*$ (Dubois et al., 2012a). Note that the two radiation efficiency parameters previously mentioned are the only ones which are tuned in the HORIZON simulations, in the sense that the other parameters (associated with the sub-grid models of star formation and stellar feedback) were not allowed to vary in order to obtain a better match to bulk galaxy properties. For instance, even though the star formation efficiency choice ensures that galaxies will fall on the Kennicutt observational law by construction, it does not automatically guarantee that they will have the correct stellar or gas mass, and/or size at any epoch.

7.2.6 Mass categories for galaxies

To facilitate the presentation of results, the sample of galaxy twins is split into three sub-samples, distinguished by the stellar mass of the H-AGN galaxy. *Small* galaxies are defined as twins with stellar masses $M_*^{\rm H-AGN} < 10^9$ M$_\odot$ in H-AGN, *medium* galaxies as those with 10^9 M$_\odot \leq$ M$_*^{\rm H-AGN} \leq 10^{11}$ M$_\odot$, and *large* galaxies as those with $M_*^{\rm H-AGN} > 10^{11}$ M$_\odot$. See Table 7.1 for the number of twins in each mass category. As a visual guidance, these mass categories will be annotated by solid vertical lines on all relevant plots.

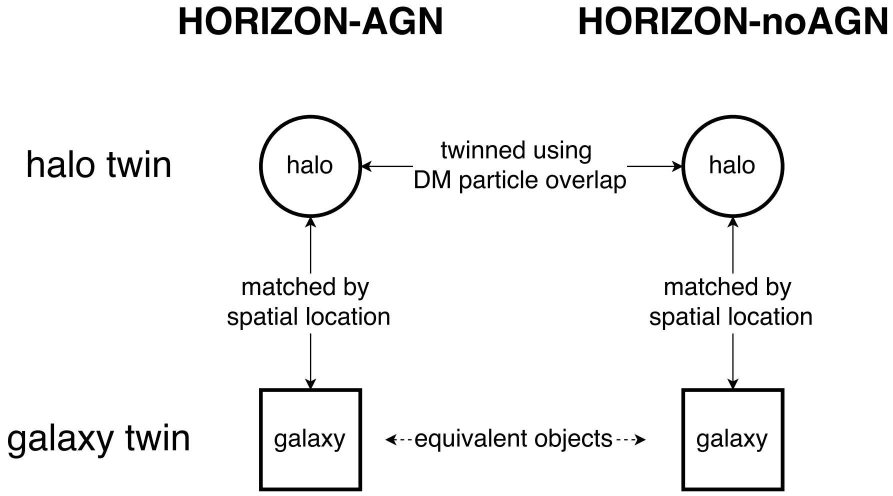

Figure 7.1: To twin galaxies across H-AGN and H-noAGN, DM (sub)halos are first identified in each simulation and (sub)halos that show a 75% or more overlap in the identities of their DM particles are twinned. Galaxies are then associated with a unique host (sub)halo in each simulation, based on how close they are to the centre of the (sub)halo (see Section 7.3 for detail). Galaxies associated with (sub)halo twins are considered galaxy twins.

7.2.7 Nomenclature

For the purpose of this chapter, "quenching" refers to any reduction in galaxy star formation rates when AGN feedback is included, compared to the case without it, not just a redshift dependent specific star formation rate threshold of $0.3/t_{\mathrm{Hubble}}$, as defined in e.g. Franx et al. (2008). The "quenching mass ratio" refers to the stellar mass ratio of galaxy twins between the cases with and without AGN feedback, i.e. $M_*^{\mathrm{H-AGN}}/M_*^{\mathrm{H-noAGN}}$.

7.3 Halo matching across simulations

7.3.1 The twinning procedure

As described in Section 3.6, the first step consists of detecting objects of interest (halos, subhalos and galaxies) in each simulation, using the ADAPTAHOP (sub)halofinder (Aubert et al., 2004; Tweed et al., 2009). Having two simulations based on identical initial conditions allows the identification of corresponding objects between the two simulations, a procedure here referred to as "twinning" (see also e.g. Geen et al. 2013). A pair of corresponding objects is called a twin and identifies two objects (one in each simulation) that have grown from the same overdensity in the initial conditions. These objects can either be (sub)halos for (sub)halo twins, or galaxies for galaxy twins. If all the algorithms implemented to describe physical processes were identical in both simulations, the twins would be identical except for minor differences introduced by the stochastic nature of the star formation algorithm. However, such seemingly innocuous differences would already prevent us from directly twinning galaxies and so the more general method of Peirani et al. (2017) is used here to perform this task, which is summarised in Figure 7.1. Only DM (sub)halos are twinned directly; to create galaxy twins, each galaxy is first associated to a host (sub)halo in the same simulation, before being twinned to the galaxy hosted by this (sub)halo's twin in the other simulation.

More specifically, as both simulations start from identical initial conditions, with uniquely identified DM particles in identical positions, particles that cluster to form gravitationally bound (sub)halos as the runs proceed can be identified. (Sub)halos that grew from the same initial overdensities in both simulations should contain a large fraction of DM particles with identical identities at any time. In practice, for two DM (sub)halos to be twinned, at least 75% of the DM particles present in a (sub)halo in H-AGN are required to also be present in the H-noAGN (sub)halo. Note that in some cases, this choice will lead to a single (sub)halo in H-noAGN being associated with several (sub)halos in H-AGN, as (sub)halos mergers lead to the formation of (sub)subhalos which are not necessarily disrupted at the same time in both simulations. In these cases, the object with the most similar mass is chosen as the twin (sub)halo, and the other matches are discarded.

Star particles stochastically form over the course of a simulation and their identifiers therefore reflect the detailed star formation history of that precise simulation, so it is not possible to identify galaxy twins directly through their star particle identities, as is done for DM (sub)halos. Instead, galaxies are considered to be twins if they are located within DM (sub)halo twins (see Figure 7.1). A host (sub)halo is assigned to each galaxy if its centre is located within a distance $R_{\mathrm{host}} = 0.05 \times R_{\mathrm{vir}}$ of the centre of the (sub)halo. The centres of these (sub)halos are computed using a shrinking sphere method (Power et al., 2003) and their precise location corresponds to the position of the most dense DM particle located in the final sphere identified with this method (Peirani et al., 2017). In case a (sub)halo contains more than one galaxy in its central region, the most massive one is selected as being hosted by this (sub)halo and other matching objects are discarded. Note that proceeding in this way biases results against so-called ‘orphan’ galaxies, i.e. galaxies whose host (sub)halo has been disrupted to the point that it falls below the particle detection threshold used by the halo

Number of (sub)halos

z	$N^{\mathrm{H-AGN}}$	$N^{\mathrm{H-noAGN}}$	$N^{\mathrm{H-noAGN}}/N^{\mathrm{H-AGN}}$	N_{twinned}
0.1	88,171	85141	0.97	75,544
1	89,930	88,488	1.02	83,176
3	53,363	53,551	1.00	53,357
5	16,042	16,065	1.00	15,818

Number of galaxies

z	$N^{\mathrm{H-AGN}}$	$N^{\mathrm{H-noAGN}}$	$N^{\mathrm{H-noAGN}}/N^{\mathrm{H-AGN}}$	N_{twinned}
0.1	75,651	73,450	0.97	63,220
1	80,588	82,238	1.02	72,713
3	34,128	47,428	1.39	31,685
5	6,255	7,757	1.24	4,575

Table 7.2: Number of galaxies, (sub)halos and twins identified in H-AGN and H-noAGN. The sample includes all (sub)halos resolved by at least 500 DM particles and all galaxies resolved by at least 100 star particles that are hosted by these (sub)halos.

finder (see Section 3.6 for details). However, such orphans are quite rare (less than 1% of the sample at any redshift) and almost exclusively belong to the category of small galaxies ($M_* < 10^9$ M$_\odot$), so the conclusions are unaffected by this bias. Relying on DM (sub)halo host twinning to twin galaxies, rather than directly measuring the galaxy orbital properties, is a more robust process than matching orbital parameters. The latter are sensitive both to internal changes in galaxy properties (in particular stellar mass) and host (sub)halo density profiles, which can differ quite significantly between simulations with and without AGN feedback (see e.g. Peirani et al. (2017)).

7.3.2 Matched fractions

As this analysis is most interested in the most massive objects, the full sample considered here includes all (sub)halos resolved by at least 500 DM particles and all galaxies hosted in these (sub)halos that contain at least 100 star particles. Note that this latter criterion, contrary to the one for DM (sub)halos, does not correspond to a strict stellar mass threshold as star particles can have different masses (integer multiples of the minimal stellar mass). Across all redshifts, H-AGN and H-noAGN contain a comparable, albeit not identical number of both galaxies and (sub)halos (see Table 7.2). For this reason, the following analysis considers the H-AGN sample as the reference to analyse the effectiveness of the twinning algorithm.

Figure 7.2 shows that at high redshift, over 98% of (sub)halos present in H-AGN, corresponding to 15,818 out of 16,042 at redshift $z = 5$, are twinned successfully. The twinned galaxies are evenly distributed across all mass bins. At lower redshift,

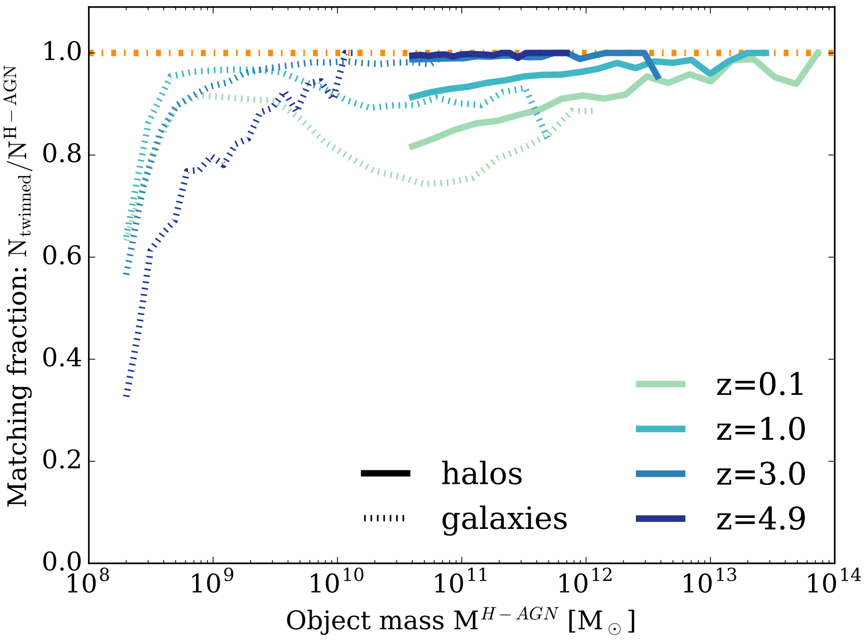

Figure 7.2: Fractions of matched (sub)halos (solid lines) and galaxies (dashed lines) in the H-AGN and H-noAGN simulations. The dashed-dotted line represents the ideal case where all objects in H-AGN are uniquely associated with a corresponding object in H-noAGN. More than 86% of (sub)halos are twinned, but galaxy twinning is less efficient, particularly at stellar masses $M_*^{\mathrm{H-AGN}} < 10^9$ M$_\odot$ for all redshifts but also, albeit less significantly, at redshift $z \leq 1$ for galaxies with higher stellar masses.

the fraction of matched (sub)halos decreases as the more and more different merger histories in the two simulations introduce larger discrepancies between individual objects. This fraction also decreases with (sub)halo mass, as (sub)halos with smaller particle numbers are more sensitive to these merger history changes. However, with 75,544 (sub)halos twinned at redshift $z = 0.1$ out of a sample of 88,171, i.e. an average matched fraction of 86%, the resulting sample remains statistically representative.

The overall rates for matched galaxies are much lower than for (sub)halos, as they require three steps to establish the twin link (galaxy to host (sub)halo, host (sub)halo to host (sub)halo twin, host (sub)halo twin to galaxy twin, see Figure 7.1), with a number of objects dropping out of the sample at each step. Identifying a galaxy with its host (sub)halo can be challenging, especially in dense environments and at high redshift where interactions are more common. For example, increasing the size of the region within which a galaxy is associated with its host (sub)halo, R_{host}, from $R_{\mathrm{host}} = 0.05 \times R_{\mathrm{vir}}$ to $R_{\mathrm{host}} = 0.10 \times R_{\mathrm{vir}}$, increases the fraction of matched galaxies at high redshift (see e.g. Chisari et al., 2017). More specifically, for redshift $z = 3$, the total number of galaxies with at least 100 star particles identified with a host (sub)halo with the same number of DM particles, would rise from 47,656 to 67,301, out of 76,887 galaxies in total, i.e. an increase in matched fraction from 61% to 87%. Selecting amongst these galaxies those hosted in (sub)halos containing more than 500 DM particles (as done in this work), further reduces the numbers to 34,128 (given in Table 7.2) and 48,354 galaxies respectively, out of 76,887. Whilst the number of galaxies excluded by the strict position and mass criteria employed for twinning does represent a significant fraction of the sample, especially at high redshift, relaxing them hardly alters the quantitative results presented in this work. This can be intuitively understood as the mass cuts chosen only eliminate low mass galaxies from the sample. Low mass galaxies are both (i) the most numerous and (ii) virtually unaffected by AGN feedback, at any redshift. This can be seen in Figure 7.3, where the entire population (i.e. all 76,887 objects for H-AGN at $z = 3$), as opposed to only twinned galaxies, is plotted. Similarly, the use of a stricter position criterion mostly affects low mass galaxies, as these have longer dynamical friction times and are more easily dislodged from the centre of their host (sub)halos during gravitational interactions. In short, the galaxy sample as defined in this section is statistically robust enough, at all redshifts and galaxy masses, to draw conclusions about the impact of AGN feedback in the H-AGN simulation, while getting rid of virtually all mismatch errors in the galaxy twinning process.

7.3.3 The effect of AGN feedback on halos

Figure 7.3 shows that no matter the redshift, the DM (sub)halo mass functions (HMF) for H-AGN and H-noAGN are so similar that they are indistinguishable on the plot. Directly comparing the DM masses of (sub)halo twins (See Figure 7.4) shows that (sub)halos with masses below $M_{\rm vir}^{\rm H-AGN} < 10^{11}$ $M_\odot$ have identical masses in both simulations, at all redshifts. The small spread in masses is mainly caused by variations in shape, which the structure finding algorithm translates into a small variation in virial mass. At redshifts of $z = 1$ and below, (sub)halos with masses above $M_{\rm vir}^{\rm H-AGN} >$ 10^{11} $M_\odot$ can have a dark matter mass up to 5% percent lower in H-AGN. This is due to the fact that in the presence of AGN feedback, the baryon content of these massive (sub)halos is strongly reduced (see Section 7.5), which translates into a reduced total (sub)halo mass, as the reduced gravitational pull slows down the cosmic inflow rate. As the (sub)halos in H-AGN systematically exhibit lower masses, it makes sense to require that 75% of DM particles from H-AGN be present in the H-noAGN (sub)halo

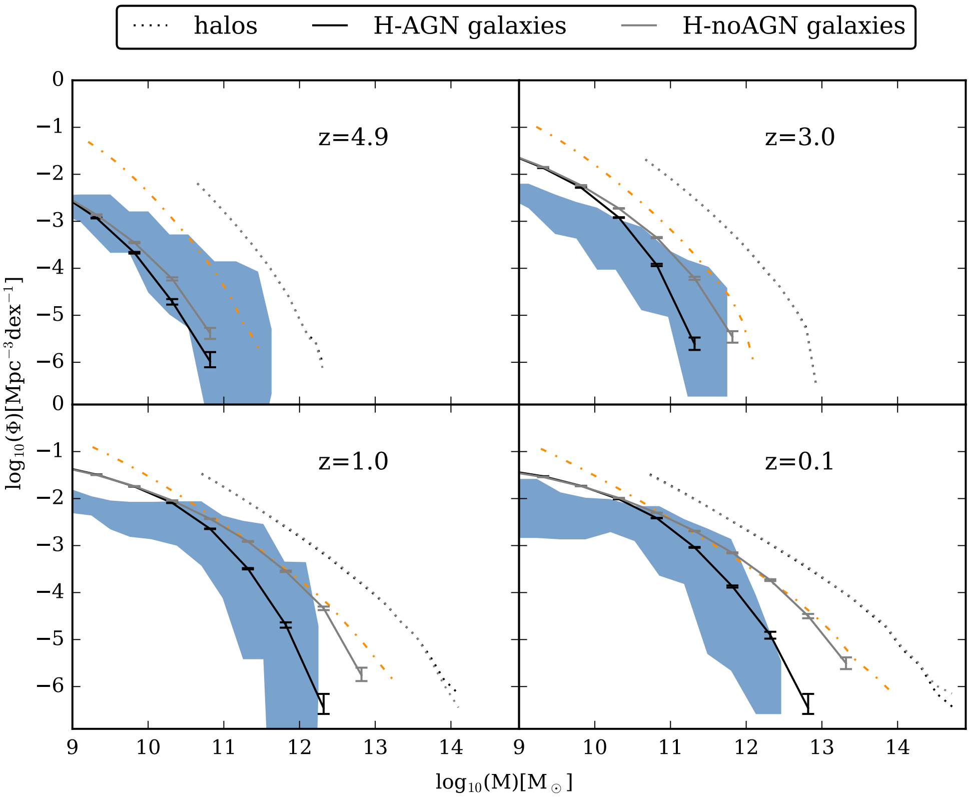

Figure 7.3: DM HMF (dotted) and corresponding GSMF (solid) for H-AGN (black) and H-noAGN (light grey). Combined observations from González et al. (2011); Bielby et al. (2014); Daddi et al. (2002); Moustakas et al. (2013); Tomczak et al. (2014); Bernardi et al. (2013); Song et al. (2016); Davidzon et al. (2017) are plotted as the shaded blue area. The expected GSMF if all the baryons within (sub)halos (using a fraction of 16.5% of the total mass) are converted into stars is shown as the dot-dashed line in yellow. A detailed comparison of individual observational datasets to the GSMF in H-AGN can be found in Kaviraj et al. (2017). Errorbars on the simulated mass functions are Poisson, with those on the HMF not plotted as they are even smaller that the GSMF ones in the mass range where they overlap. Both HMF for H-AGN and H-noAGN are plotted, but are so similar as to be indistinguishable. Note that this figure shows mass functions for the *full* sample of objects in H-AGN and H-noAGN, as opposed to the sample of twinned objects (see text for definition). At redshifts $z \leq 3$, the GSMFs for H-AGN lie within the region supported by observations for galaxies above stellar masses $M_* > 5 \times 10^{10}$ M$_\odot$. H-AGN and H-noAGN overproduce the number of smaller galaxies by up to a factor of 3, partially due to a lack of magnitude cut applied to the GSMF from simulations at low stellar masses but predominantly because of the low efficiency of the stellar feedback model implemented (see text for detail).

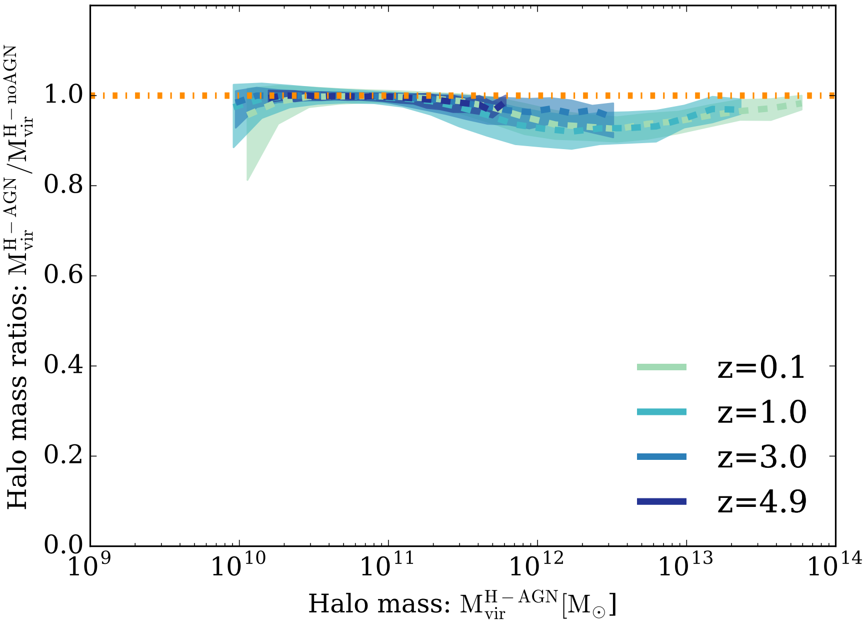

Figure 7.4: Twinned (sub)halo DM mass ratios in H-AGN and H-noAGN at different redshifts, with quartile error regions around the median (shaded region). Note that the median twin (sub)halo with $M_{\rm vir}^{\rm H-AGN} < 10^{11}$ $\rm M_\odot$ has an identical DM mass with and without AGN feedback, whatever the redshift. Median intermediate mass (sub)halos 2×10^{11} $\rm M_\odot < M_{vir}^{H-AGN} < 2 \times 10^{13}$ $\rm M_\odot$ see their virial mass reduced by up to 5% depending on mass and redshift.

and not the reverse. A more detailed analysis as to how AGN feedback affects the inner structure of DM (sub)halos is carried out in Peirani et al. (2017).

7.4 AGN feedback and stellar mass

7.4.1 The galaxy stellar mass function

Comparing the galaxy population in H-AGN and H-noAGN at the various redshifts presented in this chapter shows that AGN feedback is instrumental in bringing the high mass end of the GSMF in agreement with observations[1]. As Figure 7.3 demonstrates, AGN feedback is able to suppress star formation in galaxies with masses $M_* \geq 5 \times 10^{10}$ $\rm M_\odot$ at $z = 3$, allowing the simulation to match the number of galaxies at and above the knee of the GSMF. It is important to note that H-AGN was *not* tuned to reproduce this result. As previously mentioned, the only tuning done on

---

[1] A more detailed comparison of the GSMF in H-AGN to individual observational datasets can be found in Kaviraj et al. (2017).

global galaxy properties in the simulations involves the radiative efficiency of the AGN feedback modes, which were set to reproduce the local $M_{\mathrm{SMBH}} - M_*$ and $M_{\mathrm{SMBH}} - \sigma_*$ relations.

In the absence of AGN feedback, the GSMF in H-noAGN agrees well with predictions that the uniform baryon mass fraction of $\Omega_b/\Omega_m = 0.165$ (dotted-dashed lines, Komatsu et al., 2011) is entirely converted into stars in the galaxy mass range $5 \times 10^{10}\ \mathrm{M}_\odot < \mathrm{M}_* < 10^{12}\ \mathrm{M}_\odot$ by $z = 1$. For galaxies with masses $M_* \geq 5 \times 10^{12}\ \mathrm{M}_\odot$, a discrepancy between the GSMF in the absence of feedback (solid grey line) and expectation values from the cold dark matter model (dotted line) starts to appear because gas cooling times in host (sub)halos harbouring such massive galaxies become comparable to the Hubble time when the halos assemble. Therefore, not all the baryons enclosed have yet been able to cool and form stars. Cooling is further hampered by the fact that in the absence of AGN feedback, heavy elements do not get distributed effectively throughout the halo but remain close to the central galaxy.

The simulations systematically overproduce the number of galaxies with masses below $M_* \leq 5 \times 10^{10} M_\odot$ at $z \leq 3$. This is partly caused by the fact that observed mass functions are derived from magnitude-limited data, whereas the GSMF presented here for H-AGN and H-noAGN are raw stellar masses extracted from the simulation, with no completeness, surface brightness or luminosity cut applied. Indeed, comparing the GSMFs presented in Figure 7.3 to those plotted in Figure 7 of Kaviraj et al. (2017), which are based on the *same* simulation, H-AGN, but include a magnitude cut to match observations, one realises that the effective number of galaxies with masses $M_* = 10^9 M_\odot$ is reduced by about 0.1-0.2 dex depending on redshift. Galaxies with masses $M_* \geq 5 \times 10^{10} M_\odot$ are completely unaffected. This does somewhat flatten the simulated GSMFs at the faint end, bringing them in better agreement with the data. However, the remaining discrepancy of about 0.3 dex with the data for galaxies with stellar masses $M_* \leq 10^{10} M_\odot$ can probably be attributed to the implementation of an insufficiently energetic stellar feedback model, coupled with numerical resolution effects (Kaviraj et al., 2017)[2], although how efficient such a feedback can realistically be is still a matter of debate. Having said that, as AGN feedback, through the comparison of H-AGN and H-noAGN, is measured to have no effect on the low mass end of the GSMFs (as clearly visible in Figure 7.3) and the limitations discussed above are present in both simulations, they likely have a very limited impact on the work presented in this chapter. Still, the few absolute measurements, such as

---

[2] Note that a stronger stellar feedback is likely to affect black hole masses as well (Dubois et al., 2015; Habouzit et al., 2017).

outflow rates, have to be examined bearing in mind that stellar feedback is probably underestimated in the simulations and thus that values derived for galaxies with stellar masses $M_* \leq 5 \times 10^{10}$ $M_\odot$ are very likely too low.

Finally, both simulations systematically underproduce massive ($M_* \geq 5 \times 10^{10}$ $M_\odot$) galaxies at redshift $z \geq 5$ (Figure 7.3 top left panel). Given that the HMF multiplied by the universal baryon fraction (dot-dashed curve on the figure) seems to describe the observed data fairly well for galaxies in this mass range, this suggests that an inability to resolve the progenitors of halos early enough leads to star formation being artificially postponed. Obviously, since the gas content of these halos is still correctly estimated, galaxies will eventually catch up: their star formation rate will be slightly higher than expected, as long as more gas is present. However, at high redshift, galaxy star formation timescales cannot be considered small in comparison to the time elapsed since their host halo formed, so their stellar masses can be significantly underestimated. Note that this resolution effect has completely vanished, at least for massive galaxies, by $z = 3$. Moreover, as this issue affects H-AGN and H-noAGN in the same way, it cancels out in the comparative analysis of the two simulations performed in this chapter.

7.4.2 Quenching

Instead of having to rely on statistical averages, such as those presented in the mass functions in Figure 7.3, the twinning procedure described in Section 7.3 allows for a direct comparison of the stellar masses of each individual galaxy, with and without AGN feedback. Figure 7.5 shows the results of this comparison for a range of redshifts. As expected from the local GSMF in Figure 7.3, the most massive galaxies are the most strongly quenched at all redshifts (Figure 7.5). However the amount of quenching does not vary linearly with galaxy mass, with the function tailing off for both strongly quenched large galaxies, and barely affecting small galaxies.

There is also a redshift dependence in the maximum amount of quenching observed, ranging from 40% for galaxies with stellar masses in H-AGN of $M_*^{\mathrm{H-AGN}} > 10^{10}$ $M_\odot$ at redshift $z = 5$, up to a maximum of over 80% for the largest galaxies at redshift $z = 0.1$, suggesting that galaxy quenching is a continuous process, active throughout the merging history of galaxies. In general, the shape of the distribution is driven by small galaxies that show little influence of AGN feedback at any redshift, and a tailing off for massive galaxies. Large galaxies appear to converge to a constant quenching mass ratio $M_*^{\mathrm{H-AGN}}/M_*^{\mathrm{H-noAGN}} = 0.2$, as they grow from $M_* = 10^{11}$ $M_\odot$

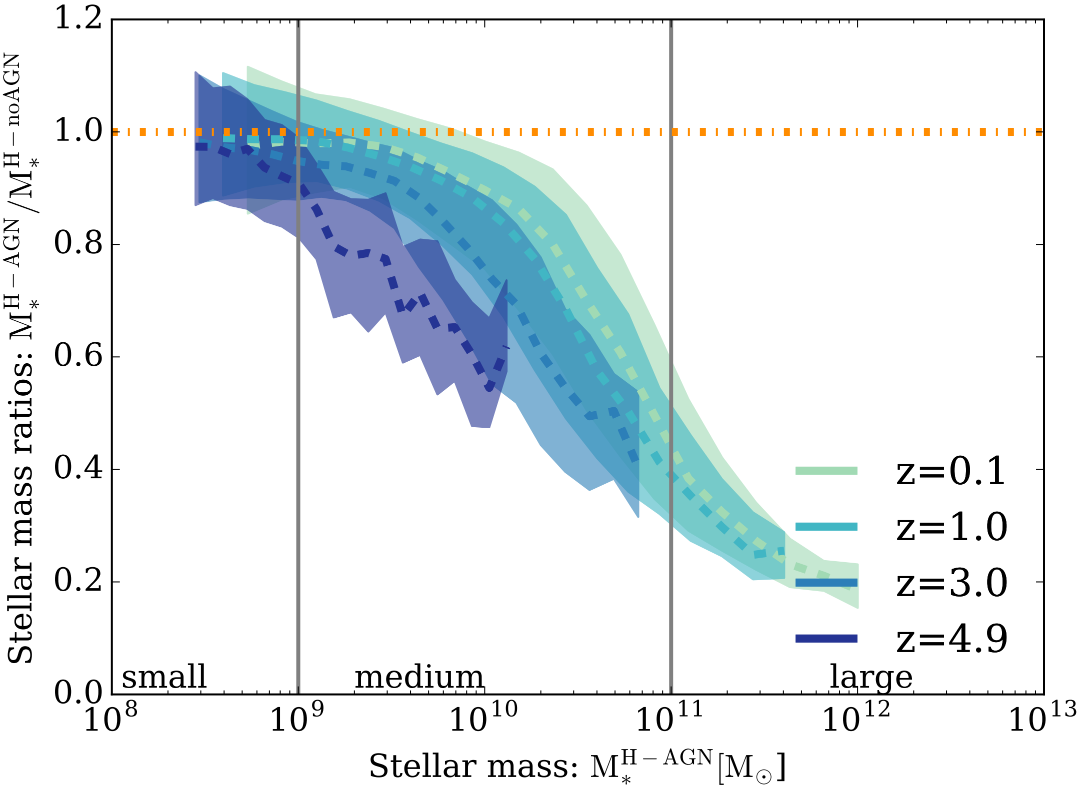

Figure 7.5: Quenching mass ratios, i.e. stellar mass ratios of twinned galaxies in H-AGN and H-noAGN. The twins are mass binned by stellar mass in H-AGN, $M_*^{\mathrm{H-AGN}}$, with the median value (dashed lines) and quartile error ranges (shaded region) plotted for each bin. If AGN feedback had no effect, the quenching mass ratio would lie on the horizontal dotted-dashed line, which denotes twin galaxies having identical mass. Note the non-linear evolution of the quenching mass ratio with galaxy mass, with small galaxies ($M_*^{\mathrm{H-AGN}} < 10^9$ M$_\odot$) barely affected, and large galaxies ($M_*^{\mathrm{H-AGN}} > 10^{11}$ M$_\odot$) being most strongly quenched. This plot also shows a redshift evolution, with smaller galaxies more affected at higher redshift.

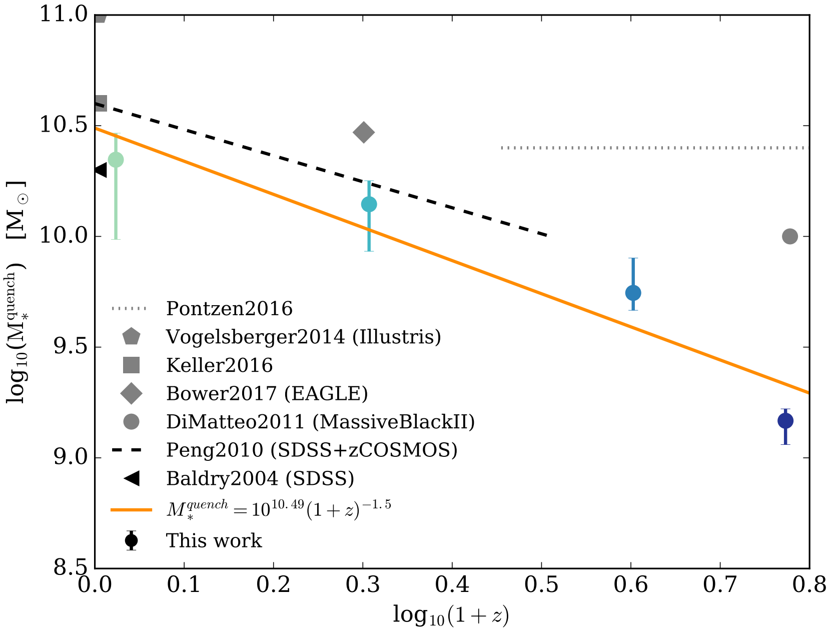

Figure 7.6: Transition galaxy stellar mass at which AGN feedback becomes important, measured where the quenching mass ratio equals 0.85 ± 0.05, as well as observational data from Peng et al. (2010); Baldry et al. (2006) (black), and results from from a range of simulations Bower et al. (2017); Di Matteo et al. (2012); Keller et al. (2015); Pontzen et al. (2017) (grey). The values for Di Matteo et al. (2012) and Pontzen et al. (2017) are converted from the halo mass quoted in the paper, to stellar mass plotted here, using the $M_{\mathrm{vir}} - M_*$ relation derived by Moster et al. (2013).

to $M_* = 10^{12}$ $\mathrm{M}_\odot$ both in H-AGN and H-noAGN. This is not due to any fundamental change in the impact of AGN feedback for galaxies with stellar masses above $M_* > 10^{11}$ $\mathrm{M}_\odot$ but rather reflects the fact that, even in the absence of AGN feedback, the GSMF in H-noAGN steepens due to long cooling times for massive objects (see Figure 7.3 and Section 7.4.1) which lead to reduced star formation rates. Therefore, the constant quenching mass ratio for large galaxies is not driven by less effective AGN feedback but rather by less effective cooling for galaxies in the absence of feedback. Note that this is not a selection effect either, as no mass ratio cut was applied to the galaxy sample.

Particularly noticeable for redshifts above $z > 3$, the minimum mass to experience quenching decreases with redshift: a typical 10^9 $\mathrm{M}_\odot$ galaxy at redshift $z = 5$ already has its stellar mass quenched by 10%, whereas a 10^9 $\mathrm{M}_\odot$ galaxy at redshift $z = 1$ shows a median reduction in stellar mass of less than 1%.

A transition mass between star forming and quenched galaxies is somewhat difficult to define in this context, as the quenching mass ratio is a cumulative measure, not an instantaneous one such as the star formation rate or the stellar mass growth timescale often used to separate star forming and quiescent populations. However, based on Figure 7.5, it seems natural to define the mass at which quenching due to AGN feedback becomes important to be the point where the quenching mass ratio equals 0.85 ± 0.05, as it corresponds to the location of the sharp break in the quenching mass ratio versus galaxy stellar mass relation. As can be seen in Figure 7.6, this quantity shows a redshift evolution, decreasing from $\log(M_*^{quench}/[\mathrm{M}_\odot]) = 10.35^{0.12}_{0.36}$ at $z = 0.1$ to $\log(M_*^{quench}/[\mathrm{M}_\odot]) = 9.17^{+0.05}_{-0.11}$ at $z = 5$. This result is in good agreement with the value of 10.3 found for a sample of Sloan Digital Sky Survey (SDSS) galaxies by Baldry et al. (2004). At higher redshift, $z \approx 1$, there is good agreement with the transition mass quoted by Peng et al. (2010), who used a large sample of SDSS and zCOSMOS galaxies to study the stellar mass at which "mass quenching", which includes the impact of AGN, becomes important. Note that the redshift evolution reported also suggests that the trend they observe can be extrapolated to $z = 2$ at least. However, mostly driven by results above $z > 4$, the best fit power law for the redshift evolution found here,

$$M_*^{quench}(z) = 10^{10.49}(1+z)^{-1.50}, \tag{7.2}$$

has a steeper slope than that reported in Peng et al. (2010). Comparing to other simulations (Bower et al., 2017; Di Matteo et al., 2012; Keller et al., 2015; Pontzen et al., 2017), the value reported here is consistently lower, which is partially due to how the transition mass is defined in each work. The value in Pontzen et al. (2017), for example, is defined to be the value at which only AGN feedback quenches the galaxy. It is not surprising that this is higher than the definition of M_*^{quench} used here, which is the transition mass at which AGN feedback begins to be dominant, but stellar feedback might still play a role. The consistently low transition masses are evidence that quenching is a long term process, whose effect accumulates over the evolution history of a galaxy. A galaxy whose star formation rate is reduced by 20% would not sufficiently change colour to be counted among the "red and dead" population but could over time show a noticeably reduced quenching mass ratio. The transition mass based on the cumulative star formation history is therefore necessarily lower than that based on a more instantaneous measure, such as the stellar mass growth timescale (Bower et al., 2017) or the star formation rate (Pontzen et al., 2017).

As the evolution in the quenching mass ratio (and the minimum mass of a galaxy affected) reported here is purely driven by AGN feedback, one would expect the

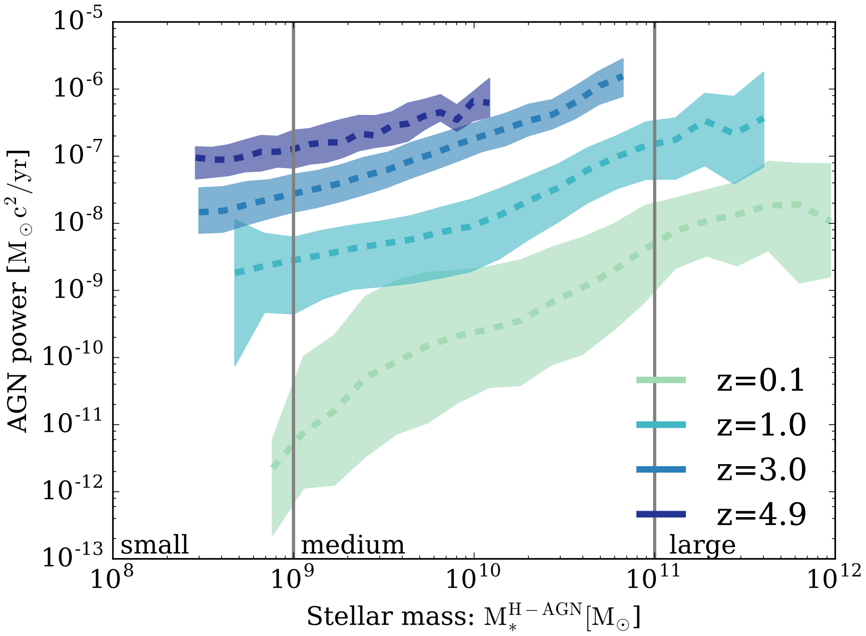

Figure 7.7: Median AGN power $\dot{E}_{\rm AGN}$ (see Equation 3.26) for each galaxy stellar mass bin in H-AGN, with the shaded region showing the samples quartile error ranges. At higher redshift, a galaxy of the same mass is subject to stronger feedback.

impact of AGN feedback to evolve with time. The evolution of the instantaneous AGN power, plotted in Figure 7.7 confirms this conjecture. Galaxies of the same stellar mass are subject to different amounts of feedback at different points in cosmic history. The median AGN power for a galaxy of a given stellar mass decreases strongly with redshift, with a galaxy with $M_*^{\rm H-AGN} = 10^9$ $M_\odot$ at $z = 5$ subject to a feedback three orders of magnitude stronger than a galaxy of equivalent stellar mass at $z = 0$.

As the sound speed of the gas, and its relative velocity with respect to the SMBH, do not vary systematically with redshift, the feedback power of an AGN, calculated according to equation 3.26 is mainly a function of local gas density in the vicinity of the black hole, the Eddington ratio, which determines the radiative efficiency of each mode of feedback, and black hole mass squared[3]. Therefore, the decreasing importance of AGN feedback in the evolution of small galaxies could be due to (i) the decreasing gas fractions associated with galaxy evolution, (ii) a shift in AGN feedback mode from quasar (high redshift) to radio dominated (low redshift), or (iii) an evolving black hole population (different $M_{\rm SMBH}$ vs M_* relation). Each of these

---

[3] The case of black holes accreting at the Eddington limit is rare and very short lived in the simulations, see e.g. Dubois et al. (2012a) or Section 7.4.3 below.

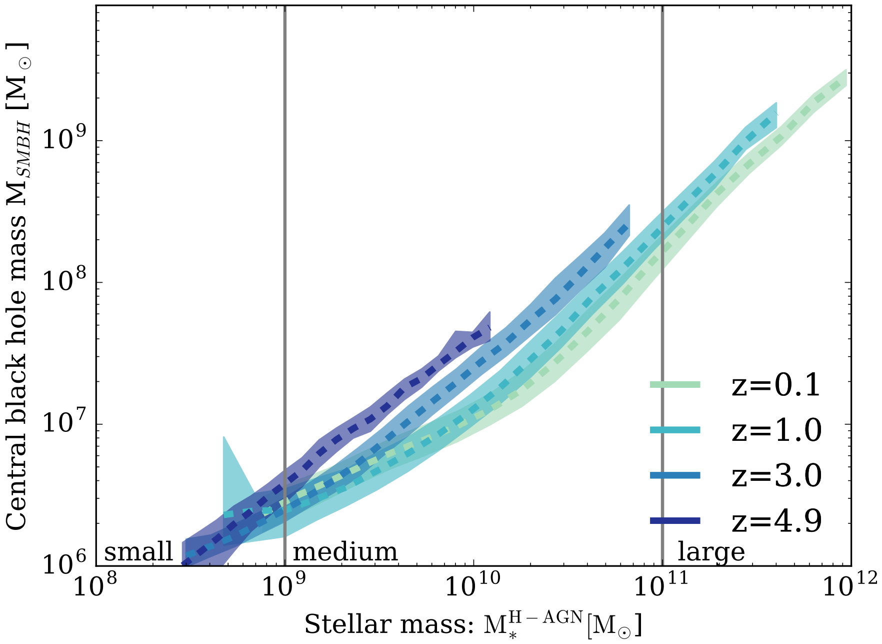

Figure 7.8: Median black hole mass (dashed lines) evolution as a function of galaxy stellar mass in H-AGN, as well as quartile ranges for the black hole mass distribution (shaded region). The plot shows a clear redshift evolution, with smaller galaxies hosting larger black holes at higher redshift[5]. Note that small galaxies with stellar masses below $M_*^{\mathrm{H-AGN}} < 10^9$ $\mathrm{M}_\odot$ are to be treated with some caution, as they are quite sensitive to resolution effects.

three options is examined in turn in the following section.

7.4.3 The coevolution of SMBHs and their hosts

The total AGN power is a function of the accretion rate onto the central SMBH, which in turn, in the Bondi regime used here, depends on the SMBH mass squared[4]. Figure 7.8 shows that the median mass of the central black hole undergoes a redshift evolution between $z = 0$ and $z = 5$, with galaxies of a given stellar mass hosting a more massive black hole at higher redshift[5]. For example, a galaxy with $M_*^{\mathrm{H-AGN}} = 10^{10}$ $\mathrm{M}_\odot$ at

---

[4] Or only of mass when accreting at the Eddington limit, but these episodes are rare and short lived in simulations with AGN feedback as shown later in this section; see Equation 3.26, using Equation 3.18 or Equation 2.5

[5] The results presented here are consistent with Figure 10 of Volonteri et al. (2016), which reports no redshift evolution in the $M_* - M_{\mathrm{SMBH}}$ relation. This is because the redshift evolution is mainly driven by low mass black holes at low redshift, a sample excluded by these authors who apply a cut in host halo mass of $M_{\mathrm{halo}} = 8 \times 10^{10}$ $\mathrm{M}_\odot$. By comparison, the sample analysed here includes all black holes identified within a host galaxy in halos with $M_{\mathrm{halo}} > 4 \times 10^{10}$ $\mathrm{M}_\odot$. The second difference between these two pieces of work concerns the statistical analysis chosen: while Volonteri

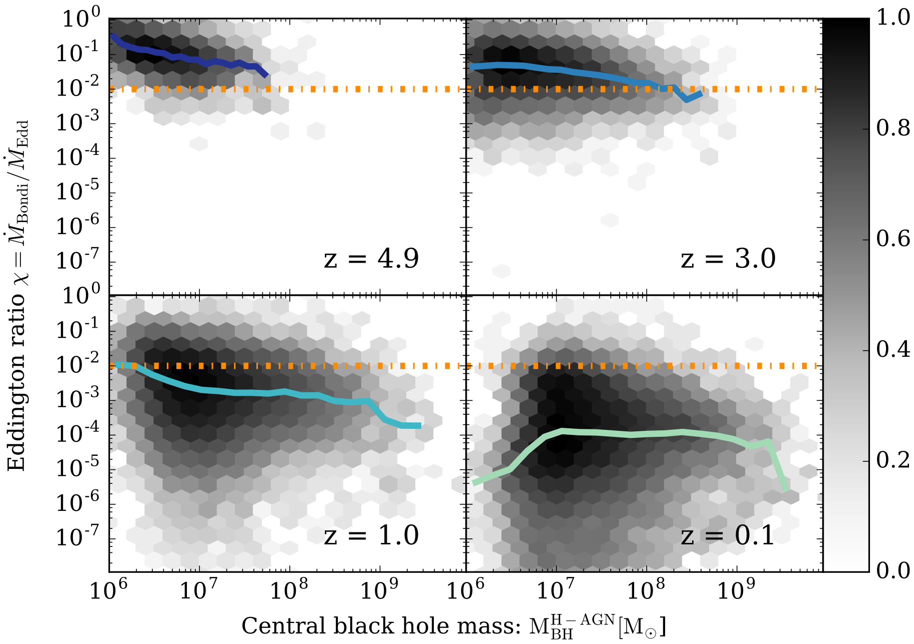

Figure 7.9: Evolution of AGN feedback mode for the whole sample of SMBHs at different redshifts, with objects above the dotted line in quasar mode and objects below in radio mode. The solid lines denote the median Eddington ratio. There is a clear redshift evolution of the feedback mode, with 95.7% of SMBHs in quasar mode at $z = 5$, in comparison to only 2.3 % at $z = 0.1$. The colourbar in each panel is normalised to the bin with the maximum number of objects at the given redshift.

redshift $z = 0.1$ typically hosts a SMBH with $M_{\rm SMBH} = 1.1\times10^{7}$ $\rm M_\odot$, whereas a galaxy with the same stellar mass at $z = 5$ hosts a SMBH with a median mass of $M_{\rm SMBH} = 4.2\times10^{7}$ $\rm M_\odot$. A similar evolutionary trend is reported observationally (Merloni et al., 2010; Decarli et al., 2010) and in other large scale cosmological simulations (Khandai et al., 2015; Sijacki et al., 2015). A detailed discussion of the shape of the $M_{\rm SMBH} - M_*$ relation can be found in Volonteri et al. (2016).

Looking at the evolution of AGN feedback mode with redshift, as determined by the Eddington ratio χ, Figure 7.9 reveals that the SMBH population transitions from quasar mode to radio mode between redshifts $z = 3$ and $z = 1$, as χ falls below 0.01. At high redshift, the vast majority of AGNs are found in quasar mode, namely 95.7% of the sample at $z = 5$ and 85.5% at $z = 3$. At lower redshift, the population is overwhelmingly in radio mode across all mass bins, with only 19.1% and 2.3% found in quasar mode at $z = 1$ and $z = 0.1$ respectively.

---

et al. (2016) employ a linear fit, as used in observational studies, here a median black hole mass is presented, which allows for a non-linear correlation between black hole and galaxy stellar masses.

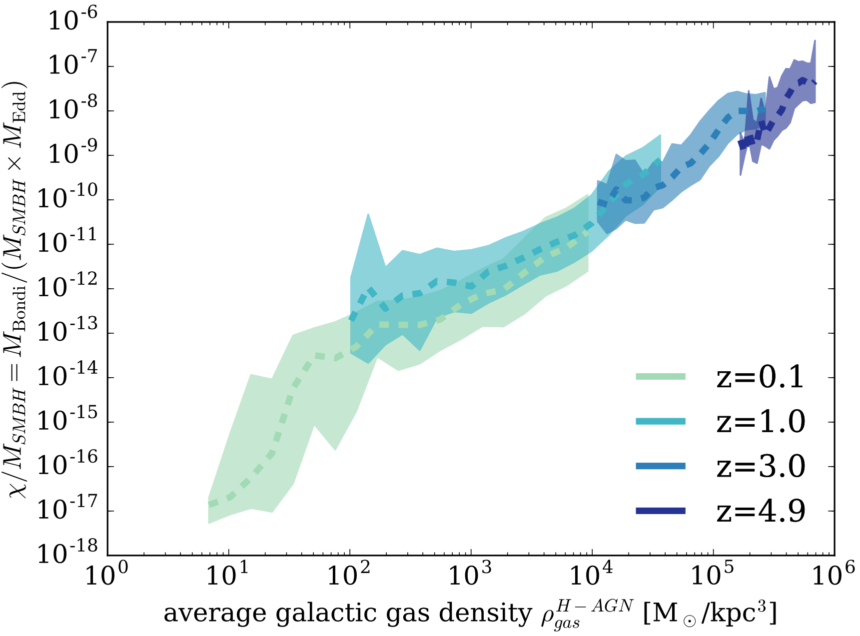

Figure 7.10: Specific accretion rate onto the central SMBH, in Eddington units, vs average gas density of the host galaxy for various redshifts. The measured decrease in average specific accretion rate for SMBHs of any mass, as redshift decreases, is driven by a decreasing gas density in the host galaxy.

As the Eddington ratio is a measure of how efficiently a black hole of a given mass is accreting, the high Eddington ratios at redshift $z > 3$ explain the evolution in the $M_{\rm SMBH} - M_*$ relation in Figure 7.8. Indeed, whilst a black hole with $M_{\rm SMBH} = 10^7$ M$_\odot$ at redshift $z = 5$ accretes with a mean Eddington ratio of $\chi = 6.27 \times 10^{-2}$, a SMBH with the same mass at redshift $z = 0.1$ accretes at only $\chi = 1.2 \times 10^{-4}$ Eddington. This means that the latter grows about two orders of magnitude more slowly than the former.

Figure 7.9 also clearly shows that except for a few outlying objects, the bulk of the population is not accreting in an Eddington limited fashion at $z \leq 5$. This evolution in the median Eddington ratio reflects an underlying evolution in the gas density of galaxies, as can be seen in Figure 7.10. χ is divided by the SMBH mass to calculate the specific accretion rate of these SMBHs in Eddington units, i.e. the dependence on black hole mass has been removed to be able to compare black holes across the whole mass range. The story which emerges from this plot is that since high redshift galaxies are more gas rich, they fuel their central black holes more efficiently, regardless of their masses. As the gas supply is depleted, accretion onto the central

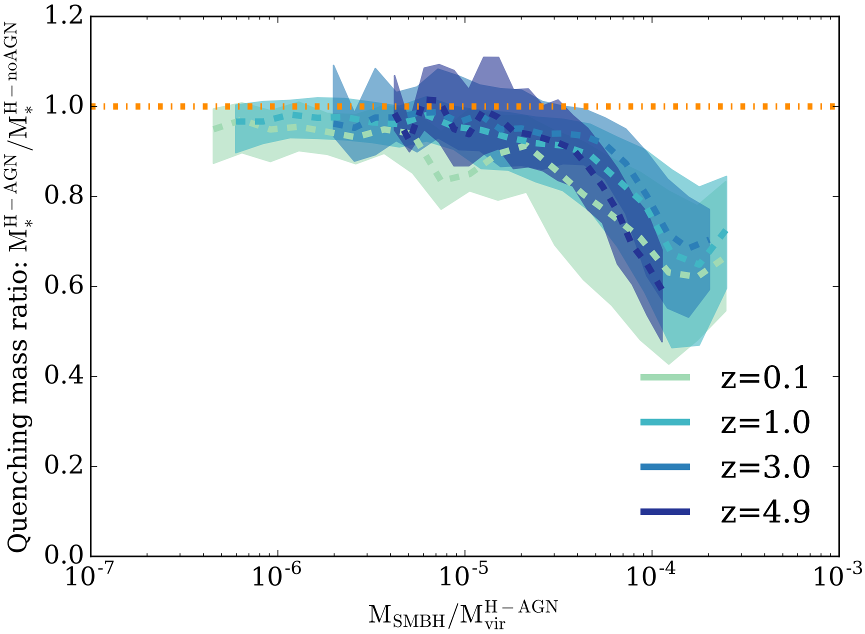

Figure 7.11: Stellar mass ratio between H-AGN and H-noAGN galaxy twins against ratio of central SMBH mass to virial mass of the host (sub)halo, instead of stellar mass as in Figure 7.5. The redshift evolution is erased in that case because the feedback energy deposited is directly connected to the central black hole mass (see Section 7.4.3 for detail).

black hole slows down and AGN feedback transitions from a quasar to a radiative feedback mode around $z = 2$. Due to the different radiative efficiencies employed for the two feedback modes, $\epsilon_f = 0.15$ for the quasar mode at $\chi \geq 0.01$ and $\epsilon_f = 1.0$ for the radio mode at $\chi < 0.01$ (see Section 7.2.5), a larger percentage of accretion energy is converted into feedback at redshifts below $z < 2$, but nevertheless the amount of energy available for feedback declines.

Summarising the impact of all three effects, the decreasing AGN power for a galaxy of a given stellar mass is driven by the cumulative effect of a proportionally larger central SMBH and the decreasing gas supply in the galaxy, for which the increasing efficiency of the feedback mode at $z < 2$ is unable to compensate. Not only do existing black holes of a given mass accrete less efficiently in the gas poor galaxies at $z < 2$, such that $\dot{M}_{SMBH}(M_{SMBH}, z > 2) > \dot{M}_{SMBH}(M_{SMBH}, z < 2)$ for all M_{SMBH}, but any galaxy of a given stellar mass also hosts a significantly smaller black hole than its equivalent counterpart at $z > 2$, ie $M_{SMBH}/M_*(z < 2) < M_{SMBH}/M_*(z > 2)$. The two effects combine together to produce the redshift evolution of AGN power seen in Figure 7.7, despite the shift to a more efficient feedback mode around $z = 2$.

How efficiently an AGN of a given power is able to remove gas from a galaxy is dependent on the depth of the gravitational potential it has to overcome in the process. Repeating the analysis of the quenching mass ratios, but plotting it against $M_{\mathrm{SMBH}}/M_{\mathrm{vir}}^{\mathrm{H-AGN}}$ instead of the galaxy stellar mass, shows that the redshift evolution in the quenching mass ratio is truly driven by the evolution of the SMBH population and its feedback power (Figure 7.11, in comparison to Figure 7.5). Indeed, whilst the relation scatter increases, the different redshift curves now overlap within the quartile error bars: the redshift dependence in Figure 7.5 has been erased. In other words, independently of redshift, black holes with a black hole to halo mass ratio of less than $M_{\mathrm{SMBH}}/M_{\mathrm{vir}}^{\mathrm{H-AGN}} \leq 4 \times 10^{-5}$ quench their host galaxy by less than 20%, whereas black holes with ratios only a factor 2-3 larger than that suppress their host galaxy stellar masses by up to 50%.

Although there exists a clear transition of the sample from one AGN feedback mode to the other, no significant difference was found when analysing the quenching mass ratio (such as in Figure 7.11) by splitting the sample into quasar or radio mode galaxies. All results shown here can be reproduced by assuming that the entire sample can be found in quasar mode at redshifts $z > 2$, and in radio mode otherwise. The only notable discrepancy between the simulation data and this simplified model is that the scatter is somewhat reduced, which is expected as objects found in the opposite feedback mode to the majority of the population are statistical outliers. However, this analysis is based on an instantaneous measure of feedback mode at a specific redshift and does not capture the accretion history of a particular object.

7.5 AGN feedback and gas flows

Figure 7.8 shows that SMBHs make up much less than 1% of the mass of their host galaxy, so the mass of baryons accreted by the SMBHs is negligible compared to the reduction in stellar mass caused by quenching. The effect of AGN feedback on the cold gas supply of the galaxy must therefore be profound to suppress star formation by up to an order of magnitude over the evolution of the galaxy. There are three possible channels through which AGN feedback can affect the gas content of the galaxy: (i) it can drive powerful outflows, emptying the reservoir of gas available in the ISM of the galaxy; (ii) it can prevent cosmic inflows from replenishing the gas supply in the galaxy or (iii) it can heat existing gas of the ISM and circumgalactic medium (CGM) to prevent cooling flows and the associated star formation. In this section, the relative importance of these three feedback channels is investigated.

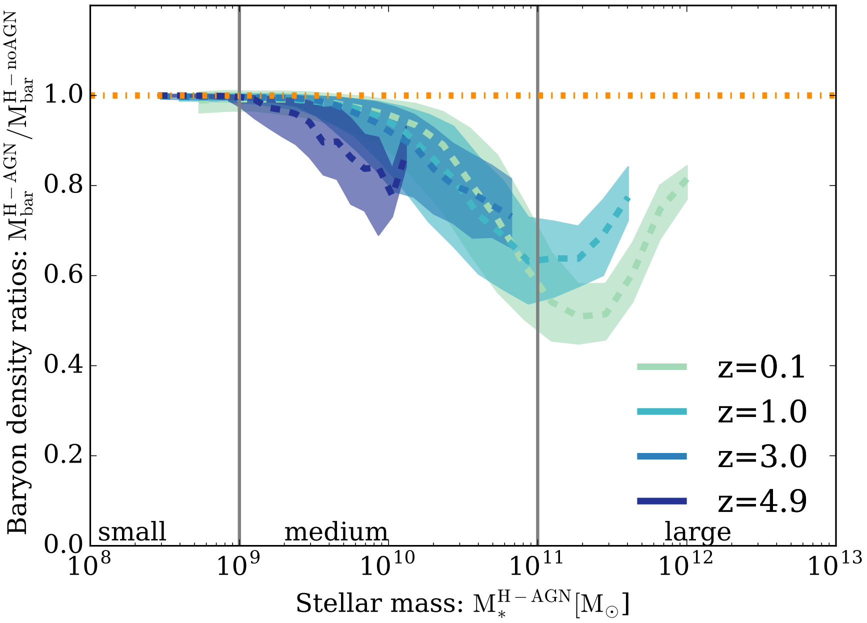

Figure 7.12: Ratio between the average baryon density $\overline{\rho}_{\rm bar}^{\rm H-AGN}/\overline{\rho}_{\rm bar}^{\rm H-noAGN}$ within the virial radius of the galaxy host halo twins: $\overline{\rho}_{\rm bar} = (M_{*} + M_{\rm SMBH} + M_{\rm gas})/(\frac{4\pi}{3}R_{\rm vir}^{3})$. The dotted line represents identical baryon mass in both host halos and the shaded regions show the quartile ranges of the sample. AGN feedback partially acts on star formation by reducing the total baryon mass in the halo, particularly for galaxies with stellar masses around $M_{*}^{\rm H-AGN} = 10^{11}$ $\rm M_{\odot}$ in H-AGN.

7.5.1 The evolution of the baryon content

Should AGN feedback primarily suppress star formation through heating the existing gas in the halo, one would expect twinned halos to have the same total baryon mass, with that in H-AGN showing a much higher gas fraction as less gas is being turned into stars. Figure 7.12 shows that AGN feedback directly lowers the baryon (gas + stars + black hole) content embedded within the virial radius of DM halos: the average baryon density ratio versus galaxy stellar mass relation follows a shape reminiscent of that of the quenching mass ratio previously discussed. Unfortunately all three major channels, through which AGN feedback is expected to affect star formation, lower the baryon density of the galaxy. Boosted outflows drive existing gas out of the galaxy, slowed inflows prevent accretion in the first place, and heating causes the gas to expand, lowering the average density. The average baryon density, as opposed to the total mass within the halo, is compared here to correct for the small differences in halo mass shown in Figure 7.4, which translate into a difference in virial radius and therefore a difference in the volume over which the gas mass in the halo is measured.

Two features stand out in comparison to the quenching mass ratio. First, the average baryon density within the halo of small galaxies is unaffected by AGN feedback at all redshifts. These galaxies do however show a reduced stellar mass, particularly at redshift $z = 5$, where their stellar mass shows a median reduction of 10%, as seen in Figure 7.5. This suggests that feedback affects the star formation efficiency of these galaxies more than it alters their gas supply. Efficiency is reduced either by locally heating the gas, redistributing the gas within the halo, or by destroying dense, star-forming clumps, but not by driving outflows or preventing gas inflows through the halo virial sphere. Secondly, for large galaxies with stellar masses above $M_* > 10^{11}$ M$_\odot$, the baryon content in both simulations becomes increasingly comparable again with increasing mass, despite the fact that these galaxies see a reduction in their stellar mass of around 80% for redshifts between $z = 1$ and $z = 0.1$. In this case, the deepening gravitational potential of the halo makes it difficult for the AGN to affect gas flows at the halo virial radius. However a much smaller fraction of the existing gas is converted into stars, because of effects (gas heating or redistribution) similar to those which plague small galaxies. These translate into the significantly reduced galaxy masses presented earlier in Figure 7.5. For medium galaxies at all redshifts, AGN feedback acts through depleting the gas reservoir at the halo scale, which reduces the supply of gas available for star formation. This means AGN feedback also directly influences the inflows and/or outflows of the galaxy.

7.5.2 The effect on inflows and outflows

There are two ways to decrease the total baryon mass of a galaxy: by reducing inflows or by boosting outflows. In this work, all flows are measured at two different radii: halo scales, also called R95, which correspond to a radius of $R95 = 0.95 \times R_{\mathrm{vir}}$, and galaxy scales, also called R20, which correspond to a radius of $R20 = 0.2 \times R_{\mathrm{vir}}$. Flows are measured through spherical surfaces located at these radii, centred on the halo. Flow masses are calculated for all cells within a narrow shell, centred on the radius in question, where $\dot{M}_{\mathrm{gas}} = \sum_i \rho \Delta x_i^3 \bar{v}_i \cdot \bar{r}_i / \omega$, where ρ is the gas density, Δx is the cell size, $\bar{v}_i$ is the gas velocity, $\bar{r}_i$ is the unit vector of the cell centre relative to the halo centre and $\omega = 2kpc$ is the width of the shell. M_{outflow} includes all cells with $\bar{v}_i \cdot \bar{r}_i > 0$ and M_{inflow} includes all cells for which $\bar{v}_i \cdot \bar{r}_i \leq 0$.

A first comparative look at the flow patterns for galaxies in H-AGN and H-noAGN (see Figure 7.13) suggests that AGN feedback drives outflows in medium and large galaxies, particularly at halo scales. There are also some large scale pseudo flows

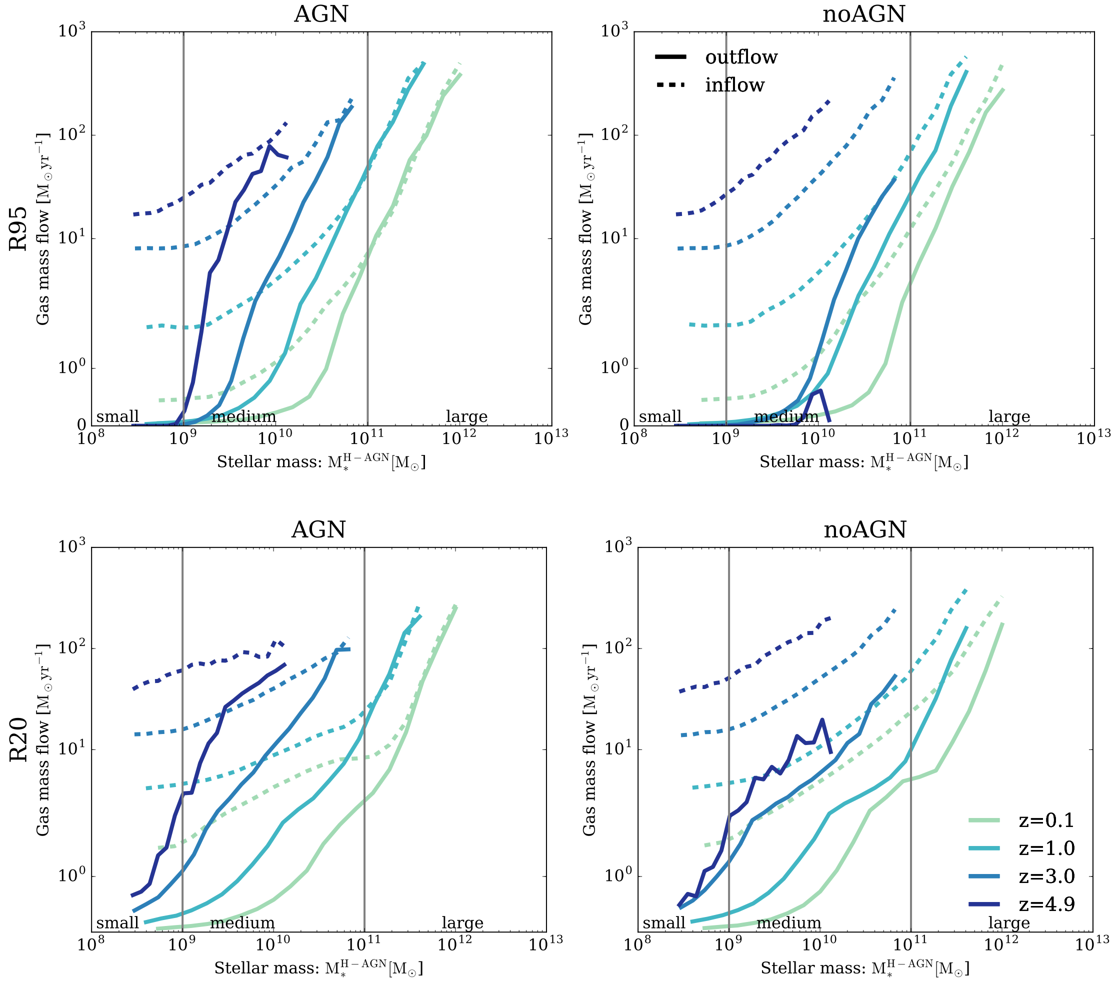

Figure 7.13: Gas inflows (dashed lines) and outflows (solid lines) at halo (R95, measured at $R = 0.95R_{\mathrm{vir}}$) and galaxy (R20, measured at $R = 0.2R_{\mathrm{vir}}$) scales for both H-AGN and H-noAGN, for a range of redshifts. To simplify the comparison with other plots, flow values for both H-AGN and H-noAGN are plotted against H-AGN galaxy stellar masses. Note that both outflows and inflows represent the median value for a given mass bin and do not necessarily belong to the same object in H-AGN and H-noAGN. The overall inflows are reduced in the presence of AGN feedback, which also drives outflows at halo scales for medium size galaxies.

present for large galaxies. These pseudo flows appear in Figure 7.13 because the algorithm used to extract the absolute flow values presented here assumes the halo can be accurately represented by a sphere. However, if the halo is non-spherical, pseudo flows are created. When rotating a non-spherical object through a spherical surface across which absolute mass flows are measured, parts of the object passing out of the sphere will register as outflows, while parts passing in will register as inflows. However, these contributions are not mass flows in the common sense and cancel out when calculating net mass flows.

Small galaxies undergo no outflows at halo scales, with or without AGN feedback, which matches the conclusion from Figure 7.12 that the baryon mass of their halos is identical in H-AGN and H-noAGN. AGN feedback reduces inflows for medium and large galaxies, at both halo and galaxy scales, but the effect is more pronounced at the latter. A more quantitative analysis of the outflows driven by AGN feedback is presented in Figure 7.14, where residual flow values for the two simulations are plotted. Residual flows are defined as the mass flow rates in H-AGN relative to those of their twin galaxies in H-noAGN, i.e $\dot{M}_{\rm gas}^{\rm residual} = \dot{M}_{\rm gas}^{\rm H-AGN} - \dot{M}_{\rm gas}^{\rm H-noAGN}$. This approach has the advantage of isolating the effect of AGN feedback and subtracting out any effects present in both simulations, such as the supernova driven outflows for small and medium galaxies at halo scales and the pseudo flows seen for large objects at both radii.

The residual gas flows shown in Figure 7.14 demonstrate that AGN feedback has an approximately equal and opposite effect on outflows and inflows. Galaxies with stellar masses $M_*^{\rm H-AGN} \leq 10^{11}$ M$_\odot$ see a similar amount of gas carried away by AGN driven outflows as that depleted in inflows. Apart for small galaxies at high redshift ($z = 5$), where AGN feedback seems able to heat up the gas in the vicinity of the galaxy which is causing gas to pile up and trigger larger inflows in H-AGN, the differences between flows at halo and galaxy scales are rather modest. There is a weak trend for boosted outflows to be less dominant at galaxy scales (especially at $z \leq 1$) than at halo scales and conversely for the suppression of inflows to be more relevant on small scales (especially for galaxies which are less massive than $M_*^{\rm H-AGN} \leq 10^{10}$ M$_\odot$), but overall material is neither being significantly swept up in the halo and kicked out, nor preferentially deposited there. Rather, these two effects combine to reduce the net median inflow for medium galaxies into both the halo and the galaxy by up to 60% for the most massive objects, as the direct comparison of net inflows for twins in Figure 7.15 shows. This results in the reduction in baryon mass seen for the same galaxies in Figure 7.12. In agreement with the same figure, small galaxies show no change in

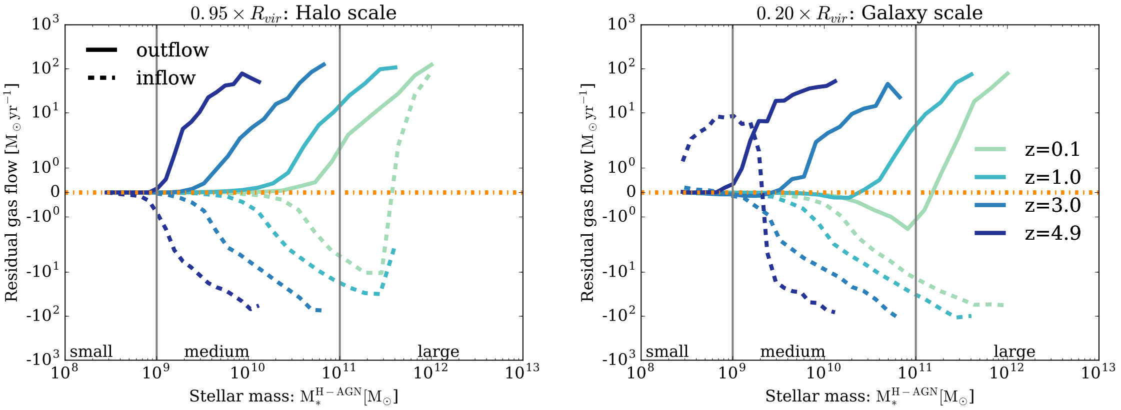

Figure 7.14: Residual gas flows $\dot{M}_{\rm gas}^{residual} = \dot{M}_{\rm gas}^{\rm H-AGN} - \dot{M}_{\rm gas}^{\rm H-noAGN}$, plotted against the twin's galaxy stellar mass in H-AGN. Data is presented for flows at two different distances from galaxies, as indicated above each panel, and a range of redshifts. Higher mass flow rates in H-AGN appear above the dash-dotted line, which denotes identical flows in both simulations and lower mass flow rates sit below this line. Note that both outflows and inflows represent the median value for that mass bin and do not necessarily belong to the same twin. In general, AGN feedback causes a boost in outflows and an approximately equal but opposite reduction in inflows.

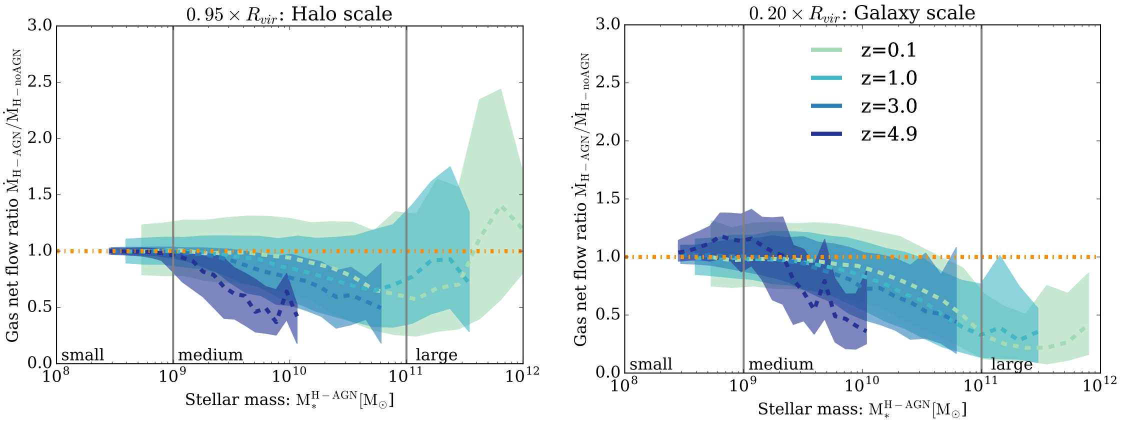

Figure 7.15: Ratio of net flow rates at halo and galaxy scales as a function of galaxy stellar masses in H-AGN, $M_*^{\rm H-AGN}$. The dash-dotted horizontal line at a ratio of 1.0 corresponds to identical net flows in both H-AGN and H-noAGN. Data is presented for flows at two different distances from galaxies, as indicated above each panel and at a range of redshifts. AGN feedback has no effect on flows in small galaxies, produces a reduction in median net flow rates of up to 60% for medium size galaxies and drives a bursty flow pattern leading to increased net flows at halo scales for large galaxies. Positive net flows are defined to be falling into the object.

net flows at halo scales, which matches their identical baryon mass in the presence and absence of AGN feedback. These results are in agreement with work by Dubois et al. (2013a), who compare high resolution zoom simulations with and without AGN feedback of a single galaxy of $M_*(z = 6) = 6.2 \times 10^9$ M$_\odot$. The authors find strong evidence for the fact that the AGN significantly heat the gas at halo scales, driving a hot super-wind and destroying cold flows. Pontzen et al. (2017) also report AGN boosted outflows in their hydrodynamical merger simulation and emphasize that long term quenching requires the inflows to be suppressed.

For large galaxies, the situation is different, particularly at low redshift. At halo scales, the simulation with AGN feedback actually shows boosted inflows carrying an amount of mass similar to that in boosted outflows. This means the outflows are being recycled, as AGN feedback becomes unable to gravitationally unbind the gas from the halo. It is important to note that the values plotted here represent the median value for a given mass bin, so the outflow and inflow values do not necessarily belong to the same object. It is therefore not necessarily correct that the two curves cancel out to produce no change in the net flow. Indeed, a comparative analysis of net inflows for each twin across H-AGN and H-noAGN (Figure 7.15) shows that at halo scales, the overall inflow is boosted by up to 50% for the most massive galaxies in the presence of AGN feedback.

At galaxy scales, the gas flow patterns for the most massive objects, $M_*^{\mathrm{H-AGN}} > \times 10^{11}$ M$_\odot$ at $z = 0.1$), become harder to predict. In these galaxies, AGN fall into maintenance feedback mode (see Section 7.4.3) at redshifts below the peak of star formation at $z = 2$. This produces very bursty outflows, as SMBHs go through cycles of being fed, which triggers strong feedback episodes. The latter drive out the gas, starving the black hole and the feedback abates until enough gas becomes available again for the SMBH to go through another accretion event. These cycles take place on timescales much shorter than the time interval between the redshift outputs considered in this work. Furthermore, the number of such galaxies is quite limited (≈ 1000). Figure 7.16, which shows that at redshift $z = 0.5$ the median outflow depends quite sensitively on the exact point in time at which the distribution of galaxies is sampled demonstrates the impact of this. As the SMBHs cycle rapidly through a wide variety of active and quiescent states, the distribution of residual outflows spans several orders of magnitude. In comparison, the outflows produced by the larger population of smaller galaxies are more steady on similar timescales and therefore their sampling is more robust (see left panel of Figure 7.16). The variation

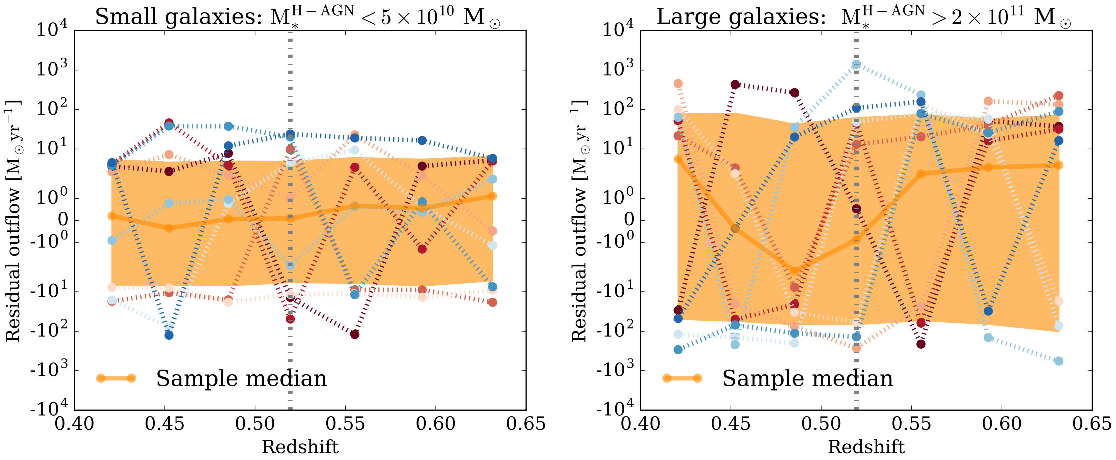

Figure 7.16: Residual outflows due to AGN feedback at galaxy scales, $R20 = 0.2 \times R_{\mathrm{vir}}$, plotted for 10 randomly selected galaxies within the small and large galaxy mass bins, between redshifts $0.64 > z > 0.42$. The solid line represents the median sample output, with the shaded region covering quartile error ranges. The vertical dotted line represents redshift $z = 0.52$, for comparison with Figure 7.14. This shows that outflows from large galaxies (right panel) with stellar masses $M_*^{\mathrm{H-AGN}} > 2 \times 10^{11}$ M$_\odot$ at redshifts around $z = 0.52$ (dotted vertical line) vary on very short timescales, so the median outflow value, (solid line, with quartile error ranges represented by the shaded regions), is very sensitive to the exact time at which it is sampled. By comparison, the outflows for small galaxies with stellar masses $M_*^{\mathrm{H-AGN}} < 5 \times 10^{10}$ M$_\odot$ (left panel) vary much less on these timescales and the median value is less sensitive to variations in time.

timescales of the large scale outflows studied here do not necessarily reflect the duty cycle of the SMBH, as each burst can be driven by a series of feedback events.

It is interesting to note that the lack of excess outflows for low mass galaxies ($M_*^{\mathrm{H-AGN}} < 5\times10^{10}$ M$_\odot$ at $z = 0.1$) in Figure 7.14 does not mean that the outflows for any given matching galaxy pair in H-AGN and H-noAGN are identical. As the panel on small galaxies in Figure 7.16 demonstrates, individual objects show a variety of residual outflows (and inflows). The build-up of small differences in galaxy properties not necessarily induced by AGN feedback (e.g. stochastic star formation algorithm, seeding and growth of the central SMBH) can result in temporarily diverging residual outflow histories at low redshift. In other words, even though the precise amount of residual outflow from any specific twinned pair of galaxies does depend on the timescale of the outflow and so is sensitive to the redshift at which it is measured, the lack of any marked systematic difference in the gas flow pattern due to AGN feedback registers as a median residual outflow of zero for the whole sample.

Another note of caution concerns the residual flows in Figure 7.14, which likely underestimate the effect of AGN on inflows and outflows. Particularly for massive galaxies at low redshift, where the H-noAGN twin has a stellar mass $\sim 5\times$ that of the H-AGN one, the H-noAGN twin has stronger stellar flows that obscure some of the AGN driven effects when calculating residual gas flows as $\dot{M}_{\mathrm{gas}}^{residual} = \dot{M}_{\mathrm{gas}}^{\mathrm{H-AGN}} - \dot{M}_{\mathrm{gas}}^{\mathrm{H-noAGN}}$. In light of the previous remarks about outflow timescales, the asymmetry between residual inflows and outflows can be safely attributed to meaning that AGN feedback does preferentially suppress inflows in the vicinity of all galaxies rather than eject gas from them, except perhaps at the very high mass end of the galaxy stellar mass function, $M_*^{\mathrm{H-AGN}} > \times10^{11}$ M$_\odot$ at low redshifts ($z < 0.1$).

Overall, flow patterns due to AGN feedback, made up in roughly equal parts of boosted outflows and reduced inflows at halo scales, comfortably explain the non-linear distribution of baryon masses plotted in Figure 7.12, which combines with a reduced star formation efficiency across all galaxy mass bins to produce the quenching mass ratios shown in Figure 7.5.

7.6 Discussion and Conclusions

The work presented in this chapter has isolated the effect of AGN feedback on stellar quenching in massive galaxies by comparing two cosmological simulations, H-AGN and H-noAGN, which were run with and without AGN feedback respectively. More specifically, by twinning individual DM halos and galaxies across the two simulations, the effect of feedback on individual objects throughout cosmic time has been quantified. In agreement with a large body of previous work (such as Springel et al., 2005; Bower et al., 2006; Sijacki et al., 2007; Ciotti & Ostriker, 2007; Cattaneo & Teyssier, 2007; Dubois et al., 2013a; Genel et al., 2014; Schaye et al., 2014; Pontzen et al., 2017, among others) the results show that AGN feedback is instrumental in quenching the massive end of the GSMF. Whilst the stellar mass of galaxies without AGN feedback closely follows predictions based on the assumption that all baryons contained in dark matter halos end up forming stars, galaxies subject to the influence of AGN feedback end up with masses distributed according to a GSMF that shows a characteristic exponential steepening at the high mass end, in line with observations (Bielby et al., 2014; Daddi et al., 2002; Moustakas et al., 2013; Tomczak et al., 2014; González et al., 2011; Song et al., 2016).

The importance of AGN feedback has been emphasised in all recent large-scale simulations of galaxy evolution, but the results differ in the details. Similar to results presented here, galaxies in the MassiveBlackII simulation exhibit signs of relatively strong quenching early on but then see a reduction in the impact of AGN feedback at lower redshifts (Khandai et al., 2015). A similar issue is reported by Illustris, who found that despite aggressive AGN feedback that produces unrealistically low gas fractions in DM halos at low redshift, star formation is not suppressed strongly enough and the simulation overproduces massive galaxies (Genel et al., 2014; Vogelsberger et al., 2014a,b). As opposed to the dual AGN feedback model used in H-AGN and Illustris, the EAGLE simulation only employs a single feedback mode, tuned to reproduce the GSMF at $z = 0$. They do also see good agreement with the GSMF at redshift $z > 2$, which supports the conclusion that the shift in feedback mode plays a subordinate role in the importance of AGN feedback (Schaye et al., 2014). Overall, regardless of the hydrodynamics scheme employed and the detail of the subgrid model implementation of AGN feedback, all four recent large-scale cosmological simulations of galaxy evolution agree that AGN feedback must play a crucial role in regulating the evolution of massive galaxies.

A closer comparison of the stellar mass for individual objects in H-AGN and H-noAGN reveals a non-linear dependence of quenching mass ratio on mass, with the most massive galaxies being the most strongly quenched and the smallest galaxies mostly unaffected by AGN feedback. This leads to a characteristic shape for the mass ratio $M_*^{\mathrm{H-AGN}}/M_*^{\mathrm{H-noAGN}}$, which shows a linear dependence on $\log(M_*^{\mathrm{H-AGN}})$ for medium sized galaxies with $10^9\ \mathrm{M}_\odot \lesssim \mathrm{M}_*^{\mathrm{H-AGN}} \lesssim 10^{11}\ \mathrm{M}_\odot$ (the exact values depend on redshift), but tails off at both low and high mass ends. The most massive galaxies, with $M_* > 10^{11}$ $\mathrm{M}_\odot$ are most strongly quenched and contain only 20% of the stellar mass in the presence of AGN feedback at $z = 0.1$, in comparison to the case without feedback.

There is also a significant redshift evolution for the smallest galaxy mass to be affected by AGN feedback, with smaller galaxies being more quenched at higher redshift. This transition mass at which AGN feedback becomes important, evolves from 2.24×10^{10} $\mathrm{M}_\odot$ at $z = 0.1$, to 1.48×10^{9} $\mathrm{M}_\odot$ at $z = 4.9$. Such a redshift dependence seems in good agreement with observations (Baldry et al., 2004; Peng et al., 2010) but systematically leads to values at high redshift which are lower than those reported by a range of hydrodynamics simulations (Di Matteo et al., 2012; Keller et al., 2015; Pontzen et al., 2017; Bower et al., 2017). These discrepancies reflect the fact that quenching is very likely a cumulative process that builds up over the entire history of the galaxy, not just a one-off event that shuts down star formation forever when the galaxy reaches a particular stellar mass. If correct, the consequence is that differences induced by AGN feedback in star formation rates can be small, particularly for objects near the transition mass, the effect of such feedback only becoming apparent for galaxies once its integral represents a significant fraction of their total stellar mass.

In the case presented here, this evolution is caused by the median central black hole mass of galaxies of a given stellar mass, M_{SMBH}/M_*, increasing with redshift by a factor of a few between $z = 0$ and $z = 5$, a result for which there exists observational support (Merloni et al., 2010; Decarli et al., 2010). However, it has been suggested that the observation trend might be biased (Lauer et al., 2007; Volonteri & Stark, 2011; Bongiorno et al., 2014). A similar trend has been reported in other large scale simulations, such as MassiveBlackII (Khandai et al., 2015) and Illustris (Sijacki et al., 2015), but not in all of them (Rosas-Guevara et al., 2016). This evolution in black hole masses combines with higher accretion rates at high redshift, due to gas-rich galaxies, such that a galaxy of the same stellar mass is subject to AGN feedback up to three orders of magnitude stronger at redshift $z = 5$ than at $z = 0.1$. There is also a shift in feedback mode, with at least 85.5% of AGN at redshift $z = 3$ and above in quasar mode, compared with a maximum of 19.1% at lower redshifts. The increasing radiative efficiency associated with this shift is unable to offset the trend of lower M_{SMBH}/M_* and lower accretion rates.

A comparative analysis of the baryon content of halos reveals that AGN feedback quenches star formation through a combination of reducing the total gas supply within the halo by driving outflows and preventing accretion of fresh gas by curbing inflows. Small galaxies with $M_*^{\mathrm{H-AGN}} \leq 10^9$ $\mathrm{M}_\odot$ show nearly identical baryon masses with and without AGN feedback, as the flows at halo-scales remain chiefly unaffected by feedback. Note that this is not true at galaxy scales and high redshift, where gas inflows can be somewhat enhanced by feedback for these small objects. On the other hand, medium size galaxies 10^9 $\mathrm{M}_\odot \leq \mathrm{M}_*^{\mathrm{H-AGN}} \leq 10^{11}$ $\mathrm{M}_\odot$ experience a significant reduction in baryon mass, caused by an approximately equal contribution from AGN-driven gas outflows and a reduction of cosmic inflows. At galaxy scales, the reduction of inflows dominates. Finally, for large galaxies (10^{11} $\mathrm{M}_\odot \leq \mathrm{M}_*^{\mathrm{H-AGN}}$) the baryon mass rises again, as inflows at halo scales are swelled by gas expelled by AGN feedback in the inner regions, which remains gravitationally bound and falls back into the halo. At galaxy scales, outflows for the most massive objects vary on very short timescales, as the AGN enter a bursty maintenance mode. Thus the gas mass has a tendency to increase on average (as compared to medium size galaxies), even though the stellar mass does not, given the long characteristic timescale of star formation.

The picture that emerges seems consistent with high resolution work (20 pc instead of 1 kpc) by Dubois et al. (2013a), who use zoom simulations of individual objects with and without AGN feedback and find that AGN drive hot super-winds which disrupt cold inflows on halo scales. The two effects combine to reduce the baryon content of galaxies by up to 30%. Similar conclusions were reached by early semi-analytic work by e.g. Benson et al. (2003), who argue that strong, hot winds are necessary to reproduce the observed luminosity function and recent simulations by Pontzen et al. (2017) and Taylor & Kobayashi (2015) that also report AGN boosted outflows, concluding that long term quenching requires gas inflows to be suppressed.

Looking at cosmic accretion specifically, Nelson et al. (2015) find that the simulation without feedback (neither AGN driven nor stellar) see much higher levels of smooth accretion onto the galaxy than the ones with (AGN and stellar) feedback. However, contrary to the results presented here, they find no evidence for recycling of gaseous material at the halo boundary. While all current large scale cosmological simulations include AGN feedback as an integral part of their galaxy evolution model, some authors contend that processes in massive galaxies that do not rely on AGN feedback can reproduce galaxy mass functions through e.g. cosmic quenching (Feldmann & Mayer, 2014) and that stellar feedback super-bubble feedback can drive powerful outflows (Keller et al., 2015). Gabor & Bournaud (2014) and Roos et al. (2015), running idealised galaxy simulations, find AGN-driven outflows consistent with those reported here but lower impact on the star formation rate of galaxies.

In conclusion, AGN feedback provides an effective mechanism to reproduce the distribution of galaxies at and above the knee of the GSMF, over a redshift range spanning 90% of the age of the Universe. For local galaxies, AGN feedback plays an important role in stifling star formation in objects above a transition mass of $M_* \geq 2 \times 10^{10}$ $\mathrm{M}_\odot$. AGN feedback acts by reducing the stellar content of galaxies by up to 80% (for the most massive objects) through a mixture of increased outflows and reduced inflows, combined with a decreased star formation efficiency of in situ gas. The results presented in this chapter predict that the influence of AGN feedback should already be noticeable by redshift $z = 5$, for galaxies with relatively modest stellar masses ($M_* \approx 2 \times 10^{9}$ $\mathrm{M}_\odot$) by current epoch standards, as these objects are close to the top end of the GSMF at these redshifts. This is exciting news as the James Webb Space Telescope should be able to test this prediction in the near future.

This completes the results reported in this thesis. The next chapter will briefly summarise Chapters 4 to 7, before giving an outlook where the work presented in this thesis will lead in the future.

Chapter 8

Conclusions

Observed as early as 700 Myr into the history of the Universe, and as recently as 26,000 years ago, the evolution history of supermassive black holes (SMBHs) spans more than 13 billion years. After forming in rapidly evolving proto-galaxies at redshifts above $z = 10$, massive black holes have been part of galaxies since their beginning and the observed tight correlations between black hole mass and the mass of different components of their host galaxy in the local Universe suggest that the evolution histories of SMBHs and galaxies remain intrinsically linked. However, their origin and early evolution, including the rapid gas accretion required to reach observed masses by redshift $z = 6$, remains an open question. The same applies for the origin of the low redshift mass correlation relations. This thesis explored the coevolution between SMBHs and their host galaxies on a range of scales, using hydrodynamical simulations.

The first billion years of the Universe are an era for which few observations are available, particularly on the pc and sub-pc scales on which early black holes interact with their environment. Studying their origin using hydrodynamical simulations is a challenging task, as the range of relevant timescales and length scales is vast, and the complex interplay between different physical processes means the problem is highly non-linear and requires a set of robust numerical tools. This thesis tested two classic sub-grid algorithms of black hole physics, Bondi-Hoyle-Lyttleton (BHL) accretion and an analytic drag force, for their suitability in this regime (Chapter 4) and introduced an algorithm to extend the range of scales resolved in the black hole environment (Chapter 5). These tools were then employed to investigate the first 200 Myr of evolution of a supermassive black hole progenitor in a full cosmological context, in which gas is tracked from the cosmic web through the host galaxy and into the gravitational potential of the black hole (Chapter 6).

At redshift $z < 5$, SMBHs are frequently observed in their active state as active galactic nuclei (AGN). Despite their small physical extent compared to the size of their massive host galaxies, observations of tight correlations between black hole mass and various galactic properties hint at connection between them. The energy emitted by AGN provides one possible link, as interactions of gas, radiation, magnetic fields and gravity can extract energy from black hole accretion and deposit it at galactic or circumgalactic scales. If this energy is able to effectively couple to the gas supply of the galaxy, it could influence the evolution history of the host galaxy as a whole. How this energy shapes the galaxy's star formation history was investigated in Chapter 7.

8.1 Summary of this thesis

There are three ways in which a black hole exchanges energy and momentum with gas: accretion, feedback and dynamical friction. The length-scales on which these interactions physically take place have so far remained below the resolution limit of galaxy evolution simulations in a full cosmological context and they are therefore included as sub-grid algorithms that combine an analytic or empirical model with information available in the simulation. Two models commonly used for black hole physics are the BHL model for accretion (Hoyle & Lyttleton, 1939; Bondi & Hoyle, 1944) and an analytic model for the drag force due to dynamical friction by Ostriker (1999). Both models are based on quantities far from the gravitational influence of the black hole and have been very successful in hydrodynamical simulations where the cell size exceeds the gravitational length scale of the black hole. Chapter 4 investigated whether these two algorithms are still appropriate choices when the gravitational potential of a black hole is resolved.

Systematically exploring the space of Mach numbers and accretor radii, simulations of an isolated black hole in a uniform density distribution in Chapter 4 showed that the BHL accretion formula overestimates the accretion rate onto a point mass by over an order of magnitude for adiabatic flow and Mach numbers of 1.5 or higher (Section 4.3.2). This is caused by the onset of accoustic-advective instabilities when the subsonic region between bowshock and accretor is resolved, which dominate the flow for sufficiently small accretors and high Mach numbers. Lower Mach numbers show consistently stable flow patterns even at the highest resolutions and have accretion rates within a factor of 2 of the analytic BHL solution.

Simulations in Chapter 4 also showed that the BHL accretion algorithm in RAMSES, where gas quantities are evaluated as mass weighted averages within an accretion region, remains a good choice even when local conditions do not fulfil the assumptions of BHL accretion. At sufficiently high resolution, when the scale radius of the gravitational potential approximately equals the local cell size, the accretion algorithm naturally transitions to flux limited accretion (FLA). In this regime, accretion onto the black hole is determined by the mass flux into the accretion region and is therefore able to handle a wide range of resolutions and complex flow configurations (see Section 4.3.3). This algorithm will naturally converge to the correct Newtonian solution when the size of the accretion region approaches the physical size of the black hole, its Schwarzschild radius.

By contrast, the sub-grid algorithm for dynamical friction performs poorly with increasing resolution (see Section 4.4). Local gas properties diverge from analytic values when measured deep in the black hole's potential. When used in the model by Ostriker (1999), the resulting force on the black hole has significant errors in magnitude and acts in the opposite direction to the force based on the bulk flow of the gas. The smaller the accretion region compared to the size of the potential, the more severe the problem becomes. Dynamical friction is self-consistently handled by the simulation once the gravitational scale radius equals the smallest cell size, as gravitational focusing produces a wake downstream of the sink. The sub-grid algorithm therefore not only becomes obsolete but actively harmful when the black hole's gravitational potential is resolved and must be disabled.

Chapter 5 introduces a Super-Lagrangian refinement algorithm that makes resolving the gravitational potential of a black hole numerically affordable within an evolving galaxy. The zoom-within-zoom algorithm embeds a sphere of fixed, user-defined refinement within an adaptively refined context, which tracks the black holes trajectory throughout the simulation (see Section 5.3 for details). The zoom-within-zoom algorithm is reliably able to reproduce gas properties of a fully adaptively refined simulation at the target resolution, including the transition to FLA, for a refinement radius of 4 cells or more (see Section 5.4).

Continuing the work on testing current black hole sub-grid algorithms in high-resolution simulations, Chapter 5 included a detailed investigation into sink particle in FLA in spherically collapsing clouds. Unlike in Chapter 4, the black hole was free to move and the simulations included cooling, self-gravity, a background gravitational potential and momentum conservation during accretion. A comparison of simulations at a range of resolutions showed that, for a black hole with FLA, evaluating the local gas properties using a mass- and kernel-weighted average over the cloud particles has to be treated with caution. When spherical symmetry gets broken in the accretion region, either by marginally asymmetric accretion or because the sink is moving, the relative velocity between black hole and the bulk flow of the gas is overestimated, and gas properties measured by the cloud particles fluctuate (see Section 5.4.2). This effect is confined to cells within the black holes accretion region and purely numerical. Local gas properties should therefore be evaluated by summing only over the outermost cloud particles, which are located outside the accretion region when in FLA.

One important parameter in simulations of early black hole evolution is the choice of black hole seed mass, particularly in evolving galaxies. Using as much information available within the simulation as possible is a particularly difficult requirement for zoom-within-zoom, as some of the information is only revealed by the extra refinement added after the sink particle has been inserted. The outcome of simulations can critically depend on this choice, as the mass evolution of black holes below a minimum seed mass fail to converge when dynamical friction is under-resolved. Section 5.5 demonstrated that this can be avoided using a low seed mass and a maximum drag force until dynamical friction is resolved reliably. This combination produces a black hole evolution history that reflects the internal structure of the collapsing cloud in which the black hole forms, independently of seed mass. Section 5.6 showed that resolving the black hole environment using zoom-within-zoom provides a wealth of additional information about local gas properties, at affordable numerical cost. In the case presented in Section 5.6, it changes the accretion duty cycle of massive black holes from chaotic to episodic, and captures the emergence of a nuclear gas disc around the black hole.

Building on work from Chapters 4 and 5, Chapter 6 presented the first 200 Myr of a supermassive black hole progenitor's evolution, which formed as the equivalent of a stellar remnant at $z = 19.3$. The chosen host halo is a progenitor to a halo with a number density of 1 Gpc^{-3} at $z = 0$. Its evolution was modelled using a hydrodynamical simulation with sub-pc resolution in the black hole vicinity and pc resolution within the host galaxy. During the first 30 Myr of evolution, the black hole mass growth proceeded primarily via early super-Eddington accretion fuelled by interacting with dense pc size clumps. Rotation played an important role in the evolution of the compact host galaxy, which formed the bulk of its stellar mass within the central 10 pc. The black hole remained attached to the gas disc on a 10 pc orbit for several dynamical times, slowing the inflow of gas into its potential and limiting the black holes interaction with the dense nuclear clump. Late accretion therefore primarily proceeded via slow gravitational draining of the galactic gas disc, even though the immediate environment of the black hole was dominated by rapidly evolving sub-pc scale gas features. If star formation and black hole accretion follow the mass evolution of their host halo, the black hole will reach 3.9×10^9 $\mathrm{M_\odot}$ by $z = 6$, hosted by a galaxy with a stellar mass of 5.1×10^{10} $\mathrm{M_\odot}$ (see Section 6.6). On the other hand, an average Eddington ratio of $\chi = 7.3 \times 10^{-3}$ over the last 10 Myr up to $z = 11.1$ makes it unlikely that the black hole mass will exceed 10^9 $\mathrm{M_\odot}$ by $z = 6$, as star formation is providing strong competition to black hole accretion. However, it is possible that mergers or disc instabilities in the galactic gas disc between $z = 11$ and $z = 6$ lead to black hole accretion bursts that allow the black hole to reach 10^9 $\mathrm{M_\odot}$.

This study demonstrated the ability of zoom-within-zoom to shed light on the early evolution of massive black holes in a cosmological context, but also highlighted several areas for improvement. Omitting feedback from black holes and stars was intended to provide an upper limit on gas accretion but, as discussed in Section 6.7, the interaction between feedback, gas supply and black hole accretion at high redshift is complex and the gas supply within the galaxy might actually be decreased by the absence of feedback. The overly smooth gas distribution in the absence of feedback also limits the number of clumps, which make an important contribution to black hole accretion. The work in Chapter 6 raised questions about the assumption that preventing star formation within the zoom-within-zoom region has little impact on the total stellar mass of the galaxy. This will be particularly important if the black hole settles into the centre of the galactic gas disc, as suggested by black holes of comparable mass and resolution in Chapter 5.

Finally, Chapter 7 presented results on a different regime of SMBH and galaxy coevolution: the impact of AGN feedback on the star formation of its host galaxy from $z = 5$ to $z = 0$. This work is based on the HORIZON suite (Dubois et al., 2014b), which consists of a set of large-scale hydrodynamical simulations that cover the evolution history of a 100 $\mathrm{Mpc\,h^{-1}}$ box at 1 kpc resolution. For this work, the impact of AGN was isolated by comparing two simulations from the HORIZON suite, which are identical except that one also includes SMBHs and related feedback models. This allows individual galaxies to be cross-identified between simulations and the effect of AGN feedback on their properties to be quantified, including on stellar mass and gas outflows (see Section 7.3). I found that massive galaxies ($M_* \geq 10^{11}$ $\mathrm{M_\odot}$) are quenched by AGN feedback to the extent that their stellar masses decrease by up to 80% at $z = 0$ (see Section 7.4). SMBHs affect their host halo through a combination of outflows that reduce their baryonic mass, particularly for galaxies in the mass range 10^9 $\mathrm{M_\odot} \leq M_* \leq 10^{11}$ $\mathrm{M_\odot}$, and a disruption of central gas inflows, which limits in-situ star formation. As a result, net gas inflows onto massive galaxies, with $M_* \geq 10^{11}$ $\mathrm{M}_\odot$, drop by up to 70% (see Section 7.5). There is a redshift evolution in the stellar mass ratio of twin galaxies with and without AGN feedback, with galaxies of a given stellar mass showing stronger signs of quenching earlier on. This evolution is driven by a progressive flattening of the $M_{\mathrm{SMBH}} - M_*$ relation with redshift, particularly for galaxies with $M_* \leq 10^{10}$ $\mathrm{M}_\odot$. M_{SMBH}/M_* ratios decrease over time, as falling average gas densities in galaxies curb SMBH growth (see Section 7.4.3).

In conclusion, the coevolution between a SMBH and star formation in its host galaxy spans their entire joint evolution history, linked by the galactic gas supply. At high redshift, competition for the existing gas supply in the host galaxy impacts both star formation and black hole accretion, even in the absence of feedback. At low redshift, AGN feedback plays an integral role in the star formation history of its host galaxy as its cumulative effect can prevent up to 90% of the host galaxy's stellar mass from forming. The interplay between host galaxy gas supply, black hole accretion and feedback links massive galaxies and massive black holes, as has been demonstrated in several different contexts in this thesis. Neither of them can be understood without the context of the other, and much exciting work remains to be done on further exploring their connection.

8.2 Future outlook

Both the work on the origin of black holes (Chapter 6) and on the impact of AGN feedback (Chapter 7) offers plenty of scope for future work, whether in the long term or by building directly on the results presented in this thesis. Several projects are already underway, as described in this section.

8.2.1 Early evolution at high redshift

The focus of investigations into the origin of SMBH will remain on creating an object that matches observations at redshift $z = 6$. This is a challenge yet to be met with sub-pc resolution, particularly in as full a physical context as possible. Work presented in Chapter 6 has shown that understanding the origin of SMBHs will require a more holistic approach, which has to include both the infant black hole and its emerging host galaxy. Building on the work presented in this thesis, which has been focused on the black hole itself, the next important step will be to improve star formation and cooling models in the host galaxy to better represent conditions in the early Universe. Eventually, these simulations will provide an important framework to interpret observations at redshifts $z > 10$, expected to be supplied from the 2030s onwards by two new observatories: the James-Webb Space Telescope (JWST) (Gardner et al., 2006) will supply observation of the first galaxies and mini-quasars at redshifts above $z > 10$ and eLisa (Barausse et al., 2015) will supply complementary insights from gravitational waves into black hole mergers in the mass range $10^3 - 10^7$ $\mathrm{M}_\odot$ out to redshift $z = 10$.

A model for stellar feedback will require a choice of initial mass function (IMF). One difficulty with high-resolution simulations of the form presented here is that the mass of a given star particle is at the upper end of the expected IMF for the first generation of stars. It is therefore not clear whether a star particle represents a population of stars, as is usually assumed in low-resolution simulations, or an individual star. Existing feedback models, such as the multi-phase model by Kimm et al. (2015), will have to be rethought accordingly. Inclusion of feedback will also provide an opportunity to track the metal enrichment of gas and potentially improve the cooling physics to reflect the rapid chemical evolution at high redshift. Another open question concerns the importance of turbulence, which could have a lasting impact on the angular momentum transport, gas clumping statistics and star formation within the host galaxy (Yoshida et al., 2008; Latif et al., 2013a) and therefore have important consequences for early black hole accretion.

The long term plan for this work involves developing a model for black hole feedback for use in zoom-within-zoom simulations at redshift $z > 10$, possibly using radiative transfer, to understand how the energy emitted by black holes couples to gas on different scales in a cosmological context. Work on this topic has begun in collaboration with Rebekka Bieri. The pilot study in Chapter 6 showed that including several zoom-within-zoom black holes in the same simulation could be numerically affordable which will raise questions about the technical implementation as well as the impact of black hole mergers: is there a last-pc problem at high redshift (Colpi et al., 2007; Nixon et al., 2011)?

8.2.2 Coevolution at low redshift

The HORIZON simulations will continue to provide a treasure trove of information concerning the impact of AGN feedback on galaxy evolution. The investigation into outflows and their consequences for star formation in Section 7.5.2 has shown that AGN feedback has a profound impact on the gas supply of the host galaxy, over a large range of scales. Many open questions remain as to the consequences of this interaction: where is the gas preferentially deposited, and what impact does slowing down inflows have on the large-scale structure of gas? Does the impact of AGN depend on whether they primarily drive hot or cold outflows (Morganti et al., 2009)? How does the baryon conversion efficiency of simulated galaxies compare to observations (Barnes et al., 2017; Schreiber et al., 2016)? How does AGN feedback influence the distribution of metals in the Universe (Mannucci et al., 2010; Werner et al., 2013; D'Odorico et al., 2016)?

I am currently involved in two ongoing projects based on AGN in the HORIZON simulation. The first project, in collaboration with Mark Richardson, aims at determining the typical extent to which metals permeate the intergalactic medium (IGM) and how this is impacted by AGN. Quantifying the importance of AGN in the enrichment of the circumgalactic medium (CGM) will put important constraints on the sample selection for both observational and numerical work and determine how the mass-metallicity relation (Mannucci et al., 2010) changes under the influence of AGN feedback. Early results show that the presence of an AGN flattens the peak but increases the extent of galactic metallicity profiles. The HORIZON simulations are ideally suited for this work as the large sample of galaxies in both H-AGN and H-noAGN means that it is possible to identify samples of "equivalent" galaxies that have comparable dark matter and stellar mass, as well as a set of twins, thereby reducing the number of free parameters in the comparison. This is particularly important at $z = 0$, where twinned galaxies can have significantly different stellar masses, which affect the total amount of metals produced.

The second project is led by Elisa Chisari and investigates the impact of baryons on the large scale structure of the universe. After collaborating on investigating the impact of baryons on the alignment of dark matter halos and galaxies (Chisari et al., 2017), the focus of our project has now shifted to the impact of baryons in general, and AGN specifically, on the matter power spectrum. Early results show a redshift evolution in the extent to which the power spectrum is suppressed by AGN winds, which matches the decreasing ability of AGN to remove baryons from massive halos at low redshift, reported in Chapter 7. This work will be important to correct the matter power spectrum extracted from N-body simulations for comparison with observations.

One strength of simulations is their ability to provide evolution histories of individual objects, which can then be used to link observations at a range of redshifts. After studying the average quenching of the whole sample in Chapter 6, there are plans for a complementary analysis that focuses on the star formation history of individual galaxies. This work will shed light on quenching timescales and determine whether quenching is a one-off event, a slow continuous process or possibly temporary or cyclic (Bahe & McCarthy, 2014; Cicone et al., 2014; Barro et al., 2016; Pontzen et al., 2017). Could this be a question of environment (Pimbblet et al., 2002), or of galaxy type (Schawinski et al., 2014)? This question will be particularly interesting in the context of "New HORIZON", a high-resolution zoom simulation of a sub-volume of H-AGN currently under development.

In the long term, studies into the coevolution of massive galaxies and SMBH will have to look beyond the existing datasets and beyond hydrodynamics. Three crucial large-scale processes missing from hydrodynamical simulations are cosmic rays, magnetic fields and radiative transfer. All three are computationally expensive but could play a vital role in shaping galaxies. Cosmic rays, in combination with magnetic fields, will be particularly interesting for the study of large-scale outflows (Everett et al., 2008; Girichidis et al., 2016; Salem & Bryan, 2014), and possibly provide an alternative to AGN driven quenching of star formation. In this context, it will also be important to improve the current models for AGN feedback. My own research will turn towards understanding how AGN feedback couples to gas on galactic and intergalactic scales, with a particular focus on large-scale jets in cool-core clusters.

In conclusion, studying the coevolution between galaxies and black holes and their impact on the Universe at large, using simulations, will remain an active and exciting field of research for many years to come.

Appendix A

Uniform initial conditions and small accretors for the Hoyle-Lyttleton problem

The simulations presented in Chapter 5 use uniform initial conditions to simplify and homogenise the setup of different flow configurations. After an initial period, during which the flow settles, the simulations are assumed to have erased the memory of their initial conditions and evolved to their natural quasi-steady state solution. As can be seen in the top row of Figure A.1, small accretors can cause shocks during the initial settling phase which generate density perturbations that remain present even after many dynamical timescales. The method presented in this appendix is designed to minimise this spurious effect.

The first precaution taken was to introduce the analytic gravity field gradually, with $F^G = GM_{\rm sink}/r^2 \times (t/t_{\rm full})$ while $t < t_{\rm full}$, where $t_{\rm full} = 2$. For large accretors ($N = 65$), the simulations show some instability for $t < 5$, visible in the profile plots of Figure A.2, particularly when the bowshock detaches. The flow pattern then settles into a stable and converged solution. However for sufficiently small accretors, i.e. when $N > 100$, simulations of any Mach number show a shock forming in front of the sink and travelling upstream, driven and supported by eddies (see top row of Figure A.1 for an example). This 'circularisation' of the gas flow breaks the symmetry of the problem and the perturbations behind this shock fail to settle during more than 10 dynamical times. This phenomenon is only observed for sufficiently small accretors, as the reverse shock is caused by the strongly peaked density profile around the sink.

In order to minimise the effect of the initial conditions for small accretors, all simulations in Chapter 5 with $N > 100$ were run using a reduced maximum level of refinement while $t < 3$, which is equivalent to an effective accretor size $N_{\rm eff}(t < 3) \approx$

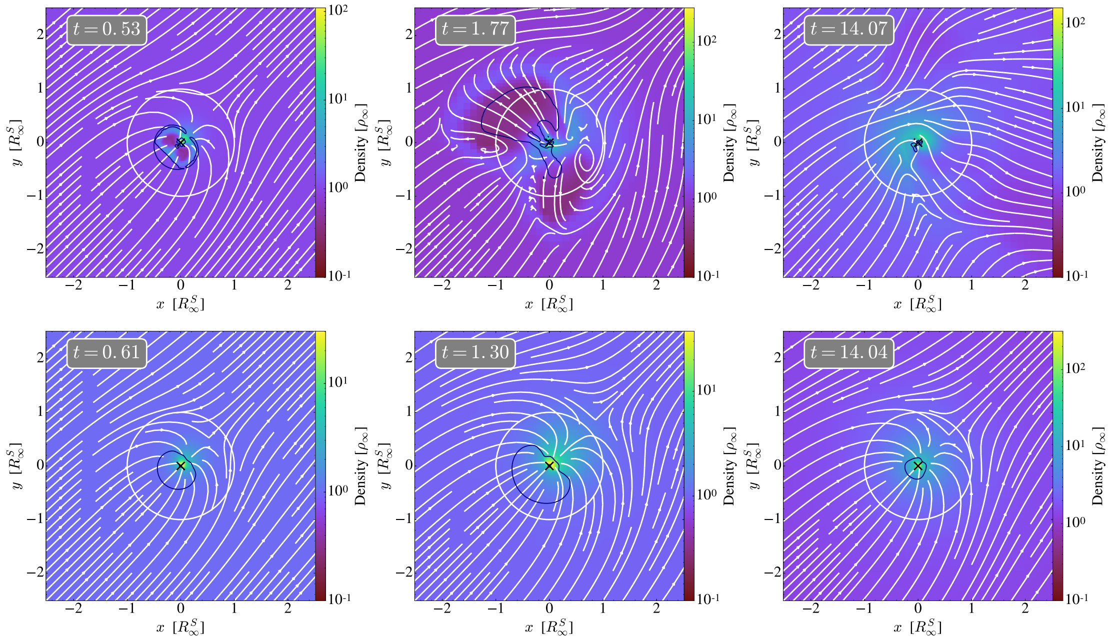

Figure A.1: Flow patterns for m0.5n262a_A with all refinement levels available from the beginning (top row) and refinement levels triggered gradually (bottom row). The sink is marked by a black cross and R_∞^S is shown as a white circle. Streamlines are annotated in white and sonic contours in dark blue. Both simulations have identical final grid configurations but the shock is avoided with gradual level releasing.

90. This avoids shocking the gas and allows the flow to settle. Around $t \simeq 3$, the extra levels of refinement are added in a staggered fashion to minimise the impact of changing the grid structure: first the size of the accretor is decreased and then an extra level of refinement is added per timestep until the target resolution is reached.

Figure A.2 shows that just after adding refinement levels, the gas "sloshes" at the bottom of the gravitational well, causing high frequency fluctuations in the gas properties $\rho_\bullet$, $v_\bullet$ and $c_{s,\bullet}$. After a few dynamical times, the solution settles into the stable flow pattern observed for larger accretors (see bottom row in Figure A.1). A test run using $N = 65$ (m0.5n65a), which is naturally stable, shows that both the run with the full resolution available from the beginning (see n65_t0 in Figure A.2), and the run where the full resolution is reached at $t = 3$ (see n65_t3 in Figure A.2) converge to $\left\langle \dot{M}/\dot{M}_\infty^{\rm BHL} \right\rangle = 0.89$, which is also recovered for the small accretor with staggered refinement, n263_t3. By contrast, the simulation in which the small accretor is fully resolved from the start, n262_t0, displays a lack of convergence, high frequency variations in the accretion rate and a much lower time averaged value of $\left\langle \dot{M}/\dot{M}_\infty^{\rm BHL} \right\rangle = 0.28$. To ensure that simulations have sufficient time to settle, all time averaged values in Chapter 5 are measured for $t > 10$.

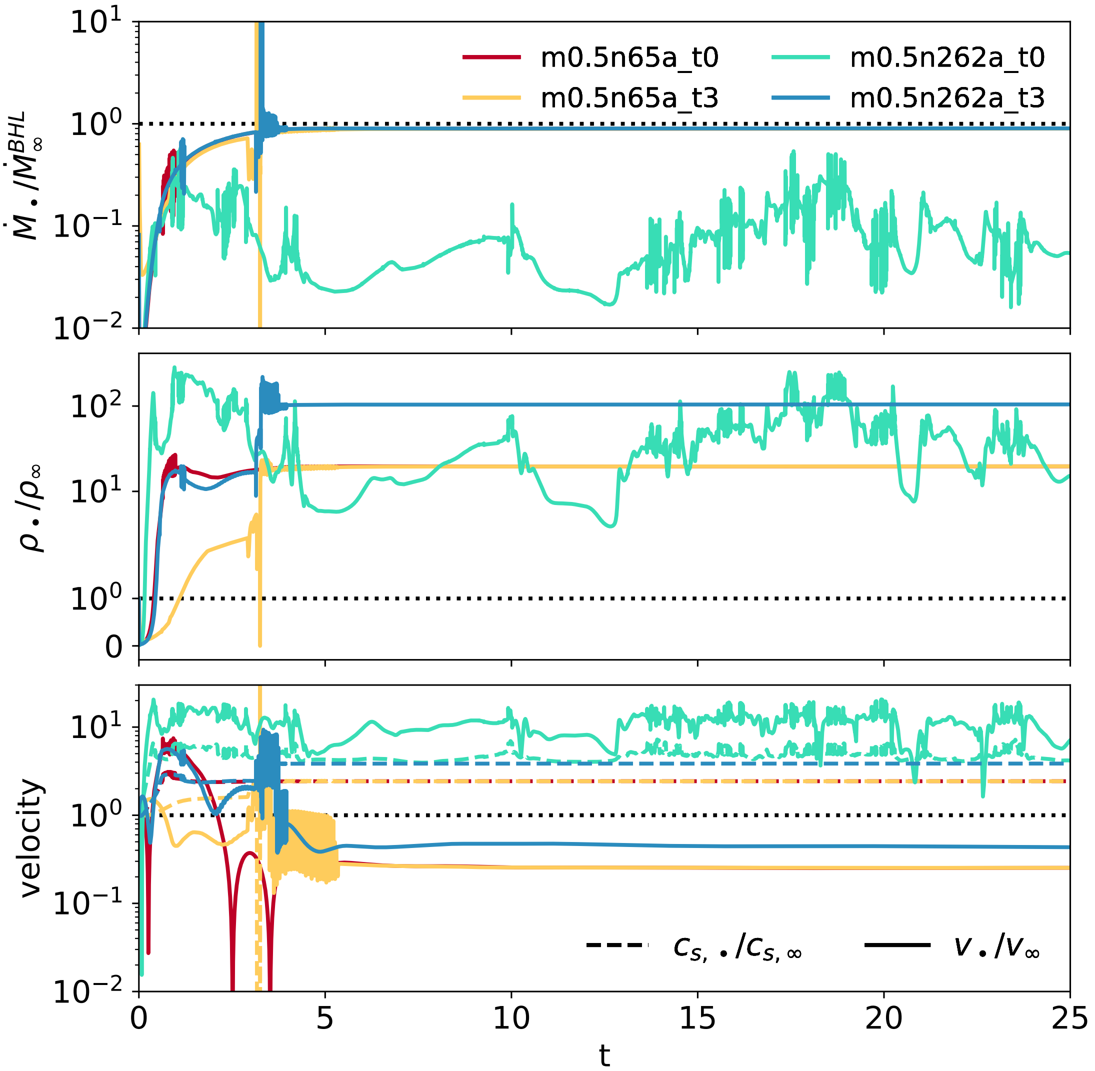

Figure A.2: Accretion rates and mass averaged sink properties for simulations with all refinement levels available from the beginning (m0.5n65a_t0 and m0.5n262a_t0) and simulations where the last two levels of refinement are added at $t = 3$ (m0.5n65a_t3 and m0.5n262a_t3).

Bibliography

Abbott B. P., et al., 2016a, Phys. Rev. Lett., 116, 061102

Abbott B. P., et al., 2016b, Phys. Rev. Lett., 116, 241103

Abel T., 2002, Science, 295, 93

Abel T., Bryan G. L., Norman M. L., 2000, ApJ, 540, 39

Abramowicz M. A., Fragile P. C., 2013, Living Rev. Relativ., 16, 1

Abramowicz M. a., Czerny B., Lasota J. P., Szuszkiewicz E., 1988, ApJ, 332, 646

Ade P. A. R., et al., 2014, A&A, 571, A16

Agis-Gonzalez B., et al., 2014, MNRAS, 443, 2862

Ahn C. P., et al., 2017, ApJ, 839, 72

Alexander T., Natarajan P., 2014, Science, 345, 1330

Alvarez M. A., Wise J. H., Abel T., 2009, ApJ, 701, L133

Anglés-Alcázar D., Özel F., Davé R., 2013, ApJ, 770, 5

Arav N., et al., 2001, ApJ, 561, 118

Arun K. G., et al., 2009, Class. Quantum Gravity, 26, 094027

Aubert D., Pichon C., Colombi S., 2004, MNRAS, 352, 376

Azadi M., et al., 2015, ApJ, 806, 187

Baade W., Minkowski R., 1954, ApJ, 119, 206

Bahe Y. M., McCarthy I. G., 2014, MNRAS, 447, 969

Baldassare V. F., Reines A. E., Gallo E., Greene J. E., 2015, ApJ, 809, L14

Baldry I. K., Glazebrook K., Brinkmann J., Ivezić Ž., Lupton R. H., Nichol R. C., Szalay A. S., 2004, ApJ, 600, 681

Baldry I. K., Balogh M. L., Bower R. G., Glazebrook K., Nichol R. C., Bamford S. P., Budavari T., 2006, MNRAS, 373, 469

Ballesteros-Paredes J., Hartmann L. W., Pérez-Goytia N., Kuznetsova A., 2015, MNRAS, 452, 566

Barausse E., Bellovary J., Berti E., Holley-Bockelmann K., Farris B., Sathyaprakash B., Sesana A., 2015, J. Phys. Conf. Ser., 610, 012001

Barger A. J., Cowie L. L., Mushotzky R. F., Yang Y., Wang W.-H., Steffen A. T., Capak P., 2005, Astron. J., 129, 578

Barnes D. J., et al., 2017, MNRAS, 471, 1088

Barro G., et al., 2016, ApJ, 820, 120

Bate M. R., Bonnell I. A., Price N. M., 1995, MNRAS, 277, 362

Beck R., Wielebinski R., 2013, in , Planets, Stars Stellar Syst.. Springer Netherlands, Dordrecht, pp 641–723

Beckmann V., Shrader C. R., 2013, in An INTEGRAL view of the high-energy sky (the first 10 years).

Begelman M. C., 1979, MNRAS, 187, 237

Begelman M. C., 2014, Lect. notes, pp 1 –18

Begelman M. C., Volonteri M., Rees M. J., 2006, MNRAS, 370, 289

Bennett D. P., et al., 2002, ApJ, 579, 639

Bennett C. L., et al., 2013, ApJS, 208, 20

Benson A. J., 2010, Phys. Rep., 495, 33

Benson A. J., Bower R. G., Frenk C. S., Lacey C. G., Baugh C. M., Cole S., 2003, ApJ, 599, 38

Bernardi M., Meert A., Sheth R. K., Vikram V., Huertas-Company M., Mei S., Shankar F., 2013, MNRAS, 436, 697

Bertone G., Hooper D., Silk J., 2005, Phys. Rep., 405, 279

Bielby R. M., et al., 2014, A&A, 568, A24

Bieri R., Dubois Y., Silk J., Mamon G. A., 2015, ApJ, 812, L36

Biernacki P., Teyssier R., Bleuler A., 2017, MNRAS, 469, 295

Binney J., Tabor G., 1995, MNRAS, 276, 663

Birnboim Y., Dekel A., 2003, MNRAS, 345, 349

Blandford R. D., Znajek R. L., 1977, MNRAS, 179, 433

Bleuler A., Teyssier R., 2014, MNRAS, 445, 4015

Bleuler A., Teyssier R., Carassou S., Martizzi D., 2015, Comput. Astrophys. Cosmol., 2, 5

Blondin J. M., Pope T. C., 2009, ApJ, 700, 95

Blumenthal G. R., Faber S. M., Primack J. R., Rees M. J., 1984, Nature, 311, 517

Bondi H., 1952, MNRAS, 112, 195

Bondi H., Hoyle F., 1944, MNRAS, 104, 273

Bongiorno A., et al., 2014, MNRAS, 443, 2077

Booth C. M., Schaye J., 2009, MNRAS, 398, 53

Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, MNRAS, 370, 645

Bower R. G., Schaye J., Frenk C. S., Theuns T., Schaller M., Crain R. A., McAlpine S., 2017, MNRAS, 465, 32

Brandt W. N., Alexander D. M., 2010, Proc. Natl. Acad. Sci. U. S. A., 107, 7184

Bridle A. H., Perley R. A., 1984, Annu. Rev. Astron. Astrophys., 22, 319

Broderick A. E., Johannsen T., Loeb A., Psaltis D., 2014, ApJ, 784, 7

Bromm V., Loeb A., 2003, ApJ, 596, 34

Bromm V., Coppi P. S., Larson R. B., 1999, ApJ, 527, L5

Bromm V., Yoshida N., Hernquist L., 2003, ApJ, 596, L135

Brook C. B., Stinson G., Gibson B. K., Shen S., Maccio A. V., Obreja A., Wadsley J., Quinn T., 2014, MNRAS, 443, 3809

Bullock J. S., Dekel A., Kolatt T. S., Kravtsov A. V., Klypin A. A., Porciani C., Primack J. R., 2001, ApJ, 555, 240

Buta R. J., 2013, in , Planets, Stars Stellar Syst.. p. 189

Cano-Díaz M., Maiolino R., Marconi A., Netzer H., Shemmer O., Cresci G., 2012, A&A, 537, L8

Capelo P. R., Volonteri M., Dotti M., Bellovary J. M., Mayer L., Governato F., 2015, MNRAS, 447, 2123

Cappellari M., 2016, Annu. Rev. Astron. Astrophys., 54, 597

Cardamone C. N., Megan Urry C., Schawinski K., Treister E., Brammer G., Gawiser E., 2010, ApJ, 721, L38

Carniani S., et al., 2016, A&A, 591, A28

Caroll B., Ostlie D., 2008, An Introduction to Modern Astrophysics, 3nd edn

Carrasco E. R., Conselice C. J., Trujillo I., 2010, MNRAS, 405, 2253

Cattaneo A., 2001, MNRAS, 324, 128

Cattaneo A., Teyssier R., 2007, MNRAS, 376, 1547

Cattaneo A., et al., 2009, Nature, 460, 213

Cen R., Riquelme M. A., 2008, ApJ, 674, 644

Ceverino D., Dekel A., Tweed D., Primack J., 2015, MNRAS, 447, 3291

Ceverino D., Glover S. C. O., Klessen R. S., 2017, MNRAS, 470, 2791

Chabrier G., 2003, Publ. Astron. Soc. Pacific, 115, 763

Chapon D., Mayer L., Teyssier R., 2013, MNRAS, 429, 3114

Chartas G., Brandt W. N., Gallagher S. C., Proga D., 2007, Astron. J., 133, 1849

Cheung E., et al., 2016, Nature, 533, 504

Chisari N. E., et al., 2017, MNRAS, 472, 1163

Choi J.-H., Shlosman I., Begelman M. C., 2013, ApJ, 774, 149

Choi J.-H., Shlosman I., Begelman M. C., 2015, MNRAS, 450, 4411

Christensen C. R., Davé R., Governato F., Pontzen A., Brooks A., Munshi F., Quinn T., Wadsley J., 2016, ApJ, 824, 57

Cicone C., et al., 2014, A&A, 562, A21

Ciotti L., Ostriker J. P., 2007, ApJ, 665, 1038

Cole S., Aragon-Salamanca A., Frenk C. S., Navarro J. F., Zepf S. E., 1994, MNRAS, 271, 781

Colpi M., Callegari S., Dotti M., Kazantzidis S., Mayer L., 2007. pp 705–714

Conselice C. J., 2014, Annu. Rev. Astron. Astrophys., 52, 291

Costa T., Sijacki D., Trenti M., Haehnelt M. G., 2014a, MNRAS, 439, 2146

Costa T., Sijacki D., Haehnelt M. G., 2014b, MNRAS, 444, 2355

Costa T., Sijacki D., Haehnelt M. G., 2015, MNRASL, 448, L30

Cowie L. L., 1977, MNRAS, 180, 491

Croton D. J., et al., 2006, MNRAS, 365, 11

Curtis M., Sijacki D., 2015, MNRAS, 454, 3445

Curtis M., Sijacki D., 2016a, MNRASL, 457, L34

Curtis M., Sijacki D., 2016b, MNRAS, 463, 63

D'Odorico V., et al., 2016, MNRAS, 463, 2690

Daddi E., et al., 2002, A&A, 384, L1

Davidzon I., et al., 2017, A&A, 605, A70

Davies M. B., Coleman Miller M., Bellovary J. M., 2011, ApJ, 740, L42

De Rosa G., et al., 2014, ApJ, 790, 145

DeBuhr J., Quataert E., Ma C.-P., 2010, MNRAS, 412, 1341

DeGraf C., Di Matteo T., Khandai N., Croft R., 2012, ApJ, 755, L8

DeGraf C., Di Matteo T., Treu T., Feng Y., Woo J.-H., Park D., 2014, MNRAS, 932, 17

DeGraf C., Dekel A., Gabor J., Bournaud F., 2017, MNRAS, 466, 1462

Decarli R., Falomo R., Treves A., Labita M., Kotilainen J. K., Scarpa R., 2010, MNRAS, 402, 2453

Dekel A., Birnboim Y., 2006, MNRAS, 368, 2

Dekel A., Burkert A., 2014, MNRAS, 438, 1870

Dekel a., Silk J., 1986, ApJ, 303, 39

Delvecchio I., et al., 2014, MNRAS, 439, 2736

Devecchi B., Volonteri M., 2009, ApJ, 694, 302

Di Matteo T., Springel V., Hernquist L. E., 2005, Nature, 433, 604

Di Matteo T., Colberg J., Springel V., Hernquist L., Sijacki D., 2008, ApJ, 676, 33

Di Matteo T., Khandai N., DeGraf C., Feng Y., Croft R. A. C., Lopez J., Springel V., 2012, ApJ, 745, L29

Dijkstra M., Ferrara A., Mesinger A., 2014, MNRAS, 442, 2036

Doeleman S. S., et al., 2008, Nature, 455, 78

Dokuchaev V., 1964, Sov. Astron., 9, 23

Dressler A., 1980, ApJ, 236, 351

Du P., et al., 2015, ApJ, 806, 22

Dubois Y., Teyssier R., 2008, A&A, 477, 79

Dubois Y., Devriendt J., Slyz A., Teyssier R., 2010, MNRAS, 409, 985

Dubois Y., Devriendt J., Slyz A., Teyssier R., 2012a, MNRAS, 420, 2662

Dubois Y., Pichon C., Haehnelt M., Kimm T., Slyz A., Devriendt J., Pogosyan D., 2012b, MNRAS, 423, 3616

Dubois Y., Pichon C., Devriendt J., Silk J., Haehnelt M., Kimm T., Slyz A., 2013a, MNRAS, 428, 2885

Dubois Y., Gavazzi R., Peirani S., Silk J., 2013b, MNRAS, 433, 3297

Dubois Y., Volonteri M., Silk J., 2014a, MNRAS, 440, 1590

Dubois Y., et al., 2014b, MNRAS, 444, 1453

Dubois Y., Volonteri M., Silk J., Devriendt J., Slyz A., Teyssier R., 2015, MNRAS, 452, 1502

Dunn J. P., et al., 2010, ApJ, 709, 611

Dutton A. A., Maccio A. V., 2014, MNRAS, 441, 3359

Edgar R., 2004, New Astron. Rev., 48, 843

Efstathiou G., 2000, MNRAS, 317, 697

Eggen O. J., Lynden-Bell D., Sandage A. R., 1962, ApJ, 136, 748

Einasto J., 1965, Tr. Astrofiz. Instituta Alma-Ata, 5, 87

Einstein A., 1916, Ann. Phys., 354, 769

El Mellah I., Casse F., 2015, MNRAS, 454, 2657

Elling V., 2009, Acta Math. Sci., 29, 1647

Emsellem E., Renaud F., Bournaud F., Elmegreen B., Combes F., Gabor J. M., 2014, MNRAS, 446, 2468

Everett J. E., Zweibel E. G., Benjamin R. A., McCammon D., Rocks L., Gallagher J. S., 2008, ApJ, 674, 258

Fabian A., 2012, Annu. Rev. Astron. Astrophys., 50, 455

Fabian A. C., Sanders J. S., Taylor G. B., Allen S. W., Crawford C. S., Johnstone R. M., Iwasawa K., 2006a, MNRAS, 366, 417

Fabian A. C., Celotti A., Erlund M. C., 2006b, MNRASL, 373, L16

Fall S. M., 1979, Nature, 281, 200

Fan X., et al., 2001, Astron. J., 122, 2833

Fan X., et al., 2004, Astron. J., 128, 515

Fan X., et al., 2006, Astron. J., 131, 1203

Farrah D., et al., 2012, ApJ, 745, 178

Feldmann R., Mayer L., 2014, MNRAS, 446, 1939

Feldmann R., Carollo C. M., Mayer L., Renzini A., Lake G., Quinn T., Stinson G. S., Yepes G., 2010, ApJ, 709, 218

Fender R., Belloni T., 2012, Science, 337, 540

Ferrara A., Salvadori S., Yue B., Schleicher D., 2014, MNRAS, 443, 2410

Ferrarese L., Merritt D., 2000, ApJ, 539, L9

Foglizzo T., 2009, ApJ, 694, 820

Foglizzo T., Galletti P., Ruffert M., 2005, A&A, 435, 397

Forman W., et al., 2007, ApJ, 665, 1057

Förster Schreiber N. M., et al., 2014, ApJ, 787, 38

Fraley G. S., 1968, Astrophys. Space Sci., 2, 96

Franx M., van Dokkum P. G., Schreiber N. M. F., Wuyts S., Labbé I., Toft S., 2008, ApJ, 688, 770

Gabor J. M., Bournaud F., 2014, MNRAS, 441, 1615

Gabor J. M., Davé R., 2012, MNRAS, 427, 1816

Gabor J. M., Capelo P. R., Volonteri M., Bournaud F., Bellovary J. M., Governato F., Quinn T. R., 2015, A&A, 592, A62

Gaibler V., Khochfar S., Krause M., Silk J., 2012, MNRAS, 425, 438

Galametz A., et al., 2009, ApJ, 694, 1309

Galli D., Palla F., 2013, Annu. Rev. Astron. Astrophys., 51, 163

Gan Z., Yuan F., Ostriker J. P., Ciotti L., Novak G., 2014, ApJ, 789, 150

Gardner J. P., et al., 2006, Space Sci. Rev., 123, 485

Gaspari M., Ruszkowski M., Oh S. P., 2013, MNRAS, 432, 3401

Gebhardt K., et al., 2000, ApJ, 539, L13

Geen S., Slyz A., Devriendt J., 2013, MNRAS, 429, 633

Genel S., et al., 2014, MNRAS, 445, 175

Genzel R., Eisenhauer F., Gillessen S., 2010, Rev. Mod. Phys., 82, 3121

Genzel R., et al., 2014, ApJ, 796, 7

Gillessen S., Genzel R., Eisenhauer F., Ott T., Trippe S., Martins F., 2007, Proc. Int. Astron. Union, 3, 466

Gillessen S., Eisenhauer F., Trippe S., Alexander T., Genzel R., Martins F., Ott T., 2009, ApJ, 692, 1075

Girardi L., Bressan A., Bertelli G., Chiosi C., 2000, Astron. Astrophys. Suppl. Ser., 141, 371

Girichidis P., et al., 2016, ApJ, 816, L19

Glover S., 2005, Space Sci. Rev., 117, 445

Glover S. C. O., 2015, MNRAS, 451, 2082

Gnedin N. Y., 2000, ApJ, 542, 535

Godunov S. K., 1962, Russ. Math. Surv., 17, 145

Goicovic F. G., Sesana A., Cuadra J., Stasyszyn F., 2017, MNRAS, 472, 514

Gong H., Ostriker E. C., 2013, ApJS, 204, 8

González V., Labbé I., Bouwens R. J., Illingworth G., Franx M., Kriek M., 2011, ApJ, 735, L34

Governato F., Willman B., Mayer L., Brooks A., Stinson G., Valenzuela O., Wadsley J., Quinn T., 2007, MNRAS, 374, 1479

Grassi T., Bovino S., Schleicher D. R. G., Prieto J., Seifried D., Simoncini E., Gianturco F. A., 2014, MNRAS, 439, 2386

Greene J. E., 2012, Nat. Commun., 3, 1304

Greif T. H., Johnson J. L., Klessen R. S., Bromm V., 2008, MNRAS, 387, 1021

Greif T. H., Glover S. C. O., Bromm V., Klessen R. S., 2010, ApJ, 716, 510

Guillet T., Teyssier R., 2011, J. Comput. Phys., 230, 4756

Gültekin K., Cackett E. M., King A. L., Miller J. M., Pinkney J., 2014, ApJ, 788, L22

Guo Q., White S., Li C., Boylan-Kolchin M., 2010, MNRAS, 404, 1111

Haardt F., Madau P., 1996, ApJ, 461, 20

Habouzit M., Volonteri M., Latif M., Dubois Y., Peirani S., 2016, MNRAS, 463, 529

Habouzit M., Volonteri M., Dubois Y., 2017, MNRAS, 468, 3935

Hahn O., Abel T., 2011, MNRAS, 415, 2101

Haiman Z., Loeb A., 2001, ApJ, 552, 459

Harrison C. M., Alexander D. M., Mullaney J. R., Swinbank A. M., 2014, MNRAS, 441, 3306

Harten A., 1983, J. Comput. Phys., 49, 357

Hasinger G., Miyaji T., Schmidt M., 2005, A&A, 441, 417

Heckman T. T. M., Best P. N. P., 2014, Annu. Rev. Astron. Astrophys., 52, 589

Heckman T. M., Kauffmann G., 2006, New Astron. Rev., 50, 677

Heckman T. M., Thompson T. A., 2016, A Brief Galactic Winds and the Role Played by Massive Stars. Springer International Publishing, Cham, pp 1–24

Heckman T. M., Lehnert M. D., Strickland D. K., Armus L., 2000, ApJS, 129, 493

Heger A., Woosley S., Baraffe I., Abel T., 2001, in , Lighthouses of the Universe: The Most Luminous Celestial Objects and Their Use for Cosmology. Springer-Verlag, Berlin/Heidelberg, pp 369–375

Hilbert D., 1891, Math. Ann., 38, 30

Hobbs A., Nayakshin S., Power C., King A., 2011, MNRAS, 413, 2633

Hopkins P. F., Quataert E., 2011, MNRAS, 415, 1027

Hopkins P. F., Bundy K., Murray N., Quataert E., Lauer T. R., Ma C.-P., 2009, MNRAS, 398, 898

Hopkins P. F., Kereš D., Oñorbe J., Faucher-Giguère C.-A., Quataert E., Murray N., Bullock J. S., 2014, MNRAS, 445, 581

Hoyle F., Lyttleton R. a., 1939, Math. Proc. Cambridge Philos. Soc., 35, 405

Hu C., Wang J., Ho L. C., Chen Y., Zhang H., Bian W., Xue S., 2008, ApJ, 687, 78

Inoue Y., Doi A., Tanaka Y. T., Sikora M., Madejski G. M., 2017, ApJ, 840, 46

Jiang L., et al., 2016, ApJ, 833, 222

Jiao C.-L., Mineshige S., Takeuchi S., Ohsuga K., 2015, ApJ, 806, 93

Johnson J. L., Bromm V., 2007, MNRAS, 374, 1557

Johnson J. L., Whalen D. J., Li H., Holz D. E., 2013, ApJ, 771, 116

Just A., Kegel W., 1990, A&A, 232, 447466

Kara E., Miller J. M., Reynolds C., Dai L., 2016, Nature, 535, 388

Katz N., Quinn T., Bertschinger E., Gelb J. M., 1994, MNRAS, 270, L71

Kauffmann G., White S. D. M., Guiderdoni B., 1993, MNRAS, 264, 201

Kaviraj S., et al., 2017, MNRAS, 467, 4739

Keller B. W., Wadsley J. W., 2017, ApJ, 835, L17

Keller B. W., Wadsley J., Couchman H. M. P., 2015, MNRAS, 453, 3500

Kennicutt, Jr. R. C., 1998, ApJ, 498, 541

Kent S. M., 1986, Astron. J., 91, 1301

Kerr R. P., 1963, Phys. Rev. Lett., 11, 237

Khandai N., Di Matteo T., Croft R., Wilkins S., Feng Y., Tucker E., DeGraf C., Liu M.-S., 2015, MNRAS, 450, 1349

Khokhlov A., 1998, J. Comput. Phys., 143, 519

Kim H., Kim W.-T., 2009, ApJ, 703, 1278

Kimm T., Cen R., 2014, ApJ, 788, 121

Kimm T., Cen R., Devriendt J., Dubois Y., Slyz A., 2015, MNRAS, 451, 2900

King A., 2003, ApJ, 596, L27

King A., Pounds K., 2015, Annu. Rev. Astron. Astrophys., 53, 115

Kitayama T., Yoshida N., 2005, ApJ, 630, 675

Koide S., 2002, Science, 295, 1688

Kollmeier J. A., et al., 2006, ApJ, 648, 128

Komatsu E., et al., 2011, ApJS, 192, 18

Kormendy J., Bender R., 2012, ApJS, 198, 2

Kormendy J., Freeman K. C., 2004, Astronomy, p. 19

Kormendy J., Ho L. C., 2013, Annu. Rev. Astron. Astrophys., 51, 511

Krumholz M. R., Tan J. C., 2007, ApJ, 654, 304

Krumholz M. R., McKee C. F., Klein R. I., 2004, ApJ, 611, 399

Lanzuisi G., et al., 2017, A&A, 602, A123

Latif M. A., Ferrara A., 2016, Publ. Astron. Soc. Aust., 33, e051

Latif M. A., Schleicher D. R. G., Schmidt W., Niemeyer J., 2013a, MNRAS, 430, 588

Latif M. A., Schleicher D. R. G., Schmidt W., Niemeyer J., 2013b, MNRAS, 433, 1607

Lauer T. R., et al., 2007, ApJ, 662, 808

Leitherer C., Robert C., Drissen L., 1992, ApJ, 401, 596

Leitherer C., et al., 1999, ApJS, 123, 3

Leitherer C., Ortiz Otálvaro P. A., Bresolin F., Kudritzki R.-P., Lo Faro B., Pauldrach A. W. A., Pettini M., Rix S. A., 2010, ApJS, 189, 309

Levine R., Gnedin N. Y., Hamilton A. J. S., Kravtsov A. V., 2008, ApJ, 678, 154

Levine R., Gnedin N. Y., Hamilton A. J. S., 2010, ApJ, 716, 1386

Lora-Clavijo F. D., Guzman F. S., 2013, MNRAS, 429, 3144

Lu R.-S., Broderick A. E., Baron F., Monnier J. D., Fish V. L., Doeleman S. S., Pankratius V., 2014, ApJ, 788, 120

Lukić Z., Heitmann K., Habib S., Bashinsky S., Ricker P. M., 2007, ApJ, 671, 1160

Lupi A., Colpi M., Devecchi B., Galanti G., Volonteri M., 2014, MNRAS, 442, 3616

Lupi A., Haardt F., Dotti M., 2015, MNRAS, 446, 1765

Lupi A., Haardt F., Dotti M., Fiacconi D., Mayer L., Madau P., 2016, MNRAS, 456, 2993

Lutz D., et al., 2008, ApJ, 684, 853

Mac Low M.-M., Ferrara A., 1999, ApJ, 513, 142

MacArthur L. A., Courteau S., Bell E., Holtzman J. A., 2004, ApJS, 152, 175

MacLeod M., Ramirez-Ruiz E., 2015, ApJ, 803, 41

Magorrian J., et al., 1998, Astron. J., 115, 2285

Maio U., Dolag K., Ciardi B., Tornatore L., 2007, MNRAS, 379, 963

Maiolino R., et al., 2012, MNRASL, 425, L66

Mandelbaum R., Wang W., Zu Y., White S., Henriques B., More S., 2016, MNRAS, 457, 3200

Mannucci F., Cresci G., Maiolino R., Marconi A., Gnerucci A., 2010, MNRAS, 408, 2115

Marconi A., Hunt L. K., 2003, ApJ, 589, L21

Matteucci F., Greggio L., 1986, A&A, 154, 279

McConnell N. J., Ma C.-P., 2013, ApJ, 764, 184

McConnell N. J., Ma C.-P., Gebhardt K., Wright S. A., Murphy J. D., Lauer T. R., Graham J. R., Richstone D. O., 2011, Nature, 480, 215

McGreer I. D., et al., 2013, ApJ, 768, 105

McKee C. F., Ostriker E. C., 2007, Annu. Rev. Astron. Astrophys., 45, 565

McNamara B., Nulsen P., 2007, Annu. Rev. Astron. Astrophys., 45, 117

McNamara B. R., Kazemzadeh F., Rafferty D. A., Bîrzan L., Nulsen P. E. J., Kirkpatrick C. C., Wise M. W., 2009, ApJ, 698, 594

Merloni A., Heinz S., Di Matteo T., 2003, MNRAS, 345, 1057

Merloni A., et al., 2010, ApJ, 708, 137

Miller L., 1988, Proc. NATO Adv. Res. Work., 229, 205

Minniti D., et al., 2015, ApJ, 810, L20

Misner C. W., Thorne K. S., Wheeler J. A., 1973, Gravitation. W. H. Freeman and Company

Moe M., Arav N., Bautista M. A., Korista K. T., 2009, ApJ, 706, 525

Morgan C. W., Kochanek C. S., Morgan N. D., Falco E. E., 2010, ApJ, 712, 1129

Morganti R., Holt J., Tadhunter C., Oosterloo T., 2009, Proc. Int. Astron. Union, 5, 429

Mortlock D. J., et al., 2011, Nature, 474, 616

Moss D., Stepanov R., Arshakian T. G., Beck R., Krause M., Sokoloff D., 2012, A&A, 537, A68

Moster B. P., Naab T., White S. D. M., 2013, MNRAS, 428, 3121

Moustakas J., et al., 2013, ApJ, 767, 50

Murray N., Quataert E., Thompson T. a., 2005, ApJ, 618, 569

Naab T., Ostriker J. P., 2017, Annu. Rev. Astron. Astrophys., 55

Nakamura F., Umemura M., 2001, ApJ, 548, 19

Nandra K., et al., 2007, ApJ, 660, L11

Narayan R., Yi I., 1994, ApJ, 428, L13

Nardini E., Risaliti G., Salvati M., Sani E., Imanishi M., Marconi A., Maiolino R., 2008, MNRASL, 385, L130

Navarro J. F., White S. D. M., 1994, MNRAS, 267, 401

Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493

Negri A., Volonteri M., 2017, MNRAS, 467, 3475

Nelson D., Genel S., Vogelsberger M., Springel V., Sijacki D., Torrey P., Hernquist L., 2015, MNRAS, 448, 59

Nixon C. J., Cossins P. J., King A. R., Pringle J. E., 2011, MNRAS, 412, 1591

Nomoto K., Saio H., Kato M., Hachisu I., 2007, ApJ, 663, 1269

Okamoto T., Gao L., Theuns T., 2008, MNRAS, 390, 920

Omma H., Binney J., Bryan G., Slyz A., 2004, MNRAS, 348, 1105

Omukai K., Tsuribe T., Schneider R., Ferrara A., 2005, ApJ, 626, 627

Oppenheimer J. R., Snyder H., 1939, Phys. Rev., 56, 455

Oppenheimer B. D., Davé R., Kereš D., Fardal M., Katz N., Kollmeier J. A., Weinberg D. H., 2010, MNRAS, 406, 2325

Oser L., Ostriker J. P., Naab T., Johansson P. H., Burkert A., 2010, ApJ, 725, 2312

Ostriker E. C., 1999, ApJ, 513, 252

Pacucci F., Ferrara A., 2015, MNRAS, 448, 104

Page M. J., et al., 2012, Nature, 485, 213

Paragi Z., Frey S., Kaaret P., Cseh D., Overzier R., Kharb P., 2014, ApJ, 791, 2

Parker M. L., et al., 2017, Nature, 543, 83

Pasquali A., Nachname V., 2015, Astron. Nachrichten, 336, 505

Peano G., 1980, Math. Ann., 36, 156

Pedlar A., Ghataure H., Davies R. D., Harrison B., Perley R., Crane P., Unger S., 1990, MNRAS, 246, 447

Peery K., Imlay S., 1988, in 24th Jt. Propuls. Conf.. Reston, Virigina

Peirani S., et al., 2017, MNRAS, 472, 2153

Pelupessy F. I., Di Matteo T., Ciardi B., 2007, ApJ, 665, 107

Peng Y.-j., et al., 2010, ApJ, 721, 193

Penna R. F., 2014, Phys. Rev. D - Part. Fields, Gravit. Cosmol., 89, 104057

Penzias A. A., Wilson R. W., 1965, ApJ, 142, 419

Pimbblet K. A., Smail I., Kodama T., Couch W. J., Edge A. C., Zabludoff A. I., O'Hely E., 2002, MNRAS, 331, 333

Pogorelov N. V., Ohsugi Y., Matsuda T., 2000, MNRAS, 313, 198

Pontzen A., Tremmel M., Roth N., Peiris H. V., Saintonge A., Volonteri M., Quinn T., Governato F., 2017, MNRAS, 465, 547

Powell L. C., Slyz A., Devriendt J., 2011, MNRAS, 414, 3671

Power C., Navarro J. F., Jenkins A., Frenk C. S., White S. D. M., Springel V., Stadel J., Quinn T., 2003, MNRAS, 338, 14

Power C., Nayakshin S., King A., 2011, MNRAS, 412, 269

Press W. H., Schechter P., 1974, ApJ, 187, 425

Prunet S., Pichon C., Aubert D., Pogosyan D., Teyssier R., Gottloeber S., 2008, ApJS, 178, 179

Quintero A. D., et al., 2004, ApJ, 602, 190

Quirk J. J., 1994, Int. J. Numer. Methods Fluids, 18, 555

Rafferty D. A., McNamara B. R., Nulsen P. E. J., Wise M. W., 2006, ApJ, 652, 216

Rafferty D. A., Brandt W. N., Alexander D. M., Xue Y. Q., Bauer F. E., Lehmer B. D., Luo B., Papovich C., 2011, ApJ, 742, 3

Rasera Y., Teyssier R., 2006, A&A, 445, 1

Recchia S., Blasi P., Morlino G., 2017, MNRAS, 470, 865

Rees M. J., Ostriker J. P., 1977, MNRAS, 179, 541

Reeves J. N., O'Brien P. T., Ward M. J., 2003, ApJ, 593, L65

Reeves J. N., et al., 2009, ApJ, 701, 493

Regan J. A., Johansson P. H., Haehnelt M. G., 2014a, MNRAS, 439, 1160

Regan J. A., Johansson P. H., Wise J. H., 2014b, ApJ, 795, 137

Regan J. A., Visbal E., Wise J. H., Haiman Z., Johansson P. H., Bryan G. L., 2017, Nat. Astron., 1, 0075

Rephaeli Y., Salpeter E. E., 1980, ApJ, 240, 20

Rodriguez-Gomez V., et al., 2016, MNRAS, 458, 2371

Roos O., Juneau S., Bournaud F., Gabor J. M., 2015, ApJ, 800, 19

Rosario D., et al., 2012, A&A, 545, A45

Rosario D. J., et al., 2013, ApJ, 771, 63

Rosas-Guevara Y. M., et al., 2015, MNRAS, 454, 1038

Rosas-Guevara Y., Bower R. G., Schaye J., McAlpine S., Dalla Vecchia C., Frenk C. S., Schaller M., Theuns T., 2016, MNRAS, 462, 190

Rosdahl J., Schaye J., Dubois Y., Kimm T., Teyssier R., 2017, MNRAS, 466, 11

Ruderman M. A., Spiegel E. A., 1971, ApJ, 165, 1

Ruffert M., 1995a, A&A, 113, 133

Ruffert M., 1995b, A&A, 331, 817

Ruschel-Dutra D., Rodríguez Espinosa J. M., González Martín O., Pastoriza M., Riffel R., 2016, MNRAS, 9

Sądowski A., Gaspari M., 2017, MNRAS, 468, 1398

Saez C., Chartas G., Brandt W. N., 2009, ApJ, 697, 194

Salem M., Bryan G. L., 2014, MNRAS, 437, 3312

Salpeter E. E., 1955, ApJ, 121, 161

Sánchez-Blázquez P., et al., 2014, A&A, 570, A6

Santini P., et al., 2012, A&A, 540, A109

Sbarrato T., Padovani P., Ghisellini G., 2014, MNRAS, 445, 81

Schawinski K., Thomas D., Sarzi M., Maraston C., Kaviraj S., Joo S.-J., Yi S. K., Silk J., 2007, MNRAS, 382, 1415

Schawinski K., et al., 2014, MNRAS, 440, 889

Schaye J., et al., 2014, MNRAS, 446, 521

Schleicher D. R. G., Spaans M., Glover S. C. O., 2010, ApJ, 712, L69

Schmidt M., 1959, ApJ, 129, 243

Schober J., Schleicher D. R. G., Klessen R. S., 2013, A&A, 560, A87

Schreiber C., Elbaz D., Pannella M., Ciesla L., Wang T., Koekemoer A., Rafelski M., Daddi E., 2016, A&A, 589, A35

Schwarzschild K., 1916, Sitzungsberichte der Königlich Preuss. Akad. der Wissenschaften zu Berlin,, Phys.-Math, 424

Sesana A., Volonteri M., Haardt F., 2007, MNRAS, 377, 1711

Shakura N. I., Sunyaev R. A., 1973, A&A, 24, 337

Shang C., Bryan G. L., Haiman Z., 2010, MNRAS, 402, 1249

Shankar F., et al., 2009, ApJ, 690, 20

Shapiro S. L., 2005, ApJ, 620, 59

Shaw L. D., Weller J., Ostriker J. P., Bode P., 2006, ApJ, 646, 815

Shen Y., Greene J. E., Strauss M. A., Richards G. T., Schneider D. P., 2008, ApJ, 680, 169

Shima E., Matsuda T., Takeda H., Sawada K., 1985, MNRAS, 217, 367

Shlosman I., Choi J.-H., Begelman M. C., Nagamine K., 2016, MNRAS, 456, 500

Sijacki D., Springel V., Di Matteo T., Hernquist L., 2007, MNRAS, 380, 877

Sijacki D., Vogelsberger M., Genel S., Springel V., Torrey P., Snyder G. F., Nelson D., Hernquist L., 2015, MNRAS, 452, 575

Silk J., 2005, MNRAS, 364, 1337

Simmons B. D., Urry C. M., Schawinski K., Cardamone C., Glikman E., 2012, ApJ, 761, 75

Simmons B. D., et al., 2013, MNRAS, 429, 2199

Smethurst R. J., et al., 2016, MNRAS, 463, 2986

Smethurst R. J., Lintott C. J., Bamford S. P., Hart R. E., Kruk S. J., Masters K. L., Nichol R. C., Simmons B. D., 2017, MNRAS, 469, 3670

Smidt J., Whalen D. J., Johnson J. L., Li H., 2013, in , The First Galaxies. pp 177–222

Soker N., 1990, Astron. J., 99, 1869

Somerville R. S., 2002, ApJ, 572, L23

Song M., et al., 2016, ApJ, 825, 5

Springel V., Hernquist L., 2003, MNRAS, 339, 289

Springel V., Di Matteo T., Hernquist L., 2005, MNRAS, 361, 776

Sturm E., et al., 2011, ApJ, 733, L16

Sutherland R. S., Dopita M. A., 1993, ApJS, 88, 253

Tanaka T., Haiman Z., 2009, ApJ, 696, 1798

Taylor P., Kobayashi C., 2014, MNRAS, 442, 2751

Taylor P., Kobayashi C., 2015, MNRASL, 452, L59

Teyssier R., 2002, A&A, 385, 337

Teyssier R., 2013, Ramses bitbucket repository, `https://bitbucket.org/rteyssie/ramses`

Teyssier R., Moore B., Martizzi D., Dubois Y., Mayer L., 2011, MNRAS, 414, 195

Thomas D., Maraston C., Bender R., de Oliveira C. M., 2005, ApJ, 621, 673

Tombesi F., Cappi M., Reeves J. N., Palumbo G. G. C., Yaqoob T., Braito V., Dadina M., 2010, A&A, 521, A57

Tomczak A. R., et al., 2014, ApJ, 783, 85

Tomczak A. R., et al., 2016, ApJ, 817, 118

Toro E. F., Spruce M., Speares W., 1994, Shock Waves, 4, 25

Trakhtenbrot B., Lira P., Netzer H., Cicone C., Maiolino R., Shemmer O., 2017, ApJ, 836, 8

Tremonti C. a., Moustakas J., Diamond-Stanic A. M., 2007, ApJ, 663, L77

Truelove J. K., Klein R. I., McKee C. F., Holliman II J. H., Howell L. H., Greenough J. a., 1997, ApJ, 489, L179

Turk M. J., Abel T., O'Shea B., 2009, Science, 325, 601

Tweed D., Devriendt J., Blaizot J., Colombi S., Slyz A., 2009, A&A, 506, 647

Urry C. M., 2003, in AGN Phys. with SDSS. p. 49

Van Wassenhove S., Volonteri M., Walker M. G., Gair J. R., 2010, MNRAS, 408, 1139

Vassiliadis E., Wood P. R., 1993, ApJ, 413, 641

Veilleux S., Cecil G., Bland-Hawthorn J., 2005, Annu. Rev. Astron. Astrophys., 43, 769

Vogelsberger M., Genel S., Sijacki D., Torrey P., Springel V., Hernquist L., 2013, MNRAS, 436, 3031

Vogelsberger M., et al., 2014a, MNRAS, 444, 1518

Vogelsberger M., et al., 2014b, Nature, 509, 177

Volonteri M., 2010, Annu. Rev. Astron. Astrophys., 18, 279

Volonteri M., Bellovary J., 2012, Reports Prog. Phys., 75, 124901

Volonteri M., Rees M. J., 2005, ApJ, 633, 624

Volonteri M., Stark D. P., 2011, MNRAS, 417, 2085

Volonteri M., Sikora M., Lasota J.-P., Merloni A., 2013, ApJ, 775, 94

Volonteri M., Dubois Y., Pichon C., Devriendt J., 2016, MNRAS, 460, 2979

Weil M. L., Eke V. R., Efstathiou G., 1998, MNRAS, 300, 773

Weiner B. J., et al., 2009, ApJ, 692, 187

Werner N., Urban O., Simionescu A., Allen S. W., 2013, Nature, 502, 656

Whalen D. J., Fryer C. L., 2012, ApJ, 756, L19

Whalen D., van Veelen B., O'Shea B. W., Norman M. L., 2008, ApJ, 682, 49

Whalen D., Hueckstaedt R. M., McConkie T. O., 2010, ApJ, 712, 101

White S. D. M., Frenk C. S., 1991, ApJ, 379, 52

White S. D. M., Rees M. J., 1978, MNRAS, 183, 341

Williams R. J., Maiolino R., Krongold Y., Carniani S., Cresci G., Mannucci F., Marconi A., 2017, MNRAS, 467, 3399

Yang X., Mo H. J., van den Bosch F. C., Pasquali A., Li C., Barden M., 2007, ApJ, 671, 153

Yoshida N., Omukai K., Hernquist L., 2008, Science, 321, 669

Zolotov A., et al., 2015, MNRAS, 450, 2327

Zubovas K., Bourne M. A., 2017, MNRAS, 468, 4956

Zuo W., Wu X.-B., Fan X., Green R., Wang R., Bian F., 2015, ApJ, 799, 189 de Blok W. J. G., 2010, Adv. Astron., 2010, 1

de Kool M., Arav N., Becker R. H., Gregg M. D., White R. L., Laurent-Muehleisen S. A., Price T., Korista K. T., 2001, ApJ, 548, 609

van Dokkum P. G., et al., 2010, ApJ, 709, 1018