

## Supermassive Black Hole Feeding in Galactic Nuclei

## Alexander Paul Hobbs

MA(Oxon) MSc MRes

Submitted in fulfilment of the requirements of the degree of Doctor of Philosophy

September 30th, 2010

Theoretical Astrophysics Group Department of Physics and Astronomy University of Leicester

### Abstract

# Supermassive Black Hole Feeding in Galactic Nuclei

#### by

## Alexander Paul Hobbs

MA(Oxon) MSc MRes

In this thesis we present numerical and analytical models of supermassive black hole (SMBH) feeding, via deposition of gas, in galactic nuclei. Through simulations, we consider the environment of galactic centres, starting at sub-parsec scales within our own Milky Way, and moving upwards in scale and outwards in generality to scales of hundreds of parsecs in typical galaxies and finally to dark matter halos within which galaxies reside. We find that the stellar features observed in our own Galactic centre are likely explained by a collision between two molecular clouds at a distance of a few parsecs from the central black hole, Sgr A<sup>\*</sup>. The amount of gas transported to small radii is large, occurring on a timescale close to dynamical. The disordered nature of the flow leads to the formation of a gaseous disc around Sgr A<sup>\*</sup> that in some cases remains small-scale, undergoing complex, time-varying evolution in its orientation. Such a disc would efficiently feed the SMBH, if replenished from larger scales.

We develop a model for ballistic accretion onto an SMBH at the centre of a typical galaxy, from scales of ~ hundreds of parsecs. We invoke turbulence in the gas, assumed to be driven by feedback from supernovae, as the means to create such a flow. The accretion mode is again dominated, soon after the initial turbulent kick, by the dynamical timescale for the gas in the angular momentum loss-cone, resulting in an accretion rate at or near Eddington,  $\gtrsim 1 \, M_{\odot} \, yr^{-1}$ .

At the largest scale, we critically evaluate the current state-of-the-art prescription for SMBH growth in cosmological simulations, finding that in general it lacks a physically consistent basis. We propose an alternative, motivated by our analytical estimates and numerical simulations, that is based on the free-fall time.

## Acknowledgements

The successful completion of this thesis is the culmination of many years of work but also many years of friendships, love and support, without which it may never have happened. I would therefore like to thank everyone who has been a valued part of my life along the way. I mention some of these people below, but it is by no means an exhaustive list; if you are reading this and are not on it but have contributed even a small amount to my (positive) experiences over the last several years then consider yourself thanked!

To start with I would like to say thank you to my fellow batcave residents for many amusing, productive and unproductive times spent both in the office and in the pub. To Peter Cossins (for his innate ability to fix absolutely anything, be it mechanical or mathematical), Patrick Deegan (for a daily dose of northerness), Jalpesh Sachania (for 'dealing' sci-fi and thesis stylefiles), Lee Cullen (for being the SPH guru), Fergus Wilson (for calling a spade a spade, never fixing something if it wasn't broken, and mutual back scratching), Chris "definitely not useless" Nixon (for his many photogenic facebook friends and for making my sense of humour seem gentlemanly by comparison), Dave Cole, Kastytis Zubovas, and Fabrizio Pedes.

My thanks to the postdocs (honorary batcave/batcorridor residents) - Chris "Powerlaw" Power (to whom I owe my sanity and a large part of my success in staying in academia; in particular for his encouragment and advice - "take a step back" - and many hours spent proof reading job applications), Celine Combet (for being French, being female, and for not suing me), Andreas Koch and Seung-Hoon Cha; and to all the other members of the theory group, past and present, for providing an intellectually stimulating atmosphere during my time at Leicester - Richard Alexander, Martin Beer, Jim Dale, Walter Dehnen, Andrew King, Giuseppe Lodato, Dean McLaughlin, Justin Read, Ugur Ural, Mark Wilkinson and Graham Wynn. To the IT support team - Gary Gilchrist, Chris Rudge, Stuart Poulton among others; and finally a special mention to Lisa Brant, for her fantastic administrative skills and superb organisation of multiple 'jollies' to lands afar!

To the members of XROA/SRC for making tea-times worth venturing out of the theory cave for - thank you to Amy Scott, Richard Owen, Owen Littlejohns, Katherine Lawrie, Victoria Heard, Lucy Heil, Charly Feldman, James Duke, Chris Cottis and Dave Baker.

A huge thank you to my supervisor, Sergei Nayakshin, for all his help and advice over the course of my PhD, and for putting up with what must have seemed like substantial ignorance on my part for the first couple of years!

On the non-academic side, thank you to the Leicester university dancesport team, for providing a much needed distraction and helping to keep me \*relatively\* same when the geekdom of an astrophysics PhD threatened to destroy what little 'street-cred' I had left. To every member of the team that I have had the pleasure of teaching and/or dancing alongside over the last four years, thank you - it has been one of the most rewarding experiences of my life. A special mention to Amy Cory, for being easily impressed, helping me to become more enthusiastic about my research, and (somehow!) persuading me to take part in the Leicester offbeat 2010 - \*one\* of the highlights of my dancing career.

To my housemates, Ally Joyce and Damian Maguire, I owe a great deal of thanks for all the enjoyable (and often informative) times spent discussing the more philosophical points of theoretical physics, as well as the inner workings of the human body to a somewhat alarming level of detail. Thanks in particular to Ally for putting up with my untidyness and bizarre working hours!

A heartfelt thank you to my dance partner, Tamara Thompson, for some of the best times that I have ever had; for all the successful competitions, demos, and enjoyable practice sessions and in particular for being beautiful, wonderful and incredibly talented.

My sincere thanks and appreciation to my parents for their love and unwavering support; my Mum, for all her hard work and sheer dedication to give me the opportunities that have got me where I am today - I am eternally grateful. My Dad, for always being there to offer advice and for being willing to drop everything on a moment's notice if I needed him.

My thanks to my brother, Chris, for his support and assistance throughout my time at university, both academic and monetary - it was much appreciated!

Finally I would like to say a special thank you to Charlotte ("Clive") Woolley, to whom this thesis is dedicated, and without whom I would never have believed in myself enough to make the plunge into the world of academia. For much needed encouragement, inspiration, and understanding while asking for nothing in return - from the bottom of my heart, thank you.

# Contents

A	Abstract				
$\mathbf{A}$	ckno	wledge	ements	ii	
1	Intr	oduct	ion	1	
	1.1	Prolog	gue	2	
	1.2	Times	scales	5	
	1.3	Our G	Balactic centre	7	
	1.4	Evide	nce for supermassive black holes (SMBHs)	9	
		1.4.1	Stellar vs. hydrodynamical evidence	10	
		1.4.2	Stellar proper motions in the Milky Way	11	
		1.4.3	Stellar rotation curve in M31	12	
		1.4.4	Water maser emission in NGC 4258	13	
		1.4.5	Other evidence	14	
	1.5	Active	e galactic nuclei (AGN)	17	
		1.5.1	Taxonomy of AGN	18	
		1.5.2	Unification of AGN	21	
		1.5.3	Outflows from AGN	23	
	1.6	The g	rowth of SMBHs	25	
		1.6.1	SMBH growth through gas accretion	26	
		1.6.2	SMBHs in the early Universe	27	
	1.7	Feedir	ng SMBHs	27	
		1.7.1	Feeding through an accretion disc	28	
			1.7.1.1 Onset of self-gravity $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$	29	
			1.7.1.2 Randomly-oriented accretion episodes $\ldots \ldots \ldots \ldots$	29	
		1.7.2	Feeding through a bar instability	30	
			1.7.2.1 Orbital resonances $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$	31	
		1.7.3	Feeding through a collisional stellar cluster	33	
	1.8	The S	MBH/host galaxy connection	35	
		1.8.1	$M_{\rm bh} - \sigma$ relation	35	
		1.8.2	$M_{\rm bh} - M_{\rm bulge}$ relation $\ldots \ldots \ldots$	37	
		1.8.3	The starburst-AGN connection	38	

<b>2</b>	Theoretical background							
	2.1	Theore	etical astrophysics	42				
	2.2	2 Black holes						
		2.2.1	Formation of a black hole	43				
		2.2.2	Structure of a black hole	43				
		2.2.3	Particle dynamics in the vicinity of a black hole	44				
			2.2.3.1 Newtonian approximation $\ldots \ldots \ldots \ldots \ldots \ldots \ldots$	44				
			2.2.3.2 General relativistic approach	46				
	2.3	Accret	ion onto compact objects	50				
		2.3.1	The Eddington limit	51				
		2.3.2	Spherically symmetric accretion	53				
		2.3.3	Bondi-Hoyle accretion	62				
		2.3.4	Accretion discs	63				
			2.3.4.1 Basic equations $\ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots$	64				
			2.3.4.2 Radial disc components	66				
			2.3.4.3 Vertical disc components	67				
			2.3.4.4 Azimuthal disc components	70				
			2.3.4.5 Evolution of the surface density	73				
			2.3.4.6 Accretion disc viscosity	73				
			2.3.4.7 Steady-state solutions	75				
			2.3.4.8 Gravitational instability and fragmentation	79				
3	Nui	umerical methods						
	3.1	Numerical astrophysics						
	3.2	2 Smoothed particle hydrodynamics (SPH)						
	-	3.2.1 Choice of smoothing kernel						
		3.2.2	'Gather' vs. 'scatter' approach	91				
		3.2.3 Fluid equations						
		3.2.3.1 Conservation of mass						
		3.2.3.2 Conservation of momentum						
		3.2.3.3 Conservation of energy	96					
3.2.3.4 Equation of state			3.2.3.4 Equation of state	97				
		Adaptive smoothing lengths	97					
		3.2.4.1 Fully conservative formulation	98					
		Dissipation and shock capturing	100					
		3251 Artificial viscosity						
	3.3	3 Gravity solvers						
	0.00	3.3.1 Force calculation						
			3.3.1.1 Direct summation	102				

			3.3.1.2 Particle-mesh						
			3.3.1.3 Tree algorithm $\ldots \ldots 103$						
			3.3.1.4 falcON						
		3.3.2	Gravitational softening						
	3.4	Time i	ntegration $\ldots \ldots \ldots$						
		3.4.1	Euler method						
		3.4.2	Leapfrog integrators						
			3.4.2.1 Drift-Kick-Drift (DKD)						
			3.4.2.2 Kick-Drift-Kick (KDK)						
		3.4.3	Timestepping						
			3.4.3.1 Adaptive timesteps $\ldots \ldots 110$						
	3.5	GADC	ET						
		3.5.1	Collisionless dynamics						
		3.5.2	Hydrodynamics						
		3.5.3	Self-gravity						
		3.5.4	Time evolution						
4	CL	1 . 11	· · · · · · · · · · · · · · · · · · ·						
4	<b>UI0</b>	ua con Introd	collisions in the Galactic centre 116						
	4.1	Numor	$\frac{120}{120}$						
	4.2	1 9 1	Cooling proscription						
		4.2.1	Cravitational potential						
		4.2.2	Star formation and SMBH accretion using sink particles						
	13	4.2.0 Initial	conditions						
	4.0	Rosult	120						
	4.4		$C_{\text{as dynamics}} \qquad $						
		1.1.1	$4411  \text{Bulk dynamics} \qquad \qquad 129$						
			4.4.1.2 Redistribution of angular momentum $134$						
		449	Star formation 130						
		1.1.2	4421 Star formation histories 139						
			4.4.2.1 Star formation instortes						
			4.4.2.2 Considered statisfication of stars 143						
		4.4.3	Inner disc						
		1.1.0	4 4 3 1 Ring instability 145						
			4 4 3 2 Midplane rotation 146						
		444	Feeding Sgr A* 148						
		4.4.5	Summary of results						
	4.5	Discus	sion						
	2.0	4.5.1	Comparison to the Galactic centre						

			4.5.1.1 Estimate of the cooling time during the cloud collision 152
		4.5.2	Alternative models
		4.5.3	Convergence tests
	4.6	Conclu	usion
<b>5</b>	Feed	ding S	MBHs through supersonic turbulence 160
	5.1	Introd	uction
	5.2	Comp	utational method $\ldots$
	5.3	Initial	conditions $\ldots \ldots 165$
		5.3.1	Turbulent velocity field
	5.4	Overv	iew of main results $\ldots \ldots \ldots$
		5.4.1	No initial turbulence
		5.4.2	Initial turbulence
		5.4.3	Turbulence and accretion $\ldots \ldots 173$
	5.5	"Lami	nar" accretion without turbulence
		5.5.1	Analytical estimates
		5.5.2	Shock mixing of gas
	5.6	Efficie	nt accretion with turbulence $\ldots \ldots 176$
		5.6.1	Analytical expectations
			5.6.1.1 Weak turbulence $\ldots \ldots 176$
			5.6.1.2 Strong turbulence: ballistic accretion
		5.6.2	Detailed analysis of results
			5.6.2.1 No net angular momentum case $\ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots 183$
			5.6.2.2 The highest rotation velocity case
			5.6.2.3 Rotation and turbulence parameter space
			5.6.2.4 Accretion and the size of the shell $\ldots \ldots \ldots \ldots \ldots \ldots \ldots 186$
			5.6.2.5 Trend with accretion radius
			5.6.2.6 Visualisation of the circumnuclear disc $\ldots \ldots \ldots \ldots \ldots 187$
			5.6.2.7 Radial structure of the disc $\ldots \ldots 187$
			5.6.2.8 Vertical structure
	5.7	Accret	ion inside the capture radius
	5.8	Discus	sion
		5.8.1	Summary of the results
		5.8.2	Feeding through ballistic accretion
		5.8.3	A star formation link to AGN?
		5.8.4	Cosmological cold stream SMBH feeding?
		5.8.5	Convergence tests
	5.9	Conclu	usions

6	On the modelling of accretion on to SMBHs 20					
	6.1 Introduction					
	6.2	Analy	tical arguments	204		
		6.2.1	Classical Bondi-Hoyle accretion	204		
		6.2.2	When is Bondi-Hoyle accretion applicable?	205		
	6.3	The so	onic point	207		
			6.3.0.1 Power law profile	207		
			6.3.0.2 NFW profile	209		
			6.3.0.3 Hernquist profile	209		
			6.3.0.4 Jaffe profile	210		
	6.4	Nume	rical tests	213		
		6.4.1	Initial conditions	213		
		6.4.2	Results	213		
			6.4.2.1 Free-fall rate	214		
	6.5	Discus	ssion and future work	214		
7	Dise	cussion	and conclusion	217		
	7.1 Summary					
	7.2	Future	e work	221		
$\mathbf{A}$	Stability of particle orbits in Kerr spacetime 224					
В	Der	ivatior	n of gravitational potential for cloud collisions in the GC	227		
С	C Derivation of tidal shear density for a BH 23					
D	Der	ivatior	n of Bremsstrahlung cooling function	233		
Re	efere	nces		<b>250</b>		

# List of Figures

1.1	Looking towards the Galactic centre with the VLA
1.2	Adaptive optics imaging of the S-stars around Sgr A* $\hfill\hfil$
1.3	Image and illustration of M31 and its core
1.4	Using maser emission to determine the rotation curve of NGC4258 $\ldots$ 14
1.5	AGN unification scheme
1.6	A model for AGN outflows
1.7	High resolution HST image of the barred spiral galaxy NGC 1300 31
1.8	The $M - \sigma$ relation
1.9	The merger of the Antennae galaxies
2.1	Effective potential of a particle in Newtonian gravity
2.2	Effective potential of a particle in a Schwarzschild geometry 48
2.3	Mach number as a function of dimensionless radius for Bondi accretion 60
2.4	Dispersion relation for axisymmetric perturbations in a thin disc 83
3.1	Gather and scatter approaches to SPH interpolation
3.2	Recursive subdivision of a quad-tree
3.3	Constructing the tree
3.4	Walking the tree
3.5	Schematic of the leapfrog algorithm over 3 timesteps
3.6	Individual timestep scheme for KDK algorithm
3.7	Two steps in the subdivision of an oct-tree
4.1	The young stars in the central parsec of the Galactic centre
4.2	Gas cloud collapsing to form a first core
4.3	Schematic of the initial conditions for simulation S1
4.4	Gas surface density and stellar locations in S1
4.5	Projected gas densities and stellar positions for S1 and S2
4.6	Projected gas density and stars for the inner two arcseconds of S1 and S2 $$ . 133 $$
4.7	Comparison of projected gas densities for S1 and S3
4.8	Projected gas density and stars for S5 and S6
4.9	Distribution of SPH particles over circularisation radius for S1 135
4.10	Gas eccentricity defined on radial shells
4.11	Angular momentum direction of gas orbits

4.12	Distribution of SPH particles over circularisation radius for S2 $\ldots$ .	140
4.13	Angular momentum orientations for gas and stars in S1 and S2	140
4.14	Projected gas density and stars for simulations S3 and S4	142
4.15	Angular momentum distribution of gas and stars for S3 and S4 $\ldots$	142
4.16	Stellar surface density vs. projected radius for S1, S2, S3 & S5 $\ldots \ldots$	144
4.17	Inner 1.5 arcseconds of the simulation S4 $\ldots \ldots \ldots \ldots \ldots \ldots \ldots$	147
4.18	Accretion rate on to the SMBH for S1, S2, S3 & S4	148
4.19	Convergence tests on artificial viscosity and resolution for S1 $\ldots \ldots \ldots$	157
5.1	Edge-on and top-down projections of gas flow in S30	171
5.2	Projected gas column densitiy in S35	172
5.3	Edge-on and top-down projections of gas flow in S35 $\ldots$	173
5.4	Accretion rate versus time for simulations S30–S37 $\ldots \ldots \ldots \ldots \ldots$	174
5.5	Distribution of gas over $l_z$ in S30 at three different times	177
5.6	Geometry of a rotating, isotropic wind $\ldots \ldots \ldots \ldots \ldots \ldots \ldots$ .	180
5.7	Accreted mass and accretion rate vs. time for zero net rotation $\hdots$	184
5.8	Accreted mass and accretion rate vs. time for high net rotation	185
5.9	Accreted mass trend with rotation velocity $\hfill \ldots \hfill \ldots \hf$	186
5.10	Scaling of accreted mass with $r_{out}$ and $r_{acc} \ldots \ldots \ldots \ldots \ldots \ldots \ldots$	187
5.11	Face-on projection of the disc in S30 and S35 $\ldots$	188
5.12	Slices of the CND in S30 and S35	188
5.13	Surface density vs. radius for S50-S57	190
5.14	Vertical disc thickness vs. radius for S50-S57	192
5.15	Accreted angular momentum direction for S30 and S37	193
5.16	Projected gas density and velocity field for inner region of S37 $\ldots \ldots \ldots$	194
5.17	Convergence tests for the viscosity parameter	200
6.1	Potential energies for a selection of dark matter halos	208
6.2	The sonic radius for a variety of halo profiles $\ldots \ldots \ldots \ldots \ldots \ldots \ldots$	211
6.3	Effect of the adiabatic index on the sonic point for an NFW halo $\hdots$	212
6.4	Mass flux through radial shells from the simulations	215
C.1	Supporting diagram for derivation of tidal shear density	231

# List of Tables

1.1	AGN types and their properties	18
4.1	Initial conditions for colliding cloud simulations	129
5.1	Initial conditions for turbulent feeding simulations	166

Dedicated to Charlotte

# Introduction

"Physicists like to think that all you have to do is say, these are the conditions, now what happens next?"

Richard Feynman

#### 1.1 Prologue

The formation, structure, and evolution of galaxies is one of the most widely studied fields in astrophysics. Galaxies constitute the largest building blocks of the known universe, and since the mid-eighteenth century have captivated astronomers and laypeople alike. The philosopher Immanuel Kant was the first to reason that the faint diffuse patches of light seen against the backdrop of stars and comets were in fact 'island universes', similar to our own Milky Way but viewed at incredibly large distances; some 250 years on, we know a great deal about these objects and their constituent elements, although a great deal more remains to be discovered.

Their formation is thought to have occurred as a consequence of the density perturbations in the early Universe. Quantum fluctuations, imprinted in the fabric of spacetime at the Big Bang, were amplified to much larger scales by inflation (Peebles, 1993), providing the seeds for structure formation through gravitational collapse. Baryonic matter began to condense within dark matter overdensities, forming the first stars and protogalaxies that re-ionized the Universe. Many of the galaxies that we now observe grew rapidly in the early Universe through mergers and accretion of smaller mass structures.

The driving force for galaxy formation is generally accepted to be gravitational attraction from dark matter (see, e.g., Padmanabhan, 2002), but galactic evolution on smaller scales is a far more complex process, involving (but not limited to) gas processes, feedback processes, star formation, and magnetic fields. Understanding how this complexity affects the ontogenesis<sup>1</sup> of galaxies is vital for the development of reliable, self-consistent theoretical models. The role of galaxy formation research is to connect the large-scale gravitational processes that form galaxies to the smaller scale, 'gastrophysical' processes that determine their evolution; this is not an easy job. The picture is further complicated by the variety of components that a galaxy can possess; over a century of observation has revealed the presence (see, e.g., Binney and Merrifield, 1998) of thin stellar discs (with or without spiral arms), thick stellar discs, thin gaseous discs, bars, rings, central stellar bulges, dark matter halos, stellar clusters, and, more recently, supermassive black holes. Some of these features are relatively passive, contributing little to the continued evolution of the galaxy; some, however, are important drivers of this evolution that determine a galaxy's fate within the wider cosmology.

One such driver is the supermassive black hole (SMBH). The largest type of black hole known, SMBHs boast a mass of the order of  $10^6 - 10^9 \,\mathrm{M_{\odot}}$ . The formation mechanism for such extreme objects is currently under debate (for a review see Volonteri, 2010) but it is likely that the black hole was born at a significantly lower mass - formation through

<sup>&</sup>lt;sup>1</sup> from the greek word *génesis*, for *birth* and the present participle *ont*-, for *existence*; namely, *formation* and *evolution* 

direct stellar collapse suggests a maximum limit<sup>2</sup> of  $\sim 100 \,\mathrm{M}_{\odot}$  (Madau and Rees, 2001) and grew through accretion of surrounding material to its eponymous size. SMBHs are found in the nuclei of galaxies, where they are surrounded by dense gas and fast-moving stars that allow us to infer both their presence and their mass. It is thought that most, if not all, galaxies harbor an SMBH at their centre (Melia, 2007).

The presence of the black hole creates an exotic environment within the galactic nucleus. The strong tidal shear of the SMBH's gravity tends to rip apart gas clouds and stars that venture too close, and anything that is unfortunate enough to pass into the event horizon is consumed (Melia, 2003). The gaseous structures that are able to survive are therefore very dense, and in fast rotation around the central object. Molecular clouds, the sites of star formation in a galaxy, are generally unable to survive in this neighbourhood<sup>3</sup>, and so the 'normal' star formation process (see, e.g., McKee and Ostriker, 2007a) is prevented from occurring. Instead, stars are thought to form in flat, rotationally-supported accretion discs, in which fast cooling competes against strong viscous heating processes to fragment the discs into high-mass, short-lived stars (Nayakshin and Sunyaev, 2005; Bonnell and Rice, 2008) that burn brightly and drive powerful stellar winds, often exploding as supernovae. The inner parts of the accretion disc around the SMBH are heated to high temperatures through intense friction, becoming ionized and generating strong magnetic fields that may accelerate jets of material to relativistic speeds.

Galactic centres are therefore an exciting area of study. From an observational perspective they are incredible powerhouses of activity, often outshining the rest of the galaxy put together (see Section 1.5). Furthermore, in many galaxies the mass of the central black hole correlates extremely well with a number of properties of the surrounding bulge (see Section 1.8) which implies a fundamental connection between the SMBH and the evolution of the host galaxy. In order to understand galaxy formation we must therefore understand galactic nuclei.

The aim of this thesis is to investigate the means by which galactic nuclei and SMBHs are 'fed' by material from larger scales. One of the currently unsolved problems in astronomy and astrophysics is how SMBHs grew through accretion of gas to the masses that we observe. It does not seem feasible, given the standard paradigms about black holes and the manner in which they accrete matter, that SMBHs could have formed in the early Universe, only a billion years after the Big Bang. We discuss this problem and proffer a suggestion for a tentative solution. More generically, we discuss the modelling of galactic centres and SMBH feeding in numerical simulations, starting first with the central parsec

 $<sup>^{2}</sup>$ but see Volonteri et al. (2008) and Mayer et al. (2010) for an alternative viewpoint concerning seed formation from an early phase of gaseous accretion at the centre of a forming galaxy, driven by mergers and instabilities

 $<sup>^3</sup>a$  typical molecular cloud has a mean density of  $10^3-10^4$  atoms per cubic centimeter (Williams et al., 2000)

of our own Galactic centre and then moving upwards in scale, investigating accretion models in galaxies at hundreds of parsecs and finally within entire dark matter halos.

Numerical simulations provide us with the only practical means of 'experimenting' with the Universe. The limitations of computational resources, however, mean that it is not possible to model all of the required physics over the full range of scales. This is particularly true in the case of simulations on the largest, *cosmological* scale, which seek to track the hiearchical formation of structure from seed perturbations in the early Universe. Numerically modelling a cosmological volume suffers acutely from the problem of finding a sub-resolution (or 'sub-grid') link between gas accretion at large (megaparsec) scales and the eventual accretion, at small (sub-parsec) scales, on to the SMBH. At present there is a strong dichotomy between detailed study of these small-scale regions, where the relevant structures are often modelled using idealised geometries without establishing a definitive model for their formation, and large-scale simulations, where gas and feedback processes below the resolution limit are treated with simplistic prescriptions that fail to capture the correct physical behaviour of the system. It is of vital importance that the correct physical processes on 'intermediate' scales are explored, as it is these processes that will strongly influence the mode of accretion within a galactic nucleus.

The modelling of intermediate scales is a focus of this work. Exploring the physics of accretion at these scales will lead to the development of better, more physically selfconsistent sub-grid models that can be embedded into cosmological simulations. At the same time, these accretion models may help to clarify the means by which SMBHs grow and influence the evolution of their host nuclei.

The layout of this thesis is as follows. In Chapter 1 (the current chapter) we introduce the relevant topics for galactic centres and SMBH feeding that equip the reader to place the subsequent science chapters in their proper context. In Chapter 2 (the 'maths' chapter) we outline the theoretical concepts that underpin this work. Chapter 3 describes the computational method that was used to carry out the simulations presented in this thesis. Chapter 4 is the first science chapter, in which we consider a progenitor model for the stellar features in our very own galactic centre, as well as the effect of this model on the feeding of our galaxy's SMBH. In Chapter 5 we discuss a potential SMBH feeding mechanism through supernova-driven turbulence at scales of ~ 100 pc, marking a first step towards an improved intermediate scale model for cosmological simulations. In Chapter 6 we discuss modelling SMBH feeding in these type of simulations, critically evaluating the current state-of-the-art and suggesting ways in which it can be improved. Finally in Chapter 7 we present our conclusions and suggest some directions for future work.

#### 1.2 Timescales

Physical processes in the Universe tend to occur on characteristic timescales. Knowledge of the relevant timescales for a particular system not only helps us to simplify the analysis of that system (for example, if a quantity changes over a timescale far longer than the effect we are interested in, we can neglect this change and still trust that our results are reliable) but can also shed light on the physical behaviour of that system. Here we list some timescales that are of importance in astrophysics.

- The Hubble time,  $t_{\text{Hubble}}$ , is a timescale derived from Hubble's law (Hubble, 1929), which states that the velocity at which a galaxy is receding from us is proportional to its distance. The law can be expressed mathematically by  $v = H_0 D$ , where v is the velocity and D is the distance of the galaxy, and  $H_0$  is known as the Hubble constant connecting the two. This constant has units of time<sup>-1</sup>, and so it is straightforward to define the Hubble time as  $1/H_0$ . Physically it represents the *e*-folding time of the expansion of the Universe, if  $H_0$  were to have remained constant since the big bang (which it has not). Nonetheless, it is often used as synonymous with the age of the Universe, a reasonable approximation since  $t_{\text{Hubble}} \sim 10^{10}$  years.
- The free-fall time,  $t_{\rm ff}$ , reflects the timescale over which a change in one part of a body can be communicated to the rest of that body. It can be defined as the time taken for a uniform density spherical cloud, i.e., with  $\rho(r) = \text{const.}$ , to collapse, and is given by the expression

$$t_{\rm ff} = \sqrt{\frac{3\pi}{32G\rho}} \tag{1.1}$$

where G is the gravitational constant.

• The orbital time,  $t_{orb}$ , is the time for a particle at radius r to complete one circular Keplerian orbit of an enclosed mass M. It is given by

$$t_{\rm orb} = 2\pi \sqrt{\frac{r^3}{GM}} \tag{1.2}$$

• The dynamical time,  $t_{\rm dyn}$ , is the time taken for a particle to cross a particular system of mass M and length r. The typical velocity is assumed to be the circular velocity,  $v_{\rm circ} = (GM/r)^{1/2}$ , so that

$$t_{\rm dyn} = \frac{r}{v_{\rm circ}} = \sqrt{\frac{r^3}{GM}}$$
(1.3)

which we can see is simply the orbital time without the numerical factor of  $2\pi$ . The dynamical time is also sometimes referred to as the *crossing time*.

• The Salpeter time,  $t_{\rm S}$ , is the timescale over which a compact object (be it a star,

white dwarf, neutron star, or black hole) can grow if accretion is assumed to proceed at the Eddington limit (see Chapter 2, Section 2.3.1). It has the form

$$t_S = M/\dot{M}_{\rm Edd} \tag{1.4}$$

where M is the mass of the object and  $\dot{M}_{\rm Edd}$  is the Eddington limited accretion rate.

• The viscous time,  $t_{\rm visc}$ , is a relevant quantity for accretion discs that we shall derive in Chapter 2, Section 2.3.4. It describes the timescale over which material can be transported through a disc under the action of viscous torques. It is given by the expression

$$t_{\rm visc} = \frac{R^2}{\nu} \tag{1.5}$$

where R is the cylindrical radius and  $\nu$  is the kinematic viscosity.

• The relaxation time,  $t_{\rm relax}$ , is a measure of the extent to which a particle is moving not in a smooth background potential but in one that is composed of discrete elements, namely, the other particles in the system. By mutual interactions each particle is perturbed slightly from the path that it would follow in a smooth potential. The relaxation time is defined as the time over which these accumulated (gravitational) interactions, on average, alter the trajectory of an object by 45°. This is equivalent to saying that the average change in velocity is equal to the velocity itself. It given via the approximate expression (Bertin, 2000):

$$t_{\rm relax} \simeq \frac{v_{\rm typ}^3}{8\pi G^2 m^2 n \ln \Lambda} \tag{1.6}$$

where  $v_{\text{typ}}$  is the typical velocity of a particle, n is the number density of particles of mass m and  $\ln \Lambda$  is the Coulomb logarithm, given by  $\ln \Lambda = \ln(b_{\text{max}}/b_{\text{min}})$ .  $b_{\text{max}}$  defines the largest impact parameter in a system, while  $b_{\text{min}}$  defines the smallest. The Coulomb logarithm is therefore set by the dynamic range of a particular system; it is interesting to notice that its logarithmic form means that even large changes in this dynamic range make little difference to its value. For the majority of galactic and stellar systems in the Universe it is reasonable to assume  $\ln \Lambda \sim 10$  (Binney and Tremaine, 1987).

The relaxation time divides self-gravitating systems into two categories: collisionless, when  $t_{\text{relax}} \gg$  the lifetime of the system,  $t_{\text{system}}$ , and the mutual interactions of the components can be approximated by a smooth potential; and collisional, when  $t_{\text{relax}} \ll t_{\text{system}}$  and the smoothness of the potential breaks down. Knowing which category a region of interest falls into allows one to decide on the best way to model that particular system.

#### **1.3** Our Galactic centre

In order to study SMBHs and galactic nuclei, we first turn to our own Galactic centre (GC). The rotational centre of the Milky Way is a mere 8 kiloparsecs from our solar system (Eisenhauer et al., 2003), with a resolution of 1 arcsecond corresponding to only 0.04 parsecs ( $\approx 1.2 \times 10^{17}$  cm), offering us an unparalleled opportunity to probe its highly dynamic environment. As we are firmly embedded in the galactic disc, with our (optical) view blocked by large amounts of intervening matter, we must make use of alternative wavelengths to image the central regions. Radio waves in particular are largely unaffected by the obscuring dust particles and so the majority of the evidence we have for the various structures in the GC has come from images produced by radio telescopes such as the Very Large Array (VLA) in New Mexico (Melia, 2003). Figure 1 shows such an image; the galactic plane stretching across the sky, taken with a wavelength of  $\lambda = 1$  m.

The panorama of activity is clearly seen in the centre of Figure 1.1, coinciding with the direction of the centre of the galaxy. We can zoom in to smaller scales by using radio waves of progressively decreasing wavelength, together with measurements in the infra-red, X-ray, and  $\gamma$ -ray regions of the spectrum to resolve as much structure as possible. Many interesting features are seen on the journey into the Galactic centre; as we progress from scales of hundreds of light years to sub-light year regions we pass by: Sgr A East, a large expanding shell of gas from a powerful supernova-like explosion (~ 25 ly across); Sgr A West<sup>4</sup>, a spiral feature of hot, ionized gas (~ 10 ly across) surrounded by a circumnuclear disc of shocked, molecular gas; IRS 16, a cluster of about 2 dozen OB stars<sup>5</sup> (within the central parsec) emitting a powerful stellar wind; IRS 7, a red giant star with a cometary tail believed to have formed from the effect of the wind from the IRS 16 cluster; a cusp of old stars exerting a spherical gravitational potential on the objects around it; and finally Sgr A<sup>\*</sup>, a radio source that coincides with the central SMBH, surrounded by a multitude of fast-moving stars arranged into two disc-like structures within the central parsec<sup>6</sup>.

Even closer to Sgr A<sup>\*</sup> (within the central arcsecond, i.e., < 0.04 pc) spectroscopic observations have been made of isotropically oriented, early main-sequence B-type stars (Ghez et al., 2003a; Eisenhauer et al., 2005) with approximate masses of  $M \sim 15 M_{\odot}$ . Some of these stars pass close enough to the central black hole for extremely accurate mass determinations of the latter, and will also provide astronomers with the means to look for general relativistic effects in the vicinity of the SMBH, perhaps for the first time yielding a measure of a black hole's spin.

Of course, the identification of an SMBH at the centre of the Galaxy should not be made so lightly. Many decades of research have gone into this determination (see, e.g.

<sup>&</sup>lt;sup>4</sup>otherwise known as the minispiral

 $<sup>^5{\</sup>rm massive}$  stars with  $T\sim 30,000~{\rm K}$ 

 $<sup>^{6}</sup>$ The evidence for all this comes from a multitude of sources; for a more detailed review and primary references the reader is referred to Melia and Falcke (2001)



Figure 1.1: Looking towards the Galactic centre with the VLA, at a wavelength of 1 m. The galactic plane runs diagonally through the image, showing the radio emission from hot gas and supernovae remnants. Image courtesy of N. E. Kassim, D. S. Briggs, T. J. W. Lazio, T. N. LaRosa, J. Imamura (NRL/RSD).

Lynden-Bell and Rees, 1971; Ghez et al., 1998; Genzel, 2010), and indeed it is of no small importance to the science in this thesis that it is the correct one. In fact the ubiquity of SMBHs in galactic nuclei is a underlying assumption of this work, but it is only relatively recently that it has become a mainstream assumption. With this in mind, the next section reviews some of the main lines of evidence for SMBHs.

#### 1.4 Evidence for supermassive black holes (SMBHs)

The paradigm of galactic nuclei harboring SMBHs (Lynden-Bell, 1969) emerged out of the discovery of quasars (Schmidt, 1963) and subsequent development in the theory of accretion on to compact objects as being the only viable mechanism for powering such phenomena (Zel'Dovich, 1964; Salpeter, 1964). Since this time the case for SMBHs has become increasingly watertight; with the development of high-resolution instruments such as the Hubble Space Telescope (HST) and the adaptive optics (AO) facilities of the Keck Observatory and the Very Large Telescope (VLT) (Albrecht et al., 1998) in the last decade we have been able to image the central regions of galaxies like never before. In this section we discuss some of the ways in which the presence of an SMBH in a galactic nucleus can be inferred, and list a few of the galaxies for which these techniques have been employed.

The presence of an SMBH at the centre of a galaxy will have an effect on the surrounding matter. This can be seen in a variety of ways, from dynamical effects to the high-energy radio lobes ejected by active galactic nuclei (AGN). Before embarking on a quick tour of these ideas we must mention a small caveat: all of the evidence for SMBHs is in actuality merely evidence of a large mass, or in some cases merely a large luminosity, confined within a small volume; it does not provide absolute proof that what is contained there is a black hole. Indeed, black holes are somewhat extreme objects and so invoking them to explain a particular set of observations should be treated with caution. The standard argument against the conclusion of a black hole is that the enclosed mass is in the form of a cluster of stellar remnants (for of course, if it was a cluster of luminous stars, we would see them). However, this model requires an unreasonably high density of objects (Maoz, 1998); packing the required number of brown dwarfs, white dwarfs, or neutron stars into such a small region almost certainly means that the stellar remnants will quickly either merge into luminous high mass stars (in the case of brown dwarfs), merge and explode as supernovae (in the case of white dwarfs), or collide and explode as gamma-ray bursts (in the case of neutron stars).

In three cases - the nuclei of M31, NGC 4258 and our own Milky Way - the constrained volume is small enough that such a 'dark' cluster of stars can be ruled out and an SMBH seen as the only viable candidate (Kormendy, 2004; Ghez et al., 2005a). For a number of other galaxies the case is less convincing but still strong. Moreover, the perceived link

between SMBHs and AGN adds credence to this idea; as was demonstrated by Soltan (1982), there is a clear correlation between the total mass density of supposed SMBHs and the mass-equivalent energy density of the quasar light that fills the Universe. We will discuss this argument further in Section 1.7.

#### 1.4.1 Stellar vs. hydrodynamical evidence

In general, when angular resolution is low, evidence for an SMBH is stronger when it comes from observations of gas motions. Since interstellar gas behaves as a fluid, and is able to easily radiate away energy and cool, the random speeds of the atoms/molecules thus tend to be small in comparison to the bulk velocity of the gas. Furthermore, the velocity distribution function of a gas is isotropic, a consequence of the extremely small relaxation time of a collisional system on molecular scales<sup>7</sup>. This means that for a smooth (non-clumpy) gaseous feature we can characterise the motion of the bulk gas flow entirely by its position, bulk velocity, density and temperature.

As we shall discuss in Chapter 2, Section 2.3.4, it is generally difficult for gas in a spherical potential to lose its angular momentum with respect to the centre of the galaxy/halo. A smooth gaseous structure orbiting around a large central mass therefore tends to be flattened (as in the case of an accretion disc), particularly when the gas is cold. By analysing the rotation curve of such a gaseous disc it is possible to tell whether the gravitational potential is dominated by a central, massive object and infer what the mass of this object might be, assuming of course that the gravitational forces are dominant over gas processes such as shocks, magnetic fields, etc. In this case the rotation speed of material in orbit of a (Keplerian) point mass potential follows a radial profile of  $(GM/r)^{1/2}$ .

When the evidence comes from observation of stellar motions, however, it is harder to construct a rotation curve. Unlike gas, stars cannot radiate<sup>8</sup> their kinetic energy through dissipative processes. Random velocities are therefore very high, and as a result a mean velocity measurement is not a useful indicator of the presence of a central mass, unlike in the gaseous case. The stellar velocity dispersion can tell us some information about the enclosed mass through the virial theorem<sup>9</sup> - for a spherical distribution of mass Mwithin a radius r this comes out to first order<sup>10</sup> as  $\sigma^2 \simeq GM/r$  - and has indeed been used to infer the masses of central black holes across a range of scales, from globular clusters

$$2K + V = 0 \tag{1.7}$$

where K and V are the kinetic and potential energies of the particles

 $^{10}$  there are, in fact, more accurate methods of determining the enclosed mass; namely, by taking moments of the *Jeans equation* - see Binney and Tremaine (1987)

<sup>&</sup>lt;sup>7</sup>note that here the term 'collisional' refers to direct hydrodynamical encounters, rather than gravitational encounters

 $<sup>^{8}</sup>a$  quote from my advisor; we mean of course in a dynamical sense

<sup>&</sup>lt;sup>9</sup>the virial theorem for a system of gravitationally interacting particles states that



Figure 1.2: Adaptive optics imaging of the S-stars (left) and schematic of the orbit of S2 around Sgr A\* (right). The measurements indicated on this diagram only go up to 2002; data collected since then completes the orbit. Figure credit: Schödel et al. (2002) and the European Southern Observatory

(Baumgardt et al., 2009) to giant galaxies (Magorrian and Ballantyne, 2001). However this is not a particularly strong constraint, as there is some degeneracy between a central, Keplerian mass and an extended mass distribution.

On the other hand, if individual stars near an SMBH can be resolved, and their orbits followed (through proper motions and radial velocity measurements), then such evidence can be very convincing, as is the case for our own Galactic centre (see, e.g., Genzel, 2010).

#### 1.4.2 Stellar proper motions in the Milky Way

Studies of the S-stars (refer back to Section 1.3) in the Galactic centre have quite recently provided us with almost incontrovertible proof that our Galaxy does indeed contain a supermassive black hole. The proper motions<sup>11</sup> of S2, a star which comes as close as 17 light-hours to Sgr A\* at pericentre (Ghez et al., 2003b) have now been measured for nearly two decades using the adaptive optics facilities of the Keck telescope (Campbell et al., 2008). Combined with the radial velocities (Genzel et al., 1996), S2 has, as of 2008, been observed to complete one full orbit (Gillessen et al., 2009b). As shown in Figure 1.2, this (Keplerian) orbit is elliptical, with a period of 15.2 years, which puts the mass of the central object<sup>12</sup> at  $(4.3 \pm 0.41) \times 10^6 M_{\odot}$ . Moreover, the volume within which it must

<sup>&</sup>lt;sup>11</sup>the angular change in position as seen from the centre of mass of the Solar system

<sup>&</sup>lt;sup>12</sup>note that the error bars on this measurement originate from the uncertainty in the distance to the Galactic centre (and therefore in converting angular changes in position to actual distances) and the



Figure 1.3: Image and illustration of M31 and its core, showing the double nucleus structure and the compact disc of blue stars in orbit of the suspected SMBH. Figure credit: in image

be contained is actually somewhat smaller than the pericentre passage of S2, as there is another member of the S-stars cluster that passes closer; S16, which has been observed to pass within  $\simeq 6$  light-hours ( $\simeq 45$  A.U.) without showing any evidence of tidal disruption or accretion (Ghez et al., 2005b). There is currently no known object that can contain the required amount of mass within this volume, other than an SMBH.

#### 1.4.3 Stellar rotation curve in M31

The Andromeda (M31) galaxy is the nearest galaxy of approximate Milky Way size, and displays the somewhat unusual feature of possessing a double nucleus (as can be seen in Figure 1.3, top right panel). The currently accepted explanation for this is that the double nucleus is in fact a single eccentric disc of old stars, the brighter object in the image

uncertainty in our reference frame (the 'local standard of rest', see e.g., Coskunoglu et al. (2010)). There is also some uncertainty in the assumption of the orbit as Keplerian (Gillessen et al., 2009a)

corresponding to the apocentre region of the disc and the fainter object being synonymous with the pericentre region (Tremaine, 1995). Supporting this idea is the HST discovery of a smaller-scale cluster of hot, young stars (Chang et al., 2007) that appear to lie inside the fainter region (which coincides with the true centre of the galaxy) the inference being that this 'blue' cluster is inside the inner radius of the eccentric, 'red' disc. The blue cluster has an approximate physical size of ~ 0.2 pc.

In addition to broad emission lines implying large velocities ( $\sigma \sim 1000 \text{ km s}^{-1}$ ) from random motions that are observed in the stellar spectra, this level of resolution allows us to construct a rotation curve for the blue stars. A model consisting of a circular disc out to a radius of 0.8 pc, inclined at  $\simeq 56^{\circ}$  to our line-of-sight and surrounding a black hole of  $1.4 \times 10^8 \text{ M}_{\odot}$  is consistent with the data (Bender et al., 2005).

#### 1.4.4 Water maser emission in NGC 4258

Molecules typically have a complex quantum structure, with a wide range of possible excitation mechanisms via rotational and vibrational transitions in addition to the usual electronic ones (Rybicki and Lightman, 1986). This makes them susceptible to 'anomalous' over-population of excited states via radiative and collisional pumping. If a region of molecular gas can be pumped enough to undergo a population inversion<sup>13</sup> then the probability of stimulated emission is enhanced. This may give rise to 'maser' emission (akin to a 'laser' but with a wavelength longer than 1 mm). However, a population inversion is a necessary but not sufficient condition for this process; along the direction of masing there must be reasonable velocity coherence so that Doppler shifts do not significantly perturb the frequencies of the emitted photons and prevent them from coupling with inverted states further down the line (Kylafis and Pavlakis, 1997).

A region that fulfills these conditions will be clearly visible as a maser 'spot' that possesses a clear spectral emission line. These are often present (see, e.g., Greenhill, 2005) in observations of accretion discs (particularly if the disc is undergoing star formation), and by analysing the Doppler shift of the particular line at various radii the rotation curve can be extracted, allowing a very precise determination of the central mass.

This method has been used to good effect in the nucleus of NGC 4258, a spiral galaxy in the constellation Canes Venatici at a distance of ~ 23 million light-years (Tonry et al., 2001). NGC 4258 shows some evidence of AGN activity (specifically that of a Seyfert II galaxy - refer to Section 1.5 for details on this type) but by far the most convincing evidence for the presence of an SMBH at its centre comes from observations of maser emission from H<sub>2</sub>O molecules in a sub-parsec accretion disc. The rotation curve of this disc is Keplerian to within high accuracy over a radial range of 0.13 - 0.26 pc (Miyoshi et al.,

 $<sup>^{13}{\</sup>rm when}$  the number of molecules in an excited state exceeds that of the number of molecules in the ground state



Figure 1.4: Artist's impression (top) and cartoon (bottom) of using maser emission to determine the rotation curve of the accretion disc in NGC4258. The left side of the disc is redshifted, implying rotation away from us, and the right is blueshifted, implying rotation toward. By studying how the strength of these shifts in the maser light (shown as 'spots' in the bottom figure) varies with radius, the rotation curve of the disc can be constructed and the mass of the central object determined. Figure credit: John Kagaya, NRAO.

1995; Herrnstein et al., 2005; Pastorini et al., 2007) and indicates a central black hole mass of  $3.6 \times 10^7 \, M_{\odot}$ .

#### 1.4.5 Other evidence

• Doppler shift (and broadening) of spectral lines

We can put a constraint on an enclosed mass by analysing the broad emission lines emanating from the central region. Spectral lines can be broadened beyond their natural<sup>14</sup> width, as well as beyond their thermal width<sup>15</sup>, by special relativistic and general relativistic effects (Padmanabhan, 2000); in the former case a finite velocity

<sup>&</sup>lt;sup>14</sup>set by the uncertainty principle,  $\Delta E \Delta t = \hbar/2$ 

<sup>&</sup>lt;sup>15</sup>set by the temperature of the emitting gas through random motions of individual atoms or molecules

dispersion in the emitting region results in different Doppler shifts, the cumulative effect of which is to broaden the line profile. The extent of this broadening traces the magnitude of the velocity of the gas in this region, and through virial arguments this allows us to estimate the mass that is contained within the observed volume. At higher resolution the *shift* in the line may also be useful. An excellent example of using spectral lines to infer the presence of an SMBH is observations of the iron  $K\alpha$  line (a transition with an energy of 6.38 keV, placing it firmly in the X-ray) in the Seyfert I galaxy MCG-6-30-15, where a width corresponding to a velocity of ~ 100,000 km s<sup>-1</sup> is seen. The line also possesses an asymmetry in the red (high wavelength) direction of the spectrum. The best interpretation of these features is the combined action of Doppler shifts and gravitational redshift from material in an accretion disc around an SMBH (Fabian et al., 1995).

• Reverberation mapping of broad-line regions

When the measurements of line widths in the broad-line region are combined with a technique known as *reverberation mapping*, the size of the emitting region can also be determined (see, e.g., Peterson and Horne, 2004). This can make the difference between evidence for simply a large mass in the central region of the galaxy (which could perhaps simply be a dark cluster of stellar remnants) and evidence for a supermassive black hole.

The technique makes the assumption that there is continuum emission from the central source that illuminates the broad-line region. It is the continuum photons that are then absorbed and re-emitted as line photons from this region. This picture implies a time lag between the variation of the broad-line emission and the continuum emission that corresponds to the light-travel time across the broad-line region. Measurements of this time lag can then be used to determine the size of the region. This method is useful because it does not require extremely high angular resolution, but it can require a large number of observations that are regularly-spaced in time over a long period in order to map the variability of each component (Peterson and Horne, 2004).

• Stellar relaxation

Although the process of stellar relaxation in a galactic core takes place over a long timescale, it can in fact provide us with a test for the presence of an SMBH at the centre of a stellar cluster that is based on the radial density distribution of the stars. This problem was first studied by Peebles (1972); Bahcall and Wolf (1976) and then later by Lightman and Shapiro (1977) in reference to the slow, gradual diffusion of stars inward due to repeated small-angle gravitational encounters. The initial papers considered only the one-dimensional change in the binding energy, with respect to

the SMBH, of the stellar distribution as a result of this process, finding that within the SMBH's radius of influence,  $GM/\sigma^2$ , the equilibrium distribution reached is that of a  $\rho_* \propto r^{-7/4}$  power law (outside this radius the stars are unbound and the distribution is largely unaffected by the presence of the hole). The later paper took this treatment one step further, moving to a 2D approach that took account of the loss-cone<sup>16</sup> effects; the authors find once again that within the black hole sphere of influence the gradient of the density is  $\simeq -7/4$  but flattens slightly further towards the centre, reaching  $\simeq -8/5$  at a radius  $\sim 10 r_t$ , where  $r_t$  is the typical radius at which stars are removed from the dynamical system by tidal disruption or coalescence. An observation of such a density profile at the centre of a galaxy is therefore indicative of the presence of a central condensation of mass, potentially an SMBH. It should be noted, however, that in general the SMBH sphere of influence must be resolved for this to be a viable method.

• Radio emission from relativistic jets

Jets are a form of highly collimated, ionized outflow (see Section 1.5.3) that (in the case of AGN jets) travel at relativistic speeds. Such emission is observed to originate from an extremely small region at the centre of a galaxy; the high energy and velocity of the outflow suggest that it emerged from a deep potential well, providing evidence for an SMBH as its progenitor.

Originally, the model for a relativistic jet employed pressure gradients (Krolik, 1999), whereby external (dense) gas pressure confines the hot outflowing gas into a 'de Laval nozzle' through which a sonic point could be reached<sup>17</sup> and the hot gas channelled outwards supersonically. However, this picture would require the confining gas to be highly pressurised and therefore either be very hot, or very dense, and cool rapidly; the subsequent luminosity would be clearly visible. Since we do not see any evidence of this, there must be an alternative explanation.

The favoured model is that the outflow is collimated by the twisting of magnetic field lines that arise within the inner radii of an accretion disc around a black hole. Since

$$J_{\min} = \left[2\left(E + \frac{GM}{r_{\rm pe}}\right)\right]^{1/2} r_{\rm pe}$$
(1.8)

where E is the specific orbital energy.

<sup>&</sup>lt;sup>16</sup>the loss-cone for an accretion radius  $r_{\rm acc}$  is defined in terms of the minimum specific angular momentum that sets  $r_{\rm pe} = r_{\rm acc}$ , where  $r_{\rm pe}$  is the pericentre of the orbit. By considering the specific orbital energy and setting  $v_r = 0$  this is given by (Binney and Tremaine, 1987)

<sup>&</sup>lt;sup>17</sup>a de Laval nozzle has a lot in common with the Bondi solution for accretion/outflow that we shall discuss in Chapter 2, Section 2.3.2. Fluid that is passing through a constriction in the nozzle is accelerated due to mass conservation, reaching the sound speed at the throat. Pressure decreases as the fluid is accelerated due to the Bernoulli effect. Subsequently, the fluid is allowed to expand, decreasing the pressure further and accelerating the fluid to supersonic velocities

the frictional stresses at the centre of an accretion disc can be very large, the gas is heated to high temperatures and becomes a plasma, in which the mix of accelerating charges gives rise to strong magnetic fields. Due to the differential rotation of the disc the field lines connecting the charged particles can become twisted, creating a vertical column stretching above and below the disc along the axis of rotation. The jet is driven out through these magnetic columns, reaching near-luminal speeds (Komissarov, 2010).

The streams of plasma can often be seen extending out to Mpc scales, and will shine through inverse Compton scattering<sup>18</sup> and synchrotron radiation<sup>19</sup>, the latter being the source of the strong radio emission that is associated with these structures. The orientation of the jet is believed to be correlated with either the spin axis of the SMBH or the angular momentum axis of the accretion disc, although the former case cannot currently be inferred from observations.

#### 1.5 Active galactic nuclei (AGN)

As we have seen, SMBHs can be detected in a variety of ways. It is not surprising, then, that the 'behaviour' of these objects can differ. Some, like the SMBH at the centre of our Galaxy, are dormant, merely content to exert a gravitational pull on the surrounding stars and gas without any further interaction. Others are more attention seeking, and swallow large quantities of material, causing them to grow in size and mass, whilst strongly influencing their environment in the form of prodigious amounts of electromagnetic radiation. These latter SMBHs are the central engines of what are termed *active galactic nuclei*, or AGN, which are often so luminous that they outshine the rest of the galaxy put together.

The first documented recognition of a class of galaxies that showed unusually bright, nuclear emission was by Carl Seyfert in 1943 (Seyfert, 1943). These objects displayed strong emission lines (in the optical region of the spectrum) of ionized gas, specifically hydrogen, helium, nitrogen and oxygen. The other notable property of these 'Seyfert galaxies', as they became to be known, was that the emission lines showed significant Doppler broadening, equivalent to velocities of  $\sim 10^3$  km s<sup>-1</sup>. The fact that these broadened lines came from a region of very small angular (and physical) size was evidence for the presence of a large amount of mass within this small volume, the high velocities (and no indication of mass being expelled from the region) being due to the Keplerian-like rotation of an accretion disc at the very centre of the galaxy.

Today, AGN are an extremely active<sup>20</sup> area of research across a range of fields, and in

<sup>&</sup>lt;sup>18</sup>scattering of a photon by a fast-moving electron, resulting in a gain of energy for the photon and a loss of energy for the electron; this will occur when the radiation is cooler than the gas kinetic temperature

 $<sup>^{19}{\</sup>rm synchrotron}$  radiation occurs due to relativistic electrons spiralling around magnetic field lines  $^{20}{\rm excuse}$  the pun

Type	Narrow lines	Broad lines	Radio	Broad-band	Variability
Seyfert I	yes	permitted only	quiet	yes	yes
Seyfert II	yes	no	quiet	yes	yes
LINER*	yes	no	quiet	no	yes
Quasar/QSO	yes	yes	loud/quiet	yes	yes
BL Lac	no	weak	loud	yes	yes
OVV	no	yes	loud	yes	yes
BLRG/NLRG	yes	yes/no	loud	yes/no	no

Table 1.1: AGN types and their properties

\*at the time of writing, the case on LINERs belonging to the AGN club is still open (Peterson, 1997).

particular are of interest to cosmologists, as they show strong evolutionary characteristics over cosmic time. Quasars in particular are far more common in the early Universe, a fact which contributed to the physical interpretation of AGN as accreting SMBHs.

#### 1.5.1 Taxonomy of AGN

The AGN that have been discovered vary across a range of different types, characterised by their observational properties. Here we briefly list the typical classifications and detail some relevant information about each type. Table 1.1 summarises the important identification characteristics of each class.

• Seyfert galaxies

Today, Seyfert galaxies are divided into two categories (Krolik, 1999). Seyfert I galaxies are characterised by narrow emission lines from low-density ionized gas, corresponding to  $v_{\rm typ} \sim 100 \rm \ km \ s^{-1}$  and broad emission lines (permitted transitions only) from high-density ionized gas, corresponding to  $v_{\rm typ}$  up to  $10^4 \rm \ km \ s^{-1}$ . The fact that the broad lines show an absence of forbidden<sup>21</sup>-line emission is evidence for the high densities, as the non-electric-dipole transitions are suppressed through collisions.

Seyfert II galaxies are distinguished from the type Is by the fact that there is no evidence for broad emission lines at all; only the narrow lines are present. This suggests that the typical velocities in the nucleus are typically at the lower end of what is expected in an AGN.

 $<sup>^{21}</sup>$ a forbidden transition is one which cannot occur in first-order quantum mechanical perturbation theory, in other words through the most efficient, *electric dipole* route. The term is misleading, however, as such transitions can indeed take place but depend on higher-order effects (electric quadrupole, magnetic multipoles, etc.) and have a much lower probability

Both types also show strong emission in the infrared, ultraviolet, and particularly the X-ray region of the spectrum. The latter is thought to arise from the hot, inner regions of an accretion disc at the very centre of the source.

• Quasars / QSOs

The term 'quasar' is a shortened form of 'quasi-stellar radio source' and refers to the discovery of these objects in early optical images as little more than points of light, indistinguishable from stars except for their unusually blue optical colour and strong emission in the radio region of the spectrum. Indeed, quasars were originally identified by the first radio surveys of the sky in the 1960s, the first quasar to be found was designated 3C 273 (Schmidt, 1963). The most puzzling aspect of the spectra of these objects was the presence of broad optical emission lines at anomalous frequencies; these were later discovered to be the Balmer-series emission lines of hydrogen (again, see Schmidt, 1963), redshifted by an unprecedented amount that had never before been seen (out to z = 0.15). Using Hubble's law for the distance put these quasars at ~ 1.5 billion light years. This raised a further question, which was how such a distant object could exhibit such a large brightness; 3C 273 has B = 13.1 in apparent magnitude (Oke, 1963). Assuming the distance estimate was correct, this implied a luminosity that was far above anything previously discovered.

What was also detected was strong time variability of the flux. This is yet another important characteristic of these objects and applies to every waveband, in both the continuua and the broad emission lines. Indeed, variability became one of the first well-studied aspects of quasars (see e.g., Wachter et al., 1988). It was found that the luminosity of the point-like sources varied on extremely short timescales, some of the order of weeks or even days. In order for such a variation in luminosity to be coherent, it would be necessary for each part of the emitting region to be in contact with the other parts on a similar timescale; this led to the astounding conclusion that a nucleus comparable in size to the Solar System was emitting at hundreds of times the luminosity of an entire galaxy. As a result it became clear that astronomers were dealing with a mode of power generation that was an order of magnitude more efficient than nuclear fusion (the most efficient energy generation mechanism known at the time, converting 0.7% of rest mass into energy). The solution to this puzzle was to invoke accretion on to a compact object - a supermassive black hole - as the means to power such an efficient energy source (refer to Chapter 2, Section 2.3).

Once a large enough sample of quasars was identified, their salient properties were defined (Schmidt, 1969) as:

- 1. Point-like sources with radio emission
- 2. Strong nuclear X-ray emission

- 3. Broad emission lines
- 4. Large redshifts
- 5. UV excess
- 6. Time-variable flux

The 'UV excess' refers to the comparison between a quasar spectrum, which is largely flat across the optical/near-UV region (the U-B band according to the photometric system; see, e.g., Binney and Merrifield, 1998), and a typical stellar spectrum, in which the U band is in the tail of the spectral distribution (Peterson, 1997). In fact, this property is indicative of a far more general characteristic of AGN; that of broad-band continuum emission. In contrast to a typical galaxy, which is essentially a composite of many stellar (blackbody) spectra and contains the majority of its power within a single decade in frequency, an AGN spectrum is far broader and has power distributed smoothly across multiple bands, from the optical/UV all the way into the hard X-ray (in other words, a *non-thermal* spectrum). This is often the means by which AGN are detected (Krolik, 1999), through color distinctions between AGNs and ordinary galaxies (or stars).

After about 1970 it was realised, through the greater number of methods available in identifying quasars, that in fact fewer than 10% of these objects possess strong radio emission (Peterson, 1997). Thus, although the first quasars were identified on this very basis, these actually belong to a small subset of the entire class. As a result, the nomenclature has become somewhat confused. The correct definition for the radio-quiet majority is 'quasi-stellar objects' (QSOs), with the radio-loud minority referred to as quasars; however, the term 'quasar' is often used to encompass both sets (and indeed sometimes to refer to AGN in general). Since the first detections of these objects, quasars and QSOs have been discovered out to extremely high redshifts, well beyond that of 3C 273. The highest redshift quasars known to date have redshifts of  $z \gtrsim 6$  (Fan et al., 2003), which means that we are observing them when the Universe was less than a billion years old.

• Blazars

'Blazar' is a somewhat unofficial term (and was originally intended as a joke) designed to unite the classes of BL Lac objects and OVVs (optically violent variables). The term 'BL Lac' originates from the prototype of this class called BL Lacertae (Hoffmeister, 1929). Both of these objects show extremely short timescale variations in their luminosities, in some cases of the order of days to hours. In this respect they are similar, but more compact, than quasars. The other defining property of a blazar is strong radio emission. There are differences between the two classes, however, in that OVVs possess the broad (but not the narrow) emission lines seen in most Seyferts, quasars, and QSOs while in the BL Lac objects these lines are very weak and in some cases non-existent. OVVs also tend to be stronger in the radio (Krolik, 1999).

• Radio galaxies

We refer to radio galaxies in terms of two types, broad-line radio galaxies (BLRGs) and narrow-line radio galaxies (NLRGs). In actuality both possess narrow emission lines, but differ on the presence of the broad lines. As their name suggests, they are characterised by strong radio emission. In effect, BLRGs and NLRGs are the radio-loud analogs of Type I and II Seyferts respectively. It should also be noted that NLRGs show somewhat more extended emission (in other words they are less point-like) than other AGN, suggesting that the nucleus does not dominate over the rest of the galaxy to quite the same extent.

• LINERs

Low ionisation nuclear emission line regions (LINERs) are low-luminosity, Seyfert II-like sources whose dominant emission lines result from neutral or weakly-ionized gas (as opposed to the strongly-ionized species that is characteristic of Seyferts). There is some controversy over whether these objects belong in the AGN class, as LINER-like emission can also arise from star forming regions (Shields, 1992). It should be noted that over a third of all spiral galaxies can be classified in terms of a LINER spectrum.

#### 1.5.2 Unification of AGN

Since the discovery of many and varied AGN, both in the local Universe and out to high redshift, much work has gone into identifying a single, unified model that might explain the seemingly disparate types. There is clear evidence that AGN possess an axisymmetric rather than spherical structure, and that in the radio-loud AGN the radio axis is aligned with the axis of symmetry (Krolik, 1999; Peterson, 1997), and so a picture has emerged that consists of an optically thick, obscuring torus surrounding the central accretion disc and the SMBH. The torus is large enough to obscure (partly, or otherwise) a broad emission line region, but does not occlude a narrow emission line region which extends considerably further out. The inner regions of the accretion disc are hot enough to emit X-rays<sup>22</sup>, explaining the nuclear high-energy emission. Completing this picture

<sup>&</sup>lt;sup>22</sup>thought to emerge from the hot *corona* of the disc through inverse Compton scattering (Stern et al., 1995), since the standard disc equations (see Chapter 2, Section 2.3.4) suggest an effective (blackbody) temperature of only  $\sim 10^5$  K in the optically thick midplane (Shakura and Sunyaev, 1973)



Figure 1.5: AGN unification scheme (left) and artist's impression of the obscuring torus, accretion disc, and relativistic jet (right). Figure credit: scholarpedia.org, Chandra X-ray Observatory, Urry and Padovani (1995)

is the possible presence of a relativistically-beamed jet along the spin axis of the central SMBH that is the origin of the radio emission.

The unification of the various classes of AGN is achieved by positing that each class is simply an observation along a different line of sight relative to the axisymmetry of the torus/accretion disc system. Unification schemes are themselves divided into two types, depending on the presence of the jet.

• Radio-quiet unification

From the differences in the (radio-quiet) Seyfert types it is clear that a scheme whereby the type Is are seen along a largely unobstructed line-of-sight into the broadline region (partially face-on), and the type IIs are seen with the torus obscuring the broad-line region (edge-on) is consistent with the observations. In this case there is no radio jet along the spin axis, and as a result little radio emission is seen. Add to this the radio-quiet quasars (QSOs) as viewing the central broad-line region and accretion disc entirely face-on and all the relevant types obey the simple unification model. LINERs, too, fit into this picture in the same manner as the Seyfert IIs (i.e., edge-on).

• Radio-loud unification

The radio-loud picture is similar but with the addition of the relativistic jet that forms a collimated, narrow beam heading outwards from the centre of the AGN along the axis of the SMBH. When the line-of-sight is directly coincident with this jet the object that is seen is a blazar, in reality a radio-loud AGN with a completely unobstructed view to the very central regions. The difference between the sub-types of the class, BL Lacs and OVVs, is then simply that the latter objects are more powerful in terms of their flux. When seen at an angle to the jet, but still with a clear view into the centre, these AGN appear as quasars, or as BLRGs. We can obtain the observational characteristics of NLRGs by viewing the nucleus edge-on, so that the broad-line region is obscured by the torus.

All of this information is summarised graphically in Figure 1.6, which divides a schematic of the standard AGN paradigm into radio-loud and radio-quiet modes. We now take a look at the various lines of evidence for the fueling of AGN through accretion on to supermassive black holes.

#### 1.5.3 Outflows from AGN

One final feature that is exhibited by some QSOs but that is in general not included in the unified model of AGN is that of *absorption troughs* corresponding to particular spectral emission lines. These troughs are located on the blue side of the emission features, providing evidence that the absorbing material is moving towards us in the form of an *outflow*. From the blueshift of the absorption trough relative to the emission line we can infer the velocity at which the material is ejected; this can be many thousands of kilometers per second along our line of sight, which is more than sufficient to escape the host galaxy entirely (Everett, 2007).

Outflows from AGN consist of hot, partially-ionized<sup>23</sup> gas that is thought to have been accelerated by radiation pressure from the central, luminous regions of the accretion disc around the SMBH (see, e.g., Arav et al., 1995; Arav, 1996). It is clear that the radiation field must provide (at least some) of the momentum imparted to the wind, as the absorption features demonstrate that energy and momentum has been removed from the radiation on interaction with the gas.

Outflows naturally have some properties in common with jets, but are far less collimated, less ionized, and tend to be significantly slower. However, recent evidence has been emerging for outflows with velocities  $\sim 0.1c$  (e.g., Pounds et al., 2003; O'Brien et al., 2005), suggesting that some unification is possible. Both are nonetheless *feedback* processes, driven by accretion close to the SMBH but able to affect the gas on larger scales, potentially slowing it down and halting the infall of material. Our understanding of accretion processes, jets, and outflows is therefore linked; to each other and to the evolution of the host galaxy as a whole.

<sup>&</sup>lt;sup>23</sup>only partially since some atoms with bound electrons must be present in order to produce spectral lines


Figure 1.6: An artist's impression of the model for AGN outflows. Material falls in from large scales within the galaxy, forming a luminous accretion disc (yellow) around the central SMBH. The radiation from this accretion disc accelerates some of the gas out of the plane of the galaxy (typical trajectories shown in white). The regions of the outflowing gas along our line-of-sight (shown in blue) absorb some of the radiation at wavelengths that correspond to the chemical composition and velocity of the outflow, creating an absorption trough in the AGN spectrum (inset. Not shown in this graph is the corresponding emission line which would be to the red of this trough). Figure credit: Daniel Zukowski

# 1.6 The growth of SMBHs

At a fundamental level, simply powering the enormous luminosities observed in quasars  $(\sim 10^{46} \text{ erg s}^{-1})$  must require a significant amount of material to accrete on to the central SMBH, growing it in mass. We will deal with the physics of accretion on to compact objects in Chapter 2, Section 2.3, but for now it suffices to look at a simple estimate for the required feeding rates of SMBHs. For matter falling into a black hole, the maxmium energy per unit mass that can be channelled into a luminosity is the binding energy per unit mass of the innermost stable circular orbit (ISCO). Within this orbit the matter is accreted beyond the event horizon on the free-fall time, and thus radiates very little. For a quick order of magnitude calculation, we consider an infalling mass m through an annulus, for which the rate of conversion of potential energy, U = GMm/r, into radiation is given by

$$L \sim \frac{\mathrm{d}U}{\mathrm{d}r} = \frac{GM}{r} \frac{\mathrm{d}m}{\mathrm{d}t} \tag{1.9}$$

which we write in terms of a mass accretion rate  $\dot{M}$  as

$$L \sim \frac{GMM}{r} \tag{1.10}$$

The 'size' of a black hole is essentially its event horizon, or Schwarzschild radius  $R_s$ , which is given by (e.g., Shapiro and Teukolsky, 1983)

$$R_s = \frac{2GM}{c^2} \tag{1.11}$$

where c is the speed of light. Putting the ISCO at  $\sim 5R_s$ , we find for the potential energy<sup>24</sup> that

$$U = \frac{GMm}{5R_s} = \frac{GMm}{10\,GM/c^2} \tag{1.12}$$

and so we see that  $U \sim 0.1mc^2$ . From the form of this it is clear that the conversion efficiency of rest mass into energy is  $\eta_{\rm acc} \sim 0.1$ . For powering a quasar, then, we find that a mass accretion rate of  $\dot{M} \sim 2 \,\rm M_{\odot} \, yr^{-1}$  is needed. This is not so huge; the Eddington limit, which defines the maximum rate of (spherically symmetric) accretion on to an object (see Chapter 2, Section 2.3), is slightly above this for a black hole of  $10^8 \,\rm M_{\odot}$ . Furthermore, if one assumes a flattened geometry rather than a spherically symmetric one, with the majority of the radiation emitted perpendicular to the gas supply, the maximum rate becomes higher still. However, as we shall see later in this section, there are some serious difficulties in getting the required amount of mass close enough to the SMBH to actually

 $<sup>^{24}</sup>$  there are of course numerical factors of a few floating around that we have ignored; the correct value of the ISCO (derived in Chapter 2, Section 2.2.3, taking account of the orbital energy of the particle rather than just the potential energy, and the relativistic effects

accrete at these rates.

#### 1.6.1 SMBH growth through gas accretion

The current paradigm for SMBH growth has its roots in the Soltan (1982) argument, which presents a convincing exposition for the proposal that SMBHs have grown primarily through AGN activity. The treatment demonstrates that the total rest mass energy of all known SMBHs at the centres of galaxies is consistent with the radiation energy of the Universe if the black hole masses grew by luminous accretion.

Using quasar/QSO counts over a volume of 1 Gpc<sup>3</sup>, and integrating over the bolometric luminosity across all redshifts, the total radiant energy due to quasars/QSOs can be found<sup>25</sup>. This puts an estimate on the total mass of their central engines by assuming that they conversion of rest mass into radiant energy has an efficiency of ~ 10%. This estimate is then compared to the total observed masses of the SMBHs, allowing us to determine to within what certainty they agree.

However, obtaining the actual masses of the SMBHs from observations is somewhat non-trivial. Originally, Soltan (1982) employed stellar relaxation measurements; in this case, the departure from the standard King profile that arises from steady-state solutions of the Fokker-Planck equation<sup>26</sup> with a finite cut-off velocity that corresponds to the escape velocity of the cluster (King, 1966). However, this approach was flawed (Dehnen, W., private communication). The resolution at the time was poor and the central regions of the stellar distributions in galaxies most likely appeared as a King profile (which is essentially a core, with a hint of an isothermal power law at the edge of the cluster) due to the blurring of the stars. Furthermore, deviations from such a profile, even if they could be detected, would not be a good indicator of a central mass as there is no standardised model for the unperturbed distributions in nature. Fortunately, this approach has since been superseded by a method that employs the  $M_{\rm bh} - \sigma$  relation (see Section 1.8) to extract SMBH masses from observations of velocity dispersions; this has been performed over a large sample of  $\sim 9000$  nearby galaxies from the Sloan Digital Sky Survey (SDSS) (Yu and Tremaine, 2002), and in addition to matching the total observed SMBH masses to the total radiant energy from QSOs, these authors also compared the BH mass function to the QSO luminosity function, finding good agreement. It should be noted of course that this method relies on the calibration of the  $M_{\rm bh} - \sigma$  relation, which is performed on the *local* Universe by measuring SMBH masses through more direct methods such as those described in Section 1.4; the crucial assumption is therefore that this relation holds ubiquitously across the Universe rather than simply locally.

 $<sup>^{25}</sup>$  we do not go into the mathematical detail here; for the full treatment the reader is directed to (Peterson, 1997)

 $<sup>^{26}</sup>$ the Fokker-Planck equation describes the evolution of the stellar distribution function subject to an *encounter operator* that takes into account the two-body relaxation process in collisional systems

From these arguments it is clear that SMBHs must assemble their mass since formation through periods of AGN activity. However, the question remains as to how the gas accreted. Before we consider the possibilities, it is worth putting this growth into context; how quickly should we *expect* the black hole to gain mass? Or, more pertinently, how quickly do we *need* it to gain mass to explain the SMBHs that we see in the Universe?

## 1.6.2 SMBHs in the early Universe

Quasars are observed with masses up to ~  $10^9 \,\mathrm{M_{\odot}}$  out to very high redshift,  $z \sim 6$  (Kurk et al., 2007). At this epoch the Universe was only about a billion years old, suggesting that a 'seed' black hole of around a solar mass would have had to grow at an average rate of  $\gtrsim 1 \,\mathrm{M_{\odot}} \,\mathrm{yr^{-1}}$  for a sustained period of time. Unfortunately this is very difficult to achieve, as angular momentum in the gas provides a centrifugal barrier to accretion. The magnitude of the specific angular momentum of a particle in a circular orbit is given by  $J = (GMr)^{1/2}$ , where M is the enclosed mass. To transfer gas on circular orbits from kiloparsec scales (say, the solar orbit at 10 kpc, inside of which the enclosed mass is ~  $10^{11} \,\mathrm{M_{\odot}}$ ) to scales where accretion can occur (at the ISCO, ~ 1 A.U. for a  $10^8 \,\mathrm{M_{\odot}}$  black hole), all but approximately one part in  $10^6$  of the angular momentum must be lost. This is far too high a demand on the level to which angular momentum transport can occur, especially when it must be maintained at a high rate for ~  $10^9$  years.

It is of course possible that the seed black holes that formed in the early Universe were considerably more massive than would occur today. The collapse of a hypothetical zero-metallicity Population III star, for example, may have left a BH remnant of up to  $\sim 100 \,\mathrm{M}_{\odot}$  (Madau and Rees, 2001), decreasing the amount that it needed to grow to reach SMBH mass. However the very existence of these seed BHs is very uncertain, and even with this model the required growth rate would still have been Eddington-limited<sup>27</sup>; the limiting luminosity scales with the mass of the accretor, and so imposes a much harsher limit in the earliest stages of the growth when the black hole was still a seed.

It is therefore crucial to the study of galaxy evolution that the growth of SMBHs is understood. The next section outlines the main mechanisms by which this might occur.

# 1.7 Feeding SMBHs

The amount of gas required to create a typical SMBH, or to power the more luminous quasars, can only have come from (significantly) larger scales. There is certainly enough fuel within a typical spiral galaxy,  $\sim 10^8 - 10^{10} \,\mathrm{M_{\odot}}$  in HI gas (Shlosman et al., 1989b) to achieve this, and invoking a major merger between galaxies (particularly if they are

 $<sup>^{27}</sup>$  this term is defined in Chapter 2, Section 2.3.1; suffice to say that the limit is on the *luminosity* of the gas as it accretes so that the radiation pressure does not halt the process of accretion

gas-rich) offers even greater quantities. It is however, far from obvious how to transport this gas from galactic scales to feed the SMBH, particularly at the rates required.

## 1.7.1 Feeding through an accretion disc

Gas accreting on to what is effectively a point mass will always possess some angular momentum, and so the natural first step is to assume that the gas was brought down to the last stable orbit of the SMBH (refer to Chapter 2, Section 2.2) through a gaseous, rotating 'accretion' disc. A detailed treatment of this is dealt with in Chapter 2, Section 2.3.4. The relevant timescale for accretion through a thin gaseous disc is known as the viscous time, which for a standard ' $\alpha$  viscosity' (Shakura and Sunyaev, 1973) is given by

$$t_{\rm visc} = \frac{R^2}{\alpha c_s H} \tag{1.13}$$

where R is the disc radius,  $c_s$  is the sound speed, and H is the typical disc height, which for a non self-gravitating disc with an angular velocity  $\Omega_k$  is given by  $H = c_s/\Omega_k$  (refer to Chapter 2, Section 2.3.4.3). AGN discs are very thin, and so the latter is found (from detailed modelling, see e.g., Collin-Souffrin and Dumont, 1990) to be ~  $10^{-3} R$  in the nonself-gravitating regime. The value of the dimensionless viscosity parameter  $\alpha$  is uncertain but for a standard accretion disc is unlikely to be larger than unity (Frank et al., 2002). While the innermost parts of the disc (out to a few tens of Schwarzchild radii) are very hot and likely provide the high energy emission seen in AGN, putting in a typical sound speed<sup>28</sup> of  $10^5 \text{ cm s}^{-1}$  (corresponding to  $T \sim 100 \text{ K}$ , assuming atomic/molecular hydrogen), with  $\alpha \lesssim 1$ , returns a viscous time of the order of the Hubble time when the disc extends beyond R = 10 pc. From this simple but powerful estimate it is clear that accretion is prohibitively slow for a disc that extends to parsec scales. With the assumption of a single-valued angular momentum vector (a disc that remains rotating in the same sense and in the same plane) for the majority of the flow, a disc of typical mass around an SMBH can easily grow to this size and beyond.

Invoking a higher temperature can bring  $t_{\rm visc}$  down to a more reasonable value, but simple heating and cooling arguments<sup>29</sup> at  $R \gtrsim 1$  pc preclude a temperature any higher than  $T \sim 10^3$  K (Shlosman and Begelman, 1987).

 $<sup>^{28} \</sup>rm typical$  (although perhaps a little low) for an AGN disc outside  $\sim 1~\rm pc$ 

 $<sup>^{29}</sup>$ assuming a standard dust-to-gas ratio and a dominant quantity of H<sub>2</sub>. In the first case the interior of the disc will be shielded from scattered AGN radiation, and in the second case the cooling will be efficient since molecular hydrogen is able to cool through vibrational and rotational transitions down to low temperatures

#### 1.7.1.1 Onset of self-gravity

At parsec scales there is a further complication. As is discussed in Chapter 2, Section 2.3.4.8, the outer parts of AGN discs that are sufficiently massive and able to efficiently cool are likely to fragment through gravitational instability and form stars, using up the majority of the gas. Since stars cannot transport angular momentum through viscous processes, this deprives the SMBH of its fuel. In particular, the critical mass accretion rate above which fragmentation occurs, i.e., when the mass being transported through a given annulus is too large for the stabilising forces of pressure and rotation to support it against gravity (refer to Chapter 2, Section 2.3.4.8), is given by (see, e.g., equation 2.186 or Lodato, 2007):

$$\dot{M} \simeq 5 \times 10^{-4} \left(\frac{\alpha}{1.0}\right) \left(\frac{T}{100 \text{K}}\right)^{3/2} \text{ M}_{\odot} \text{yr}^{-1}$$
 (1.14)

and so a disc with a mass accretion rate sufficient for fueling an AGN  $(\dot{M} \sim 1 \,\mathrm{M_{\odot}yr^{-1}})$ would be unstable to self-gravity.

The radius at which self-gravity typically occurs therefore depends on the temperature profile of the disc. Goodman (2003) considered thin, self-gravitating discs around QSOs, where the effective temperature is determined from the balance between viscous dissipation and radiative diffusion within the disc and is therefore  $\propto R^{-3/4}$  (refer to Chapter 2, Section 2.3.4.7). With standard parameters for a QSO, and considering both the gas pressure dominated and total pressure (gas + radiation) regimes, the author finds that the selfgravity radius lies at  $R_{\rm sg} = 0.01 - 0.1$  pc. AGN discs are therefore not likely to extend much beyond this.

We note too that there is observational evidence for the self-gravitating nature of an AGN disc beyond  $R_{sg}$ . Rotation curves of AGN are often significantly non-Keplerian (see, e.g., Lodato and Bertin, 2003), and the presence of disc-like stellar features in the GC and the nucleus of M31 (Rafikov, 2009; Levin and Beloborodov, 2003; Hopkins and Quataert, 2010) at a distance that corresponds approximately to the expected self-gravity radius lends credence to this model.

## 1.7.1.2 Randomly-oriented accretion episodes

Yet another issue is the dependance of the black hole spin parameter on the radiative efficiency of accretion; a rapidly spinning BH has a smaller ISCO and subsequently is able to release a larger amount of energy per unit mass (see Chapter 2, Section 2.2). This has the effect of allowing the BH to reach its Eddington limit at a lower accretion rate, and as a result, growth will be suppressed. A disc accreting on to the BH will naturally impart angular momentum to the hole, both in magnitude and direction. To keep the spin low it is therefore important to feed the BH with a low time-averaged angular momentum flow (Volonteri et al., 2005).

What we therefore seek to avoid is accretion of gas with single-valued angular momentum, as this will (a) create a disc that is too large in radius for viscous transport to proceed quickly, (b) lead to the onset of fragmentation at large radii, and (c) 'spin up' the black hole and slow its growth. A possible solution to this problem lies in a form of randomised, or 'stochastic', feeding through individual accretion events that have no preferred orbital direction. In this picture the gas flow that accretes on to the SMBH is not part of a larger, planar geometry but rather the result of a chaotic process e.g., interactions and collisions between molecular clouds, large-scale turbulence, repeated merger events, etc. This mode of accretion has been suggested by a number of authors, including King and Pringle (2007), Nayakshin and King (2007), and Hopkins and Hernquist (2006), and has recently been found to arise in cosmological simulations with adaptive resolution (Levine et al., 2010).

In particular, King et al. (2008) point out that due to the tendency of AGN discs to fragment beyond the self-gravity radius, the feeding is unlikely to be steady over cosmic time but will instead be separated into a series of accretion *episodes*. The picture here is one in which a gas disc is built up from an inflow that results from a major merger, and is capped at the self-gravity radius, within which it viscously accretes on to the SMBH. The timescale for the merger, and thus for the accretion from larger scales, will be orders of magnitude longer than the viscous time at the outer edge of the disc, and so the low-mass disc 'empties' on to the hole before being replenished by the merger inflow. There is no requirement here for each of the accretion episodes to be correlated in terms of orientation, and so they are not, providing the SMBH with stochastic, efficient growth at a low spin parameter.

The model of randomised accretion is of particular relevance to this thesis. The work on the Galactic centre (Chapter 4), whilst tailored to reproducing the stellar structures that reside there from a collision between two molecular clouds, also looks at the accretion rate on to the SMBH, with the resulting flow forming an accretion disc that rapidly changes its angular momentum and sense of rotation. The chapter on turbulent feeding (Chapter 5) is the directly relevant work, penning a mode of ballistic accretion on to a galactic nucleus through the action of supersonic turbulence within the bulge.

## 1.7.2 Feeding through a bar instability

On larger scales, a standard paradigm often invoked for feeding the nucleus is that of the non-axisymmetric, *bar* instability that arises in a galactic disc as a response to a gravitational perturbation by an external body (Binney and Tremaine, 1987). Through resonances<sup>30</sup> some of the circular orbits within the disc are driven slightly eccentric, and if there

 $<sup>^{30}</sup>$ a resonance is a regular, periodic interaction between orbiting bodies, or in this case an interaction between an orbit within the galactic disc and the external perturber



Figure 1.7: High resolution HST image of the barred spiral galaxy NGC 1300. Figure credit: HST (NASA)

is a sufficient supply of surrounding material to remove angular momentum, grow more eccentric, scattering more and more particles into their elongated, often self-intersecting orbits as the bar rotates. In the gas component, the rotation of the bar and the orbital intersections cause shocks and direct angular momentum cancellation, transporting some of the gas inwards.

However, although it can assist in driving gas towards the SMBH, the stellar bar instability cannot be the sole mechanism by which AGN are fed, for two reasons; (i) the scales over which a galactic bar can transport gas ( $\sim$  a kpc) are far larger than the scales on which SMBHs accrete (Shapiro and Teukolsky, 1983), and (ii)  $\sim 50\%$  of observed spiral galaxies show bar-like features but only  $\sim$  a few % are host to AGN (Shlosman et al., 1989a). Nonetheless, it is an important starting point for the process of gas inflow, and much research has been devoted to a variety of bar-driven feeding models. These models are loosely divided into two categories, depending on the presence of an inner *Lindblad resonance* (see below) inside the co-rotation radius of the bar.

## 1.7.2.1 Orbital resonances

In the 1960s it was proposed that the spiral structure of galaxies could be explained by the 'density-wave' theory whereby the spiral arms were not composed of separate material (and in which case would be subject to winding up as a result of the differential rotation) but were instead the result of stars, gas, and dust moving through a density wave that compresses the material as it passes through, making it visible as a denser region against the background of the rest of the galaxy (Lin and Shu, 1966). It was further proposed that the gravitational attraction of the material as it was passing through the arms could actually maintain the spiral pattern, provided that the latter was located between what were termed the inner and outer Lindblad resonances (Binney and Tremaine, 1987). The mathematical description of this is relatively straightforward; one defines the angular velocity of the arms to be  $\Omega_p$ , namely the pattern speed, and the co-rotation radius (the radius at which the stars and the arms have identical angular velocities,  $\Omega(R) = \Omega_p$  to be  $R_c$ . Inside of  $R_c$ , the stars move with  $\Omega(R) > \Omega_p$ , and outside with  $\Omega(R) < \Omega_p$ . Physically, then, the spiral structure rotates as a solid body, and in a reference frame that is rotating with the spiral, the stars rotate with  $\Omega_{\rm rel} = \Omega(R) - \Omega_{\rm p}$ . For a spiral with m arms, these orbits close when  $\Omega_{\rm rel} = \pm \kappa/m$ , where  $\kappa$  is the epicyclic frequency<sup>31</sup>. The outer and inner Lindblad resonances are defined as the radii at which  $\Omega(R) = \Omega_{\rm p} + \kappa/m$  for the OLR and  $\Omega(R) = \Omega_{\rm p} - \kappa/m$  for the ILR. Within these radii the self-gravity of the stars in the spiral arms enforces the density perturbation, causing the (skewed) elliptical orbits to move in sync.

The ILR is of particular relevance to a barred potential. Depending on the density profile and the value of  $\Omega_p$ , this can have either 0, 1, or 2 ILRs. In general the epicylic frequency of an orbit is related to the angular velocity by a constant factor, the value of which depends on the density profile. For a constant density (harmonic) core, for example, orbits have  $\kappa = 2\Omega(R)$  so the ILR is found when  $\Omega(R) = \Omega_p/2$ . In this case however the form of the rotation curve has a stationary point (a maximum), and if  $\Omega_p/2$  lies above this, there will be no ILRs. If  $\Omega_p/2$  lies below the maximum then we will have two ILRs; and inner (IILR) and an outer (OILR). For a cusp or a Keplerian potential where there is no stationary point, one ILR exists.

#### 1. With ILRs

If one or more ILRs are present, they may prevent the gas from falling all the way in, as material within the co-rotation radius (CR) will be re-distributed into a thin oval ring near the ILRs. However, through analogy with a damped oscillator (the inner elliptical gas ring) subject to a periodic driving force (the stellar bar) it can be shown that the closed gas orbits inside of the IILR incline to the major axis of the bar in a *leading* sense (Wada, 1994) and so the gas tends to lose angular momentum and fall inwards. A ring that is self-gravitating due to having accumulated a suffi-

$$\kappa^{2} = \frac{2\Omega}{R} \frac{\mathrm{d}}{\mathrm{d}R} \left( R^{2} \Omega \right) \tag{1.15}$$

<sup>&</sup>lt;sup>31</sup>the epicylic frequency is the radial frequency of an eccentric orbit, given by

where R is the radius. It should be noted that for the special case of a Keplerian disc,  $\kappa = \Omega$ .

cient amount of gas will be more enlongated (Wada and Habe, 1993) and the loss of angular momentum will occur more rapidly. If the ring is sufficiently massive it may fragment into clumps, which can collide and lose angular momentum through shocks or sink further in through dynamical friction.

 $2. \ Without \ ILRs$ 

If ILRs are not present, the gas inside the CR is simply shocked by the leading edge of the bar and transported inwards on a dynamical timescale, accumulating in a circumnuclear region within  $\sim 1$  kpc. Transport beyond this is not efficient, however, as in general a stellar bar potential becomes weak at  $\sim$  a tenth of the bar outer radius (Shlosman et al., 1990).

In each case, the material can be transported to  $r \sim 100$  pc on a dynamical timescale, a fact that is consistent with the central 100 pc region of spiral galaxies often being gas-rich (Sakamoto et al., 1999). However neither model is able to bring the gas in much further (Wada, 2004) and the real issue for AGN feeding models is a sufficient inflow mechanism within this radius. A model of 'bars within bars' has been proposed by Shlosman et al. (1989b) for getting the gas down to  $\sim 10$  pc, involving a large-scale stellar bar to deposit gas into a disc within 100 pc, which subsequently becomes unstable and forms a smaller scale, gaseous bar, repeating the process down to smaller radii. However, gaseous bar-like instabilities are harder to create, since gas discs possess fewer degrees of freedom than stellar discs. This model therefore favours an inhomogenous gas disc composed primarily of molecular clouds that behave more like stars (but necessarily still collide and shock, leading to inward transport of gas).

The role of nested bars in enabling gas to accumulate in the centres of galaxies has been explored as a model by a number of authors (see, e.g., Barnes and Hernquist, 1991; Wada and Habe, 1992) and observations suggest that this is a promising scenario (see, e.g. Colina and Wada, 2000; Maiolino et al., 2000) but in general this process is relatively contrived, and in any case thought to only be effective down to tens of parsecs (Wada and Fukuda, 2001); its relevance to the fueling of AGN is therefore not clear.

#### 1.7.3 Feeding through a collisional stellar cluster

A final suggestion that we briefly consider is that the fuel for an SMBH might come from the accretion of stars via the collapse of a stellar cluster, centered on the black hole. To investigate this possibility, we draw on our definition of the *relaxation time* from Section 1.2, which gives us the typical time that it takes for gravitational interactions between the stars to significantly perturb their orbits. A system that has undergone relaxation is said to be *collisional*, in that interactions between individual stars become important. Such a system will display the bizarre property of a negative specific heat capacity (Lynden-Bell and Wood, 1968), whereby a loss in the overall energy of the system leads to an increase in the random motions of the stars (a dynamical 'heating'). This process can be understood by the virial theorem, which says that

$$2K + V = 0 (1.16)$$

where K and V are the kinetic and potential energies respectively. For a system with N stars of individual mass m,  $K = 1/2M\overline{v^2}$ , where M = Nm and  $\overline{v^2}$  is the mean-square velocity. We can define a mean temperature,  $\overline{T}$ , for the system via  $K = 3/2Nk_B\overline{T}$ . The *total* energy of the system is simply E = K + V, and so we see that K = -E, with V = 2E. The total energy is therefore

$$E = -\frac{3}{2}Nk_B\overline{T} \tag{1.17}$$

giving us a heat capacity for the system as

$$C \equiv \frac{\mathrm{d}E}{\mathrm{d}\overline{T}} = -\frac{3}{2}Nk_B \tag{1.18}$$

which is negative. A decrease in E therefore leads to an increase in the kinetic energy of the system, with a corresponding decrease (although twice as much, thus conserving energy) in the potential energy. This property of collisional systems<sup>32</sup> leads to what is known the *gravothermal catastrophe*, whereby a hot cluster core that loses heat (and therefore energy) to a cooler halo through scattering processes will, through its negative specific heat, both shrink and heat up. If the cooler halo is collisionless, it will possess a positive specific heat capacity and therefore heat up also, but since it is in general far larger than the core this heating will proceed at a much slower rate. The temperature gradient between the two is driven steeper, and so the (now hotter) core will lose even more heat to its surroundings, shrinking the core even further. A runaway process of collapse ensues; this is known as *core collapse*.

In general the collapse time is considerably longer than the relaxation time; as a rule of thumb we put  $t_{\text{collapse}} = 100 t_{\text{relax}}$  (Binney and Tremaine, 1987). Putting typical numbers (as measured by HST for a typical large galaxy) into equation 1.6 we find that for a stellar

 $<sup>^{32}</sup>$ which is not restricted to just stellar dynamics, but is a consequence of any system described by the virial theorem, i.e., where self-gravity between the particles is important. In fact, the stability of nuclear burning in the cores of stars can be understood by this process. Runaway nuclear burning begins to set in when the reaction rates become too fast for the heat to be conducted away, but through the property of negative specific heat the increase in energy in the core leads to an expansion and a *cooling* which brings the rates back into equilibrium (Binney and Tremaine, 2008)

cluster at the centre of a galaxy

$$t_{\rm collapse} \simeq 9.5 \times 10^{12} \,{\rm yr} \, \left(\frac{v_{\rm typ}}{200 {\rm km \ s}^{-1}}\right)^3 \left(\frac{10^6 \,{\rm pc}^{-3}}{n}\right) \left(\frac{1 \,{\rm M}_{\odot}}{m}\right)^2$$
(1.19)

where we have used  $\ln \Lambda = 15$ . It is clear then for the majority of galaxies that  $t_{\text{collapse}}$  is longer than  $t_{\text{hubble}}$ , and indeed even the smallest galaxies have  $t_{\text{collapse}} \sim 10^{10}$  yr (Rosswog and Bruggen, 2003).

We can therefore conclude that a collisional nuclear stellar cluster is not a viable means for feeding an SMBH at the rates required for the observed growth or AGN activity.

## 1.8 The SMBH/host galaxy connection

Finally in this chapter, we address a fundamental question in galaxy evolution; that of the effect of the SMBH on the larger host system. Understanding the growth of SMBHs is important not only to explain the observed quasars at high redshift, but also to shed light on what appears to be a fundamental connection between the SMBH and its host spheroid. Furthermore, the apparent co-evolution of star-forming galaxies and active galaxies lends support to a causal relationship between nuclear activity and galactic activity.

## 1.8.1 $M_{\rm bh} - \sigma$ relation

It was discovered at the turn of the millenium (Gebhardt et al., 2000; Ferrarese and Merritt, 2000) that there is a strong empirical correlation between the stellar velocity dispersion,  $\sigma$  of a galactic bulge, and the mass,  $M_{\rm bh}$ , of the SMBH residing at its centre. Somewhat unusually for observational correlations, this relationship is relatively tight, with very little scatter (see Figure 1.8). It was clear once this was discovered that the  $M - \sigma$  relation implies a connection between the formation and evolution of a galaxy bulge and the evolution of the central black hole; where previously the effort involved in detecting SMBHs was purely a numbers game, the focus after this discovery became one of understanding the nature of the SMBH/host galaxy connection<sup>33</sup>.

It is clear too from the tightness of the relation that it cannot be motivated by galactic formation alone, as the spread in formation mechanisms (major/minor mergers, gradual gas accretion) would translate to a similar spread in the data points. As a result, theoretical models that seek to explain  $M - \sigma$  invoke a *feedback* mechanism that can be applied ubiquitously across a large sample of galaxies. The currently preferred picture by many authors (see, e.g., Silk and Rees, 1998; King, 2003; Sazonov et al., 2005) is one in which

<sup>&</sup>lt;sup>33</sup>it should be noted however that the  $M_{\rm bh} - \sigma$  relation is subject to a *selection effect*, since detections of SMBHs naturally favour those at the high mass end of the distribution; in general therefore the relation should be viewed as more of an upper limit and there may be a population of galaxies with 'underweight' black holes



Figure 1.8: The  $M - \sigma$  relation for dynamical mass detections of SMBHs in a large sample of galaxies. Colors refer to Hubble type, shapes refer to detection method. The label 'S0' refers to lenticular galaxies. Figure credit: Gültekin et al. (2009)

an outflow, or wind, driven by gas accretion (and therefore in keeping with the Soltan argument) inhibits growth of the black hole above a critical mass, where the outflow sweeps away the ambient gas. So far the most promising model (in terms of optimum fit and minimal free parameters) appears to be that of King (2003, 2005), in which an initially momentum-driven wind (rather than thermally-driven, since for the dense gas in the region around the SMBH the cooling time (through inverse Compton scattering) is longer than the flow time) produced by Eddington-limited accretion and so possessing a velocity of  $\sim 0.1c$  (King and Pounds, 2003), sweeps up a shell of gas within an isothermal sphere of dark matter (which has  $M(R) \propto R$ ) that expands rapidly initially. As the shell gains mass, it slows down, eventually stalling and forming stars. Some of the gas may fall back and supply the SMBH with more fuel, beginning the process over again. The black hole grows (given an adequate mass supply) on the Salpeter timescale, and indeed the shell is able to slowly expand (since it is being driven by continuous accretion) on this timescale also. The growth of the black hole is shut off, however, once the shell reaches a critical radius  $R_c$  where the gas is no longer able to cool efficiently (this is ~ 1 kpc; see, e.g. King, 2003). The momentum-driven shell then becomes thermally-driven, and the extra pressure accelerates the shell and allows it to escape the galaxy. The mass that the SMBH has grown to by this point is given by

$$M_{\rm bh} = \frac{f_g \kappa}{\pi G^2} \sigma^4 = 1.5 \times 10^8 \left(\frac{\sigma}{200 \text{ km s}^{-1}}\right)^4 \text{ M}_{\odot}$$
(1.20)

where  $f_g$  is the baryon fraction, and  $\kappa$  is the inverse Compton scattering opacity. This formula reproduces well the slope and normalisation of the  $M-\sigma$  relation (Tremaine et al., 2002) such as is shown in Figure 1.8.

## 1.8.2 $M_{\rm bh} - M_{\rm bulge}$ relation

There is a further correlation that connects the mass of the SMBH to the stellar mass in the bulge. Originally penned by Magorrian et al. (1998), it has since been revised through more accurate determinations of black hole masses by ~ a factor of 10 in normalisation (Häring and Rix, 2004). The  $M_{\rm bh} - M_{\rm bulge}$  relation has slightly more scatter than  $M_{\rm bh} - \sigma$ , although the correlation is still clear. The slope is slightly above unity,  $M_{\rm bh} \propto M_{\rm bulge}^{1.12}$ , with the normalisation factor at ~ 1/1000 (Häring and Rix, 2004). In general therefore the bulge mass in a galaxy is expected to be about 1000 times the black hole mass. Explaining this relationship is difficult, but we note that the King (2003, 2005) model for  $M_{\sigma}$  can also be invoked to obtain a scaling of  $M_{\rm bh} \sim 10^{-3} M_{\rm bulge}^{1.25}$ , if the amount of mass swept up in the shell is turned into bulge stars.

## 1.8.3 The starburst-AGN connection

Certain galaxies have an unusually high amount of recent or ongoing star formation, evidenced by either the blue optical and UV colours in their spectra, or the IR signature of obscured UV emitting (and therefore star-forming) regions. These galaxies are termed starburst galaxies, and are found out to  $z \sim 4$  (Thompson, 2005)

It can be difficult (particularly at high redshifts) to separate the signature of a starburst from that of an AGN. In these cases a spectroscopic approach can help. Stellar atmospheres have particular emission lines in the optical/IR corresponding to the Balmer series of hydrogen, and neutral calcium, and in the UV corresponding to a ionized helium, silicon and carbon (Robert et al., 1993). The star formation rate (SFR) of a starburst galaxy is generally  $\sim 10 - 100 \,\mathrm{M_{\odot}yr^{-1}}$ , orders of magnitude higher than a typical gas-rich galaxy such as the Milky Way, which has an SFR of  $\sim 1 \,\mathrm{M_{\odot}yr^{-1}}$ .

Triggering a starburst is generally thought to involve a perturbing event that can concentrate a large amount of H<sub>2</sub> in a small volume, normally  $\sim$  a kpc. This can be the result of supernovae explosions, but on a large scale is mostly likely due to galactic encounters. Indeed, galaxies that have recently merged (or are in the process of merging) show a disproportionately high fraction of starbursts (Muxlow et al., 2006). One such merging pair of galaxies is shown in Figure 1.9. These are local, at  $z \simeq 0.004$ . The blue colours indicate young stellar populations, yet few supernovae have been detected in the merging regions and so these galaxies provide strong evidence for the merger-starburst paradigm.

Direct, head-on mergers, where the gas undergoes strong shocks and loses a large amount of angular momentum, are naturally the easiest explanation for a strong inflow of gas to the centre of the potential well. These may not be necessary in every case, however, as interactions with a large impact parameter can excite a bar instability (refer back to Section 1.7.2) that transports large amounts of molecular gas inwards, giving rise to a starburst. Indeed, observations of CO in starburst galaxies (Lo et al., 1987; Ishizuki et al., 1990; Kawabe et al., 1991) have shown that a significant amount of star-forming gas has large non-circular motions and is distributed in an elongated ring, lending support to this model.

With the antecedents of starbursts so similar to the agents of AGN, it is not surprising that there is a correlation between them: in large samples of galaxies, starbursts often coexist with AGN activity (see, e.g., Bushouse et al., 2002; Heckman, 2004). Furthermore, the strength of the starburst and the luminosity of the AGN are often correlated as well (Veilleux, 2008). From a cosmological perspective, the rate at which SMBHs grew by AGN accretion has a trend which peaks in the early Universe and falls off to the present day, and a similar trend is seen in the star formation rate (SFR) of galaxies (Heckman,



Figure 1.9: Ground-based view of the merger of the Antennae galaxies (left) and high resolution HST image of the burst of new star formation triggered as a result (right). The bright blue regions are newly-formed star clusters. Figure credit: Brad Whitmore (STScl), HST (NASA)

2004). This implies that there is a certain amount of co-evolution between star-forming galaxies and SMBHs that cannot be ignored.

Unfortunately, the starburst-AGN connection is something of a 'chicken and the egg' scenario. Some theoretical models suggest that, during the process of a merger and gas inflow, the burst of star formation came first and eventually the gas reached the SMBH to trigger an accretion event (see, e.g., Sanders et al., 1988; Kauffmann et al., 2003). Others maintain that the AGN activity came first and powerful outflows from the nucleus compressed the gas in the galaxy, triggering gravitational instability and star formation (see, e.g., Rees, 1989; Daly, 1990). Further to this divide there are those who have proposed more complex models, such as the 'first pass' in a merger causing an initial inflow that fuels star formation but does not reach the SMBH, with the final merging phase fuelling the SMBH as well as a second starburst (Mihos and Hernquist, 1994), and those who invoke an AGN event to shut-off an ongoing starburst (Schawinski, 2010). Finally there are those who believe the two events are in actuality decoupled, and simply coexist because they possess similar origins (Lin et al., 1988).

The symbiotic relationship between star formation and AGN activity remains a key question in the search for a complete picture of galaxy formation and evolution. It is one of the questions that we address in this work; specifically, Chapter 4, in which feedback from star formation in the bulge is invoked as a driver for supersonic turbulence that enhances the growth of the SMBH.

# 2 Theoretical background

"We live on an island surrounded by a sea of ignorance. As our island of knowledge grows, so does the shore of our ignorance."

John Archibald Wheeler

# 2.1 Theoretical astrophysics

The field of theoretical astrophysics seeks to understand the behaviour of the Universe by means of theoretical *models*, which take account of observations of the phenomena of interest and attempt to construct a complete picture of these phenomena in a mathematically and physically consistent manner. A theoretical model can be both *predictive*, in suggesting behaviour that should be looked for, and *explanatory*, in providing an answer to why an observational result is seen.

Among the tools available to a theorist, analytical models in particular are a vital resource that can give a great deal of insight into the problem being studied. In this chapter we outline the main analytical models and theoretical topics that are relevant to the work published in this thesis.

# 2.2 Black holes

Black holes are one of the more intriguing predictions of Einstein's 1915 theory of general relativity, but for a long time were not accepted as mainstream in the field of theoretical physics, despite being introduced in a conceptual fashion over a century previously. The central paradigm to general relativity is the notion of curvature by both matter and energy in a four-dimensional *spacetime*; a black hole is the most extreme example of this, with spacetime reaching infinite curvature in the centre. Particles that orbit in the vicinity of a black hole follow paths that increasingly depart from standard Newtonian theory as their distance from the hole decreases, finally reaching a radius within which all possible paths lead further into the potential well. This property of black hole applies to photons as well as massive<sup>1</sup> particles, making these objects truly 'dark' in the sense that no light can escape.

The radius at which this occurs is known as the *event horizon*, which quite simply delineates the region of spacetime that cannot communicate with (i.e., from which events cannot affect) an outside observer. The behaviour of particles inside this radius is therefore unknown, although Einstein's equations predict that all of the mass accreted by the black hole will eventually end up at an infinitesimal central point known as a *singularity*. Such a concept is at odds with the quantum theory of matter, however, and efforts to marry the two theories have been a large part of theoretical physics research over the last century. A unifying theory of quantum gravity would determine the eventual fate of matter inside a black hole from quantum mechanical arguments, removing the need for singularities which represent a breakdown in physical laws (Rovelli, 2000).

<sup>&</sup>lt;sup>1</sup>i.e., possessing finite rest mass

## 2.2.1 Formation of a black hole

The end state in the life-cycle of a massive star  $(M \gtrsim 20 \,\mathrm{M_{\odot}})$  is believed to be a black hole (Shapiro and Teukolsky, 1983). Theoretical work on the upper limit to the mass of a white dwarf by Chandrasekhar (1931) implied that a star with a large enough mass would eventually collapse beyond electron degeneracy pressure, with nothing to prevent the collapse from continuing and eventually forming a singularity. This was since superseded by the discovery of a stable configuration at a denser state through neutron degeneracy pressure (Oppenheimer and Volkoff, 1939), known as a *neutron star*. The maximum mass of this stable state is  $\gtrsim 3 \,\mathrm{M_{\odot}}$  (Cameron, 1959), and so once this limit is exceeded, collapse will inevitably lead to the formation of a black hole.

The process of gravitational collapse of a massive star can create stellar mass black holes with  $M \sim 3 - 10 \,\mathrm{M_{\odot}}$ , but is not a viable mechanism for the production of intermediate  $(M \sim 10^2 - 10^4 \,\mathrm{M_{\odot}})$  or supermassive black holes  $(M \gtrsim 10^5 \,\mathrm{M_{\odot}})$ , unless the progenitor star was significantly more massive than our understanding of stellar structure (and observations of the local Universe) suggests. The collapse of a hypothetical population III star - a zero-metallicity star formed in the early Universe, believed to have a mass of  $\sim 10^3 \,\mathrm{M_{\odot}}$  (Madau and Rees, 2001) - may provide a means of forming a black hole in the intermediate mass range, but is still not sufficient to produce an SMBH. Current thinking is therefore that black holes grew to larger masses through accretion of matter, although it should be noted that our understanding of this topic is still somewhat lacking, since although there is a great deal of observational evidence for stellar mass black holes and SMBHs, so far no reliable detection of an intermediate mass black hole has been made.

#### 2.2.2 Structure of a black hole

If collapse proceeds in an entirely spherically symmetric fashion, a *Schwarzchild* black hole will be formed. The general relativistic solution for the spacetime around such a spherical body was derived from Einstein's field equations by Karl Schwarzchild in 1916, although at the time the mathematical treatment was not realised to be applicable to the external gravitational field of a black hole. This formation scenario is idealised, however, as all bodies in the Universe tend to possess some angular momentum; the Schwarzchild solution is therefore a limiting case. A more general solution was found later, by Kerr (1963) for a *rotating* black hole.

Both of these cases can be characterised very simply, as for an uncharged black hole only two observable (classical) parameters are required: mass and spin. The 'no-hair' theorem (as it was coined by J. A. Wheeler) postulates that all other information is lost within the event horizon and cannot be extracted. For this to be the case, any higher order multipole moments that a star possesses as it collapses into a black hole must be lost, so that the black hole can be described entirely by a monopole (i.e., spherically symmetric). It is thought that this information is radiated away in the form of gravitational waves during the last stages of collapse (Shapiro and Teukolsky, 1983).

Although it does not possess a hard surface, a black hole has a characteristic size that is determined by its event horizon; this is located at the Schwarzchild radius, given by

$$r_S = \frac{2GM}{c^2} \tag{2.1}$$

where M is the mass of the black hole, and c is the speed of light. Since this is  $\propto M$ , low-mass black holes have a greater density<sup>2</sup> than their more massive counterparts. Far from the Schwarzschild radius,  $r \gg r_S$ , the gravitational field is described well by a standard Newtonian inverse-square law; however as  $r \to r_S$ , general relativistic effects due to spacetime curvature become increasingly important. For  $r < r_S$ , all matter must end up at the singularity.

In the case of a rotating, Kerr black hole, an additional boundary exists within which the underlying spacetime is itself rotating. Matter within this horizon is unable to 'stand still' and instead must possess some angular momentum. The region of spacetime within which this occurs is known as the ergosphere, and leads to an effect known as *framedragging* that can, through the Bardeen-Petterson effect (Bardeen and Petterson, 1975), act to co-align or counter-align an orbiting accretion disc with the equatorial plane of the black hole (King et al., 2005).

#### 2.2.3 Particle dynamics in the vicinity of a black hole

The purpose of this theoretical section is to determine where *accretion*, namely the adding of mass to a black hole, actually takes place. We therefore proceed to derive the radius at which this occurs, and determine what it means for the radiative efficiency of accretion on to a black hole, as was discussed in Chapter 1. We start by considering the dynamics of particles in Newtonian gravity.

#### 2.2.3.1 Newtonian approximation

In Newtonian gravity, a particle at position r moving with speed v in orbit of a body with mass M has a specific Hamiltonian (the total energy per unit mass) of

$$E = \frac{1}{2} \left(\frac{\mathrm{d}r}{\mathrm{d}t}\right)^2 + V(r) \tag{2.2}$$

 $<sup>^{2}</sup>$ although it is unclear how relevant the term *density* is in this context, since all of the mass is presumed to be located at an infinitesimal point at the centre



Figure 2.1: Effective potential (red) of a particle with specific angular momentum 1/5 of that of the Keplerian value at r = 1 kpc, moving in the Newtonian gravitional potential (black) of a black hole with  $M_{\rm bh} = 10^8 \,\mathrm{M_{\odot}}$ . The lowest energy orbit (marked by a cross) is single valued for V(r) = r, corresponding to a circle.

where V(r) is the *effective potential* that takes into account the angular momentum of the particle in addition to the potential energy:

$$V(r) = \frac{J^2}{2r^2} - \frac{GM}{r}$$
(2.3)

where J is the specific angular momentum. The effective potential of a particle moving under Newtonian assumptions is plotted in Figure 2.1, along with the potential corresponding to zero angular momentum. We can characterise the orbits of a particle with finite angular momentum based on the solution(s) of V(r) in terms of r.

- Bound orbits have V(r) < 0. There are two solutions to V(r) = r, corresponding to an apocentre,  $r_{\rm ap}$ , and a pericentre,  $r_{\rm pe}$ , of an elliptic orbit. The exception to this is the singular point at the bottom of the effective potential well where V/r = 0, which defines a *circular* orbit where  $r_{\rm ap} = r_{\rm pe}$ .
- Unbound orbits have V(r) > 0. There is only one solution to V(r) = r, corresponding to a pericentre. These orbits are hyperbolic, and escape the system.
- Marginally bound orbits have V(r) = 0. Technically there are two solutions to V(r) = r, but the apocentre solution is at infinity, giving a parabolic orbit. A particle moving out on a such an orbit is said to have the minimum energy to escape, while a particle coming in has the maximum energy to be captured.

As we can see from Figure 2.1, finite angular momentum acts as a potential *barrier*, preventing a particle from reaching r = 0. This concept holds for Newtonian gravity, but is not the case when general relativity is taken into account, as we will now see.

#### 2.2.3.2 General relativistic approach

In general relativity the motion of a particle is modified by the fact that the *energy* contributes to the gravitating mass of the system, and thus to the potential. While this is always the case, it only produces noticeable effects once the gravitational potential energy becomes of the order of the rest mass energy,  $GMm/r \sim mc^2$ . In this limit the Newtonian approximation breaks down. The following analysis is based on the corresponding section in Hobson et al. (2006), to which the reader is directed for more detail.

The Hamiltonian, or 'energy' equation for a particle under the influence of general relativistic effects in a Schwarzschild geometry is given by

$$\frac{1}{2} \left(\frac{\mathrm{d}r}{\mathrm{d}\tau}\right)^2 + \frac{J^2}{2r^2} \left(1 - \frac{2GM}{c^2r}\right) - \frac{GM}{r} = \frac{c^2}{2} \left(\frac{E^2}{m^2c^4} - 1\right)$$
(2.4)

where J is again the specific angular momentum of the particle, E is the total (nonspecific) orbital energy, and m is the rest mass. We can define an effective potential V(r) such that

$$V(r) = \frac{J^2}{2r^2} - \frac{GM}{r} - \frac{GMJ^2}{c^2r^3}$$
(2.5)

where we see that there is an additional term  $\propto 1/r^3$  in comparison to the Newtonian approximation. This term tends to zero as  $c \to \infty$ , i.e., in the non-relativistic limit.

The effective potential is plotted in Figure 2.2 for different values of the specific angular momentum, J. The curves are similar to those for the Newtonian effective potential (Figure 2.1) at large radii, with a minimum corresponding to a stable circular orbit, but take an entirely different form as  $r \to 0$ . We find in fact that a *maximum* occurs at small radii, which corresponds to an *unstable* circular orbit. The extrema of the effective potential are found at

$$r = \frac{J}{2GM} \left( J \pm \sqrt{J^2 - \frac{12G^2M^2}{c^2}} \right)$$
(2.6)

and are plotted in Figure 2.2. The red curve is the particular case of  $J = 2\sqrt{3} GM/c$ , where there is only one solution to the above equation and therefore only one extremum. This defines the *innermost stable circular orbit*, or ISCO, which is eventually reached as the gas orbiting the central SMBH loses its angular momentum. Beyond the ISCO, no stable orbits are found, and the gas falls in on a dynamical time, disappearing beneath  $r_s$ .

Recalling our discussion in Chapter 1, Section 1.6, we are now in a position to calculate the radiative efficiency of accretion on to a Schwarzschild black hole precisely. We require the gravitational binding energy (the orbital energy E) that is "lost" in bringing a particle from infinity to the ISCO; this we assume to have been radiated away (beneath the ISCO the particle falls beyond  $r_S$  too quickly to radiate further). The efficiency is the ratio of this radiated energy to the rest mass energy  $mc^2$ , namely,

$$\eta_{\rm acc} = \frac{mc^2 - E}{mc^2} = 1 - \frac{E}{mc^2} \tag{2.7}$$

where we can obtain the last term from equation 2.4. Substituting in for  $r = 6GM/c^2$  and  $J = 2\sqrt{3} GM/c$ , with  $\dot{r} = 0$  for a circular orbit we find that

$$\frac{E}{mc^2} = \frac{2\sqrt{2}}{3} \tag{2.8}$$

and so the radiative efficiency of accretion is

$$\eta_{\rm acc} \approx 1 - 0.943 = 0.057 \tag{2.9}$$

a figure that is slightly lower than our order of magnitude prediction in Chapter 1, but nonetheless far higher than the efficiency of nuclear burning in a star ( $\eta_{\text{nuclear}} \approx 0.007$ ). However, we note that so far we have only considered the ideal case of a *non-rotating* black



Figure 2.2: Effective potential of a particle with  $J = 2\sqrt{7} GM/c$  (blue),  $J = 2\sqrt{3} GM/c$  (red), and J = 0 (black; a standard Newtonian potential) moving in the general relativistic potential of a black hole with  $M_{\rm bh} = 10^8 \,\mathrm{M_{\odot}}$ , where the axes are normalised so that  $r_S = 1$ . For the two orbits with finite angular momentum there are minima in V(r) (marked by a diagonal cross) corresponding to a (stable) circular orbit for each, as well maxima (marked by an upright cross) that indicate *unstable* circular orbits. When the two extrema coincide, as is the case for the red curve, we find the ISCO, at  $r_{\rm ISCO} = 3 r_S$ .

hole; in reality any matter that is accreted will possess some angular momentum, which will be imparted to the black hole and thus to the underlying spacetime. To this end we must re-do our analysis with a Kerr, or rotating, black hole.

Once again we define the Hamiltonian for a massive particle, which in a Kerr geometry (assuming the motion is in the equatorial plane) is given by (Kerr, 1963),

$$\frac{1}{2}\left(\frac{\mathrm{d}r}{\mathrm{d}\tau}\right)^2 + \frac{J^2 - a^2c^2(k^2 - 1)}{2r^2} - \frac{GM(J^2 - ack)^2}{c^2r^3} - \frac{GM}{r} = \frac{c^2}{2}\left(k^2 - 1\right)$$
(2.10)

where we have written  $k \equiv E/mc^2$ . It is important to note that a new parameter, a (the angular momentum per unit mass of the black hole) has been introduced; this describes the *spin* of the hole, and in the limit  $a \to 0$ , we recover the Schwarzschild solution as above. We write an effective potential V(r) as

$$V(r) = \frac{J^2 - a^2 c^2 (k^2 - 1)}{2r^2} - \frac{GM(J^2 - ack)^2}{c^2 r^3} - \frac{GM}{r}$$
(2.11)

where we see that in the limits  $a \to 0$  and  $c \to \infty$  we recover the Newtonian approximation. It should be pointed out here that it is somewhat incorrect to identify this expression as an effective potential, since it itself depends on the orbital energy, E. Nonetheless it does correctly describe the stability of circular orbits in the same way as above, and so we proceed to use it here.

While the analysis follows the same route as in the Schwarzschild case, it is considerably more involved algebraically, and so we confine it to Appendix A. The ISCO for a rotating black hole is found from the solution to the equation

$$r^{2} - \frac{6GMr}{c^{2}} - 3a^{2} \mp 8a \left(\frac{GMr}{c^{2}}\right)^{1/2} = 0$$
(2.12)

where the upper sign corresponds to an orbit that is *counter*-rotating with respect to the black hole spin parameter, and the lower sign to an orbit that is *co*-rotating. In the limit that  $a \to 0$  we have  $r^2 = 6GMr/c^2$ , giving us back the Schwarzschild ISCO. In the opposite limit that the spin parameter takes the maximum possible value<sup>3</sup> of  $a = GM/c^2$ , we have a quartic in  $\sqrt{r}$ , which we solve by inspection to give  $r_{\rm ISCO} = 9GM/c^2$  for the counterrotating case and  $r_{\rm ISCO} = GM/c^2$  for the co-rotating case. If we consider the radiative efficiency of the more bound of the two (namely, for maximal spin and a co-rotating orbit), we find that

$$\eta_{\rm acc} = 1 - \frac{1}{\sqrt{3}} \approx 0.42$$
 (2.13)

which is far larger than the Schwarzschild value. In reality neither extreme (zero spin vs.

<sup>&</sup>lt;sup>3</sup>according to the cosmic censorship conjectore (see, e.g., Penrose, 1973)

maximal spin) is likely, and so the radiative efficiency of accretion will lie somewhere in between,  $0.057 < \eta_{\rm acc} < 0.42$ , making our order of magnitude estimate in Chapter 1 a reasonable one. Nonetheless, an important conclusion here is that a more rapidly rotating black hole will possess a significantly larger conversion efficiency of rest mass to luminosity as matter is accreted.

# 2.3 Accretion onto compact objects

Accretion in astrophysics refers to the release of gravitational energy by infalling matter as it accumulates on to a central object. The amount of energy yielded from accretion is dependent on the 'compactness' of the accreting object,  $M_*/R_*$ , where  $M_*$  is the mass of the accretor and  $R_*$  is the radius. The gravitational energy that is released by the accretion of a mass m on to a compact accretor is given by

$$\Delta E_{\rm acc} = \frac{GM_*m}{R_*} \tag{2.14}$$

where G is the gravitational constant. Putting in some typical numbers, for, say, a neutron star - with  $M_* \sim 1 \,\mathrm{M}_{\odot}$  and  $R_* \sim 10^5 \,\mathrm{cm}$  - can help us to see that the yield from accretion on to a sufficiently compact object can be significantly higher than any other process of energy generation. Nuclear fusion, for example, has an efficiency of conversion of rest mass into energy of 0.7%, giving an energy release of  $\Delta E_{\rm acc} = 0.007 mc^2$ , where c is the speed of light. Per unit mass, this corresponds to  $\simeq 6 \times 10^{18} \,\mathrm{erg g^{-1}}$ . The accretion on to a neutron star with these approximate numbers yields  $\simeq 10^{20} \,\mathrm{erg g^{-1}}$ , about twenty times that of nuclear fusion. However this comparison is not so favourable for more extended accretors. White dwarfs, on the other hand, are less compact than neutron stars (but of similar mass), having  $R_* \sim 10^9 \,\mathrm{cm}$ ; accretion on to these objects loses out to nuclear burning by factors of tens. For a main sequence star, the accretion yield is smaller still. The efficiency of accretion in releasing energy is thus a strong function of the compactness of the accretor, and can be very large provided the object that we are dealing with is sufficiently dense.

The luminosity derived from accretion at a particular mass accretion rate M can be correspondingly large. If we assume that the entirety of the kinetic energy of the infalling matter is converted into electromagnetic radiation at the surface of the accreting object, we find an accretion luminosity as

$$L_{\rm acc} = \frac{GM_*}{R_*}\dot{M} \tag{2.15}$$

which for white dwarfs and neutron stars (assuming a typical accretion rate from close binary systems) is of the order of  $10^{33}$  erg s<sup>-1</sup> and  $10^{36}$  erg s<sup>-1</sup>. For the case of a black

hole, however, the assumption of 100% efficiency in converting the rest mass energy into radiation no longer holds. As we discussed in the previous section, matter is not accreting on to a 'hard surface' but rather the black hole's event horizon, or Schwarzschild radius, within which radiation cannot escape. We therefore require an efficiency parameter,  $\eta$ , for the conversion of mass accretion into luminosity. The accretion luminosity from a black hole can then be written as

$$L_{\rm acc} = 2\eta \frac{GM_{\rm bh}}{R_{\rm bh}} \dot{M} \tag{2.16}$$

where, by putting  $R_{\rm bh} = 2GM/c^2$ , the Schwarzschild radius, we find

$$L_{\rm acc} = \eta \dot{M} c^2 \tag{2.17}$$

We discussed the realistic values that  $\eta$  is likely to take in 2.2, and found that the typical conversion efficiency for stellar mass black holes is  $\simeq 10\%$  ( $\eta \simeq 0.1$ ). This is comparable with the energy release for accretion on to a neutron star. It is clear then that accretion on to black holes or neutron stars is a likely candidate for powering some of the most luminous phenomena in the Universe, as indeed is believed to be the case; X-ray binaries<sup>4</sup>, with  $L \sim 10^{39}$  erg s<sup>-1</sup>, quasars, with  $L \sim 10^{46}$  erg s<sup>-1</sup>, and gamma-ray bursts<sup>5</sup>, with  $L \sim 10^{52}$  erg s<sup>-1</sup>.

Of course, we have not taken into account the effect of the radiation on the accreting material. At high luminosities, the momentum transferred to the infalling matter can be significant, putting an upper limit on the accretion rate. This is known as the Eddington limit, and is an important concept in astronomy and astrophysics, providing a means of characterising the stability of an accreting object.

## 2.3.1 The Eddington limit

We can derive the Eddington limit for accretion by equating the inward gravitational force on accreting matter to the outward radiation pressure force. Under the assumptions of spherical symmetry, and the infalling matter being composed of ionized hydrogen (a reasonable assumption for stellar atmospheres or accretion flows on to compact objects), the radiation pressure force is given by

$$F_{\rm rad} = \frac{L\sigma_T}{4\pi r^2 c} \tag{2.18}$$

where L is the luminosity of the accreting object, and r is the radial distance of the infalling material from the centre of the accreting object. For an ionized gas, the radiation pressure acts on the electrons, which is why the Thomson scattering cross-section,  $\sigma_T$ , appears in

 $<sup>^{4}\</sup>mathrm{discussion}$  of these systems is beyond the scope of this thesis

 $<sup>^5\</sup>mathrm{as}$  is discussion of these

this equation<sup>6</sup>. Protons have too large a mass and their scattering cross-section is therefore negligible. However, due to the Coulomb force, as the electrons are pushed out they drag the protons with them, which means that the opposing gravitational force must include the masses of both. The force due to gravity on an electron-proton pair is

$$F_{\rm grav} = \frac{GM(m_p + m_e)}{r^2} \tag{2.20}$$

where  $m_p$ ,  $m_e$  are the mass of a proton and an electron, respectively, and M is the mass of the accreting object. Equating these two forces yields

$$L_{\rm Edd} = \frac{4\pi G M m_p c}{\sigma_T} \tag{2.21}$$

where we have neglected the mass of the electron  $(m_e \ll m_p)$ . At greater luminosities than this the outward radiation pressure would exceed the gravitational attraction of the infalling material, and accretion would be halted. The Eddington limit for luminosity thus implies a corresponding limit on the mass accretion rate:

$$\dot{M}_{\rm Edd} = \frac{4\pi G M m_p}{\eta \sigma_T c} \tag{2.22}$$

where we have employed the radiative efficiency parameter  $\eta$ . This particular limit allows us to determine a timescale on which a mass can grow through accretion; if we define a timescale  $\tau$  in the usual way i.e., as an *e-folding* time, with an initial mass  $M_0$  our growth is limited to

$$M(t) \le M_0 e^{t/\tau} \tag{2.23}$$

where

$$\tau = \frac{\eta \sigma_T c}{4\pi G m_p} \tag{2.24}$$

This defines the Salpeter timescale for accretion,  $t_S = \tau = M/\dot{M}_{\rm Edd}$ , and is the typical maximum rate than one can grow an object through accretion processes alone. We should be careful here, however, as so far we have assumed accretion to proceed spherically symmetrically. In fact this will not be the case in practice, as infalling material will possess some angular momentum and most likely accrete through an accretion disc (see Section 2.3.4). Nonetheless, our spherically symmetric picture has allowed us to derive an important quantity - the Eddington limit, which may change a small factors with a change in the geometry of the accretion flow but still demonstrates that an object that is

$$\sigma_T = \frac{8\pi}{3} \left(\frac{e^2}{m_e c^2}\right)^2 = 6.65 \times 10^{-25} \text{cm}^2$$
(2.19)

where e is the charge on an electron.

 $<sup>^{6}</sup>$  the Thomson scattering cross-section is given by

at least partially stable must not be accreting (or radiating) at more than  $L_{\rm Edd}$ . We should also note that there was a further assumption that was implicitly included in the above treatment, which is that of a *steady* accretion flow. Our conclusion of Eddington-limited accretion, or luminosity was valid only assuming a steady-state; the limit can of course be exceeded by one-off events such as a sudden burst of accretion, or a supernova explosion.

Now that we have introduced the relevant concepts for accretion on to a compact object, we must go into some more detail as to how this process actually proceeds. For gas accretion the picture is complicated by fluid properties such as the pressure and the sound speed; we begin by considering the simplest possible case of Bondi (or spherically symmetric) accretion.

#### 2.3.2 Spherically symmetric accretion

Steady, spherically symmetric accretion of gas on to a point mass M was studied by Bondi (1952), to which the majority of this derivation is similar. We work in spherical polar coordinates  $(r, \theta, \phi)$ , and consider purely radial infall with velocity  $v_r = -v$ , where the negative sign indicates inward motion of material; for an outflowing wind the analysis is the same but with  $v_r > 0$ . The steady mass flux is described by the time-independent continuity equation:

$$\nabla \cdot (\rho v) = \frac{1}{r^2} \frac{\mathrm{d}}{\mathrm{d}r} (r^2 \rho v) = 0 \qquad (2.25)$$

where  $\rho$  is the mass density. Integrating this equation gives us  $r^2\rho v = \text{const.}$ , and when compared with the mass flux through a spherical surface,  $\dot{M} = 4\pi r^2 \rho v$ , we see that this integration constant must be  $\dot{M}/4\pi$ . We call the mass flux the *accretion rate*, for an infall given by

$$\dot{M} = -4\pi r^2 \rho v \tag{2.26}$$

For our analysis we require an equation of state (EQS), which we assume to be barotropic<sup>7</sup> and given by

$$P = K\rho^{\gamma} \tag{2.27}$$

where P is the pressure, and K is a constant. We shall not specify the value of the adiabatic index,  $\gamma$ , until later in the derivation. Finally we relate the radial velocity profile of the fluid to the underlying forces, via the (steady) Euler equation:

$$v\frac{\mathrm{d}v}{\mathrm{d}r} + \frac{1}{\rho}\frac{\mathrm{d}P}{\mathrm{d}r} + \frac{GM}{r^2} = 0$$
(2.28)

While the full solution for the variation of the fluid variables with r requires integrating this equation (which we shall do in a moment), we can first determine an important property

<sup>&</sup>lt;sup>7</sup>defined as where the pressure is a function only of the density

of the flow from simple re-arrangement and substitution. We use the equation of state to relate the pressure gradients to density gradients, namely

$$\frac{\mathrm{d}P}{\mathrm{d}r} = \frac{\mathrm{d}P}{\mathrm{d}\rho}\frac{\mathrm{d}\rho}{\mathrm{d}r} \tag{2.29}$$

and

$$\frac{\mathrm{d}P}{\mathrm{d}\rho} = \gamma K \rho^{\gamma - 1} \tag{2.30}$$

where we can clearly see that  $\gamma K \rho^{\gamma-1} = \gamma P / \rho$ . This is the definition of the sound speed, i.e.,

$$c_s^2 = \frac{\gamma P}{\rho} \tag{2.31}$$

which gives us

$$\frac{\mathrm{d}P}{\mathrm{d}r} = c_s^2 \frac{\mathrm{d}\rho}{\mathrm{d}r} \tag{2.32}$$

We can now substitute this into equation 2.28. We may also use equation 3.19 to relate the change in density to r and v, namely

$$\frac{1}{\rho}\frac{\mathrm{d}\rho}{\mathrm{d}r} = -\frac{1}{vr^2}\frac{\mathrm{d}}{\mathrm{d}r}(vr^2) \tag{2.33}$$

and putting both of these into the Euler equation gives us

$$v\frac{\mathrm{d}v}{\mathrm{d}r} - \frac{c_s^2}{vr^2}\frac{\mathrm{d}}{\mathrm{d}r}(vr^2) + \frac{GM}{r^2} = 0$$
(2.34)

which can be re-arranged to give

$$\frac{1}{4} \left(c_s^2 - v^2\right) \frac{\mathrm{d}}{\mathrm{d}r} \ln v^2 = \frac{GM}{r^2} \left[1 - \left(\frac{2c_s^2 r}{GM}\right)\right]$$
(2.35)

Note here that as well as v,  $c_s$  will be a function of r except for in the limiting case of an isothermal equation of state i.e.,  $\gamma = 1$ . It can, however, clearly be seen from this equation that the LHS becomes zero when  $c_s(r) = v(r)$ , defining a particular radius in the flow that is referred to as the sonic point. If such a point exists (and indeed it does not have to) then it will occur where the factor  $1 - (2c_s^2 r/GM)$  reaches zero. The sonic point is therefore defined as,

$$r_{\rm s} = \frac{GM}{2c_s^2(r_s)} \tag{2.36}$$

where  $c_s^2(r_s)$  is the value of the sound speed at  $r_s$ . We should be careful here, however, as the LHS will also vanish in the event that

$$\frac{\mathrm{d}}{\mathrm{d}r}v^2 = 0 \tag{2.37}$$

Thus at  $r = r_s$  there are two types of solution; those which are *transonic* and transition between  $v^2 > c_s^2$  and  $c_s^2 > v^2$ , and those that simply reach a stationary point in the Mach number. In fact these types of solution partition yet further depending on our assumptions for the inner and outer boundary condition; we will discuss these later when we plot the full solutions for v and  $c_s$ . We now proceed to integrate equation 2.28 to obtain Bernoulli's equation:

$$\frac{v^2}{2} + \int_{P_{\infty}}^{P} \frac{1}{\rho} \,\mathrm{d}P - \frac{GM}{r} = \text{constant}$$
(2.38)

where  $P_{\infty}$  and P are the limits of the integral in terms of the pressure at infinity and the pressure at a particular radius. Putting in boundary conditions at infinity we find the value of the constant; assuming that v = 0, and  $P = P_{\infty}$ ,  $\rho = \rho_{\infty}$  at  $r = \infty$ , we obtain

$$\frac{v^2}{2} + \int_{P_{\infty}}^{P} \frac{1}{\rho} dP - \frac{GM}{r} = 0$$
 (2.39)

Our pressure and density are related by the equation of state, which here we write as,

$$\frac{P}{P_{\infty}} = \left(\frac{\rho}{\rho_{\infty}}\right)^{\gamma} \tag{2.40}$$

By converting the integral to one in terms of density, via,

$$\frac{\mathrm{d}P}{\mathrm{d}\rho} = \frac{\gamma P_{\infty}}{\rho_{\infty}} \left(\frac{\rho}{\rho_{\infty}}\right)^{\gamma-1} \tag{2.41}$$

we can now write equation 2.39 in a different form, namely

$$\frac{v^2}{2} + \frac{\gamma}{\gamma - 1} \frac{P_{\infty}}{\rho_{\infty}} \left[ \left( \frac{\rho}{\rho_{\infty}} \right)^{\gamma - 1} - 1 \right] = \frac{GM}{r}$$
(2.42)

where we see that this is only valid for  $\gamma > 1$ . In order to find the solution to this equation we must introduce the Mach number as a variable,  $\mathcal{M} = v/c_s$ . Writing this in terms of values at infinity we have

$$\mathcal{M} = \frac{v}{c_{\infty}} \left(\frac{\rho}{\rho_{\infty}}\right)^{\frac{1-\gamma}{2}} \tag{2.43}$$

since  $c_s^2 \equiv dP/d\rho$ , and  $c_{\infty}^2 = \gamma P_{\infty}/\rho_{\infty}$ . From this we can obtain expressions for  $\rho$  and v in terms of  $\mathcal{M}$ , i.e.,

$$\rho = \rho_{\infty} \left(\frac{v}{c_{\infty}\mathcal{M}}\right)^{\frac{2}{\gamma-1}} \tag{2.44}$$

and

$$v = c_{\infty} \mathcal{M} \left(\frac{\rho}{\rho_{\infty}}\right)^{\frac{\gamma-1}{2}}$$
(2.45)

which we substitute into our expression for the accretion rate (equation 2.26) to parametrise these quantities in terms of r:

$$\dot{M} = 4\pi r^2 \rho_{\infty} \left(\frac{1}{c_{\infty} \mathcal{M}}\right)^{\frac{2}{\gamma-1}} v^{\frac{\gamma+1}{\gamma-1}}$$
(2.46)

$$\dot{M} = 4\pi r^2 \rho_{\infty} \left(\frac{1}{c_{\infty} \mathcal{M}}\right)^{\frac{2}{\gamma-1}} (c_{\infty} \mathcal{M})^{\frac{\gamma+1}{\gamma-1}} \left(\frac{\rho}{\rho_{\infty}}\right)^{\frac{\gamma+1}{2}}$$
(2.47)

so that

$$v = \left[\dot{M}^{-1} 4\pi r^2 \rho_{\infty} \left(\frac{1}{c_{\infty} \mathcal{M}}\right)^{\frac{2}{\gamma-1}}\right]^{\frac{1-\gamma}{\gamma+1}}$$
(2.48)

$$\rho = \left[\dot{M}^{-1} 4\pi r^2 \rho_{\infty} \left(\frac{1}{c_{\infty} \mathcal{M}}\right)^{\frac{2}{\gamma-1}} (c_{\infty} \mathcal{M})^{\frac{\gamma+1}{\gamma-1}}\right]^{\frac{-2}{\gamma+1}}$$
(2.49)

We substitute these expressions into equation 2.42, and note that the sound speed at infinity is given by  $c_{\infty} = \gamma P_{\infty}/\rho_{\infty}$ , such that,

$$\frac{1}{2} \left[ \dot{M}^{-1} r^2 A \right]^{\frac{2(1-\gamma)}{\gamma+1}} \mathcal{M}^{\frac{4}{\gamma+1}} + \frac{1}{\gamma-1} \left[ \dot{M}^{-1} r^2 A \right]^{\frac{2(1-\gamma)}{\gamma+1}} \mathcal{M}^{\frac{2(1-\gamma)}{\gamma+1}} = \frac{GM}{r} + \frac{c_{\infty}^2}{\gamma-1}$$
(2.50)

where  $A \equiv 4\pi \rho_{\infty} (1/c_{\infty})^{\frac{2}{\gamma-1}}$ . Multiplying through by the common factor in square brackets (and simplifying) allows us to separate the variables  $\mathcal{M}$  and r:

$$\frac{1}{2}\mathcal{M}^{4/(\gamma+1)} + \frac{1}{\gamma+1}\mathcal{M}^{\frac{2(1-\gamma)}{\gamma+1}} = \dot{M}^{\frac{2(1-\gamma)}{\gamma+1}} \left[ \left(4\pi\rho_{\infty}c_{\infty}r^{2}\right)^{\frac{2(\gamma-1)}{\gamma+1}} \left(\frac{1}{r}\frac{GM}{c_{\infty}^{2}} + \frac{1}{\gamma-1}\right) \right] \quad (2.51)$$

which takes the form of an important relation that we shall use later, namely

$$f(\mathcal{M}) = \dot{M}^{\frac{2(1-\gamma)}{\gamma+1}}g(r)$$
(2.52)

where f and g are functions of the Mach number and the radial coordinate respectively, with

$$f(\mathcal{M}) = \frac{1}{2}\mathcal{M}^{\frac{4}{\gamma+1}} + \frac{1}{\gamma+1}\mathcal{M}^{\frac{2(1-\gamma)}{\gamma+1}}$$
(2.53)

$$g(r) = (4\pi\rho_{\infty}c_{\infty}r^2)^{\frac{2(\gamma-1)}{\gamma+1}} \left(\frac{1}{r}\frac{GM}{c_{\infty}^2} + \frac{1}{\gamma-1}\right)$$
(2.54)

It is clear that identifying these functions is a crucial step in determining the solution for the accretion rate, as we must now have

$$\dot{M} = \left[\frac{f(\mathcal{M})}{g(r)}\right]^{\frac{\gamma+1}{2(1-\gamma)}} \tag{2.55}$$

As we shall see when we plot the full solution, the maximum (steady) accretion rate for our boundary conditions is attained at the stationary points of f and g. We therefore take the derivative of each, starting with

$$f'(\mathcal{M}) = \frac{2}{\gamma+1} \mathcal{M}^{\frac{3-\gamma}{\gamma+1}} - \frac{1}{\gamma+1} \mathcal{M}^{\frac{1-3\gamma}{\gamma+1}}$$
(2.56)

where the stationary point,  $f'(\mathcal{M}) = 0$ , is reached when  $\mathcal{M} = 1$ . We then have

$$g'(r) = GM \frac{5 - 3\gamma}{\gamma + 1} r^{\frac{2\gamma - 6}{\gamma + 1}} - \frac{c_{\infty}^2}{\gamma + 1} r^{\frac{3\gamma - 5}{\gamma + 1}}$$
(2.57)

with the stationary point corresponding to g'(r) = 0 at  $r = GM(5 - 3\gamma)/4c_{\infty}^2$ . Putting these back in we obtain the extremum for each function as

$$f_m = \frac{\gamma + 1}{2(\gamma - 1)} \tag{2.58}$$

$$g_m = \left(\frac{4\pi\rho_{\infty}G^2M^2}{c_{\infty}^3}\right)^{\frac{2(\gamma-1)}{\gamma+1}} \frac{1}{4}\frac{\gamma+1}{\gamma-1} \left[\frac{5-3\gamma}{4}\right]^{-\frac{5-3\gamma}{\gamma+1}}$$
(2.59)

These extrema coincide at the sonic point,  $\mathcal{M} = 1$  and  $r = r_s = GM/2c_s^2(r_s)$ , where we can see that the latter is true from equation 2.74, derived later. This maximum value of  $\dot{M}$  that we obtain from the ratio of the two functions at their stationary points is therefore the *transonic* accretion rate, and is given by

$$\dot{M}_{\max} = \left(\frac{4\pi\rho_{\infty}G^2M^2}{c_{\infty}^3}\right) \left(\frac{1}{2}\right)^{\frac{\gamma+1}{2(\gamma-1)}} \left(\frac{5-3\gamma}{4}\right)^{\frac{5-3\gamma}{2(1-\gamma)}}$$
(2.60)

It is convenient to introduce a non-dimensional parameter,  $\lambda$ , to describe this accretion rate, such that  $\lambda = \dot{M}/(4\pi\rho_{\infty}G^2M^2c_{\infty}^{-3})$ . At this juncture we consider the dependance of this parameter on the value of  $\gamma$ . It is instructive to see how  $\lambda$  can vary between the two extreme values of the adiabatic index, namely  $\gamma = 1$  (isothermal, corresponding to instantaneous cooling) and  $\gamma = 5/3$  (adiabatic, corresponding to an infinite cooling time). However as is often the case with extrema, these two limiting cases are somewhat non-trivial to solve. We begin with the isothermal condition.

• Isothermal case,  $\gamma = 1$ 

As can be seen from equation 2.60, substituting for  $\gamma = 1$  causes the solution for  $\lambda$  to blow up to infinity. We therefore employ a change of variable approach and take the *limit* as  $\gamma \to 1$ . Putting  $\alpha = \gamma - 1$ , our equation becomes,

$$\lambda_{\max} = \left(\frac{1}{2}\right)^{\frac{\alpha+2}{2\alpha}} \left(\frac{2-3\alpha}{4}\right)^{-\frac{2-3\alpha}{2\alpha}}$$
(2.61)

We seek a solution of the form,

$$e = \lim_{x \to 0} (1+x)^{1/x} \tag{2.62}$$

which, as the equation suggests, returns the base of the natural logarithm in the limit that x tends to zero. We therefore proceed to separate out the fractional exponents in our formula, such that

$$\lambda_{\max} = \left(\frac{1}{2}\right)^{1/2} \left(\frac{1}{2}\right)^{1/\alpha} \left[\frac{1}{2} - \frac{3\alpha}{4}\right]^{-1/\alpha + 3/2}$$
(2.63)

Re-arranging gives,

$$\lambda_{\max} = \left(\frac{1}{2}\right) \left[1 - \frac{3\alpha}{2}\right]^{-1/\alpha} \left[1 - \frac{3\alpha}{2}\right]^{3/2}$$
(2.64)

The last square bracket in this expression does not blow up in our limit, and so we can tentatively substitute  $\alpha = 0$  to obtain a value of unity. We are then left with

$$\lambda_{\max} = \left(\frac{1}{2}\right) \left[1 - \frac{3\alpha}{2}\right]^{-1/\alpha} \tag{2.65}$$

which we can modify to follow the form of equation 2.62 by taking a power of 3/2 out:

$$\lambda_{\max} = \left(\frac{1}{2}\right)^2 \left[ \left(1 - \frac{3\alpha}{2}\right)^{-\frac{2}{3\alpha}} \right]^{3/2}$$
(2.66)

such that via our identity we have

$$\lambda_{\max} = \frac{1}{4}e^{3/2} \approx 1.12$$
 (2.67)

• Adiabatic case,  $\gamma = 5/3$ 

This case is simpler, with the solution obtained by noting that as  $\gamma \rightarrow 5/3$ , the second bracket in equation 2.60 will tend to unity faster than it tends to zero, as the zero in the exponent is dominant over the zero inside the bracket. In the adiabatic

case we therefore have

$$\lambda_{\max} = \frac{1}{4} = 0.25 \tag{2.68}$$

The question of whether this maximum accretion rate is attained relates to particular families of solutions, as alluded to earlier. We can determine the run of  $\mathcal{M}$  with r - in other words, the full solution to the problem - by letting r vary, calculating the corresponding values of g(r), substituting those values into  $f(\mathcal{M})$  via equation 2.52, and subsequently solving for  $\mathcal{M}$ . This latter step we perform numerically, using the *Newton-Raphson* algorithm<sup>8</sup>. There are different solutions for different values of  $\lambda$ . Figure 2.3 shows a few of these solutions for  $\gamma = 4/3$ .

As can be seen from the Figure, the steady-state transonic solutions (where the flow changes from subsonic to supersonic or vice-versa at  $r_s$ ) correspond to  $\lambda = \lambda_{\text{max}}$ . The solutions that have  $\lambda < \lambda_{\text{max}}$  are either subsonic or supersonic everywhere, while those with  $\lambda > \lambda_{\text{max}}$  are double-valued in  $\mathcal{M}$  for a given r, and do not cover the full range in radius. The latter solutions are therefore non-physical, and we disregard them. Of those remaining we see that we have four types (Frank et al., 2002):

- 1. Type I:  $v(r_s) = c_s(r_s), v \to 0 \text{ as } r \to \infty$  (transonic)
- 2. Type II:  $v(r_s) = c_s(r_s), v \to 0 \text{ as } r \to 0$  (transonic)
- 3. Type III:  $v(r) < c_s(r_s), v \to 0$  in both limits,  $\frac{\mathrm{d}v}{\mathrm{d}r} = 0$  at  $r_s$  (subsonic)
- 4. Type IV:  $v(r) > c_s(r_s), v \to \infty$  in both limits,  $\frac{\mathrm{d}v}{\mathrm{d}r} = 0$  at  $r_s$  (supersonic)

The type IV solutions can be elimated by virtue of our boundary conditions; that we require the gas to be at rest at infinity. Likewise, we can discard type II, although we note that this corresponds to the Bondi solution for time-reversed accretion, or in other words, a stellar wind (Parker, 1960; Lamers and Cassinelli, 1999). We are therefore left with types I and III for spherically symmetric accretion; steady-state transonic flow and subsonic flow respectively. We can see from Figure 2.3 that the transonic type I solution gives the maximum possible accretion rate for steady (constant  $\dot{M}$ ) flow, as v is everywhere higher than any of the type III solutions, and increasing v further would only serve to make the flow follow one of the unphysical, double-valued solutions.

For a compact object with a surface, steady-state subsonic flow (type III) is a valid solution. The physical interpretation of this case can be inferred from noting that as

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} \tag{2.69}$$

and the process repeated for  $x_1$  until a root is found to a user-defined accuracy

<sup>&</sup>lt;sup>8</sup>the Newton-Raphson algorithm, sometimes referred to as *Newton's method*, finds successively better approximations to the roots of a real-valued function by extrapolating the derivative. An initial guess,  $x_0$ , to the root of a function f(x), can be improved upon (providing that the function is reasonably well-behaved) by taking


Figure 2.3: Mach number as a function of the dimensionless radius  $r/r_s$ , where  $r_s = GM/2c_s^2(r_s)$  is the sonic point (equation 2.36), for a spherically symmetric gas flow in the (Newtonian) gravitational field of a point mass, with an adiabatic index of  $\gamma = 4/3$ . A few solutions for differing values of the dimensionless accretion rate parameter  $\lambda$  are shown. The transonic solutions are shown in red, while the other lines represent the solutions that do not pass through a sonic point.

 $v \rightarrow 0$ , the second term on the LHS in equation 2.28 (the Euler equation) dominates over the first term, giving

$$\frac{1}{\rho}\frac{\mathrm{d}P}{\mathrm{d}r} \approx -\frac{GM}{r^2} \tag{2.70}$$

which is the hydrostatic equilibrium condition, and for a polytropic EQS holds for  $1 \leq \gamma \leq 5/3$ . The flow is therefore prevented from reaching a sonic point by 'back pressure' from the large density gradient that builds up as matter accumulates at the hard surface of the star (Shapiro and Teukolsky, 1983).

In the case of a black hole, however, no hard surface exists, and only the transonic solution is allowed. The full proof of this requires a general relativistic treatment of the Bondi flow solutions, which is beyond the scope of this thesis; the reader is directed to Appendix G in Shapiro and Teukolsky (1983). Type I is therefore the solution most relevant to this thesis. The accretion rate in this case is given by

$$\dot{M} = \frac{4\pi\lambda\rho_{\infty}G^2M^2}{c_{\infty}^3} \tag{2.71}$$

where  $\lambda$  is given by equation 2.61. The flow is characterised by the sonic radius,  $r_s = GM/2c_s^2(r_s)$  (equation 2.36). We can relate the sound speed at this radius to the sound speed at infinity by re-writing equation 2.42 (the Bernoulli equation) as

$$\frac{v^2}{2} + \frac{c_s^2}{\gamma - 1} - \frac{GM}{r} = \frac{c_\infty^2}{\gamma - 1}$$
(2.72)

such that at  $r_s$ , where  $v = c_s$ , we have

$$c_s^2(r_s) \left[ \frac{1}{2} + \frac{1}{\gamma - 1} - 2 \right] = \frac{c_\infty^2}{\gamma - 1}$$
(2.73)

giving us

$$c_s(r_s) = c_\infty \left(\frac{2}{5-3\gamma}\right)^{1/2} \tag{2.74}$$

Likewise, we can obtain the value of  $\rho$  at  $r_s$  in terms of  $\rho_{\infty}$  through equation 2.44 and  $c_s \equiv dP/d\rho$  from equation 2.41:

$$\rho(r_s) = \rho_{\infty} \left[ \frac{c_s(r_s)}{c_{\infty}} \right]^{2/\gamma - 1}$$
(2.75)

It is important to point out, as can be seen from the denominator in the RHS of equation 2.74, in the limiting case of a purely adiabatic EQS with  $\gamma = 5/3$ , the sonic radius cannot be defined. A purely adiabatic flow cannot therefore reach a steady-state transmic solution, and the flow remains subsonic everywhere.

Finally, we note that there is a second radius that characterises the accretion flow, which can be obtained by considering the point at which the flow velocity becomes equal to the sound speed *at infinity* (rather than the sound speed at r, which defines the sonic point). This has the eponymous distinction of being known as the *Bondi* radius, and can be found from equation 2.72 by noting that at large r, the flow is subsonic and all quantities are close to their ambient values,  $\rho_{\infty}$ ,  $c_{\infty}$ , and  $v \approx 0$ . In this case equation 2.72 becomes

$$\frac{v^2}{2} = \frac{GM}{r} \tag{2.76}$$

As v increases it eventually reaches  $c_{\infty}$ , and so the Bondi radius is defined as

$$r_B = \frac{2GM}{c_\infty^2} \tag{2.77}$$

Within this radius  $\rho$  and  $c_s$  begin to increase above their ambient values, and as we have seen, eventually reach a sonic point  $r_s$ . The Bondi radius is therefore the first characteristic radius that the flow encounters as it falls from infinity, and  $r_B > r_s$ . The accretion rate is often written in terms of this fiducial parameter, by

$$\dot{M} = \pi \lambda r_B^2 \rho_\infty c_\infty \tag{2.78}$$

An alternative interpretation of the Bondi radius is that of a ratio of *energies*; it is approximately the point at which the gravitational potential energy becomes larger than the thermal energy in the gas, 3/2 kT = GM/r for a monatomic gas, where k is the Boltzmann constant and  $c_s^2 = kT/\mu m_p$ , with the assumption of  $c_s \approx c_\infty$ .

# 2.3.3 Bondi-Hoyle accretion

Bondi's spherically symmetric analysis can be modified to take account of a *moving* accretor in a stationary, uniform medium. This problem was first studied by Hoyle and Lyttleton (1939) and Bondi and Hoyle (1944), these authors finding that when the sound speed was negligible in comparison to the velocity of the accretor, and pressure effects could therefore be disregarded, the accretion flow was described by

$$\dot{M}_{\rm BHL} = \frac{2\pi\alpha G^2 M^2 \rho_{\infty}}{v_{\rm rel}^3} \tag{2.79}$$

where  $v_{\rm rel}$  is the relative velocity of the accretor with respect to the medium, and  $\alpha$  is a constant that is dependent on the exact geometry of the problem; in general  $1 \leq \alpha \leq 2$ . This accretion rate is to an order of magnitude the mass flux through the capture radius at which the flow is bound to the accretor - using  $v^2/2 = GM/r_{\rm capt}$  we see that  $\dot{M} = 4\pi r_{\rm capt}^2 \rho v \sim \dot{M}_{\rm BHL}$ . Bondi (1952) proposed that the latter case of negligible pres-

sure (namely, velocity-limited accretion) and the spherically symmetric case of negligible relative velocity (namely, temperature-limited accretion) could be combined by writing

$$\dot{M} = \frac{4\pi G^2 M^2 \rho_{\infty}}{(c_{\infty}^2 + v_{\rm rel}^2)^{3/2}}$$
(2.80)

thereby combining in a single formula the spherical accretion rate relating to flow at rest far from the accretor and the accretion rate from a body moving supersonically with respect to the medium; the limiting cases are recovered as  $v_{\rm rel} \rightarrow 0$  and  $c_{\infty} \rightarrow 0$  respectively.

As a final note in this this section we make the important point that the Bondi/Bondi-Hoyle picture of accretion is very idealised. In reality the mass that is accreting will not directly hit the accretor, but instead orbit it due to non-zero angular momentum. This is especially the case when accreting from large scales, as the accretor becomes effectively a point mass, and in this limiting case the fraction of material with a trajectory aimed directly at the accreting object is a 'subset of size zero'. The Bondi/Bondi-Hoyle approach is therefore best interpreted as a *capture* condition, where the mass that has reached the relevant radius is bound to the central object and will accrete, but not immediately.

#### 2.3.4 Accretion discs

Given then that accreting matter will tend to orbit the accreting object, we must ask how the desired goal of accretion - namely for the mass of the infalling material to actually add to the mass of the central star, or black hole - can take place. Since in an axisymmetric or spherically-symmetric potential angular momentum is conserved, we must find a way for the orbiting matter to lose its angular momentum before it can be accreted. When the infalling material is gaseous, collisions and shocks between the various parts of the gas can dissipate energy, causing the orbits to circularise (the lowest energy orbit for a given angular momentum is a circle) and the action of viscous torques can transport angular momentum<sup>9</sup>, and so in the case of a central point mass, the material spirals in through a sequence of circular Keplerian orbits. This is known as an *accretion disc*. Since, in practice, entirely radial orbits are ruled out due to finite initial angular momentum, accretion discs are universal across a wide range of scales.

The ubiquity of accretion discs across the Universe therefore makes them an extremely important area of study in the field of astrophysics. Here we discuss some of the important characteristics of these flattened geometries and how they are relevant to a number of different fields. For a detailed review the reader is directed to Lodato (2007) and references therein.

<sup>&</sup>lt;sup>9</sup>since angular momentum must be *transported*, whereas energy can be radiated

We start by reviewing the relevant equations that determine the evolution of a gaseous accretion disc. Much of this analysis is based on the seminal work by Shakura and Sunyaev (1973). For the purposes of concision, we restrict our treatment to what is known as the 'thin disc' approximation. Since the main theme of this thesis is AGN, a regime characterised by very thin discs on large scales, this is not an unreasonable omission. The thin disc approximation sets a limit on the aspect ratio of the disc,  $H/R \ll 1$ . This firstly allows us to vertically integrate our equations without fear of the loss of too much information, and secondly, to make the assumption that the sound speed in the disc is very much less than the azimuthal velocity,  $c_s \ll v_{\phi}$ . This will be shown directly in Section 2.3.4.3, but we will require it before then and so we mention it now. This is also the case with the further assumption that the accretion timescale through a gaseous disc is long, such that the radial velocity is very much less than both the sound speed and  $v_{\phi}$ . We therefore have

$$v_R \ll c_s \ll v_\phi \tag{2.81}$$

where the latter assumption will also be derived in Section 2.3.4.4.

#### 2.3.4.1 Basic equations

To describe the evolution of a typical (axisymmetric) accretion disc it is convienent to begin with the conservation of mass law; namely, the continuity equation:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 \tag{2.82}$$

where  $\rho$  is the density and **v** the velocity vector. The divergence of the velocity in cylindrical polar coordinates is

$$\nabla \cdot \mathbf{v} = \begin{pmatrix} 1/R \, \partial/\partial R \\ 1/R \, \partial/\partial \phi \\ \partial/\partial z \end{pmatrix} \cdot \begin{pmatrix} Rv_R \\ v_\phi \\ v_z \end{pmatrix}$$
(2.83)

where  $R = (x^2 + y^2)^{1/2}$ , the cylindrical radius. This gives

$$\nabla \cdot \mathbf{v} = \frac{1}{R} \frac{\partial}{\partial R} (Rv_R) + \frac{1}{R} \frac{\partial v_\phi}{\partial \phi} + \frac{\partial u_z}{\partial z}$$
(2.84)

where, due to axisymmetry the  $\partial/\partial \phi$  term is zero and from the thin disc assumption the  $\partial/\partial z$  term is zero. Vertically integrating transforms  $\rho$  into  $\Sigma$ , the surface density, so that the (one dimensional) continuity equation for an accretion disc becomes

$$\frac{\partial \Sigma}{\partial t} + \frac{1}{R} \frac{\partial}{\partial R} (R \Sigma v_R) = 0$$
(2.85)

where  $\Sigma$  is defined as

$$\Sigma = \frac{1}{2\pi} \int_0^{2\pi} \int_{-\infty}^{\infty} \rho \, \mathrm{d}z \mathrm{d}\phi \tag{2.86}$$

Next we consider the equation of motion (sometimes termed the 'momentum equation') in order to obtain an expression for the velocity  $\mathbf{v}$ . For an *inviscid* fluid the equation of motion is

$$\rho \left[ \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} \right] = -\nabla P - \rho \nabla \Phi$$
(2.87)

where P is the pressure and  $\Phi$  the gravitational potential. Essentially this equation relates the velocity (and acceleration) of the fluid to the forces acting upon it (the terms on the RHS; namely, pressure gradients and gravity). This inviscid form of the momentum equation is known as Euler's equation. However, in the case of an accretion disc an additional force must be added due to viscosity, resulting in the Navier-Stokes equation:

$$\rho \left[ \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} \right] = -(\nabla P - \nabla \cdot \sigma) - \rho \nabla \Phi$$
(2.88)

where  $\sigma$  is known as the *stress tensor*, which contains the vector component of *stress*, or force per unit area, for all surface area elements i.e., in any direction. The stress tensor in cylindrical polars is given by

$$\sigma = \begin{bmatrix} \sigma_R & \tau_{R\phi} & \tau_{Rz} \\ \tau_{\phi R} & \sigma_{\phi} & \tau_{\phi z} \\ \tau_{zR} & \tau_{z\phi} & \sigma_z \end{bmatrix}$$
(2.89)

where  $\sigma_R$ ,  $\sigma_{\phi}$ ,  $\sigma_z$  are the normal components of stress; in other words the product of the scalar *pressure* with the unit tensor. The  $\tau$  components are the *shear* stresses. Under the assumptions of a circular, thin accretion disc the only non-vanishing shear component is the  $\tau_{R\phi}$  component<sup>10</sup>. In terms of the angular velocity,  $\Omega = v_{\phi}/R$ , this component is given by

$$\tau_{R\phi} = \mu R \frac{\partial \Omega}{\partial R} \tag{2.92}$$

$$\tau = \mu \frac{\partial v}{\partial y} \tag{2.90}$$

$$\tau_{R\phi} = \mu R \frac{\partial}{\partial R} \left( \frac{v_{\phi}}{R} \right) \tag{2.91}$$

<sup>&</sup>lt;sup>10</sup>We can derive the form of this component by considering the velocity gradient of a shearing fluid. In straight, parallel flow (with 1D velocity along, say, the *x*-direction and a velocity gradient along the *y*-direction) the shear stress will be proportional to this velocity gradient,  $\tau \propto \partial v / \partial y$ . We introduce a proportionality constant,  $\mu$ , which is effectively the coefficient of shear viscosity such that

Generalising to *circular* flow, we have the same situation but this time the velocity direction is along  $\Omega = v_{\phi}/R$  and our velocity gradient is radial. We therefore have the  $R\phi$  component of the stress tensor in cylindrical coordinates (Hupfer et al., 2006) as

where  $\mu$  is an effective dynamic viscosity (often denoted with  $\mu$ ). Recall that in the case of the continuity equation, we were able to reduce the general three-dimensional form (equation 2.82) to that of a one (spatial) dimensional dependance of  $\Sigma(r,t)$ . Here, however, our equation 2.88 (the Navier-Stokes equation) so far contains contributions from the radial, vertical, and azimuthal directions and so we now proceed to address of each these separately.

#### 2.3.4.2 Radial disc components

To write our Navier-Stokes (N-S) equation in radial components we start with the LHS, where the first term (the time derivative) is straightforward but the second term requires some working; we start with the vector identity (in cylindrical polar coordinates)

$$(\mathbf{A} \cdot \nabla) \mathbf{B} = \begin{pmatrix} A_R \frac{\partial B_R}{\partial R} + \frac{A_\phi}{R} \frac{\partial B_R}{\partial \phi} + A_z \frac{\partial B_R}{\partial z} - \frac{A_\phi B_\phi}{R} \\ A_R \frac{\partial B_\phi}{\partial R} + \frac{A_\phi}{R} \frac{\partial B_\phi}{\partial \phi} + A_z \frac{\partial B_\phi}{\partial z} + \frac{A_\phi B_R}{R} \\ A_R \frac{\partial B_z}{\partial R} + \frac{A_\phi}{R} \frac{\partial B_z}{\partial \phi} + A_z \frac{\partial B_z}{\partial z} \end{pmatrix}$$
(2.93)

The radial component of the  $(\mathbf{v} \cdot \nabla)\mathbf{v}$  term is therefore

$$[(\mathbf{v}\cdot\nabla)\mathbf{v}]_R = v_R \frac{\partial v_R}{\partial R} + \frac{v_\phi}{R} \frac{\partial v_R}{\partial \phi} + v_z \frac{\partial v_R}{\partial z} - \frac{v_\phi^2}{R}$$
(2.94)

where, as we recall from our thin disc assumption, we can put  $v_R \approx 0$  and  $v_z \approx 0$ . The LHS of the radial N-S equation then simply becomes  $-v_{\phi}^2/R$ .

We now examine the RHS. The divergence of the stress tensor term vanishes since the viscous stress here is only in the shear component, which is orthogonal to the radial direction. We start then with the pressure gradient term,  $1/\rho \nabla P$ , the radial component of which is  $1/\rho \partial P/\partial R$ . In the case of a barotropic equation of state we can introduce the sound speed as

$$c_s^2 = \frac{\mathrm{d}P}{\mathrm{d}\rho} \tag{2.95}$$

so our pressure gradient term becomes

$$\frac{1}{\rho}\frac{\partial P}{\partial R} = \frac{c_s^2}{\rho}\frac{\partial \rho}{\partial R}.$$
(2.96)

Since we have from our thin disc assumption that  $c_s \ll v_{\phi}$ , this term is therefore negligible w.r.t the LHS term, and we can ignore it<sup>11</sup>. On the RHS we therefore only have the gravity

<sup>&</sup>lt;sup>11</sup>although it is more accurate to say that it this term appears as second order. In fact in making the thin disc assumption of negligible pressure, it is easy to miss an important physical process within an accretion disc; that of *migration* of solid bodies. This process is of particular interest in the field of planet formation. In we leave the pressure term in the radial equation of motion, and introduce a power law dependance on density e.g.,  $\rho \propto R^{-q}$ , we find that the rotation curve  $v_{\phi}$  is non-circular, due to a small correction to the

derivative left, and our radial equation of motion becomes

$$\frac{v_{\phi}^2}{R} \simeq \frac{\partial \Phi}{\partial R} \tag{2.99}$$

from which, if the form of the gravitational force is known, the angular velocity can be easily derived. When the gravity is dominated by a central massive compact object, we can re-arrange to get the familiar Keplerian form

$$v_{\phi} \equiv \Omega = \left(\frac{GM}{R^3}\right)^{1/2} \tag{2.100}$$

where M is the mass of the central object.

# 2.3.4.3 Vertical disc components

Since the velocity in the vertical direction is negligible, and the viscous force relates only to  $\tau_{r\phi}$  stress, we need only consider the pressure and gravitational forces in deriving an equilibrium structure for an accretion disc. We start then with the hydrostatic equilibrium condition:

$$\frac{1}{\rho}\frac{\partial P}{\partial z} = -\frac{\partial \Phi}{\partial z} \tag{2.101}$$

for which we must consider two different cases; self-gravitating, when the gravity provided by the disc is comparable to that of the central object, and non self-gravitating, when it is not. We shall first consider the latter.

## Non self-gravitating

Here we need only treat the gravitational potential of the central object, which has an angular separation  $\theta$  from a given radius R in the plane of the disc and height z above the disc plane. The length of the position vector to such a point is denoted r. Our derivative

circular case that scales with the sound speed. Namely,

$$v_{\phi} = v_{\rm circ} \left[ 1 - q \left( \frac{c_s}{v_{\rm circ}} \right)^2 \right]^{1/2} \tag{2.97}$$

where  $v_{\rm circ}$  is the perfectly Keplerian orbital velocity,

$$v_{\rm circ} = \left(R\frac{\partial\Phi}{\partial R}\right)^{1/2} \tag{2.98}$$

Pressure effects therefore become more important, and the thin disc approximation breaks down, as the sound speed increases in comparison to  $v_{\text{circ}}$ , i.e., when the disc is hot. This is a fairly obvious conclusion, but an important one. If solid bodies are present in a region of the disc where pressure effects are non-negligible, their dynamics will be entirely Keplerian and a small velocity drift will be introduced with respect to the gas. Such a velocity difference can give rise to a drag force that causes the solids to radially migrate inwards (Lodato, 2007)

w.r.t. z becomes

$$\frac{\partial \Phi}{\partial z} = \frac{GM}{r^2} \sin \theta = \frac{GM}{r^2} \frac{z}{r}$$
(2.102)

For small  $z, r \approx R$ , so

$$\frac{\partial \Phi}{\partial z} \simeq \frac{GM}{R^2} \frac{z}{R} \tag{2.103}$$

Hydrostatic equilibrium therefore becomes

$$\frac{c_s^2}{\rho}\frac{\partial\rho}{\partial z} = -\frac{GMz}{R^3} \tag{2.104}$$

where we have used the fact that  $c_s^2 = \partial P / \partial \rho$ . Solving this equation for  $\rho(z)$  we find

$$\ln \rho = -\frac{GM}{2c_s^2 R^3} z^2 + \text{const.}$$
 (2.105)

which can be written, after defining a central (midplane) density  $\rho_0$  when z = 0, as

$$\rho(z) = \rho_0 \exp\left(-\frac{GMz^2}{2c_s^2 R^3}\right) \tag{2.106}$$

This expression has the form of a Gaussian. We can write it in terms of a vertical scaleheight, H:

$$\rho(z) = \rho_0 \exp\left(-\frac{z^2}{2H^2}\right) \tag{2.107}$$

where  $H = c_s (R^3/GM)^{1/2} = c_s/\Omega_k$ , where  $\Omega_k$  is the Keplerian angular frequency. At this point we define the *aspect ratio* of the disc, which is given by

$$\frac{H}{R} = \frac{c_s}{v_k} \tag{2.108}$$

where  $v_k$  is the Keplerian circular velocity. For a thin disc we have  $H/R \ll 1$ , so that  $c_s \ll v_k$  (our assumption in Section 2.3.4). For the thin disc approximation, therefore, the local Keplerian velocity i.e., the disc rotation speed, is highly supersonic.

#### Self-gravitating

The procedure here is initially the same. We start with hydrostatic equilibrium, but now we must consider the contribution to the gravity from the disc itself. We do this via Poisson's equation in the vertical direction:

$$\frac{\partial^2 \Phi}{\partial z^2} = 4\pi G\rho \tag{2.109}$$

so that

$$\frac{\partial}{\partial z} \left( \frac{c_s^2}{\rho} \frac{\partial \rho}{\partial z} \right) = -4\pi G \rho \tag{2.110}$$

Integrating gives us an integral over  $\rho$  with z, which if we denote by a quantity  $\sigma$  allows us to construct a 2nd-order ordinary differential equation (ODE):

$$c_s^2 \frac{\partial^2 \sigma}{\partial z^2} = -4\pi G \frac{\partial \sigma}{\partial z} \sigma \tag{2.111}$$

which we solve by substituting a solution for  $\sigma$  of the form  $\sigma = a \tanh bz$ . The derivatives are therefore  $\partial \sigma / \partial z = ab \operatorname{sech}^2 bz$  and  $\partial^2 \sigma / \partial z^2 = 2ab^2 \operatorname{sech}^2 bz \tanh bz$ . Putting these in we get

$$2ab^2 \operatorname{sech}^2 bz \tanh bz = -\frac{4\pi G}{c_s^2} ab \operatorname{sech}^2 bz \ a \tanh bz \tag{2.112}$$

where the hyperbolic terms on the left- and right-hand sides cancel leaving

$$b = -\frac{2\pi G}{c_s^2}a\tag{2.113}$$

Our solution for  $\sigma$  is therefore

$$\sigma = a \tanh\left[\left(-\frac{2\pi G}{c_s^2}a\right)z\right] \tag{2.114}$$

so that we recover  $\rho(z)$  as the expression

$$\rho = -\frac{2\pi G}{c_s^2} a^2 \operatorname{sech}^2 \left[ \left( -\frac{2\pi G}{c_s^2} a \right) z \right]$$
(2.115)

As before, when z = 0 we have a midplane density of  $\rho_0$ :

$$\rho_0 = -\frac{2\pi G a^2}{c_s^2} \tag{2.116}$$

We also define a *surface* density,  $\Sigma$ , such that

$$\Sigma = \int_0^\infty \rho dz = \left[ a \tanh\left(-\frac{2\pi G}{c_s^2}az\right) \right]_0^\infty = a \tag{2.117}$$

This allows us to write the vertical density profile in the self-gravitating case as

$$\rho(z) = \rho_0 \operatorname{sech}^2\left(\frac{z}{H}\right) \tag{2.118}$$

where the scaleheight is now  $H = \Sigma/\rho_0$ . This profile is therefore not Gaussian, but is instead an inverse hyperbolic function. The disc thickness, H, is given by

$$H = \frac{c_s^2}{2\pi G\Sigma} \tag{2.119}$$

which relates the surface density  $\Sigma$  to the midplane density  $\rho_0$  by  $\Sigma = \rho_0 H$ .

We can calculate how massive the disc has to be to change become self-gravitating by simply equating the contributions to the gravitational potential from the disc and the central object. Thus we have  $GMz/R^3 = GM_{disc}/R^2$  and putting z = H to account for the entire disc we find

$$\frac{M_{\rm disc}}{M} \simeq \frac{H}{R} \tag{2.120}$$

for a central object of mass M. For a thin disc  $(H/R \ll 1)$  we can therefore see that the ratio of masses  $M_{\rm disc}/M$  can in fact be  $\ll 1$  i.e., a low-mass disc for the vertical structure of the disc to be significantly affected by the disc's contribution to the gravity.

It is instructive here to consider another means of comparing the non self-gravitating and self-gravitating cases, namely to consider the ratio of the disc heights. Recall that we had  $H_{\rm nsg} = c_s / \Omega_K$  and  $H_{\rm sg} = c_s^2 / \pi G \Sigma$ . The ratio is

$$\frac{H_{\rm sg}}{H_{\rm nsg}} = \frac{c_s \Omega_K}{\pi G \Sigma} \tag{2.121}$$

where the right-hand side in this equation is a useful quantity (for a Keplerian disc) that we will discuss in more detail later in this Section. It is known as the Toomre parameter, denoted by Q, and it describes the threshold for the onset of gravitational instability in an accretion disc. Indeed, we can see here that the disc becomes self-gravitating when

$$\frac{c_s \Omega_K}{\pi G \Sigma} \sim 1 \tag{2.122}$$

in other words, when  $Q \sim 1$ .

#### 2.3.4.4 Azimuthal disc components

We now come to the last (but certainly not least!) component of our N-S equation. We have chosen to address the azimuthal treatment last because it naturally leads on to a discussion about the form of the viscosity in an accretion disc, which is an entire sub-topic in itself.

Since we intend to link the result of this Section with our vertically-integrated continuity equation for a disc (equation 2.85), it is convienient to begin by considering a vertically-integrated form of the N-S equation:

$$\Sigma \left[ \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} \right] = -(\nabla P - \nabla \cdot \mathbf{T}) - \Sigma \nabla \Phi$$
 (2.123)

where here we have introduced **T**, the vertical integral of the stress tensor. Once again, in a circular shearing flow, only the  $R\phi$  component is relevant. We had,

$$\tau_{R\phi} = \mu R \frac{\mathrm{d}\Omega}{\mathrm{d}R} \tag{2.124}$$

therefore we now put

$$T_{R\phi} = \nu \Sigma R \frac{\mathrm{d}\Omega}{\mathrm{d}R} \tag{2.125}$$

where we define a vertically integrated kinematic<sup>12</sup> viscosity such that

$$\nu\Sigma = \int_{-\infty}^{\infty} \mu \, \mathrm{d}z \tag{2.126}$$

The divergence of a tensor in cylindrical polar coordinates is given by

$$\nabla \cdot \mathbf{T} = \begin{pmatrix} \frac{1}{R} \frac{\partial}{\partial R} (RT_{RR}) + \frac{1}{R} \frac{\partial}{\partial \phi} (T_{\phi R}) + \frac{\partial T_{zR}}{\partial z} - \frac{1}{R} T_{\phi \phi} \\ \frac{1}{R} \frac{\partial}{\partial R} (RT_{R\phi}) + \frac{1}{R} \frac{\partial}{\partial \phi} (T_{\phi \phi}) + \frac{\partial T_{z\phi}}{\partial z} + \frac{1}{R} T_{\phi R} \\ \frac{1}{R} \frac{\partial}{\partial R} (RT_{Rz}) + \frac{1}{R} \frac{\partial}{\partial \phi} (T_{\phi z}) + \frac{\partial T_{zz}}{\partial z} \end{pmatrix}$$
(2.127)

where we are interested in the azimuthal component and eliminate all terms except in  $R\phi$  to give

$$(\nabla \cdot \mathbf{T})_{\phi} = \frac{1}{R^2} \frac{\partial}{\partial R} \left( R^2 T_{R\phi} \right)$$
(2.128)

We now take the azimuthal component of equation 2.123, by considering the vector identity introduced in equation 2.93. From this we obtain the  $\phi$ -component of the  $(\mathbf{v} \cdot \nabla)\mathbf{v}$  term, giving us

$$\Sigma \frac{\partial v_{\phi}}{\partial t} + \left[ \Sigma v_R \frac{\partial v_{\phi}}{\partial R} + \Sigma \frac{v_{\phi}}{R} \frac{\partial v_{\phi}}{\partial \phi} + \Sigma v_z \frac{\partial v_{\phi}}{\partial z} + \Sigma \frac{v_{\phi} v_R}{R} \right]$$
(2.129)

$$= -\frac{1}{\rho} \frac{c_s^2}{R} \frac{\partial \rho}{\partial \phi} - \frac{1}{R^2} \frac{\partial}{\partial R} \left( R^2 \nu \Sigma R \frac{\mathrm{d}\Omega}{\mathrm{d}R} \right) - \rho \frac{\partial \Phi}{\partial \phi}$$
(2.130)

where due to the assumption of axisymmetry in the potential we can eliminate the relevant terms to leave

$$\Sigma\left(\frac{\partial v_{\phi}}{\partial t} + v_R \frac{\partial v_{\phi}}{\partial R} + \frac{v_R v_{\phi}}{R}\right) = \frac{1}{R^2} \frac{\partial}{\partial R} \left(R^2 \nu \Sigma R \frac{\mathrm{d}\Omega}{\mathrm{d}R}\right)$$
(2.131)

Multiplying through by R and writing  $v_R = \partial R / \partial t$  allows us to simplify via the product

<sup>&</sup>lt;sup>12</sup>note that in a local (non-integrated) sense, dynamic viscosity  $\mu$  and kinematic viscosity  $\nu$  are related by  $\mu = \rho \nu$ 

rule:

$$\Sigma\left(\frac{\partial}{\partial t}Rv_{\phi} + v_{R}\frac{\partial}{\partial R}Rv_{\phi}\right) = \frac{1}{R}\frac{\partial}{\partial R}\left(R^{2}\nu\Sigma R\frac{\mathrm{d}\Omega}{\mathrm{d}R}\right)$$
(2.132)

By putting  $v_{\phi} = R\Omega$ , where  $\Omega$  is the Keplerian angular velocity, we can see that we have in fact an expression in terms of the specific orbital angular momentum,  $J = R^2 \Omega$ , such that

$$\Sigma\left(\frac{\partial}{\partial t}J + v_R\frac{\partial J}{\partial R}\right) = \frac{1}{R}\frac{\partial}{\partial R}\left(R^3\nu\Sigma\frac{\mathrm{d}\Omega}{\mathrm{d}R}\right)$$
(2.133)

The LHS here is simply the convective derivative of the angular momentum, with the first term denoting the (Eulerian) temporal rate of change of J for an annulus and the second term representing the spatial rate of change of J due to its radial advection. If these terms are exactly balanced then there will be no radial evolution of the angular momentum in a disc; the fact that there is a term on the RHS indicates this is not so. In fact this (viscous) term demonstrates a transport of angular momentum with a radial velocity  $v_R$ . It is important to note here that the sign of  $d\Omega/dR$  determines the direction in which the angular momentum is transported; for Keperian rotation  $d\Omega/dR$  is negative and so an inward transport of mass ( $v_R$  negative) requires an outward transport of angular momentum ( $\partial J/\partial R$  positive), as one would expect for matter to accrete. We note too that in the case of solid body rotation ( $\Omega = \text{const.}$ ) the viscous term vanishes and there is no transport of angular momentum or material through the disc.

If we now assume a steady-state condition, the time derivative term disappears and we are left with

$$\Sigma v_R \frac{\partial J}{\partial R} = \frac{1}{R} \frac{\partial}{\partial R} \left( R^3 \nu \Sigma \frac{\mathrm{d}\Omega}{\mathrm{d}R} \right)$$
(2.134)

From this we can write down an expression for the radial velocity  $v_R$ , providing us with the rate at which material is transported through the disc due to viscous torques:

$$v_R = \frac{1}{R\Sigma} \frac{\partial}{\partial R} \left( R^3 \nu \Sigma \frac{\mathrm{d}\Omega}{\mathrm{d}R} \right) \left( \frac{\partial J}{\partial R} \right)^{-1}$$
(2.135)

This concludes our treatment of the cylindrical components of the N-S equation, subject to the assumption of a thin, highly supersonic accretion disc. Each component provided us with some useful information on the properties of such a disc; from the radial component we determined the rotation profile of the disc subject to the assumption of negligible disc self-gravity; from the vertical component we derived the density profile along z and defined the disc scaleheight; and from the azimuthal component we have found a link between the mass accretion rate and the radial rate of change of the viscous torque.

## 2.3.4.5 Evolution of the surface density

We are now in a position to derive a governing equation for the surface density of an accretion disc. We substitute for  $v_R$  in the continuity equation (equation 2.85), giving us

$$\frac{\partial \Sigma}{\partial t} + \frac{1}{R} \frac{\partial}{\partial R} \left[ \frac{\partial}{\partial R} \left( R^3 \nu \Sigma \frac{\mathrm{d}\Omega}{\mathrm{d}R} \right) \left( \frac{\partial J}{\partial R} \right)^{-1} \right] = 0 \qquad (2.136)$$

For a Keplerian disc we have  $\Omega = (GM/R^3)^{1/2}$  and  $J = (GMR)^{1/2}$ , and so we are left with

$$\frac{\partial \Sigma}{\partial t} = \frac{3}{R} \frac{\partial}{\partial R} \left[ R^{1/2} \frac{\partial}{\partial R} \left( \Sigma \nu R^{1/2} \right) \right]$$
(2.137)

which, providing that  $\Omega$  and  $\nu$  are independent of  $\Sigma$ , has the form of a *diffusion* equation<sup>13</sup>. In the case that the viscosity is constant with radius, it is possible to construct an analytical solution to this equation via Bessel functions, although this lies beyond the scope of this thesis; the reader is directed to Lodato (2007), Frank et al. (2002), or Pringle (1981) for a full treatment. If the viscosity depends upon R, however, the diffusion equation must in general be solved numerically<sup>14</sup>.

#### 2.3.4.6 Accretion disc viscosity

Equation 2.137 therefore describes the time evolution of the surface density of an accretion disc subject to a shearing viscosity  $\nu$ . We can obtain a characteristic timescale for this evolution by reducing the derivatives in this equation to quotients, effectively deriving an *e-folding* time. This gives us the 'viscous timescale', defined as

$$t_{\rm visc} = \frac{R^2}{\nu} \tag{2.139}$$

This timescale can tell us some important information about the nature of the viscosity in an accretion disc. Writing down the Reynolds number, which is essentially the ratio of the viscous and dynamical times for a given flow, we have

$$\operatorname{Re} = \frac{t_{\operatorname{visc}}}{t_{\operatorname{dyn}}} = \frac{R^2 \Omega}{\nu} \tag{2.140}$$

<sup>13</sup>i.e., of the form,

$$\frac{\partial \rho(\mathbf{r}, t)}{\partial t} = \frac{\partial}{\partial r} \left[ D(\rho, \mathbf{r}) \frac{\partial}{\partial r} \rho(\mathbf{r}, t) \right]$$
(2.138)

in 1D along the **r** vector, where  $\rho(\mathbf{r}, t)$  is the density of the diffusing material at **r** at time t, and D is a diffusion coefficient.

 $<sup>^{14}</sup>$  although it should be noted that analytic solutions for the specific case of  $\nu \propto R^{\gamma}$  exist; see Lynden-Bell and Pringle (1974)

This dimensionless number characterises two regimes in a fluid flow; *laminar*, where the viscous forces are important and the Reynolds number is low, and *turbulent*, where dynamical (inertial) forces dominate. If we were to assume that the viscosity in an accretion disc comes from the standard molecular viscosity,  $\nu_{\text{molecular}} = \lambda_{\text{coll}}c_s$ , where  $\lambda_{\text{coll}}$  is the collisional mean free path between molecules and  $c_s$  is the sound speed (the typical random velocity of the molecules), we find

$$\nu_{\text{molecular}} = \frac{1}{n\sigma_{\text{coll}}}c_s = \frac{\mu m_p}{\rho\sigma_{\text{coll}}}c_s = \frac{2\mu m_p}{\Sigma\sigma_{\text{coll}}}c_s H$$
(2.141)

where  $\sigma_{\text{coll}}$  is the cross-sectional area for collision and we have used the fact that for an accretion disc,  $\rho = \Sigma/2H$  where H is the disc height. Using this viscosity in the Reynolds number we have

$$\frac{t_{\rm visc}}{t_{\rm dyn}} = \left(\frac{\Sigma\sigma_{\rm coll}}{2\mu m_p}\right) \left(\frac{H}{R}\right)^{-2} \tag{2.142}$$

Let us now put some typical numbers into this equation; for a protostellar disc we assume a surface density of  $\Sigma \approx 10$  g cm<sup>-2</sup> with a radius of  $R \approx 50$  A.U. and a thickness of  $H/R \approx 0.1$ . The cross-section of a hydrogen molecule is ~ 10<sup>-16</sup> cm<sup>2</sup>. Our ratio of viscous to dynamical times thus becomes

$$\frac{t_{\rm visc}}{t_{\rm dyn}} \simeq 10^{11} \tag{2.143}$$

which, when we put in the dynamical timescale at such a radius, namely  $\sim$  a few years, returns a viscous time that is longer than  $t_{\text{Hubble}}$ . If we assume that accretion discs provide a significant component of the accretion on to a compact object (refer back to Section 2.3) then it is clear from this comparison that the corresponding mass accretion rate through a disc cannot be due to molecular viscosity alone.

The precise form of the viscosity that operates in an accretion disc is as yet unknown. However, a likely candidate can be readily inferred from noting that the Reynolds number of an accretion disc (the ratio of the viscous to the dynamical times that we have just found) is extremely large, implying strong *turbulent* flow. Invoking a turbulent viscosity is therefore a logical next step, as the typical lengthscale for turbulence would be many orders of magnitude larger than the mean free path for molecular collisions. We should be careful here, however, as in a standard, non self-gravitating disc, there appears to be no intrinsic instability that would drive turbulent behaviour. The Rayleigh stability criterion, which describes the onset of exponentially diverging wave modes in an axisymmetric, rotating flow (the derivation of which is similar to our approach for the gravitational instabilities in Section 2.3.4.8) requires

$$\frac{\mathrm{d}}{\mathrm{d}R}\left(R^2\Omega\right) < 0\tag{2.144}$$

for instability. For a Keplerian disc the specific angular momentm,  $R^2\Omega$ , is an increasing function of radius ( $\propto R^{1/2}$ ) so its derivative is positive and the flow is stable. The origin of the turbulence is therefore unknown, and is not likely to be purely hydrodynamic in nature. Promising suggestions include the action of strong magnetic fields in the disc, leading to a magneto-rotational instability (MRI) - (see, e.g., Balbus and Hawley, 1991), and the driving of subsonic turbulence by feedback from massive star formation in the disc (Kawakatu and Wada, 2008; Chen et al., 2009). It is also possible that the viscosity is not entirely due to turbulence; a possible candidate is a form of stable angular momentum transport as a result of the onset of disc self-gravity, which provided the disc can be maintained above the fragmentation limit (see Section 2.3.4.8) can drive a non-axisymmetric perturbation in the form of a spiral arm than can transport mass inwards and angular momentum outwards, as required (Gammie, 2001; Rice et al., 2005; Lodato and Rice, 2004).

Unfortunately, we cannot say for sure what the magnitude or characterisation of the viscosity will be. Instead, we invoke what is effectively a 'parameterisation of our ignorance' whereby a dimensionless parameter  $\alpha$  is assumed to encapsulate everything that we do not know about the form of the viscosity. What we do know (or at least it is reasonable to assume) is that the resulting stress tensor should scale with the vertically-integrated pressure,  $T_{R\phi} \sim \Sigma c_s^2$ . Our dimensionless parameter  $\alpha$  is then invoked as a proportionality factor between the two. From dimensional arguments, and substituting in the (non self-gravitating) disc scaleheight,  $H = c_s/\Omega_K$ , we find that this implies a viscosity of the form

$$\nu = \alpha c_s H \tag{2.145}$$

We can offer some constraints on this unknown  $\alpha$  quantity by making a parallel with the viscosity of kinetic theory, namely that the turbulent viscosity will have an approximate magnitude given by

$$\nu \sim v_{\rm turb} l_{\rm turb} \tag{2.146}$$

where  $v_{\text{turb}}$  is a typical turbulent velocity and  $l_{\text{turb}}$  is the lengthscale of the largest turbulent eddies. In the latter case we have an upper limit of the disc thickness, H. For the turbulent velocity we note that supersonic turbulence would decay relatively quickly due to disspation via shocks, and so it is likely that the turbulence here is subsonic. The characteristic random velocity that we need to consider is thus bounded at  $v_{\text{turb}} \lesssim c_s$ . Taken together, these upper limits imply that  $\alpha < 1$ .

#### 2.3.4.7 Steady-state solutions

Under the assumption of a steady-state, the  $\partial/\partial t$  term in the mass conservation equation (2.85) disappears and we integrate to obtain  $R\Sigma v_R = \text{const.}$  This constant is the vertically-

integrated mass flux through a cylindrical radius R:

$$\dot{M} = -2\pi R v_R \Sigma \tag{2.147}$$

where the  $2\pi$  emerges from the disc geometry. Substituting for  $v_R$  from equation 2.135 gives us

$$\dot{M} = -\frac{2\pi}{(R^2\Omega)'}\frac{\partial}{\partial R}\left(R^3\nu\Sigma\Omega'\right) \tag{2.148}$$

where the primes denote a derivative w.r.t. R. We can do the same with the angular momentum conservation equation (2.133), substituting in for the constant  $\dot{M}$  from equation 2.147 to obtain

$$\dot{M}R^2\Omega = -2\pi\nu\Sigma R^3\Omega' + \text{const.}$$
 (2.149)

where the constant of integration can be determined by the inner boundary condition. For this we assume that as material accretes on to the central object the torque vanishes (along with the radial derivative of the angular velocity,  $\Omega' = 0$ ) at a radius  $R_{\rm acc}$ . Physically this might correspond to the surface of the accreting object, if one exists. Our integration constant is then  $\dot{M}R_{\rm acc}^2\Omega_{\rm acc} = \dot{M}(GMR_{\rm acc})^{1/2}$ . Noting also that for Keplerian rotation,  $R\Omega' = -3\Omega/2$ , we can re-arrange to get

$$\nu\Sigma = \frac{\dot{M}}{3\pi} \left( 1 - \sqrt{\frac{R_{\rm acc}}{R}} \right) \tag{2.150}$$

where we see that for  $R \gg R_{\rm acc}$ , or for  $R_{\rm acc} = 0$ , the steady-state mass accretion rate through the disc is

$$\dot{M} = 3\pi\nu\Sigma \tag{2.151}$$

and so we note that the surface density of a steady-state disc is purely a function of the viscosity,  $\Sigma \propto \nu^{-1}$ .

We are now in a position to derive the temperature profile of the disc, subject to the assumption that the energy flux is provided by the viscous dissipation rate. We recall from Section 2.3.4.4 that the relevant component of the stress tensor between shearing layers was  $T_{R\phi}$ , given by equation 2.125. The (vertically-integrated) viscous *torque* is defined by

$$G(R) = \int_0^{2\pi} R^2 T_{R\phi} \,\mathrm{d}\phi \tag{2.152}$$

which over a circular boundary becomes

$$G(R) = 2\pi R^2 \nu \Sigma R \Omega' \tag{2.153}$$

The torque provides a *rate of work* that can be seen dimensionally to be

$$\dot{W} = G\Omega' \mathrm{d}R \tag{2.154}$$

over a radial width dR. Over an area (based on the two plane faces of a thin disc) given by  $4\pi R dR$ , we can write the viscous disspation as

$$D(R) = \frac{G\Omega'}{4\pi R} = \frac{1}{2}\nu\Sigma(R\Omega')^2$$
(2.155)

where we note that for solid body rotation this dissipation term vanishes. Comparing now with equation 2.150 we see that this can be written as

$$D(R) = \frac{3GM\dot{M}}{8\pi R^3} \left(1 - \sqrt{\frac{R_{\rm acc}}{R}}\right)$$
(2.156)

The temperature is determined by first assuming that the energy flux through the disc as a result of dissipation is radiative, and further assuming that this radiation cannot escape directly i.e., that the disc is optically thick and can be approximated as a blackbody. In this case we can equate the viscous dissipation to the blackbody flux,  $D(R) = \sigma T_{\text{eff}}^4$ , where  $\sigma$  is the Stefan-Boltzmann constant, giving an effective temperature profile of

$$T_{\rm eff}(R) = \left[\frac{3GM\dot{M}}{8\sigma\pi R^3} \left(1 - \sqrt{\frac{R_{\rm acc}}{R}}\right)\right]^{1/4}$$
(2.157)

which will reflect the profile in  $T_c$ , the midplane temperature of the disc, if we have assume that variation of T in z is negligible i.e., the disc is thin and is dominated by  $T_c$ . In reality there will be a temperature gradient in the z-direction,  $\partial T/\partial z$ , which will determine the radiative flux within the disc together with the opacity,  $\kappa$ . However the vertical temperature *structure* of the disc is not of interest to us here, and so we do not go into any detail on this part of the analysis. What we do concern ourselves with here is the profile for  $T_{\text{eff}}$ , which we can see for  $R \gg R_{\text{acc}}$  reduces to

$$T_{\rm eff} = T_{\rm acc} \left(\frac{R}{R_{\rm acc}}\right)^{-3/4} \tag{2.158}$$

where we have defined the temperature at the inner edge of the disc as

$$T_{\rm acc} = \left(\frac{3GM\dot{M}}{8\sigma\pi R_{\rm acc}^3}\right)^{1/4} \tag{2.159}$$

For a radiative disc with an optical depth  $\tau = \rho H \kappa_R$  the effective and midplane tempera-

tures are related by (Goodman, 2003):

$$T_c^4 = \frac{3\kappa_R \Sigma}{16} T_{\text{eff}}^4 \tag{2.160}$$

where  $\kappa_R$  is the *Rosseland mean opacity*, a frequency-averaged quantity comprised of contributions from photon *absorption* and photon *scattering*. The former is determined from Kramers' law for bound-free and free-free transitions (Frank et al., 2002)

$$\kappa_{\rm abs} \simeq 5 \times 10^{24} \ \rho T_c^{-7/2} \ {\rm cm}^2 {\rm g}^{-1}$$
(2.161)

and the latter from Thomson scattering<sup>15</sup>, which has a mean value (Shapiro and Teukolsky, 1983) of

$$\kappa_{\text{scatt}} \simeq 0.40 \text{ cm}^2 \text{g}^{-1} \tag{2.162}$$

such that the Rosseland mean opacity satisfies

$$\frac{1}{\kappa_R} = \frac{1}{\kappa_{\rm abs}} + \frac{1}{\kappa_{\rm scatt}}$$
(2.163)

where the above absorption and scattering opacities tend to kick in at  $\gtrsim 10^3$  K and  $\gtrsim 10^5$  K respectively. The reader is referred to Figure 9a in Bell and Lin (1994) for the standard opacity dependance on temperature in these regimes. We note that in the very outer parts of AGN accretion discs ( $R \gtrsim 100$  pc), the temperature is likely to be  $\lesssim 100$  K and the opacity will be dominated by dust, with  $\kappa \propto T^2$  (for more detail and the fiducial opacity plot in this temperature range the reader is directed to Thompson et al., 2005).

Finally in this section, we note that the radiative flux through the disc necessarily provides a radiation *pressure* which should, if we are to be self-consistent, be factored into our equation of state (which we have previously assumed to depend barotropically only on the gas pressure). We write the total pressure as (Frank et al., 2002)

$$P = \frac{\rho k T_c}{\mu m_p} + \frac{4\sigma T_c^4}{3c} \tag{2.164}$$

where the first term describes the gas pressure and the second term is due to the radiation pressure. By substituting in for expressions from equations 2.151, 2.157, 2.160, and 2.145 (or by use of the full Shakura-Sunyaev disc solutions in e.g., Shakura and Sunyaev, 1973; Frank et al., 2002) we can obtain a ratio between the two terms as

$$\frac{P_{\rm rad}}{P_{\rm gas}} \simeq 6 \times 10^{-2} \ \alpha^{1/10} \left(\frac{\dot{M}}{0.5 \,\rm M_{\odot} yr^{-1}}\right)^{7/20} \left(\frac{M}{10^8 \,\rm M_{\odot}}\right)^{1/8} \left(\frac{R}{0.01 \,\rm pc}\right)^{-3/8} f^{7/5} \qquad (2.165)$$

<sup>&</sup>lt;sup>15</sup>scattering of a photon by a free electron

where  $f^4 = 1 - (R_{\rm acc}/R)^{1/2}$ . For the gas composition we have assumed atomic hydrogen with  $\mu = 1$ , and for the accretion radius we have taken the ISCO for the black hole,  $R_{\rm acc} \lesssim 6GM/c^2$ . The fiducial accretion here corresponds to just below the critical  $\dot{M}$  for self-gravity (refer to equations 1.14 and 2.186) at  $T = 10^4 K$ . Although this temperature would be higher in the very inner parts of a disc, we note that if the SMBH is fed through a steady-state disc that extends beyond 0.01 pc (where T is likely to be  $\lesssim 10^4$ K), the accretion rate cannot be much above this otherwise gravitational instability and fragmentation would set in at these scales. We discuss these concepts in the next section.

With these parameters then it is interesting to see at what radius radiation pressure dominates over gas pressure; we set  $P_{\rm rad}/P_{\rm gas} \ge 1$  and obtain

$$R \lesssim 2 \times 10^{-4} f^{7/5} \,\mathrm{pc}$$
 (2.166)

which, when contrasted with the ISCO at  $R \simeq 2 \times 10^{-5}$ , demonstrates that a thin, steady accretion disc, fed from scales of ~ 0.01 pc, will remain gas pressure dominated for the majority of the flow.

#### 2.3.4.8 Gravitational instability and fragmentation

The secular evolution of an accretion disc due to the action of viscosity can be somewhat complicated by the onset of gravitational instabilities when the disc becomes sufficiently massive. In this section we discuss the role of gravitational instability in driving fragmentation.

#### Jeans instability

In the simplest case, the Jeans instability (Jeans, 1902) occurs when the pressure support of an arbitrary mass of gas is not sufficient to overcome its own internal gravitational attraction, leading the region to collapse. We can perform a brief analysis by considering a uniform, static region that is infinite in extent, with density  $\rho_0$  and constant pressure  $P_0$ . The relevant equations are 2.82 and 2.87; namely, mass conservation and momentum conservation respectively<sup>16</sup>. We consider a (linear) perturbation to the fluid variables, taking  $P = P_0 + \Delta P$ ,  $\rho = \rho_0 + \Delta \rho$ , with a change in velocity  $\Delta \mathbf{v}$ . The change to the gravitational potential is introduced via  $\Phi = \Phi_0 + \Delta \Phi$ . Conservation of mass becomes

$$\frac{\partial(\Delta\rho)}{\partial t} + \rho_0 \nabla \cdot (\Delta \mathbf{v}) = 0 \tag{2.167}$$

<sup>&</sup>lt;sup>16</sup>in actuality these equations are not consistent with the assumption of a uniform density, uniformly pressured, *static* medium. By equation 2.87 the potential  $\Phi_0$  in this case must be a constant, which of course it is not as  $\rho_0$  is non-zero. Nonetheless, we shall continue to perpetuate what is termed the *Jeans swindle* as it provides a simple derivation that is accurate enough for our purposes. The reader is referred to Kiessling (1999) for more details and a mathematically rigorous approach

since  $\rho_0$  is a constant, and retained only first-order perturbation terms. Similarly, conservation of momentum becomes

$$\frac{\partial(\Delta \mathbf{v})}{\partial t} = -c_s^2 \frac{\nabla(\Delta \rho)}{\rho_0} - \nabla(\Delta \Phi)$$
(2.168)

where we have used a barotropic EQS such that  $c_s^2 = dP/d\rho$ . The perturbed density and potential are related via Poisson's equation

$$\nabla^2(\Delta\Phi) = 4\pi G \Delta\rho \tag{2.169}$$

We consider a wave-like form for the perturbation, and write

$$\Delta \rho = \rho_1 e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)} \tag{2.170}$$

$$\Delta \mathbf{v}_1 = \mathbf{v} e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)} \tag{2.171}$$

$$\Delta \Phi = \Phi_1 e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)} \tag{2.172}$$

where  $\mathbf{k}(\mathbf{r})$  is the wavevector,  $\mathbf{r}$  the position vector, and  $\boldsymbol{\omega}$  is the angular frequency of the wave. Terms subscripted with 1 are constants. Conservation of mass is now

$$-\rho_1\omega + \rho_0 \mathbf{k} \cdot \mathbf{v}_1 = 0 \tag{2.173}$$

and conservation of momentum simplifies to

$$-\rho_0 \omega \mathbf{v}_1 = -c_s^2 \rho_1 \mathbf{k} - \rho_0 \Phi_1 \mathbf{k} \tag{2.174}$$

with Poisson's equation now written as

$$-k^2 \Phi_1 = 4\pi G \rho_1 \tag{2.175}$$

where  $k^2 \equiv |\mathbf{k}|^2$ . Eliminating  $\mathbf{v}_1$  using equations 2.173 and 2.174, and substituting for  $\Phi_1$  in equation 2.175, gives us a *dispersion relation* for the perturbed system:

$$\omega^2 = k^2 c_s^2 - 4\pi G \rho_0 \tag{2.176}$$

which is simply the standard dispersion relation for sound waves modified by the presence of gravity. For wavelike behaviour we require  $\omega^2 \ge 0$ , and so we can define a critical wavenumber for instability (exponential decay or growth) that is

$$k^2 = \frac{4\pi G\rho_0}{c_s^2}$$
(2.177)

and similarly, a critical wavelength, via  $\lambda = 2\pi/k$ , above which instability sets in:

$$\lambda_J = \frac{c_s \sqrt{\pi}}{\sqrt{G\rho_0}} \tag{2.178}$$

where  $\lambda_J$  is known as the Jeans length. The intuitive interpretation of this result is that the maximum stable wavelength is the maximum size of the region across which a pressure wave can propagate, at speed  $c_s$ , to counteract the speed of gravitational collapse, which proceeds on a free-fall time,  $1/\sqrt{G\rho_0}$ .

We can define a critical mass at which this instability sets in, by considering a sphere of diameter  $\lambda_J$  such that

$$M_J = \left(\frac{4\pi}{3}\right)\rho\left(\frac{\lambda_J}{2}\right) = \left(\frac{\pi}{6}\right)\frac{c_s^3}{G^{3/2}\rho^{1/2}}$$
(2.179)

which is known as the Jeans mass. This quantity is of no small importance to the field of star formation, since it defines the maximum mass of gas that can be contained within a molecular cloud before the region undergoes collapse, the progenitor of protostellar evolution.

#### Toomre instability

Regions where pressure support provides the only stabilising force against gravity will behave in the manner just described, but in galactic nuclei star formation often occurs in a disc, where rotational support plays a role. In this case we require an alternative dispersion relation, namely

$$\omega^2 = \kappa^2 - 2\pi G |\mathbf{k}| \Sigma + k^2 c_s^2 \tag{2.180}$$

where  $\kappa$  is the epicyclic frequency and  $\Sigma$  is the vertically integrated surface density. This, again, is the standard dispersion relation for sound waves, but this time modified both by self-gravity (in a flattened geometry) and rotation. The perturbation considered is radial, across an axisymmetric disc. The derivation of this equation is beyond the scope of this thesis, but for a detailed treatment the reader is directed to Pringle and King (2003) or, for a fuller treatment including non-axisymmetric spiral perturbations, to Cossins (2010b). Once again we require  $\omega^2 > 0$  for stability, which yields a criterion on  $|\mathbf{k}|$  such that

$$\frac{\pi G\Sigma}{c_s^2} \left( 1 - \sqrt{1 - \frac{c_s^2 \kappa^2}{\pi^2 G^2 \Sigma^2}} \right) \le |\mathbf{k}| \le \frac{\pi G\Sigma}{c_s^2} \left( 1 + \sqrt{1 - \frac{c_s^2 \kappa^2}{\pi^2 G^2 \Sigma^2}} \right)$$
(2.181)

While stable modes will always exist outside of this range, for unstable modes to exist we require that the roots to the quadratic are real (and so part of the curve passes below the  $\omega^2 = 0$  axis - see Figure 2.4). This puts a constraint on the term in the square root, namely that

$$\frac{c_s \kappa}{\pi G \Sigma} \le 1 \tag{2.182}$$

If this is not satisfied, there will be no unstable modes and the disc will not collapse. This defines the *Toomre* parameter (Toomre, 1964), denoted by Q, which is a test for stability of a disc:

$$Q = \frac{c_s \kappa}{\pi G \Sigma} \tag{2.183}$$

For Q above unity, the disc is stable, and for Q below unity it is unstable to collapse (within a certain wavenumber range) and subsequent fragmentation. For a Keplerian disc we can set  $\kappa = \Omega_k$ . For disc undergoing gravitational instability we can determine the most unstable wavelength from the minimum in the dispersion relation (see Figure 2.4) which turns out to be

$$k_0 = \frac{\pi G \Sigma}{c_s^2} \tag{2.184}$$

where we recall from equation 2.119 that this is of the order of the self-gravitating disc thickness,  $k_0 \sim 1/H_{sg}$ .

The Toomre stability criterion has been verified by numerical simulations (see, e.g., Boley et al., 2007; Forgan and Rice, 2009) and is a fundamental concept in the study of self-gravitating discs. We should note, of course, that it derives from an idealised picture in which perturbations are entirely axisymmetric and the disc is razor thin (i.e., with unphysical zero thickness). Finite thickness effects and non-axisymmetric disturbances such as low-order mode spiral arms can decrease and increase Q respectively<sup>17</sup>; therefore, the criterion is best viewed as approximate.

#### Self-regulation at $Q \sim 1$

The Toomre parameter can be understood intuitively in terms of the stabilising quantities  $c_s$  and  $\Omega$  (in the case of a Keplerian disc) and the destabilising effect of self-gravity, provided through  $G\Sigma$ . Naturally, a colder and more massive disc will be more likely to fragment. However in reality the picture is more complicated. A disc that is cooling down (via radiative losses) from  $Q \gg 1$  will become unstable to non-axisymmetric perturbations before actually fragmenting, and the spiral density waves that are excited will act to heat and *stabilise* the disc (Paczyński, 1978; Lodato and Rice, 2004) through compressions and weak shocks (Cossins et al., 2009). Under the right conditions, the disc will therefore enter into a period of *self-regulation*, maintaining Q close to unity.

It is reasonable to assume that full onset of gravitational instability vs. self-regulation will depend strongly on the ratio of the cooling time to the timescale over which the heating occurs; the latter is  $1/\Omega_p$ , where  $\Omega_p$  is the pattern speed of the spiral wave. This is related to the angular frequency and mode number of the spiral by  $\omega = m\Omega_p$ . As discussed above,

 $<sup>^{17}</sup>$  for a review see Toomre (1964) and Cossins (2010b)



Figure 2.4: Dispersion relation for linear axisymmetric perturbations in a thin disc (equation 2.180), showing unstable modes (shaded area), and stable, wave-like modes (non-shaded area). Blue symbols correspond to the roots of the dispersion relation, and the red symbol is the most unstable wavelength, which is the order of the disc thickness H. For unstable modes to be present (along with stable modes), roots must exist, requiring the Toomre parameter to be less than unity.

the most unstable wavelength occurs at  $k_0$ , which from equation 2.180 gives  $\Omega_p \sim \Omega$ , and so the heating timescale is approximately the dynamical timescale,  $t_{\rm dyn}$ . This was originally shown by Gammie (2001), who found that for  $t_{\rm cool} \ll t_{\rm dyn}$ , i.e.,  $\Omega t_{\rm cool} \ll 1$ , the disc cannot self-regulate and instead undergoes the Jeans instability and fragments, whilst for  $\Omega t_{\rm cool} \gtrsim 1$  the disc settles into a marginally stable state of *gravitoturbulence* (Rafikov, 2009) with  $Q \sim 1$ .

In fact the constraint on  $\Omega t_{\rm cool}$  for fragmentation has since been revised upwards via simulations; in the same paper as the order of magnitude argument, Gammie (2001) considered 2D shearing sheet simulations and found a refined criteron of  $\Omega t_{\rm cool} \lesssim 3$  for fragmentation, while Rice et al. (2005) found  $\Omega t_{\rm cool} \lesssim 6$  for a 3D disc with  $\Sigma \propto R^{-1}$ . Cossins et al. (2009) found an intermediate case of  $\Omega t_{\rm cool} \lesssim 4.5$  for a 3D disc with  $\Sigma \propto R^{-3/2}$ . All of these authors used a simple parameterisation of the cooling time such that  $\Omega t_{\rm cool} = \beta$ , with a term

$$\left. \frac{\mathrm{d}u}{\mathrm{d}t} \right|_{\mathrm{cool}} = -\frac{u}{t_{\mathrm{cool}}} \tag{2.185}$$

added to the energy equation, where  $\beta$  is a constant. Cossins et al. (2010) extended this analysis of the fragmentation boundary to a temperature- and opacity-dependant regime for protoplanetary discs, finding that the effective value of  $\beta$  can vary from ~ unity up to ~ several tens.

#### Fragmentation boundary

For a disc that can cool efficiently, reaching  $Q \lesssim 1$  is therefore expected to lead to fragmentation. As the disc grows in mass, the point at which this occurs is a function of the accretion rate through the disc. For an  $\alpha$ -disc we can use equation 2.151, a viscosity of  $\alpha c_s H$  and the self-gravitating scaleheight (equation 2.119) to derive

$$\dot{M}_{\rm crit} = \frac{3\alpha c_s^3}{2G} \tag{2.186}$$

for the critical accretion rate at which gravitational instability sets in. Below this the disc is stable, but above this it is likely to undergo fragmentation (again, provided that it is efficiently cooled and is not able to maintain the  $Q \sim 1$  steady-state).

# **3** Numerical methods

"A common mistake that people make when trying to design something completely foolproof is to underestimate the ingenuity of complete fools."

Douglas Adams

# **3.1** Numerical astrophysics

Theoretical studies of astrophysical phenomena before the advent of computers were limited to analytical and, at best, semi-analytical methods, restricting the range of physical systems and behaviours that could be explored. Many systems in the universe do not exhibit the required steady-state behaviour to be sufficiently modelled by analytical formulae, and so numerical methods must be employed to follow their time evolution. Computational modelling in an astrophysical context seeks to follow the interactions between objects (be they celestial bodies, or simply particles in a gas) in order to infer their dynamics. This is done through simulation, evolving the system one step at a time whilst preserving as far as possible the accuracy and the self-consistency of the method.

Naturally, the inclusion of correct physical laws into such a simulation is essential, but the scope is often limited by computational resources. As a result, many models prioritise the most relevant areas of physics for inclusion and leave out the rest; as well as allowing the simulation to run in a reasonable time this can often simplify the study of individual processes and help to determine the contribution that they make to the wider picture. An example of one such method is an N-body simulation, in which the behaviour of an astrophysical system subject only to gravity is explored through the computation of the relevant forces on each particle.

*N*-body codes are widely used in the fields of stellar dynamics and cosmology, and can provide an accurate description of the system when hydrodynamical effects are not important (such as in the modelling of dark matter). If a sufficient quantity of gas is present, however, the model must incorporate additional physics to cater for the extra complexity. The inclusion of gas dynamics is often essential to the design of an astrophysical simulation, as much of the observable Universe has been shaped by fluid processes where pressure effects and dissipation are important.

There are two main types of hydrodynamical simulation, which determine the behaviour of the system either from a static or comoving<sup>1</sup> point of view:

- Eulerian codes solve for the equations of hydrodynamics via a grid that is overlaid on the spatial domain. The (Eulerian) derivatives of the motion are calculated at each grid point, and the physical variables (e.g., pressure, density, etc.) are evolved for each cell. The grid itself can either be fixed or made to adaptively refine in particular regions of interest. Grid codes offer excellent treatment of flow discontinuities (e.g., shocks) and have seen much development in the field of computational fluid dynamics over the last 50 years<sup>2</sup>.
- Lagrangian codes work within the framework of an N-body approach. The calcula-

<sup>&</sup>lt;sup>1</sup>the term comoving here refers to moving with the *fluid*, rather than the expansion of space. <sup>2</sup>for a review see the ENZO code paper by O'Shea et al. (2004)

tion here is placed squarely on the 'bodies' in the flow. The bodies themselves are very rarely individual particles<sup>3</sup> - note that this is particularly true for an astrophysical system - but rather Monte-Carlo representations of the continuous distribution function of the particles (see Section 3.2 for more details). This method is gridless, inherently adaptive, and can be constructed to explicitly conserve momentum, making it a powerful tool for modelling non-axisymmetric phenomena where the dynamic range is large and substantial variation in resolution is required.

The difference between these two approaches lies in the calculation of time derivatives. In the Eulerian case these are determined at a fixed point in space, while in the Lagrangian picture they are evaluated in a frame of reference that is attached to a moving fluid element. The relationship between the Eulerian time derivative  $\partial/\partial t$  and the Lagrangian time derivative D/Dt operators is given by

$$\frac{\mathrm{D}}{\mathrm{D}t} = \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla \tag{3.1}$$

where  $\mathbf{v} = \mathbf{v}(\mathbf{x}, t)$  is the flow velocity from an Eulerian perspective, namely

$$\mathbf{v} = \frac{\partial \mathbf{x}}{\partial t} \tag{3.2}$$

for the position vector  $\mathbf{x}$  of a fluid element. The choice of approach depends on the system in question, and in general no single method is suited to all problems. For the purposes of the research presented in this thesis, we favour the Lagrangian viewpoint, as it adapts well to the asymmetries and dynamic ranges contained within our numerical models (cf. Chapters 4 & 5). Even within a particular approach, however, there exist a large variety of different codes available for modelling astrophysical systems. The applicability of some of these codes, both grid-based and particle-based methods, to particular astrophysical problems has been reviewed by Frenk et al. (1999) and Tasker et al. (2008).

In this chapter we outline one particular Lagrangian method, known as smoothed particle hydrodynamics (SPH). We start with an introduction to the method, proceed to derive the relevant equations, and discuss some specifics of its implementation as a numerical scheme for astrophysical flows. The rest of the chapter is divided into a section on gravitational algorithms (for both SPH and N-body), a discussion of time integration methods, and finally an outline of the specific (SPH + gravity) code that was used for the simulations presented in this thesis.

<sup>&</sup>lt;sup>3</sup>i.e., baryons, leptons, etc.

# 3.2 Smoothed particle hydrodynamics (SPH)

The SPH method was developed by Lucy (1977); Gingold and Monaghan (1977) as an alternative to grid-based finite difference schemes that were not sufficient to simulate the internal dynamics of non-spherical stars. There has been much development of the method since its conception and it is now a robust simulation technique that is widely used in cosmology and astrophysics, as well as fluid dynamics in general. In this section we give a brief outline of the method; for further details see the review by Monaghan (1992).

Whilst grid-based codes seek to discretise space, Lagrangian codes such as SPH seek to discretise mass. As the name suggests, the SPH method is built on the principle of a smooth, or *continuous*, fluid, the properties of which are calculated by interpolating over a set of disordered points (fluid elements) that move with the flow. These elements are deemed 'particles', and possess a spatial distance known as the smoothing length. The *form* of this smoothing is set by an interpolating kernel which acts over a distance equal to the smoothing length. The requirements for the kernel are that it integrates to unity over all space and shrinks to a delta function in the limit that the smoothing length goes to zero. Mathematically, we can express a scalar function defined over an arbitrary volume V by the integral

$$f(\mathbf{r}) = \int_{V} f(\mathbf{r}') \,\delta(\mathbf{r} - \mathbf{r}') \,\mathrm{d}\mathbf{r}'$$
(3.3)

where  $\mathbf{r}'$  is a dummy variable. Approximating this by an integral interpolant (assuming a symmetric kernel; see Section 3.2.1) our expression becomes

$$f(\mathbf{r}) = \int f(\mathbf{r}') W(\mathbf{r} - \mathbf{r}', h) \,\mathrm{d}\mathbf{r}' + \mathcal{O}(h^2)$$
(3.4)

where h is the smoothing length, and W is the interpolating kernel satisfying the conditions already mentioned, namely

$$\int W(\mathbf{r} - \mathbf{r}', h) \,\mathrm{d}\mathbf{r}' = 1 \tag{3.5}$$

and

$$\lim_{h \to 0} W(\mathbf{r} - \mathbf{r}', h) = \delta(\mathbf{r} - \mathbf{r}')$$
(3.6)

Evaluating  $f(\mathbf{r})$  numerically over a number of discrete points, the integral turns into a sum. If we have N points distributed over our spatial domain such that

$$n(\mathbf{r}) = \sum_{j=1}^{N} \delta(\mathbf{r} - \mathbf{r}_j)$$
(3.7)

and we consider that

$$n(\mathbf{r}) = \frac{\rho(\mathbf{r}_j)}{m_j} \tag{3.8}$$

where  $\rho$  is the density and m the mass, our function  $f(\mathbf{r})$  becomes

$$f(\mathbf{r}) \approx \sum_{j=1}^{N} \frac{m_j}{\rho(\mathbf{r}_j)} f(\mathbf{r}_j) W(\mathbf{r} - \mathbf{r}_j, h)$$
(3.9)

Equation 3.9 can then be used to compute approximations for the various physical properties of the fluid, by means of choosing  $f(\mathbf{r})$  to represent a physical quantity, such as density:

$$\rho(\mathbf{r}) = \sum_{j=1}^{N} m_j W(\mathbf{r} - \mathbf{r}_j, h)$$
(3.10)

Vector quantities of the fluid properties (such as the gradient) can also be obtained, in a similar way. The SPH formalism therefore allows a convenient way of smoothing the hydrodynamical properties of each particle over the kernel to obtain a continuous distribution throughout the fluid.

The SPH method is of course an approximation to the behaviour of a real fluid, and errors arise both through (a) representing each particle by means of a smoothing kernel and (b) interpolating over a finite set of points i.e., moving from a integral interpolant to that of a summation. To quantify these errors we can use a Taylor series expansion, namely,

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \cdots$$
(3.11)

for a function f(x) that is infinitely differentiable in the neighbourhood of a. For the error introduced by convolving  $f(\mathbf{r})$  with a particular kernel we expand the integral form of the function, which in 1D is

$$f(\mathbf{r}') = f(\mathbf{r}) + \frac{\mathrm{d}f}{\mathrm{d}\mathbf{r}} \int (\mathbf{r}' - \mathbf{r}) W(r) \,\mathrm{d}\mathbf{r}' + \frac{1}{2} \frac{\mathrm{d}^2 f}{\mathrm{d}\mathbf{r}^2} \int (\mathbf{r}' - \mathbf{r})^2 W(r) \,\mathrm{d}\mathbf{r}' + \mathcal{O}[(\mathbf{r}' - \mathbf{r})^3] \quad (3.12)$$

where  $r \equiv |\mathbf{r}' - \mathbf{r}|$ . In 3D this has the same form but is performed over all three coordinate directions (see e.g., Price, 2005). The kernel error shrinks as the smoothing length, h, decreases, as r will similarly decrease (see Section 3.2.1).

The error in the summation interpolant can be quantified by expanding our vector quantity  $f(\mathbf{r})$  from equation 3.9 around  $\mathbf{r}_i$ :

$$f(\mathbf{r}) = f(\mathbf{r}_i) \sum_j \frac{m_j}{\rho(\mathbf{r}_j)} W_{ij} + \nabla f(\mathbf{r}_i) \cdot \sum_j \frac{m_j}{\rho(\mathbf{r}_j)} (\mathbf{r}_j - \mathbf{r}_i) W_{ij} + \mathcal{O}[(\mathbf{r}_j - \mathbf{r}_i)^2]$$
(3.13)

where  $W_{ij} \equiv W(\mathbf{r}_i - \mathbf{r}_j, h)$ . There is therefore an error even if only the first term is present (for example, in the case of a constant function), as the discretisation does not guarantee

that the sum over the neighbours is unity i.e., that

$$\sum_{j} \frac{m_j}{\rho(\mathbf{r}_j)} W_{ij} = 1 \tag{3.14}$$

where we see that this error relates to the *smoothness* of the particle distribution inside the kernel; as we have mentioned, the SPH approximation is built on the premise that the fluid is smooth and infinitely differentiable everywhere, and breaks down otherwise. Incorporating a greater number of neighbours within the smoothing kernel can ameliorate this discretisation error to some degree.

It can be shown (see, e.g., Price, 2005; Cossins, 2010a) that similar errors arise in the derivatives of the relevant functions, and can be reduced in the same way; namely, by decreasing the smoothing length, and increasing the neighbour number. For values of  $N_{\text{neigh}}$  within certain ranges, however, it can be shown that the particle distribution is unstable to wave-like perturbations (Read et al., 2010) and can undergo 'clumping' instabilities (for longitudinal waves) and 'banding' instabilities (for transverse waves) that exacerbate the *smoothness* error through undersampling of the kernel. It is therefore important to use a neighbour number that ensures stability.

# 3.2.1 Choice of smoothing kernel

Aside from the requirement of satisfying equations 3.5 and 3.6, the kernel can be chosen to minimise the error introduced by the SPH approximation. If the kernel is an even function of the spatial vector,  $\mathbf{r} - \mathbf{r}'$ , the odd terms in equation 3.12 reduce to zero and the Taylor approximation becomes  $\mathcal{O}(h^2)$ , since in principle  $|\mathbf{r} - \mathbf{r}'|$  is always less than the smoothing radius. This is physically equivalent to requiring that the kernel be a function of the magnitude of the spatial vector and not the direction. In three dimensions, a kernel of this form can be written:

$$W(r,h) = \frac{1}{\pi^{3/2}h^3} f\left(\frac{r}{h}\right)$$
(3.15)

where  $r \equiv |\mathbf{r} - \mathbf{r}'|$ . A simple example of such a kernel is a Gaussian, given by:

$$W(r,h) = \frac{1}{\pi^{3/2}h^3}e^{-(r/h)^2}$$
(3.16)

which has the advantages of being smooth i.e., differentiable, and having good stability<sup>4</sup>. However, the disadvantage of this kernel is that its support is infinite, meaning that the summation is performed over the entire computational domain and becomes  $\mathcal{O}(N^2)$  in terms of cost. This is not a desirable property, as the relative contribution to the sum from neighbouring particles decreases sharply with increasing distance. It is better to

<sup>&</sup>lt;sup>4</sup>in that its first derivative is also smooth

employ a kernel with compact support and reduce the calculation to close neighbours. This leads to a significant reduction in the cost of the kernel summation, which ends up as  $\mathcal{O}(nN)$  for *n* neighbours.

A good choice for a kernel is therefore one which possesses a similar shape to a Gaussian but has compact support. The most commonly used type is based on a cubic spline (Monaghan and Lattanzio, 1985), which is of the form

$$W(r,h) = \frac{1}{\pi h^3} \begin{cases} 1 - \frac{3}{2} (\frac{r}{h})^2 + \frac{3}{4} (\frac{r}{h})^3, & 0 \le \frac{r}{h} \le 1, \\ \frac{1}{4} (2 - \frac{r}{h})^3, & 1 < \frac{r}{h} \le 2, \\ 0, & \frac{r}{h} > 2 \end{cases}$$
(3.17)

which satisfies the normalisation condition and the requirement that the kernel reduce to a delta function in the zero limit, as well as having continuous first derivatives. The (compact) support of this kernel is 2h.

# 3.2.2 'Gather' vs. 'scatter' approach

The SPH interpolation for a vector quantity, i.e.

$$\rho(\mathbf{r}_i) = \sum_{j=1}^N m_j W(|\mathbf{r}_i - \mathbf{r}_j|, h)$$
(3.18)

can be interpreted in two different ways:

- 1. The gather interpretation treats the neighbours of particle *i* as point mass particles that reside within the smoothing kernel of *i*; essentially those *j* particles whose centres fall within the kernel are 'gathered' by particle *i* into the local average (left-hand plot in Figure 3.1). The summation therefore employs  $W(|\mathbf{r}_i \mathbf{r}_j|, h_i)$ .
- 2. The scatter interpretation treats *i* as a point mass and considers the contributions from the *j* smoothing kernels at that point i.e. the smoothing volumes of *j* 'scatter' on to location  $\mathbf{r}_i$  (right-hand plot in Figure 3.1). The summation here uses  $W(|\mathbf{r}_i \mathbf{r}_j|, h_j)$ .

For constant smoothing lengths i.e.  $h_i = h_j$  these two interpretations are equivalent, but if adaptive smoothing lengths are used they are not. In the latter case an average of the smoothing length must therefore be taken to ensure that Newton's 3rd law is not violated; this will be discussed in Section 3.2.4.



Figure 3.1: The gather (left) and scatter (right) approaches to the SPH interpolation process.

# 3.2.3 Fluid equations

In this section we detail the basic fluid equations that are used in the SPH formalism. They are constructed directly from conservation laws, making SPH a first-order fully conservative method in the absence of dissipative processes.

# 3.2.3.1 Conservation of mass

For a fluid we can write the continuity equation as

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 \tag{3.19}$$

where  $\mathbf{v}$  is the particle velocity. We can see that this equation is satisfied by the SPH density estimate (equation 3.10), which we write as,

$$\rho_i = \sum_j m_j W_{ij} \tag{3.20}$$

where  $\rho_i \equiv \rho(\mathbf{r})$ , i.e. the estimate is evaluated at particle *i*. Taking the Lagrangian time derivative, namely,

$$\frac{\mathrm{D}\rho_i}{\mathrm{D}t} = \frac{\partial\rho_i}{\partial t} + \mathbf{v} \cdot \nabla_i \rho_i \tag{3.21}$$

where  $\nabla_i \equiv \partial/\partial \mathbf{r}_i$ , we have

$$\frac{\mathrm{D}\rho_i}{\mathrm{D}t} = \frac{\partial\rho_i}{\partial t} + \sum_j m_j (\mathbf{v}_i - \mathbf{v}_j) \cdot \nabla_i W_{ij}$$
(3.22)

If we convert from the summation form to continuum form, using equations 3.4 and 3.9, we end up with

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} = \mathbf{v} \cdot \nabla \rho - \nabla \cdot (\rho \mathbf{v}) \tag{3.23}$$

$$= -\rho(\nabla \cdot \mathbf{v}) \tag{3.24}$$

which is simply the Lagrangian form of equation 3.19. The SPH density estimate is therefore a discrete form of the continuity equation, demonstrating that the SPH method explicitly conserves mass.

#### 3.2.3.2 Conservation of momentum

To derive the momentum equation we start with the Lagrangian:

$$L = T - U \tag{3.25}$$

where we are ignoring the gravitational potential and considering only the kinetic energy T and the internal energy U. For a fluid with density  $\rho$  evaluated over a volume this becomes

$$L = \int (\frac{1}{2}\rho v^2 - \rho u) \,\mathrm{d}V$$
 (3.26)

where u is the specific internal energy. Discretising this over the SPH particle neighbours, we get

$$L = \sum_{j} m_j \left(\frac{1}{2}v_j^2 - u_j\right) \tag{3.27}$$

where the volume element  $\rho dV$  has been replaced by the SPH particle mass m. To obtain the equation of motion of a particle we find the extremum<sup>5</sup> of L by employing the Euler-Lagrange equation over the coordinates ( $\mathbf{r}, \mathbf{v}$ ):

$$\frac{\partial L}{\partial \mathbf{r}_i} - \frac{\mathrm{d}}{\mathrm{d}t} \left( \frac{\partial L}{\partial \mathbf{v}_i} \right) = 0 \tag{3.28}$$

The derivative with respect to the velocity comes out as simply

$$\frac{\partial L}{\partial \mathbf{v}_i} = m_i \mathbf{v}_i \tag{3.29}$$

 $<sup>^{5}</sup>$ since in the absence of external influences the particle's trajectory will be the one corresponding to the minimum energy

whilst the derivative with respect to the position is given by

$$\frac{\partial L}{\partial \mathbf{r}_i} = \sum_j \frac{\partial u_j}{\partial \rho_j} \cdot \frac{\partial \rho_j}{\partial \mathbf{r}_i} \tag{3.30}$$

The first law of thermodynamics states that:

$$\mathrm{d}U = \mathrm{d}Q - \mathrm{d}W \tag{3.31}$$

$$= T dS - P dV \tag{3.32}$$

where S is the entropy and P is the pressure. Making the assumption that the fluid is dissipationless i.e., isentropic, we have

$$\mathrm{d}U|_s = -P\mathrm{d}V\tag{3.33}$$

and the specific internal energy comes out as

$$u = -\frac{PV}{m} = -\frac{P}{\rho} \tag{3.34}$$

giving us

$$\left. \frac{\partial u_j}{\partial \rho_j} \right|_s = \frac{P}{\rho_j^2} \tag{3.35}$$

We can obtain the other derivative by substituting in for  $\rho$  from equation 3.20,

$$\frac{\partial \rho_j}{\partial \mathbf{r}_i} = \frac{\partial}{\partial \mathbf{r}_i} \sum_k m_k W_{jk} \tag{3.36}$$

where we have an additional subscript k that refers to the neighbours of particle j. This subscript is necessary for evaluating a derivative of particle j with respect to a quantity defined on particle i; since i is fixed, a third reference point is needed as a dummy subscript to maintain generality. We can employ the Kronecker delta function to relate the kparticles to i and j, namely:

$$\delta_{ij} = \begin{cases} 1, & j = i \\ 0, & \text{otherwise} \end{cases}$$
(3.37)

This allows us to take the derivative of  $W_{jk}$  with respect to  $\mathbf{r}_i$ :

$$\frac{\partial \rho_j}{\partial \mathbf{r}_i} = \sum_k m_k \nabla_i W_{jk} (\delta_{ji} - \delta_{ki}) \tag{3.38}$$

through simple vector addition. The first term in equation 3.28 can therefore be written as

$$\frac{\partial L}{\partial \mathbf{r}_i} = \sum_j m_j \frac{P_j}{\rho_j^2} \sum_k m_k \nabla_i W_{jk} (\delta_{ji} - \delta_{ki})$$
(3.39)

Using the fact that  $W_{jk}\delta_{ji} = W_{ik}$  at j = i, and  $W_{jk}\delta_{ki} = W_{ji}$  at k = i, we can say

$$\frac{\partial L}{\partial \mathbf{r}_i} = \sum_k m_i \frac{P_i}{\rho_i^2} m_k \nabla_i W_{ik} - \sum_j m_j \frac{P_j}{\rho_j^2} m_i \nabla_i W_{ji}$$
(3.40)

In this instance *i* is fixed but both *j* and *k* are independent dummy indices, and are therefore interchangeable. We can thus set  $j \equiv k$ , and, together with the fact that the kernel gradient is anti-symmetric i.e.  $\nabla_i W_{ij} = -\nabla_i W_{ji}$ , we can obtain the expression

$$\frac{\partial L}{\partial \mathbf{r}_i} = m_i \sum_j m_j \left( \frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2} \right) \nabla_i W_{ij}$$
(3.41)

The Euler-Lagrange equation (3.28) has now become

$$m_i \sum_j m_j \left(\frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2}\right) \nabla_i W_{ij} - \frac{d}{dt}(m_i \mathbf{v}_i) = 0$$
(3.42)

so our equation of motion for the acceleration of particle i is given by

$$\frac{\mathrm{d}\mathbf{v}_i}{\mathrm{d}t} = -\sum_j m_j \left(\frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2}\right) \nabla_i W_{ij} \tag{3.43}$$

for an isentropic fluid i.e. in the absence of dissipation. This equation can be used to demonstrate both conservation of momentum and conservation of angular momentum; for the former, since the kernel is anti-symmetric the equation of motion contributes an equal and opposite amount of momentum to particle *i* than it contributes to particle *j*, namely  $d\mathbf{v}_j/dt = -d\mathbf{v}_i/dt$  and momentum is thus manifestly conserved (pair-wise). To see the latter we require a little more working, and so we compute the time derivative of *J*, the total angular momentum, as

$$\frac{\mathrm{d}J}{\mathrm{d}t} = \sum_{i} m_i \left( \mathbf{r}_i \times \frac{\mathrm{d}\mathbf{v}_i}{\mathrm{d}t} \right) \tag{3.44}$$

$$=\sum_{i}\sum_{j}m_{i}m_{j}\left(\frac{P_{i}}{\rho_{i}^{2}}+\frac{P_{j}}{\rho_{j}^{2}}\right)\left(\mathbf{r}_{i}\times\nabla_{i}W_{ij}\right)$$
(3.45)

The time derivative of J is a sum over all particles and is therefore independent of both i and j. These function as interchangeable dummy indices, allowing us to set the previous
expression equal to its equivalent with the indices reversed i.e.

$$\frac{\mathrm{d}J}{\mathrm{d}t} = \sum_{j} \sum_{i} m_{j} m_{i} \left( \frac{P_{j}}{\rho_{j}^{2}} + \frac{P_{i}}{\rho_{i}^{2}} \right) \left( \mathbf{r}_{j} \times \nabla_{j} W_{ji} \right)$$
(3.46)

Due to the anti-symmetry of the kernel gradient, this further implies

$$\frac{\mathrm{d}J}{\mathrm{d}t} = -\sum_{i} \sum_{j} m_{i} m_{j} \left( \frac{P_{i}}{\rho_{i}^{2}} + \frac{P_{j}}{\rho_{j}^{2}} \right) \left( \mathbf{r}_{i} \times \nabla_{i} W_{ij} \right)$$
(3.47)

which is the negative of the expression in equation 3.45. For this to be the case, the time derivative of J must be zero, implying that the total angular momentum must remain constant for the duration of the SPH simulation.

#### 3.2.3.3 Conservation of energy

There are a number of ways in which the energetic properties of an SPH particle can be evolved. Depending on the problem at hand, one may choose to integrate the specific internal energy, the specific entropy, or the specific total energy. Here we discuss the latter method. For a review on the relative merits of the internal energy and entropy formulations the reader is directed to (Springel and Hernquist, 2002).

Integrating the specific total energy naturally implies total energy conservation. We construct the Hamiltonian,

$$H = \sum_{i} \mathbf{p}_{i} \dot{\mathbf{q}}_{i} - L \tag{3.48}$$

where (q, p) are generalised coordinates for the position and momentum respectively, and L is the Lagrangian. We can write the momentum in terms of the derivative of L with respect to the velocity,

$$p_i = \frac{\partial L}{\partial \mathbf{v}_i} \tag{3.49}$$

and using the expression for the Lagrangian as

$$L = \sum_{j} m_j \left(\frac{1}{2}\mathbf{v}_j^2 - u_j\right) \tag{3.50}$$

the total energy becomes

$$E_i = \sum_i m_i \left(\frac{1}{2}\mathbf{v}_i^2 - u_i\right) \tag{3.51}$$

Taking the time derivative:

$$\frac{\mathrm{d}E_i}{\mathrm{d}t} = \sum_i m_i \left( \mathbf{v}_i \frac{\mathrm{d}\mathbf{v}_i}{\mathrm{d}t} + \frac{\mathrm{d}u_i}{\mathrm{d}t} \right) \tag{3.52}$$

We have already calculated the acceleration term i.e., equation 3.43, and we can obtain the derivative of the internal energy from equations 3.34 and 3.22:

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = \frac{P_i}{\rho_i^2} \frac{\mathrm{d}\rho_i}{\mathrm{d}t} \tag{3.53}$$

$$= \frac{P_i}{\rho_i^2} \sum_j m_j (\mathbf{v}_i - \mathbf{v}_j) \cdot \nabla_i W_{ij}$$
(3.54)

Substituting these derivatives in, we have

$$\frac{\mathrm{d}E_i}{\mathrm{d}t} = \sum_i \sum_j m_i m_j \left(\frac{P_i}{\rho_i^2} \mathbf{v}_j + \frac{P_j}{\rho_j^2} \mathbf{v}_i\right) \cdot \nabla_i W_{ij} \tag{3.55}$$

where as per our discussion of angular momentum conservation in Section 3.2.3.2, we note that a sum over all particles allows us to reverse the indices and through the antisymmetry set the above expression equal to its negative. We therefore see that  $dE_i/dt = 0$ and so total energy is manifestly conserved in the absence of dissipation. Finally, for the derivative of the specific total energy we simply remove the summation over *i* and the mass of particle *i* from the expression, giving

$$\frac{\mathrm{d}e_i}{\mathrm{d}t} = \sum_j m_j \left( \frac{P_i}{\rho_i^2} \mathbf{v}_j + \frac{P_j}{\rho_j^2} \mathbf{v}_i \right) \cdot \nabla_i W_{ij} \tag{3.56}$$

#### 3.2.3.4 Equation of state

The continuity equation (3.19), momentum equation (3.43) and energy equation (3.56) together make up the transport equations for a non-gravitating fluid at constant entropy. This system of equations must be closed by an equation of state (EQS), which relates the pressure P to the density  $\rho$ . For an ideal gas this is

$$P_i = (\gamma - 1)u_i\rho_i \tag{3.57}$$

where  $\gamma$  is the adiabatic index and  $u_i$  is the internal energy.

# 3.2.4 Adaptive smoothing lengths

The conservation equations discussed in Section 3.2.3 were derived assuming a fixed value of the smoothing length, h, i.e. one that is constant in space<sup>6</sup>. The true Lagrangian nature of SPH however is not realised unless the smoothing length is allowed to vary, achieving high resolution in regions of high density. These adaptive smoothing lengths are normally chosen to maintain a nearly fixed number of neighbours (Attwood et al., 2007). However,

<sup>&</sup>lt;sup>6</sup>although not necessarily in time

neither the gather nor scatter formulations (see Section 3.2.2) are symmetric in i and jand so spatially varying smoothing lengths can violate the conservation of momentum and energy. It is possible to symmetrize the dynamical equations by taking an average of the smoothing kernels (Hernquist and Katz, 1989), namely,

$$W_{ij} = \frac{1}{2} [W_{ij}(h_i) + W_{ij}(h_j)]$$
(3.58)

Unfortunately, this does not completely solve the problem since taking the spatial derivative of the kernel when adaptive smoothing lengths are employed necessarily generates an extra term in  $\nabla h$ , i.e.

$$\nabla W_{ij} = \nabla W_{ij}|_h + \frac{\partial}{\partial h} W_{ij} \nabla h \tag{3.59}$$

which, whilst generally neglected (Nelson and Papaloizou, 1994), can become important if the smoothing length varies on scales less than h itself.

#### 3.2.4.1 Fully conservative formulation

A fully conservative formulation of SPH can be derived using a variational approach (Springel and Hernquist, 2002). This derivation takes the  $\nabla h$  terms into account by ensuring that h is defined such that a fixed *mass* is contained within the kernel volume, namely

$$\frac{4\pi}{3}h_i^3\rho_i = N_{\rm sph}m_{\rm sph} \tag{3.60}$$

where  $N_{\rm sph}$  is the number of neighbours and  $m_{\rm sph}$  is the particle mass. This above expression acts as a constraint on the Lagrangian, which we extremise in a similar manner to the derivation of the momentum equation in Section 3.2.3. We write the Lagrangian in terms of an entropic function A(s) and assuming an ideal gas equation of state, namely  $P = A(s)\rho^{\gamma}$ . The latter choice allows us to evolve the entropy, rather than the internal energy or the total energy (see Section 3.2.3.3). This is the choice made in Springel and Hernquist (2002) and indeed in GADGET (Springel, 2005) as it ensures conservation of entropy and appears to function better (in a standard SPH formulation) than the internal energy equation, particularly at sharp density contrasts<sup>7</sup>. We therefore write the Lagrangian as,

$$L(\mathbf{q}, \dot{\mathbf{q}}) = \frac{1}{2} \sum_{j} m_{j} \mathbf{v}_{j}^{2} - \frac{1}{\gamma - 1} \sum_{j} m_{j} A_{j} \rho_{j}^{\gamma - 1}$$
(3.61)

where  $\mathbf{q} = (\mathbf{r}_i, h_i)$  is our independant variable, and  $A_j \equiv A(s_j)$ . We require the extremum of equation 3.61 subject to the constraint 3.60, for which we use the Euler-Lagrange

<sup>&</sup>lt;sup>7</sup>note however that there does not appear to be any reason why a *conservative* internal energy formulation would be any worse than its entropy equivalent, at least that the author knows of

equation together with the method of Lagrange multipliers:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{\mathbf{q}}_i} - \frac{\partial L}{\partial \mathbf{q}_i} = \sum_j \lambda_j \frac{\partial g(\mathbf{q}_j)}{\partial \mathbf{q}_j} \tag{3.62}$$

where  $g(\mathbf{q}_j)$  is the constraining function. We first take the derivatives with respect to  $h_i$ , which reduces the equation to:

$$\sum_{i} m_i A_i \rho_i^{\gamma - 2} \frac{\partial \rho_i}{\partial h_i} - \sum_{i} \lambda_i \left( 4\pi \rho_i h_i^2 + \frac{4\pi}{3} h_i^3 \frac{\partial \rho_i}{\partial h_i} \right) = 0$$
(3.63)

where we have used the fact that  $\partial h_j / \partial h_i = 1$  at i = j but zero otherwise. Re-arranging for  $\lambda_i$  we find

$$\lambda_i = \frac{3}{4\pi} \frac{m_i}{h_i^3} \frac{P_i}{\rho_i^2} \left[ 1 + \frac{3\rho_i}{h_i} \frac{\partial h_i}{\partial \rho_i} \right]^{-1}$$
(3.64)

where we have substituted in for  $P_i$  from the equation of state. We now take the derivatives of equation 3.62 with respect to  $r_i$ , with the expression for  $\lambda_i$  that we have just calculated, giving us:

$$m_i \frac{\mathrm{d}\mathbf{v}_i}{\mathrm{d}t} = -\frac{1}{\gamma - 1} \frac{\partial}{\partial \mathbf{r}_i} \sum_j m_j A_j \rho_j^{\gamma - 1} + \sum_j \frac{P_j}{\rho_j^2} \left[ 1 + \frac{3\rho_j}{h_j} \frac{\partial h_j}{\partial \rho_j} \right]^{-1} \frac{\partial \rho_j}{\partial \mathbf{r}_i} \tag{3.65}$$

since  $\rho_j$  is independent of  $\mathbf{v}_i$  and  $\mathbf{v}_j$  is independent of  $\mathbf{r}_i$ . Simplifying we get

$$m_i \frac{\mathrm{d}\mathbf{v}_i}{\mathrm{d}t} = -\sum_j m_j \frac{P_j}{\rho_j^2} \left[ 1 + \frac{h_j}{3\rho_j} \frac{\partial\rho_j}{\partial h_j} \right]^{-1} \nabla_i \rho_j \tag{3.66}$$

From equation 3.38 we can see that

$$\nabla_i \rho_j = m_i \nabla_i W_{ij}(h_j) + \delta_{ij} \sum_k m_k \nabla_i W_{ki}(h_i)$$
(3.67)

This yields the expression for the acceleration as

$$\frac{\mathrm{d}\mathbf{v}_i}{\mathrm{d}t} = -\sum_{j=1}^N m_j \left[ f_i \frac{P_i}{\rho_i^2} \nabla_i W_{ij}(h_i) + f_j \frac{P_j}{\rho_j^2} \nabla_i W_{ij}(h_j) \right]$$
(3.68)

for a non-gravitating flow. The coefficients  $f_i$ ,  $f_j$  arise as a direct result of the spatially varying smoothing length, and are given by:

$$f_i = \left[1 + \frac{h_i}{3\rho_i} \frac{\partial\rho_i}{\partial h_i}\right]^{-1} \tag{3.69}$$

For total energy conservation we obtain the expression for the internal energy with adaptive smoothing lengths as

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = f_i \frac{P_i}{\rho_i^2} \sum_j m_j \mathbf{v}_{ij} \cdot \nabla_i W_{ij}(h_i)$$
(3.70)

# 3.2.5 Dissipation and shock capturing

The SPH equations derived so far relate to an inviscid and isentropic fluid, free of dissipation. In reality, however, this is not a good approximation to astrophysical flows of ideal gases, in which viscous processes on a molecular scale readily generate entropy and heat, converting ordered kinetic energy into random kinetic energy. This is particularly the case at shocks, where properties of the flow change very rapidly over the mean free path of a gas particle. In a numerical scheme that does not resolve this scale, entropy must be generated artifically. Moreover, shocks appear discontinuous, and must be broadened over a smoothing length for pressure changes to be resolved, as well as to prevent large errors in the numerical scheme. Both requirements are dealt with by the introduction of an *artificial* viscosity that allows for dissipation in shocks and ensures that shock fronts are spread out on a scale large enough to be properly resolved. At the same time, this helps to prevent unphysical inter-particle penetration in high Mach number collisions.

#### 3.2.5.1 Artificial viscosity

The force due to the artificial viscosity is added as an extra term in the momentum and energy equations; for example, equation 3.68 becomes

$$\frac{\mathrm{d}v_i}{\mathrm{d}t} = -\sum_{j=1}^N m_j \left[ f_i \frac{P_i}{\rho_i^2} \nabla_i W_{ij}(h_i) + f_j \frac{P_j}{\rho_j^2} \nabla_i W_{ij}(h_j) \right] - \sum_{j=1}^N m_j \Pi_{ij} \nabla_i \overline{W}_{ij} \tag{3.71}$$

which, for conservation of the total energy as per Section 3.2.3.3, equation 3.54 becomes

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = \frac{P_i}{\rho_i^2} \sum_j m_j \mathbf{v}_{ij} \cdot \nabla_i W_{ij} + \sum_j m_j \Pi_{ij} \mathbf{v}_i \nabla_i \overline{W}_{ij}$$
(3.72)

where  $\overline{W}_{ij}$  is the arithmetic average of the kernels  $W_{ij}(h_i)$  and  $W_{ij}(h_j)$ , and  $\Pi_{ij}$  is the artificial viscosity. The standard form of the latter is that of Monaghan (1992), which is given by:

$$\Pi_{ij} = \begin{cases} (\alpha \mu_{ij} \overline{c}_{ij} + \beta \mu_{ij}^2) / \overline{\rho}_{ij}, & \mathbf{v}_{ij} \cdot \mathbf{r}_{ij} < 0\\ 0, & \mathbf{v}_{ij} \cdot \mathbf{r}_{ij} \ge 0 \end{cases}$$
(3.73)

3.3. Gravity solvers

$$\mu_{ij} = \frac{h_{ij} \mathbf{v}_{ij} \cdot \mathbf{r}_{ij}}{\mathbf{r}_{ij}^2 + 0.01 h_{ij}^2} \tag{3.74}$$

where  $h_{ij}$  is the arithmetic average of the smoothing length i.e.,  $h_{ij} = (h_i + h_j)/2$ ,  $\bar{c}_{ij}$  is the average sound speed, and  $\bar{\rho}_{ij}$  is the average density, while  $\alpha$  and  $\beta$  are free parameters. Dimensionally, both terms in  $\Pi_{ij}$  have the form  $v^2/\rho$  with v as a characteristic velocity; this is based on the fact that the force in the momentum equation (eqs. 3.43 and 3.68) is  $\propto P/\rho^2$ , and through the equation of state  $P \propto c_s^2 \rho$ . In particular, we see that the form of  $\mu$  is based on the velocity divergence, which gives a measure of the expansion or contraction of the fluid at a given point. To first order,  $|\nabla \cdot \mathbf{v}_{ij}| \approx |\mathbf{v}_{ij}|/|\mathbf{r}_{ij}| = |\mathbf{v}_{ij} \cdot \mathbf{r}_{ij}|/|\mathbf{r}_{ij}|^2$ . In addition, the viscosity should be proportional to the typical length scale, h, and includes a (small) perturbation term in the denominator to ensure stability as  $\mathbf{r}_{ij} \to 0$ .

The terms prefixed by  $\alpha$  and  $\beta$  have different properties. The former is linear in  $\mu$  and includes the average sound speed as a characteristic velocity of the flow. It is derived from the linear viscous stress tensor in the Navier-Stokes equation (see, e.g., Landau and Lifshitz, 1959; Meglicki et al., 1993) and has the advantage of providing stable and correct behaviour<sup>8</sup> in weak shocks, i.e.,  $\mathcal{M} \lesssim 5$  (Monaghan, 1985), but is often not sufficient in strong shocks to prevent inter-particle penetration (Lattanzio et al., 1985). In contrast, the quadratic term in  $\mu$  (von Neumann and Richtmyer, 1950) performs far better in strong shocks but decays too rapidly in weak shocks, giving rise to numerical 'post-shock oscillations' (Rosswog, 2009). Using the two in conjuction has been shown to produce good results in the majority of cases (Monaghan, 1989) when the coefficients are set to  $\alpha = 1$  and  $\beta = 2\alpha$ .

Finally, we note that the negative  $\mathbf{v}_{ij} \cdot \mathbf{r}_{ij}$  constraint ensures that the viscosity only acts on approaching particle pairs. In such a situation, therefore, the viscosity functions as an excess pressure force with  $P_{visc} \simeq \frac{1}{2}\rho_{ij}^2 \Pi_{ij}$ .

# 3.3 Gravity solvers

One of the four fundamental forces of nature, gravity drives structure formation on cosmological scales, and is therefore a important ingredient of any self-consistent astrophysical

$$\rho_1 v_1 = \rho_2 v_2 \tag{3.75}$$

$$\rho_1 v_1^2 + P_1 = \rho_2 v_2^2 + P_2 \tag{3.76}$$

$$\frac{1}{2}v_1^2 + u_1 + \frac{P_1}{\rho_1} = \frac{1}{2}v_2^2 + u_2 + \frac{P_2}{\rho_2}$$
(3.77)

where  $u = 1/(\gamma - 1)P/\rho$  is the internal energy of the gas and the subscripts 1 & 2 refer to pre- & post-shock regions respectively

<sup>&</sup>lt;sup>8</sup>in this case *correct* behaviour refers to the satisfaction of the pre- and post-shock fluid variables at scales larger than the smoothed shock front, as set by the Rankine-Hugoniot conditions (see, e.g., Clarke and Carswell, 2003), namely

simulation. Correctly following the evolution of a group of celestial bodies subject to the laws of gravitation is known as the N-body problem, and was first presented in a mathematical form in Isacc Newton's Principia (Newton, 1999) in 1687.

Unfortunately, analytical solutions to the N-body problem only exist for  $N \leq 3$  (Sundman, 1913) and so for larger values of N the equations must be solved numerically.

# 3.3.1 Force calculation

The basic requirement for a gravity solver is to obtain the forces on each particle at a given timestep, based on the underlying gravitational potential of the particle distribution. The latter is given by Poisson's equation for an arbitrary (smooth) density distribution  $\rho$ :

$$\nabla^2 \phi = 4\pi G\rho \tag{3.78}$$

There are a number of ways to do this, depending on the relative importance of the accuracy of the algorithm compared to the speed of computation.

# **3.3.1.1** Direct summation $\mathcal{O}(N^2)$

The simplest and most accurate approach to solving the N-body problem is numerical integration. The force experienced by a particle i from N-1 other particles j (assuming Newtonian gravity) is found from

$$m_i \ddot{\mathbf{r}}_i = G \sum_{i \neq j}^N m_i m_j \frac{\mathbf{r}_i - \mathbf{r}_j}{|\mathbf{r}_i - \mathbf{r}_j|^3}$$
(3.79)

This algorithm is slow, becoming prohibitively expensive for large numbers of particles due to its scaling as  $\mathcal{O}(N^2)$ . Originally the only method available for computing the gravitational force on a particle, it has since been superseeded by more efficient methods that either (a) utilise multipole expansions of ensembles of particles or (b) map the spatial distribution on to a regular grid. Its usage has endured far longer than would have been expected against these faster algorithms, however, due to the development of specialised hardware for pair-wise interactions that scale as  $r^{-2}$  (Hut and Makino, 1999). This technology is known as GRAPE (**GRA**vity **P**ip**E**), and remains competitive with other methods up to  $N \sim 100,000$ .

#### **3.3.1.2** Particle-mesh $\mathcal{O}(N \log N)$

One approach that is efficient for large numbers of particles, and easy to implement, is the particle-mesh (PM) method, which maps the underlying density distribution of the particles on to a regularly-spaced grid, or mesh. Two sets of interpolations are used: the



Figure 3.2: Recursive subdivision of a quad-tree

particle masses are interpolated on to the mesh, the Poisson equation is then solved on the mesh, and the forces at the vertices are interpolated back to the particles. The presence of the regular mesh allows for the use of a Fourier transform, solving for the Poisson equation in the frequency domain. In fact, the motivation behind the PM method was largely based on the development of the Fast Fourier Transform (FFT) which scales in complexity as  $\mathcal{O}(N \log N)$ . The PM method therefore scales in a similar manner.

In cases where sub-mesh resolution is required, such as in non-uniform distributions with strong density contrasts, the PM method can be optionally combined with a local direct summation approach to yield the particle-particle particle-mesh (P<sup>3</sup>M) method (Hockney and Eastwood, 1981). For hydrodynamical simulations, the PM and P<sup>3</sup>M algorithms are well-suited to Eulerian grid-based codes.

## **3.3.1.3 Tree algorithm** $\mathcal{O}(N \log N)$

The most popular choice for a gravity solver is that of a tree algorithm. This method makes use of the fact that the precise form of the mass distribution becomes less important with increasing distance. The potential of remote groups of particles can therefore be approximated to reasonable accuracy by low-order multipole expansions.

The algorithm employs the multipole expansion hierarchically over the spatial domain, grouping the particles by partitioning it into a series of cubes, or nodes (Barnes and Hut, 1989). The root node encompasses the full mass distribution, and is repeatedly subdivided into 'daughter' nodes down to 'leaf' nodes that mark the end of the tree structure. There are a number of different types of tree that can be used; Figure 3.2 depicts the subdivision of the spatial domain for a *quad-tree* which recursively partitions each node into four quadrants. The forces can then be calculated by starting at the root node and considering each of the daughter nodes in turn until the multipole expansion of the node in question provides an accurate enough partial force, subject to a prescribed accuracy parameter. This process is known as 'walking' the tree. The complexity of this method



Figure 3.3: Constructing the tree, starting at the root node, subject to a constraint of one particle per leaf node. Once this constraint is reached the 'leaves' are identified (grey boxes) and the tree is ready to be walked to determine the force on each particle.

scales as  $\mathcal{O}(N \log N)$ .

The procedure for a tree algorithm is as follows:

- 1. The initial step is to construct the tree, which is done by inserting the particles oneby-one and then walking the tree once subject to a constraint that identifies when a leaf node has been reached. The standard constraint is to require that each leaf node contains only one particle, an example of which is shown in Figure 3.3. During this process, information about each node is stored; for the lowest order multipole moment (i.e. monopole) this comprises just the mass of the particles contained within the node, the centre of mass of the node and the side length of the cell<sup>9</sup>.
- 2. Once the tree has been constructed, we return to the root node and walk the tree again for each particle, subject to an 'opening' criterion that determines whether a node must be expanded into its daughter nodes or whether the force from the particles in the node can be approximated by a single force from the centre of mass. An example of this step is shown in Figure 3.4.
- 3. After the force on a particular particle has been computed, the tree is reset and walked again for the next particle. The cell opening criterion can be adjusted for accuracy and/or speed as required.

For hydrodynamics, the tree can also be used to find the neighbours of the SPH particles (Hernquist and Katz, 1989). A spherical search region is defined around the particle in question, with radius  $h_i$ , and the tree is walked with an opening criterion that determines whether there is any geometric overlap between a node and the search region. In this way the tree functions as a nearest-neighbour finding algorithm.

<sup>&</sup>lt;sup>9</sup>note that going to higher order will increase the number of variables that need to be stored.



Figure 3.4: Walking the tree subject to an opening criterion. The particle shown in red is the reference particle, and the distance to the centre of mass of the node outlined in red is calculated. In this case the criterion is not satisfied and the node is 'accepted' rather than opened, the force from the particles in the node being replaced by a single monopole force acting from the centre of mass.

# **3.3.1.4** falcON $\mathcal{O}(N)$

The force solver falcON (force algorithm with complexity O(N)) is currently the fastest gravity solver for astrophysical simulations. It employs the fast multipole method (FMM), which performs a multipole expansion of the spatial Green's function<sup>10</sup> and applies it hierarchally via a tree.

Rather than evaluating the force on a particle-by-particle basis, falcON groups close particles into cells and exploits the similarity in the force from a distant cell upon these particles (Dehnen, 2002). This results in a substantial speed up of the force calculation over standard tree-based methods.

# 3.3.2 Gravitational softening

It has already been mentioned (§3.2) that for hydrodynamical simulations employing a Lagrangian approach, particles are represented by smoothing kernels that mediate the hydrodynamical forces across the particle distribution. However, this method can (and should) also be applied to the gravitational interaction, and indeed for a pure N-body simulation a 'softening' length is used to smooth out the gravitational force between bodies. This is essential for two reasons; firstly, to avoid divergence in the force at small separations due to the infinite potential well that a point particle would represent, and secondly, because the 'particles' themselves are usually randomly sampled elements of an underlying

$$\phi(\mathbf{r}_i) = -\sum_{j \neq i} m_j g(\mathbf{r}_i - \mathbf{r}_j)$$
(3.80)

so here we have  $g(\mathbf{r}_i - \mathbf{r}_j) = G/|\mathbf{r}_i - \mathbf{r}_j|$ , where G is the gravitational constant.

<sup>&</sup>lt;sup>10</sup>A Green's function solves a partial differential equation over a continuous distribution e.g.  $\nabla^2 \phi = 4\pi G \rho$  by representing it with the underlying distribution of delta functions, making it akin to performing a Fourier analysis. For the solution to Poisson's equation, the Green's function is found from

distribution function, encompassing hundreds, if not thousands, of actual bodies<sup>11</sup>.

The use of gravitational softening can also help to prevent unphysical two-body interactions (such as those that might form binaries), effectively ensuring that the potential felt by an individual particle from all the other particles is smooth and that the system is collisionless. The specific implementation of the softening is often model-dependant, but for N-body simulations a standard choice is a Plummer sphere, which is a density profile designed to fit observations of globular clusters (Plummer, 1911). The force in this case is softened according to

$$\mathbf{F}_{i} = -\frac{Gm_{i}m_{j}}{r_{ij}^{2} + \epsilon_{i}^{2}}\hat{\mathbf{r}}_{ij}$$
(3.81)

for particle *i*, where  $\epsilon$  is the softening length. When combined with an SPH code, however, a sensible choice is to use a softening law that is identical to the smoothing kernel (cf. equation 3.17). This has an advantage over a Plummer model in that the gravity reduces to Newtonian outside  $2\epsilon$ , whereas in the Plummer case the force converges much more slowly to this value. It also allows for identical resolutions in both the gravitational and hydrodynamical calculations, provided that an  $\epsilon = h$  condition is enforced (Bate and Burkert, 1997).

# 3.4 Time integration

The implementation of a numerical simulation requires that we solve the relevant equations of motion and evolve the system forward in time. Whether our code is purely gravitational, or includes hydrodynamics, the overall behaviour of the particles is governed by a set of coupled first-order differential equations. Integrating these equations must be done numerically (cf. Section 3.3) and the accuracy to which the integration can be performed varies over the different numerical algorithms available. As with any such method, we seek a balance between accuracy and efficiency, which in general translates into a fine-tuning of the timestepping criteria for an integrator of a given order.

# 3.4.1 Euler method

The basic concept of numerical integration is best presented via the Euler method, which determines an approximate solution to an initial value problem evolved forward in time via a linear approximation. Namely, for the initial value of a function  $\mathbf{x}(t_0) = \mathbf{x}_0$ , and the time derivative  $\dot{\mathbf{x}} = f(\mathbf{x}(t), t)$  the first step of the Euler algorithm will be

$$\mathbf{x}_1 = \mathbf{x}_0 + f(\mathbf{x}_0, t_0)\Delta t \tag{3.82}$$

<sup>&</sup>lt;sup>11</sup>this estimate assumes a stellar dynamical simulation; for dark matter or indeed gas each body can represent many, many more orders of magnitude in actual particles!



Figure 3.5: Schematic of the leapfrog algorithm over 3 timesteps.

where  $\Delta t$  is the timestep. This process is repeated for each step of the integration; obtaining the n + 1th value from the *n*th value is performed via

$$\mathbf{x}_{n+1} = \mathbf{x}_n + f(\mathbf{x}_n, t_n)\Delta t \tag{3.83}$$

Euler integration is an *explicit* method, meaning that the state of the system at a later time is calculated entirely from the current state. It is also first-order, in that only the first derivative term in the Taylor expansion of  $\mathbf{x}$  is used. For the purposes of a dynamical system, it is generally considered to be a poor choice, both due to its low-order and the fact that it is notoriously unstable<sup>12</sup>.

It is possible however to construct an *implicit* solver for Euler integration known as the 'backward' Euler method. In this case the state of the system at a later time appears more than once in the calculation, requiring that an equation be solved (either analytically, or numerically via a root finding algorithm if no analytical solution exists) to obtain  $\mathbf{x}_{n+1}$ , namely

$$\mathbf{x}_{n+1} + f(\mathbf{x}_{n+1}, t_{n+1})\Delta t = \mathbf{x}_n \tag{3.84}$$

A typical choice for a root finding algorithm is that of Newton-Raphson. Whilst the backward Euler method clearly requires extra effort per timestep (namely, finding a solution to the above equation) it nonetheless displays far better stability that its explicit counterpart, allowing a larger timestep to be used without the error becoming unbounded.

## 3.4.2 Leapfrog integrators

A popular integration scheme for astrophysical simulations is that of the leapfrog (LF) method. This is a second-order algorithm that treats the positions and their derivatives (i.e., velocities) as offset from each other by half a timestep. The system is evolved as per

 $<sup>^{12}</sup>$ the definition of numerical instability is that small approximation errors are *magnified*, rather than damped, as the algorithm attempts to converge to a solution.

the equations

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \mathbf{v}_{n+1/2} \Delta t \tag{3.85}$$

$$\mathbf{v}_{n+1/2} = \mathbf{v}_{n-1/2} + \mathbf{a}_n \Delta t \tag{3.86}$$

where  $\mathbf{v} \equiv \dot{\mathbf{x}}$ , and  $\mathbf{a} \equiv \ddot{\mathbf{x}}$ . Note that the acceleration is calculated from the potential,  $\mathbf{a}_n = -\nabla \phi_n$ . Equation 3.85 is referred to as a 'drift' step, while equation 3.86 is known as a 'kick' step. The leapfrog gets its name from the way in which the velocities, evaluated at half timesteps, are always 'leaping' over the positions, which are evaluated at whole timesteps. A graphical representation is shown in Figure 3.5. Should the velocity at an integer timestep be needed, it can be calculated via

$$\mathbf{v}_n = \frac{\mathbf{v}_{n+1/2} + \mathbf{v}_{n-1/2}}{2} \tag{3.87}$$

In practice the algorithm is constructed in such a way as to evaluate all quantities at the same point in time. This is useful in preserving the Hamiltonian structure of the system, namely, obtaining the time-evolution operators for  $H_{\rm kin}$  and  $H_{\rm pot}$  at the same timestep, where

$$H = H_{\rm kin} + H_{\rm pot} \tag{3.88}$$

is the Hamiltonian. There are two main ways to do this, depending on the order in which the steps are performed:

#### 3.4.2.1 Drift-Kick-Drift (DKD)

The DKD leapfrog integrator starts by evaluating the *position* at half a timestep forward from the initial position. This is calculated from the initial velocity, so the particle has been 'drifted' but not yet 'kicked'. The kick step then evolves the velocity forward a whole timestep, from the acceleration calculated at the intermediate half timestep. It should be noted that since the acceleration depends only on the position (via  $\nabla \phi$ ), it must be evaluated at the same point in time. Finally the particle is again drifted for the remaining half timestep, using the new value of the velocity. The sequence is therefore:

$$\mathbf{x}_{n+1/2} = \mathbf{x}_n + \mathbf{v}_n \frac{\Delta t}{2},$$
$$\mathbf{v}_{n+1} = \mathbf{v}_n + \mathbf{a}_{n+1/2} \Delta t,$$
$$\mathbf{x}_{n+1} = \mathbf{x}_{n+1/2} + \mathbf{v}_{n+1} \frac{\Delta t}{2}$$

#### 3.4.2.2 Kick-Drift-Kick (KDK)

Put simply, the KDK integrator is the reverse of the one above. The velocity is evaluated at half timesteps (so in this sense it is similar to Figure 3.5), first using the initial acceleration (the first 'kick') and then later using the acceleration at the subsequent integer timestep (the last 'kick'). The drift step in the middle is essentially equation 3.85 but advanced for only half a timestep. The steps are:

$$\mathbf{v}_{n+1/2} = \mathbf{v}_n + \mathbf{a}_n \frac{\Delta t}{2},$$
$$\mathbf{x}_{n+1} = \mathbf{x}_n + \mathbf{v}_{n+1/2} \frac{\Delta t}{2},$$
$$\mathbf{v}_{n+1} = \mathbf{v}_{n+1/2} + \mathbf{a}_{n+1} \frac{\Delta t}{2}$$

Note that for a fixed timestep across the spatial domain, the DKD and KDK integrators are equivalent. Each contains a sequence of Hamiltonian operations, with the drift steps corresponding to  $H_{\rm kin}$  and the kick steps corresponding to  $H_{\rm pot}$ .

It has already been mentioned that the Leapfrog algorithm is second-order and therefore a more accurate choice than the Euler method. It is also manifestly more stable, retaining full stability for oscillatory behaviour (such as a closed orbit) provided that  $\Delta t \leq 1/\omega$ , where  $\omega$  is the angular frequency of the oscillation. It should be noted too that the Leapfrog displays excellent energy conserving properties, a consequence of each step being a canonical transformation<sup>13</sup>. This type of integrator is known as *symplectic*. It is this property which makes the Leapfrog the standard choice for dynamical models, often over higher-order methods which, whilst technically more accurate, suffer from energy drift over the course of the simulation.

## 3.4.3 Timestepping

So far our discussion of integration methods has assumed a timestep that can vary in time but is spatially fixed across the entire computational domain. However, this can severely limit either the accuracy or the speed of an algorithm. In general, the timestepping

$$\dot{\mathbf{p}} = -\frac{\partial H}{\partial \mathbf{q}}$$
$$\dot{\mathbf{q}} = \frac{\partial H}{\partial \mathbf{p}}$$

where H is the Hamiltonian and  $(\mathbf{q}, \mathbf{p})$  are the generalised coordinates of position and momentum respectively.

 $<sup>^{13}</sup>$ A canonical transformation in Hamiltonian dynamics is a change of coordinates that preserves the form of Hamilton's equations, which are given by



Figure 3.6: Individual timestep scheme, assuming the force evaluation (designated by black squares) occurs on an integer timestep i.e., as per KDK. Note that drift operations are always performed on timesteps of  $\Delta t_{\min}$ .

criterion for a particle takes a Galilean invariant form, namely

$$\Delta t \propto \left(\frac{\epsilon}{|\mathbf{a}|}\right)^{1/2} \tag{3.89}$$

where  $\epsilon$  is the softening length. Note that for SPH timestepping this would be replaced by h, the smoothing length. We need to limit this value of  $\Delta t$ , however, for reasons of numerical stability. This concept has been mentioned in the previous section; here we refer to it again in terms of an upper bound on the allowed timestep. For a simulation with a resolvable spatial distance of  $\Delta x$ , we require for stability that

$$\Delta t < \frac{\Delta x}{c_s} \tag{3.90}$$

where  $c_s$  is the sound speed, but more generically should be thought of as the speed of information propagation in the medium of interest. Note that for a dry, i.e., N-body simulation this would refer to the velocity v in the wave equation,  $\partial^2 u/\partial t^2 = v^2 \nabla^2 u$ , with  $u \equiv u(x_1, x_2, x_3, ..., x_n, t)$ . Intuitively this corresponds to the requirement that the numerical speed of information propagation does not exceed the physical one. Equation 3.90 is known as the Courant-Friedrichs-Lewy condition (or Courant condition for short), and is a necessary but not *sufficient* criterion for numerical stability in an explicit integration method.

#### 3.4.3.1 Adaptive timesteps

From equation 3.89 we can see that if the model possesses a finite dynamic range then the desired timesteps will vary spatially as well as temporally. We are then left with a choice of allowing the system to evolve either on (a) the longest timestep, (b) the shortest timestep, or (c) adaptive timesteps. The first option will compromise accuracy, while the second will be a waste of computational resources. Naturally for a simulation in which all the physical timescales are roughly the same, the first two options are largely identical and do not constitute a major deficiency in the method. However, if a system contains a large range of dynamical times we must choose option (c).

Assigning individual timesteps to particles can speed up a simulation significantly whilst still retaining reasonable integration accuracy. However, in the case of a symplectic algorithm such as the Leapfrog, there is a price. An individual timestep scheme requires that the force evaluation (the 'kick' step) be performed at different times for different particles. Unfortunately, the corresponding part of the Hamiltonian,  $H_{\text{pot}}$ , is not separable, since the potential of each particle depends on the position of every other particle. Using individual timesteps therefore introduces a perturbation to the Hamiltonian nature of the system, compromising its stability and conservation properties.

From a practical perspective, an adaptive timestep scheme is implemented in a block format, grouping particles together in timestep bins rather than using a complete spectrum. This goes some way to retaining an approximate symplectic form. Typically the blocks are defined over n levels in powers of 2, such that

$$\Delta t_n = 2^n \Delta t_{\min} \tag{3.91}$$

where  $\Delta t_{\min}$  is the smallest required timestep. The drift operations are performed on this minimum timestep for every particle, while the kicks occur less frequently for those particles in a higher block. Figure 3.6 displays a typical block timestepping procedure.

An additional side-effect of individual timesteps is that the DKD and KDK algorithms are no longer equivalent. While they still require the same amount of computational effort, the error in the energy calculation grows four times as fast for the DKD than it does for the KDK case. This can be understood by a simple argument: while the kick operation depends explicitly on the force calculation (which is now time asymmetric) it is the *drift* step where the error will be manifest, as the positions are advanced based on the kicked velocities. The DKD scheme therefore has twice the error of the KDK variant, and since  $H_{\rm kin} \propto v^2$ , the energy drift is increased by a factor of four.

# 3.5 GADGET

In this final section we detail the main features of the code used for the simulations presented in this thesis. GADGET (**GA**laxies with **D**ark matter and **G**as int**E**rac**T**) was developed as a cosmological N-body/SPH code designed to cater for both self-gravitating collisionless fluids as well as for collisional gas. It was created at the Max-Planck-Institute for Astrophysics by Volker Springel, and the first public version of the code was released in 2000 (Springel et al., 2001). The code is designed to run on massively parallel computers with distributed memory, but can also be run in serial on a single processor. In 2005 a second public version, GADGET-2, was released, which marked a nearly complete rewrite of the first version to include updated core algorithms, better communication and integration methods, and much additional physics (Springel, 2005). The outline in this section refers mostly to generic GADGET methods although some details pertain soley to GADGET-2.

## 3.5.1 Collisionless dynamics

Dark matter and stars are both modelled in GADGET as collisionless fluids, the equations of which i.e. the collisionless Boltzmann equation coupled to the Poisson equation, are solved with the N-body method. The approach is therefore one of following the dynamics of the particles under their own self-gravity, and since the code is designed largely for cosmological simulations, the Hamiltonian describing the dynamics includes a scale factor term, a(t):

$$H = \sum_{i} \frac{\mathbf{p}_{i}^{2}}{2m_{i}a(t)^{2}} + \frac{1}{2}\sum_{ij} \frac{m_{i}m_{j}\phi(\mathbf{x}_{i} - \mathbf{x}_{j})}{a(t)}$$
(3.92)

This scale factor can be set to unity in order to simplify to Newtonian space; this is usually the case for non-cosmological simulations e.g. galactic scales or less.

We recall that an important aspect of the N-body approach in this case is the use of softening in the gravitational potential at small separations between particles. The choice of softening length is important to reduce the large-angle scattering in two-body collisions whilst avoiding too much of a bias towards the softened forces. GADGET computes the single particle density distribution function by convolving the Dirac  $\delta$ -function with a normalised softening kernel, W(r); this spline-based kernel is identical in form to the smoothing equivalent employed in the SPH part of the code, and is described by

$$W(r,\epsilon) = \frac{8}{\pi\epsilon^3} \begin{cases} 1 - 6(\frac{r}{\epsilon})^2 + 6(\frac{r}{\epsilon})^3, & 0 \le \frac{r}{\epsilon} \le \frac{1}{2}, \\ 2(1 - \frac{r}{\epsilon})^3, & \frac{1}{2} < \frac{r}{\epsilon} \le 1, \\ 0, & \frac{r}{\epsilon} > 1 \end{cases}$$
(3.93)

where  $\epsilon$  is the gravitational softening length<sup>14</sup>.

<sup>&</sup>lt;sup>14</sup>note that there is a difference in the notation here with respect to the standard cubic spline kernel, in that GADGET defines the softening (smoothing) to drop to zero at a distance of  $\epsilon$  (h), not  $2\epsilon$  (2h) as equation 3.17 and the majority of the literature suggests.

# 3.5.2 Hydrodynamics

GADGET models the dynamics of gas with SPH, employing a 'gather' approach with an equivalent smoothing kernel W(r, h) to that of equation 3.93. The equation of state is set to that of an ideal gas (equation 3.57) where  $\gamma$ , the adiabatic index, is typically taken as  $\gamma = 5/3$ , its monatomic value. The thermodynamics of each fluid element is defined in terms of an entropic function A(s), where s is the entropy per unit mass (cf. §3.2.4). The pressure of an SPH particle in terms of this entropic function is therefore given by

$$P_i = A_i \rho_i^{\gamma} \tag{3.94}$$

This treatment allows the fully conservative formulation of SPH (again see Section 3.2.4) to be used. The smoothing lengths in GADGET are therefore fully adaptive, selected such that the kernel volume contains a fixed mass for the estimated density (equation 3.60), and the equation of motion for each SPH particle is given by equation 3.68.

Shocks in GADGET are resolved via an artificial viscosity, with a viscous acceleration term given by:

$$\left. \frac{\mathrm{d}v_i}{\mathrm{d}t} \right|_{visc} = -\sum_{j=1}^N m_j \Pi_{ij} \nabla_i \overline{W}_{ij} \tag{3.95}$$

where  $\overline{W}_{ij}$  is the arithmetic average of the kernels  $W_{ij}(h_i)$  and  $W_{ij}(h_j)$ . The artificial viscosity is parameterised by  $\Pi_{ij}$ , as per Section 3.2.5. There are two types of viscosity available for the code, the first being the Monaghan (1992) form (equation 3.73). The second is an upgraded method for GADGET-2, and follows a signal velocity  $v_{ij}^{\text{sig}}$  approach (Monaghan, 1997), where

$$v_{ij}^{\rm sig} = c_i + c_j - 3w_{ij} \tag{3.96}$$

and  $w_{ij}$  is the projection of the relative velocity for approaching pairs, namely

$$w_{ij} = \begin{cases} \mathbf{v}_{ij} \cdot \mathbf{r}_{ij} / |\mathbf{r}_{ij}|, & \mathbf{v}_{ij} \cdot \mathbf{r}_{ij} < 0, \\ 0, & \text{otherwise} \end{cases}$$
(3.97)

The viscosity is then given by

$$\Pi_{ij} = -\frac{\alpha}{2} \frac{v_{ij}^{\text{sig}}}{\rho_{ij}} \tag{3.98}$$

The difference between the two methods lies in the  $h_{ij}/r_{ij}$  factor, which in the Monaghan (1992) form can make the viscous force diverge for small particle pair separations. For the signal velocity approach therefore, simulations have found that the occurrence of large viscous accelerations is reduced and the time integration is more stable. However, the latter method has also been found to give unpredictable angular momentum transport



Figure 3.7: Two steps in the subdivision of an oct-tree

through gaseous discs (Alexander, R., private communication) and so the choice of the viscosity prescription depends on the problem. In Chapter 4 we use the signal velocity approach since we have an extremely high Mach number initial shock, whereas in Chapter 5 we are interested in SMBH growth due to angular momentum transport and so we employ the Monaghan (1992) form.

## 3.5.3 Self-gravity

GADGET uses a tree algorithm to evaluate gravitational forces between particles (see Section 3.3), in the form of a geometric oct-tree for 3D simulations (Figure 3.7). The algorithm uses a hierarchical multipole expansion for the spatial domain, partitioning it into a series of cubes, or nodes, as per Barnes & Hut (Barnes and Hut, 1986). The root node encompasses the full mass distribution, and is repeatedly subdivided into eight 'daughter' nodes down to 'leaf' nodes that contain single particles. The forces are then calculated in the usual way by 'walking' the tree.

For the upgraded GADGET-2, the tree approximates the partial force from unopened nodes using monopole moments, for ease of construction and low cost in terms of memory requirements. For processors with sufficient cache volume this can lead to a substantial speed-up of the algorithm, since access to the main (slow) memory block can be largely avoided. The node opening criterion in GADGET-2 is set as per

$$\frac{GM}{r^2} \left(\frac{l}{r}\right) \le \alpha |\mathbf{a}| \tag{3.99}$$

where M is the mass of the node, r is the distance of the node from the particle in question, and l is the node extension (i.e., length).  $\alpha$  is a tolerance parameter on the total acceleration **a** obtained from the previous timestep. The tolerance parameter controls the size of the force error at the expense of computational resources.

A optional upgrade to GADGET-2 from the original version is that of a hybrid tree/particle-mesh method known as TreePM. This treats the long-range and short-range

contributions to the gravitational potential separately, splitting them in Fourier space via  $\phi_k = \phi_k^{\text{long}} + \phi_k^{\text{short}}$ . The long-range part is solved for using a mesh-based FFT, while the short-range force is computed in real space via a tree. This method can deliver high accuracy in the long-range force while retaining reasonable efficiency.

# 3.5.4 Time evolution

Simulations with GADGET are evolved using a Leapfrog algorithm with adaptive timesteps that employs a kick-drift-kick strategy. For collisionless particles the timestep is set via the criterion

$$\Delta t^{\rm grav} = \min\left[\Delta t_{\rm max}, \left(\frac{2\eta\epsilon}{|\mathbf{a}|}\right)^{1/2}\right]$$
(3.100)

where  $\eta$  is an accuracy parameter. The maximum allowed timestep,  $\Delta t_{\text{max}}$ , is simply set to a small fraction of the dynamical time of the particular system being modelled.

For gas particles, an SPH scheme can only be formulated in terms of a Hamiltonian provided that there are no sources of entropy. In this case fully-reversible hydrodynamics can be achieved in a timestepping scheme by including a thermal energy term in the Hamiltonian,

$$H = H_{\rm kin} + H_{\rm pot} + H_{\rm therm} \tag{3.101}$$

with

$$H_{\rm therm} = \frac{1}{\gamma - 1} \sum_{i} m_i A_i \rho_i^{\gamma - 1}$$
(3.102)

and treating the kinematics of the SPH particles in a similar way to that of the collisionless particles. However, most hydrodynamical simulations will develop shocks, and so the SPH method is in practice always irreversible. The long-term integration properties of Hamiltonian systems are therefore not as relevant here.

For GADGET-2 the hydrodynamical timestep (assuming the signal velocity approach to the viscosity) takes a Courant-like form of

$$\Delta t_i^{\text{hyd}} = \frac{\tau_{\text{courant}} h_i}{\max_j (c_i + c_j - 3w_{ij})} \tag{3.103}$$

where the maximum in the denominator is determined over the neighbours j of particle i.  $\tau_{\text{courant}}$  is the Courant factor, which takes the range (0, 1]. A value of  $\tau_{\text{courant}} \approx 0.3$  is recommended by Monaghan (1989) for good results.

# 4

The formation of stellar discs in the Galactic centre via cloud-cloud collisions

"Life is pretty simple: You do some stuff. Most fails. Some works. You do more of what works. If it works big, others quickly copy it. Then you do something else. The trick is the doing something else."

Leonardo da Vinci

# 4.1 Introduction

Over 10 years' worth of data on stellar proper motions in the Galactic centre have shown the presence of a population of young, massive stars (with age  $\lesssim 10$  Myr,  $M \sim 30 120 \,\mathrm{M_{\odot}})$  that are close enough to the supermassive black hole (SMBH) to be within its gravitational sphere of influence (Paumard et al., 2006). Formation scenarios initially considered one of two possible methods; either the stars were formed in situ in a massive, self-gravitating accretion disc around the black hole (Levin and Beloborodov, 2003), or are the result of a massive star cluster spiralling in via dynamical friction and losing some of its population to the overpowering gravity of Sgr A<sup>\*</sup> (Gerhard, 2001). There has since been observational evidence in favour of the *in situ* scenario (Paumard et al., 2006; Nayakshin and Sunyaev, 2005), the alternative spiral-in model being all but ruled out on the basis of (i) the apparent lack of young, low-mass stars outside the  $\sim 0.5$  parsec region (Paumard et al., 2006; Bartko, 2010) which would be expected from tidal stripping of the cluster (Levin et al., 2005) (ii) the fact that the cluster would either have to be extremely dense to inspiral through dynamical friction in the required time (Kim and Morris, 2003), or be bound by an intermediate-mass black hole<sup>1</sup> (Kim et al., 2004) and (iii) the presence of a significant warp in the observed stellar disc(s), a feature that would be very difficult to achieve with an infalling cluster since the stars would not have had long enough to warp to such a degree by differential precession (Bartko et al., 2009).

Attention has therefore turned to the specifics of both the formation and structure of the stars in the Galactic centre. A number of observations (see, e.g., Paumard et al., 2006; Bartko et al., 2008) point to the presence of two distinct counter-rotating OB stellar populations, co-eval 6 million years ago (to within  $\sim 1$  million years), distributed on scales of  $\sim 0.04 - 0.5$  parsecs (Levin and Beloborodov, 2003) with a large angle of orientation with respect to each other, and possessing a top-heavy initial mass function (IMF) (Nayakshin, 2006; Alexander et al., 2006). In addition, there is evidence that the clockwise rotating stars form a well defined, relatively circular disc/ring (Lu et al., 2006), born out of a massive, self-gravitating accretion disc of gas. The required mass of this disc has been constrained to at least  $10^4 \,\mathrm{M_{\odot}}$  (Nayakshin and Cuadra, 2005). What is not yet confirmed is the nature of the second, counter-clockwise population of stars; whether it is another ring similar to the first, or is part of a larger stellar population with more isotropic orbits (Lu et al., 2006). The important features of this second structure are (i) that it is more diffuse, and thicker than the clockwise disc (ii) that it is located further out at  $\sim 0.15$  pc and (iii) that the orbits of the stars are significantly more eccentric, with  $e \sim 0.8$ . The origin of this feature is therefore less clear than the clockwise-rotating stellar population, for the self-gravitating accretion disc scenario implies more circular orbits for stars formed

<sup>&</sup>lt;sup>1</sup>there is currently no definitive evidence for the existence of intermediate-mass black holes (IMBHs; mass range:  $\sim 10^3 - 10^5 M_{\odot}$ ) and so a mechanism invoking them is generally viewed with skepticism



Figure 4.1: The young stars seen in the central parsec of the Galactic centre. Image courtesy of Keck/UCLA Galactic Centre Group.

there (Nayakshin and Cuadra, 2005). Since the two populations are co-eval, however, it seems likely that they were formed by a single progenitor event.

As well as the massive OB stars discussed above, we also see two more distinct collections of stars whose formation is still very much the subject of debate. Firstly the star cluster, IRS13E, which contains both massive young stars and some lighter ones observations of its dynamics seem to place it in the more eccentric, counter-clockwise disc, and it appears to be internally gravitationally bound despite the fact that its total visible mass is not sufficient for this to occur. Possible explanations have included the theory that there may be an intermediate mass black hole (IMBH) at the centre of the cluster (Gürkan and Rasio, 2005). Secondly, within the inner arcsecond i.e. < 0.04 pc (inside of the OB star distribution) spectroscopic observations have been made of isotropically oriented, early main-sequence B-type stars (Ghez et al., 2003a; Eisenhauer et al., 2005) with approximate masses of  $M \sim 15 \,\mathrm{M}_{\odot}$ . Since gravitational instabilities of a massive accretion disc are not expected to form stars any closer in than the 1 arcsecond edge (see Chapter 1, Section 1.7.1.1), the *in situ* formation scenario is not favoured for these (so-called) S-stars.

The explanation for these features (as well as many others) is crucial to constructing a consistent model of stellar populations in the Galactic centre both in terms of formation and evolution. There are many ideas currently being banded around the scientific community but a great deal is still not known to any acceptable precision. One important aspect to the model is the progenitor of the stellar populations. The *in situ* model has, so far, not addressed in detail the origin of the gaseous discs where the stars were born, although similar discs are believed to exist in AGN and quasars. Indeed, sub-parsec, massive gaseous discs are often invoked as means of feeding SMBHs' immense appetite for gaseous fuel (e.g. Frank et al., 2002). Sgr A<sup>\*</sup>, however, is currently not a member of the AGN club, let alone the even more powerful and elitist quasar community. Indeed, Sgr A<sup>\*</sup> bolometric luminosity is around ~  $10^{-9}$  of its Eddington limit (e.g., Narayan, 2002; Baganoff et al., 2003). The amount of ionized gas currently present in the inner parsec is estimated at perhaps a mere hundred solar masses (Paumard et al., 2004), and there are also tight observational constraints (Falcke and Melia, 1997; Cuadra et al., 2003) on the presence of an optically thin disc on smaller scales (say 0.01 pc or less).

We believe that the requisite gaseous discs could not have been assembled by viscous transport of angular momentum as is expected to be the case in other systems with planar geometries (Shakura and Sunyaev, 1973; Frank et al., 2002). The viscous timescale of a thin, marginally self-gravitating disc can be estimated as  $t_{\rm visc} = (M_{\rm bh}/M_{\rm disc})^2 (\alpha \Omega)^{-1}$ , where  $M_{\rm bh} \approx 3.5 \times 10^6 \,\mathrm{M_{\odot}}$  and  $M_{\rm disc}$  are masses of the blackhole and the disc, respectively, and  $\Omega$  is the Keplerian angular frequency (refer to Chapter 2, Section 2.3.4.6). For  $\alpha = 0.1$ and  $M_{\rm disc} \approx 10^4 \,\mathrm{M_{\odot}}$  (reasons for this choice of mass are discussed in Nayakshin et al. (2006)),  $t_{\rm visc} \sim 3 \times 10^7$  years at r = 0.03 pc, the inner edge of the discs (Paumard et al., 2006), and 10<sup>9</sup> years at 0.3 pc. These times are very long compared with the age of the stellar systems. In fact, the gaseous discs would have to evolve even faster as the two stellar systems are co-eval within one million years (Paumard et al., 2006). Moreover, had the original gaseous disc(s) extended to larger radii than the visible stellar population(s) they would have been susceptible to fragmentation through self-gravity and we would expect to observe additional stellar populations further out.

It is therefore not clear how the required mass of gas came to reside within the inner parsec in the first place. The one-off star formation event appears to be best explained by a one-off deposition of gas within the central parsec. There are several ways in which this could have happened. A giant molecular cloud (GMC) with a sub-parsec impact parameter (in relation to Sgr A<sup>\*</sup>) could have self-collided and become partially bound to the central parsec (e.g., Nayakshin and Cuadra, 2005). Alternatively, a GMC could have struck the circumnuclear disc (CND) located a few parsecs away from Sgr A<sup>\*</sup>, and then created gas streams that settled into the central parsec.

It is very hard to estimate the probability of a cloud-cloud collision in the central part of the Galaxy. Hasegawa et al. (1994) estimated the rate of GMC collisions as one per 20 Myrs in the central 100 pc, but the estimate is uncertain by at least an order of magnitude in either direction due to the lack of knowledge about the size distribution of GMCs. However, there is observational evidence that GMCs can be put on orbits significantly different from the simple circular orbits seen in the inner Galaxy. Stolte et al. (2008) recently measured the proper motions of the Arches star cluster. This cluster is  $\sim 2.5$  million years old and has a mass in excess of  $10^4 \,\mathrm{M}_{\odot}$  (e.g., Figer, 2004). It is presently about 30 pc away from Sgr A<sup>\*</sup> in projection. Its orbit must be strongly non-circular as its 3D velocity is over 200 km s<sup>-1</sup> in the region where the circular velocity is only ~ 110 km sec<sup>-1</sup>. Therefore, this suggests that massive GMCs can be on highly non-circular orbits. It does not seem implausible that one of these clouds would pass within a few parsecs of Sgr A<sup>\*</sup>.

In this Chapter we explore such a one-off collision event in a very simple setup (see Section 6.4.1). We allow two massive, uniform and spherical clouds on significantly different orbits to collide with each other one parsec away from Sgr A<sup>\*</sup>. The resulting gas dynamics, in particular the way in which gas settles into the inner parsec, is the focus of our effort here. We find that the collision forms streams of gas with varying angular momentum, both in magnitude and direction. Parts of these streams collide and coalesce to form a disc, and the remaining parts form one or more orbiting filaments. As the gas cools, it becomes self-gravitating and stars are born, usually in both the disc and the filaments. This overall picture is discussed in Section 4.4. A reader mainly interested in a comparison between the simulations and observations may find Sections 4.4.2.3, 4.4.4 and 4.5 most relevant<sup>2</sup>. In Section 4.5.1.1 we argue that our simulation results suggest that the location of the original impact was somewhat further away from Sgr A<sup>\*</sup> than we assumed here, e.g., perhaps a *few* parsecs.

# 4.2 Numerical setup

The numerical approach that we used is after Nayakshin et al. (2007) with only slight modifications. To perform the simulations, we employ GADGET-2, a smoothed particle hydrodynamic (SPH)/N-body code (Springel, 2005). The Newtonian N-body gravitational interactions of particles in the code are calculated via a tree algorithm, while an artificial viscosity is used to resolve shocks in the gas.

# 4.2.1 Cooling prescription

Gas cools according to  $du/dt = -u/t_{cool}(r)$ , where the cooling time depends on radius as

$$t_{\rm cool}(r) = \beta t_{\rm dyn}(r) \tag{4.1}$$

where  $t_{\rm dyn} = 1/\Omega$  and  $\Omega = (GM_{\rm bh}/r^3)^{1/2}$ , the Keplerian angular velocity, where  $\beta$  is a dimensionless number (refer to Chapter 2, Section 2.3.4.8). This approach is motivated by simulations of marginally stable self-gravitating gaseous discs, where  $\beta$  is expected to

<sup>&</sup>lt;sup>2</sup> for movies of the simulations the reader is directed to http://www.astro.le.ac.uk/~aph11/movies.html

be of the order of a few (Gammie, 2001; Rice et al., 2005). In the problem at hand,  $\beta$  would most likely be much smaller during the collision phase of the clouds as gas can heat up to high temperatures (see Section 4.5), but it could be much larger than unity when gas cools into a geometrically thin disc or a thin filament. In the framework of our simplified approach, and given numerical limitations, we only consider two values for  $\beta$ . The  $\beta = 1$  tests approximate the situation when cooling is inefficient, whereas  $\beta = 0.3$  cases correspond to 'rapid' cooling.

# 4.2.2 Gravitational potential

Gas moves in the gravitational potential of Sgr A<sup>\*</sup>, modelled as a motionless point mass with  $M_{\rm bh} = 3.5 \times 10^6 \,\mathrm{M_{\odot}}$  (Reid et al., 1999) at the origin of the coordinate system, and a much older relaxed isotropic stellar cusp. For the latter component we use the stellar density profile derived from near-IR adaptive optics imaging by Genzel et al. (2003):

$$\frac{\rho_*(r)}{\rho_{*0}} = \begin{cases} \left(\frac{r}{r_{\rm cusp}}\right)^{-1.4}, & r < r_{\rm cusp} \\ \left(\frac{r}{r_{\rm cusp}}\right)^{-2}, & r \ge r_{\rm cusp} \end{cases}$$
(4.2)

where  $r_{\rm cusp} = 10$  arcseconds ( $\approx 0.4 \text{ pc}$ ), and  $\rho_{*0} = 1.2 \times 10^6 \,\mathrm{M_{\odot} pc^{-3}}$ . Note that the enclosed stellar cusp mass is approximately  $6 \times 10^5 \,\mathrm{M_{\odot}}$  at  $r = r_{\rm cusp}$  and equals Sgr A\* mass at  $R \sim 1.6$  pc. We do not include the gravitational potential of the putative stellar mass black hole cluster around Sgr A\* e.g., (Morris, 1993; Miralda-Escudé and Gould, 2000), as at present its existence and exact properties are highly uncertain (Freitag et al., 2006; Schödel et al., 2007; Deegan and Nayakshin, 2007).

We derive the full expression for the gravitational potential in Appendix B. We note that the SMBH radius of influence, as defined by  $F_{\rm bh} \geq F_{\rm stellar}$ , where  $F_{\rm stellar}$  is the force exerted by the stellar cusp mass enclosed, is naturally where  $M_{\rm bh} \geq M_{\rm stellar}$ , at 1.6 pc. We can therefore see that Sgr A<sup>\*</sup> dominates the gravitational potential for the entirety of our computational domain ( $\leq 1$  pc).

#### 4.2.3 Star formation and SMBH accretion using sink particles

As gaseous haloes collapse under their own gravity, it becomes increasingly difficult to follow the time evolution of the simulation due to the high densities that are achieved. The power of adaptive smoothing lengths is a weakness in this case, with h becoming infintesimally small and the hydrodynamical timesteps following suit. The standard approach to circumvent this problem is to introduce N-body particles, known as 'sink particles', wherever the gas reaches an arbitrarily-specified density threshold.

Star formation and accretion in SPH is traditionally modelled in this way, via the sink particle formalism (Bate et al., 1995). Gravitationally contracted gas haloes are replaced

by collisionless sink particles of same mass, which can accrete gas particles either stochastically from the neighbours of the sink, or via an accretion radius  $r_{acc}$ . The gas particles accreted are then removed from the simulation, and in order to be conservative, their mass and momentum is added to the sink.

We follow this recipe both for star formation in our simulations as well as for the central SMBH. The latter is modelled as a sink particle at r = (0, 0, 0), and accretes gas particles that fall within an accretion radius of  $r_{\rm acc} \sim 0.1$  arcseconds (exact values vary between different cooling parameters, see Section 4.3). This is significantly smaller than the inner radius of the observed 'disc' population of young massive stars, i.e.,  $\sim 1''$  (Paumard et al., 2006). The mass of the gas particles accreted is added to Sgr A<sup>\*</sup> but their momentum is not, the black hole being held fixed at the origin to eliminate any unwanted numerical effects that might be caused by its motion<sup>3</sup>. We note that, although this breaks overall momentum conservation, since the SMBH mass is many orders of magnitude larger than the gas (refer to Section 4.3), this should not compromise the reliability of the simulation.

To model star formation, we introduce new sink particles when the gas density exceeds a threshold value. The choice of this threshold value is dependent on the factors affecting the collapse of a gaseous halo. Firstly, we require that the region of interest has reached the Jeans mass, at which point gravity dominates over pressure forces and collapse can start to occur (see Chapter 2, Section 2.3.4.8). This consideration becomes more complex when dealing with numerical simulations, however, as the concept of a minimum resolvable mass enters into play. In an N-body sense, this is simply that of a single SPH particle; however when dealing with pressure forces and self-gravity the situation is somewhat different. Bate and Burkert (1997) discuss the concept of mass resolution in relation to marginally self-gravitating gas clumps, and point out that when the Jeans length becomes comparable to the smoothing length, the code has reached its limit of mass resolution, as the relevant forces governing the collapse of the clump are significantly altered by numerical details. The minimum mass resolution must therefore be somewhat higher than the number of particles contained within a Jeans length. Navarro and White (1993) found that, when modelling the collapse of a gas cloud, reasonable results were obtained when the cloud was comprised of  $\sim 100$  SPH particles. Since a standard value for the number of neighbours is  $N_{\text{neigh}} \approx 50$ , as indeed it is in our code, Bate and Burkert (1997) suggested a rule of thumb of  $2N_{\text{neigh}}$  within a Jeans length for reliable collapse resolution. With this approach our minimum mass resolution is  $1.6 \,\mathrm{M}_{\odot}$ , while an individual SPH particle in our simulations has a mass of  $0.02 \,\mathrm{M}_{\odot}$ .

 $<sup>^{3}</sup>$ as a result the overall gravitational potential is 'nailed down' since the stellar cusp is at all times centered on the SMBH. We note that while there is some physical justification for a fixed SMBH, in that since we do not include a 'live' halo the process of dynamical friction that would act to suppress the SMBHs motion away from the bottom of the potential well is not modelled, the main motivation for keeping the black hole fixed is to avoid any numerical instabilities that might arise as a result of the sink particle oscillating about the origin (for a different but related numerical issue see Section 4.4.3.1)

ŀ

The Jeans mass criterion can be turned into a constraint on the density, by using the minimum resolvable mass. This marks the point at which collapse can be trusted to be *physical*, rather than dominated by numerical effects. To resolve collapse we therefore require that

$$p_{\rm J} \lesssim \left(\frac{3}{4\pi}\right) \left(\frac{5N_{\rm A}k_{\rm B}T}{2G\mu}\right)^3 \left(\frac{N_{\rm tot}}{2N_{\rm neigh}}\frac{1}{M_{\rm tot}}\right)^2$$

$$(4.3)$$

where  $N_{\rm A}$  is Avagadro's number,  $k_{\rm B}$  is the Boltzmann constant,  $\mu$  is the mean molecular weight,  $N_{\rm tot}$  is the total number of particles in the simulation and  $M_{\rm tot}$  is the total mass. Above this limit, we can no longer follow Jeans collapse as an entirely physical process, and we are (relatively) justified in replacing the gas particle with a sink.

However, this is not the whole picture. In the case of the GC, the environment is quite different to that of a 'standard' galactic molecular cloud, and the dominant requirement for star formation is overcoming the *tidal shear* density in the potential of Sgr A<sup>\*</sup>. We derive the expression for this in Appendix C, finding that

$$\rho_{\rm BH} = \frac{3M_{\rm BH}}{\pi r^3} \tag{4.4}$$

where r is the distance from the central black hole. Below this density the gas would be sheared out due to the potential of the SMBH and therefore be unable to form stars. To a certain extent this shearing occurs self-consistently within the simulation; indeed, the behaviour of the two clouds after the collision is evidence of this. However, since we do not resolve the gas down to its atomic composition, but rather at a resolution lower limit of  $\sim 1.6 \,\mathrm{M}_{\odot}$ , we could easily imagine a scenario where a single SPH particle (which cannot itself be sheared) reaches the Jeans density for collapse but is not above the tidal density. In this case turning the particle into a sink would be extremely artificial, as in reality the gas within it would be subject to strong tidal forces that would prevent the star formation process.

We therefore require a combination of the Jeans constraint and the tidal constraint in order to model sink particle creation in our simulations. We chose a relatively simple approach to this, based on that used in Nayakshin et al. (2007), which we acknowledge does not capture the full physics of the problem but nonetheless should ensure that the global properties of the stars formed are reasonably reliable. We introduce sink particles when the gas density exceeds

$$\rho_{crit} = \rho_0 + A_{\rm col}\rho_{\rm BH} \tag{4.5}$$

where  $\rho_0 = 1.67 \times 10^{-12} \text{ g cm}^{-3}$ , and  $A_{\text{col}} = 50$  (these values are based on tests performed in Nayakshin et al. (2007), which found that the results are not sensitive to the exact values of  $\rho_0$  and  $A_{\text{col}}$ , provided they are sufficiently large). The  $\rho_0$  quantity is essentially our Jeans constraint, albeit one that is fixed (note from equation 4.3 that  $\rho_J$  is a function of temperature). It corresponds to the  $\rho_{\rm J}$  limit for gas at  $T \simeq 3600$  K. In general the temperature over the majority of our computational domain is somewhat higher,  $\sim 10^4$  K. On its own, then, in these hotter regions this constraint would lead to the formation of sink particles before collapse reached the point where numerical effects began to play a role. However, we have in addition the tidal constraint, which dominates over  $\rho_0$  within  $\sim$  a few arcseconds, i.e., where the majority of the circular disc stars form (refer to Section 4.4.2). The factor of 50 is to ensure that once the tidal limit is reached, collapse has a chance to get underway before a sink is introduced.

The full details of our sink particle approach are necessarily somewhat different to that of the standard prescription in molecular clouds. Numerical simulations of the 'normal' process of star formation have shown (Larson, 1969; Masunaga et al., 1998) that as gas haloes on the scale of typical molecular clouds collapse, the first quasi-hydrostatic object to form has a characteristic size of ~ 5 A.U. =  $7.5 \times 10^{13}$  cm. This is known as a *first core*. Through self-gravity and optically thick (blackbody) cooling, the core decreases in size while also accreting mass from the surrounding envelope that is continuing to infall. When the mass of the first core reaches ~  $0.1 M_{\odot}$ , H<sub>2</sub> dissociation occurs, along with the ionisation of H, due to compressional heating. At this point the core is able to cool much more effectively through optically thin processes (refer to Appendix D) and collapses to a higher density and smaller size (Silk, 1977). A diagram of this process is shown in Figure 4.2.

In this picture, the timescale for the formation of a first core and its collapse to stellar densities is short, ~ 10<sup>3</sup> yrs, proceeding on approximately the cooling time. This is far shorter than the free-fall times,  $t_{\rm ff} \sim 1/(G\rho)^{1/2}$ , of typical molecular clouds in a galaxy. The creation of these first collapsed objects is therefore effectively *instantaneous* in a normal star-forming environment. In addition, the sizes of the first cores are far smaller than the typical cloud size of ~ 10<sup>18</sup> cm. As a result, simulations of star formation in molecular clouds needs not consider the lifetime nor the physical size of a first core.

For star formation around Sgr A<sup>\*</sup>, however, the picture is different. Since the gas is far hotter than a standard GMC, the Jeans density limit is significantly higher. We can therefore assume that the initial stages of collapse are similar to normal star formation, but starting from higher gas densities. Since the local free-fall time is  $\propto \rho^{-1/2}$ , we find that at the distance where the young, OB stars surrounding Sgr A<sup>\*</sup> are located ( $\leq 0.5$  pc), this timescale is short ( $\approx 10^2 - 10^3$  yrs for  $\rho/m_p \approx 10^8 - 10^{12}$  cm<sup>-3</sup>, the typical density range for a self-gravitating AGN disc). In fact, even at the minimum (tidal) star-forming density in the potential of Sgr A<sup>\*</sup>, the local  $t_{\rm ff}$  for the cloud is  $\sim t_{\rm dyn}$  for the system<sup>4</sup>. It

$$t_{\rm ff} \simeq \frac{1}{(G\rho_{\rm tidal})^{1/2}} = \left(\frac{2\pi r^3}{3GM}\right)^{1/2} \simeq \left(\frac{r^3}{GM}\right)^{1/2} = \Omega^{-1}$$
 (4.6)

<sup>&</sup>lt;sup>4</sup>since in a potential dominated by a point mass, putting  $\rho_{\text{tidal}} = 3M/2\pi r^3$  into  $t_{\text{ff}} \simeq 1/(G\rho)^{1/2}$  yields



Figure 4.2: Gas cloud collapsing to form a first core.

is therefore clear that the free-fall time of the collapsing cloud is comparable to or shorter than the cooling time, and thus the formation time, of the first core.

This means that we must treat first cores, or *protostars*, as having a finite lifetime. Furthermore, rather than being essentially point masses with respect to the larger gas halo, in this case the protostars are only an order of magnitude smaller than the typical scaleheight for the disc,  $H \sim 10^{15}$  cm, and should therefore be modelled as having finite size. This means, in turn, that we must consider the possibility of a first core merging with a fully-fledged sink particle, or indeed another protostar.

In all of our simulations we therefore employ two different types of sink particles for star formation, distinguishing between *first cores* and *stars*. We briefly summarise the implementation of each type below.

• First core (protostar) particles

Once the gas reaches  $\rho_{\rm crit.}$ , an SPH particle is turned into a first core with an initial mass equal to  $m_{\rm SPH}$ . The geometric size of the first core is  $r_{\rm core} = 10^{14}$  cm. Physically, first cores represent massive gaseous clumps that have not yet collapsed to stellar densities, and thus have finite sizes. First cores are assumed to merge with each other if they pass within a distance  $|r_1 - r_2| < 2f_m r_{\rm core}$ , where  $f_m$  is a gravitational focusing parameter set to unity. First cores can also grow by accretion of SPH particles, up to a mass of  $0.1 \, {\rm M}_{\odot}$ .

• Star particles

Once the first core grows (either by mergers or by accretion, or both) beyond a mass of  $0.1 M_{\odot}$ , it is designated a 'star particle'. These have zero geometrical size; in

other words they are treated as point masses. As such they cannot merge, but may accrete first cores if  $|r_1 - r_2| < 2f_m r_{\text{core}}$ , once again. Stars can grow to arbitrarily large masses by accretion of first cores or accretion of SPH particles.

The accretion of gas particles by stars and protostars differs from that of accretion on to the SMBH. Rather than employing an accretion radius approach, it is calculated using the Bondi-Hoyle formalism (refer to Chapter 2, Section 2.3.2), with the accretion rate given by,

$$\dot{M} = 4\pi\rho \frac{(GM)^2}{(\Delta v^2 + c_s^2)^{3/2}}$$
(4.7)

where M is the mass of the sink particle,  $\rho$  is the ambient gas density,  $c_s$  is the sound speed, and  $\Delta v$  is the relative velocity between the gas and the accreting particle. The accretion rate is capped at the Eddington accretion rate:

$$\dot{M}_{\rm Edd} = \frac{4\pi m_p r_* c}{\sigma_T} \,. \tag{4.8}$$

Note that this expression depends only on the size of the object,  $r_*$ . For first cores,  $r_* = r_{\text{core}}$ , which yields a very high accretion rate limit of almost a solar mass per year for  $r_{\text{core}} = 10^{14}$  cm. Since stars in our simulations do not have a physical size, we use the observational results of Demircan and Kahraman (1991); Gorda and Svechnikov (1998) to relate the size of a star to its mass, namely

$$\frac{r}{R_{\odot}} = 1.09 \left(\frac{M}{M_{\odot}}\right)^{0.969} \text{ for } M < 1.52 \,\mathrm{M_{\odot}}, \tag{4.9}$$

$$\frac{r}{R_{\odot}} = 1.29 \left(\frac{M}{M_{\odot}}\right)^{0.6035} \text{ for } M > 1.52 \,\mathrm{M}_{\odot}.$$
(4.10)

For  $r = R_{\odot}$ , this yields an Eddington accretion rate limit of  $\sim 5 \times 10^{-4} \,\mathrm{M_{\odot}}$  year<sup>-1</sup>. The actual gas particles that are then accreted by the sink particle are chosen from its neighbours via the stochastic SPH method of Springel (2005). The mass and momentum of the SPH particles accreted are added to that of the respective sink particle.

We should acknowledge here that this plausible star formation *prescription* cannot possibly reproduce the full complexity of the star formation process in nature. Therefore, the stellar mass function obtained in our simulations cannot be trusted in detail. However, our focus is on the dynamics of gas as it settles into a disc or a filamentary structure. When star formation does occur, the gas temperature is very much smaller than the virial temperature<sup>5</sup> in the simulated region ( $\gtrsim 10^6$  K). Gas orbits thus determine the resulting

$$k_B T_{\text{virial}} \simeq \mu m_p \Phi$$

$$\tag{4.11}$$

<sup>&</sup>lt;sup>5</sup>the virial temperature in a gravitational potential  $\Phi$  is found from

where  $k_B$  is the Boltzmann constant,  $\mu$  is the mean molecular weight, and  $m_p$  is the mass of a proton

stellar orbits which we compare to the observed ones. We believe that this important aspect of our simulations is reliably modelled.

The set of units used in our simulations is  $M_u = 3.5 \times 10^6 \,\mathrm{M_{\odot}}$ , the mass of Sgr A\*(e.g. (Schödel et al., 2002)),  $r_u = 1.2 \times 10^{17} \,\mathrm{cm} \approx 0.04 \,\mathrm{pc}$ , equal to 1" when viewed from the  $\approx 8 \,\mathrm{kpc}$  distance to the GC, and  $t_u = 1/\Omega(r_u)$ , the dynamical time evaluated at  $r_u$ , approximately 60 years.

# 4.3 Initial conditions

Simulations were performed for a set of initial conditions, all of which comprise a collision between two gas clouds at the edge of the inner parsec of the GC. 2,625,000 SPH particles were used in each simulation. The specific parameters for each run, labeled S1 to S6, can be found in Table 4.1. Each cloud is spherically symmetric and of uniform density, containing less than 1% of the mass of Sgr A<sup>\*</sup>. The composition of the clouds is mostly molecular hydrogen, with the mean molecular weight set to  $\mu = 2.46$ . The initial temperature of the clouds is set to 20K. A more complex model could have included a turbulent velocity and density field, but for practical purposes we limit ourselves to a smaller set of input parameters for this study.

We define the primary to be the larger cloud, with a radius of  $R_1 = 0.2$  pc and  $M_1 = 3.4 \times 10^4 \,\mathrm{M}_{\odot}$ , whilst the secondary has a radius of  $R_2 = 0.172$  pc with  $M_2 = 2.6 \times 10^4 \,\mathrm{M}_{\odot}$ . The initial positions of the cloud centres are the same for all the simulations, and are  $r_1 = (25, 0, 0)$  and  $r_2 = (22, 6, 7)$ , in dimensionless units. In all the tests the absolute velocity of cloud 1 is of the order of the Keplerian velocity at that radius. Orbits in a cusped potential cannot be solved for analytically and so a numerical method is required to calculate the eccentricity. We consider the effective potential for a Newtonian orbit (refer to Chapter 2, Section 2.2.3.1), namely,

$$V(R) = \frac{l^2}{2R^2} - \Phi(R)$$
(4.12)

where l is the specific angular momentum, and  $\Phi(R)$  is the potential energy. The latter can be obtained either from Poisson's equation (see Chapter 3, equation 3.78) or by summing the contribution from spherical shells at all radii; we show the latter method in Appendix B. Since the effective potential is equivalent to the orbital energy at a purely *tangential* velocity vector, the resulting expression for equation 4.12 is equated with the specific orbital energy to solve for the pericentre and apocentre positions. The roots of this equality must be computed numerically, for which we employ the Newton-Raphson algorithm.



Figure 4.3: Schematic of the initial conditions for simulation S1, showing the initial orbits of the clouds. The parts that do not collide are sheared out along these orbits, forming streams at approximately the orientation given in this diagram (see Figures 4.4 and 4.5).

Table 4.1: Initial conditions of the simulations presented in the chapter. The meaning of the symbols in the Table are:  $\beta$  is the cooling parameter,  $\mathbf{r}_2$ ,  $\mathbf{v}_2$  are the initial positions and velocity vectors of the secondary cloud, respectively;  $r_{2,pe}$  and  $r_{2,ap}$  are the pericentres and the apocentres of the secondary's orbit; e is the eccentricity;  $\theta$  is the angle between the orbital planes of the clouds, and b is the impact parameter.

ID	$\beta$	$\mathbf{r_2}$	$\mathbf{v}_2$	$\left  v_{1}-v_{2}\right $	$r_{2,pe}$	$r_{2,\mathrm{ap}}$	$e_2$	$\theta$ (°)	b
S1	1	(22,6,7)	(0, -0.11, -0.21)	0.37	13.5	28.2	0.35	116	3.8
S2	1	(22, 6, 7)	(0, -0.21, -0.11)	0.42	13.8	27.9	0.34	151	6.8
S3	0.3	(22, 6, 7)	(0, -0.11, -0.21)	0.37	13.5	28.2	0.35	116	3.8
S4	0.3	(22, 6, 7)	(0, -0.21, -0.11)	0.42	13.8	27.9	0.34	151	6.8
S5	1	(22, 6, 7)	(0.16, -0.11, -0.21)	0.41	21.5	33.4	0.22	120	2.4
S6	1	(22, 6, 7)	(0.16, -0.21, -0.11)	0.45	21.2	33.7	0.23	147	5.3

Orbital eccentricity is then defined via

$$\frac{1+e}{1-e} = \frac{r_{ap}}{r_{pe}} \tag{4.13}$$

where  $r_{\rm ap}$  and  $r_{\rm pe}$  are the pericenter and the apocenter of the orbit. Hyperbolic eccentricities are capped at 1. In the (BH + stellar) potential that we use here (see equation 4.2.2), the orbit of cloud 1 is therefore slightly eccentric, with pericentre and apocentre of 12.9 and 25 respectively, a velocity vector of (0, 0.2, 0), and an eccentricity e = 0.32. The initial orbit trajectory of the second cloud is varied between the tests to cover a small range of possibilities. The parameters for this trajectory in terms of pericentre, apocentre and eccentricity are given in Table 4.1. Note that both clouds are bound to the central parsec. The collision itself is highly supersonic, with a Mach number of ~ 500.

We run tests with cooling parameter  $\beta = 1$  and  $\beta = 0.3$  (see equation 4.1). These values are low enough so that fragmentation would occur if and when regions of the gas became self-gravitating (Gammie, 2001; Rice et al., 2005). Since the faster cooling runs were expected to require on average shorter timesteps the SMBH accretion radius for  $\beta = 0.3$  was set to  $R_{\rm acc} = 0.33$  whilst for  $\beta = 1$  a smaller value of  $R_{\rm acc} = 0.06$  was used.

# 4.4 Results

#### 4.4.1 Gas dynamics

#### 4.4.1.1 Bulk dynamics

In this section we briefly describe the main characteristics of the bulk gas dynamics. In all our simulations, the clouds undergo an off-centre collision at  $t \sim 10$  ( $\approx 600$  years).

• As the cooling time is longer than the collision time,  $t_{coll} \sim (R_1 + R_2)/(|\mathbf{v_1} - \mathbf{v_2}|)$ , the postshocked gas heats up significantly and hence initially expands somewhat. This

thermal expansion modifies velocities of the different parts of the clouds by giving gas thermal velocity 'kicks'. The net result is a distribution of gas velocities that is much broader than the one would get if the two clouds simply got stuck together, i.e., if the collision were inelastic.

- As the total mass of the clouds is only about a percent of Sgr A\* mass, the postcollision cloud is easily sheared by the tidal field of Sgr A\*. The collision and the resulting mixing of the clouds leads to angular momentum cancellation of some parts of the gas. Regions of gas that acquire smaller angular momentum infall to the respective circularisation radius on the local dynamical time. A small scale disc around the black hole is thus formed on this timescale. Regions of the clouds that did not directly participate in the collision are affected less and retain more of their initial angular momentum. These regions are sheared, then cool and result in filaments of length comparable to the initial sizes of the clouds' orbits.
- As time progresses, parts of the filaments collide with each other or with the inner disc if the pericentres of their orbits are small enough. The small scale disc gains mass in an asymmetric non-planar manner. The disc becomes warped and begins to change its orientation with time. Eventually, a sizable fraction of the gaseous mass makes it into the inner few arcseconds. As the mass of the disc increases, it circularises, cools and gradually becomes gravitationally unstable.

Figure 4.4 presents four snapshots of simulation S1 showing gas column density and positions of the stars viewed from the line of sight  $\theta = 0^{\circ}$ . The initial mass deposition into the central few arcseconds forms a small-scale disc that changes its orientation by about  $30^{\circ}$  with time, stabilising by  $t \sim 300$ . As a result of the small impact parameter of the collision, the secondary cloud is largely destroyed and does not collapse into a coherent filament. Instead, the diffuse gas left over from the impact gradually accretes on to the disc that forms around the black hole, growing it in mass and radius. The inner disc therefore ends up extending out to  $\sim 10''$ . The gas left over from the primary cloud collapses into a filament which orbits at a larger radius, separate from the disc.

The simulation S2 is set up identical to S1 except for the initial velocity of the secondary cloud (refer to Table 4.1). There are two important differences: (1) the angle between the orbital plane of the two clouds is greater,  $\theta \sim 150^{\circ}$ , and (2) the collision between the clouds is more grazing as the impact parameter b is greater. As a result of (2), a smaller fraction of the clouds' mass is involved in a bodily collision in S2 than S1, and so the clouds in the former will survive the initial collision better. Indeed, comparing the S1 and S2 simulations in Figure 4.5, one sees that in the latter simulation most of the secondary cloud does indeed survive the collision. This cloud forms a clockwise stream which was largely absent in simulation S1.



Figure 4.4: Gas surface density and stellar locations for snapshots from simulation S1 taken at times t = 0, 100, 250, and 1000, left to right and top to bottom, respectively. Sgr A<sup>\*</sup> is located at (0,0), and the line of sight is along the z direction.

On the other hand, the parts of the clouds involved in the direct collision experience a stronger shock in S2 than they do in S1. Indeed, in S2 the velocity vectors are almost directly opposing. Thus, despite the larger impact parameter, the stronger angular momentum cancellation in S2 creates an inner r < 1 disc as massive as that in S1. However,


Figure 4.5: Projected gas densities and stellar positions for simulations S1 (left) and S2 (right) at t = 1250. Note the survival of the secondary stream in S2.

the 'mid-range' disc, e.g., the disc at  $r \sim 5$ , is far less pronounced in S2 than in S1. The evolution and the structure of the inner gaseous disc is different in S2 from its S1 counterpart. In both simulations the angular momentum of gas infalling on to the discs evolves with time, and hence both discs undergo what we term 'midplane rotation' (refer to Section 4.4.3.2). However, the infalling gas seems to be more intermittent in S2 than in S1, possessing a larger range of angular momentum. While in S1 the accreting gas largely adds to the existing disc, creating one large warped disc, in S2 the infalling gas causes enhanced midplane rotation, and to later times ( $t \sim 400$ ), actually creates a second disc around the first at a different orientation. Figure 4.6 compares the inner discs of S1 and S2 in the process of forming.

To show the effect of the faster cooling, Figure 4.7 contrasts gas surface density plots of simulations S1 and S3 at time t = 150. The bulk dynamics of S3 are the same as in S1, but apart from the innermost few arcseconds, maximum gas densities are higher in S3 than they are in S1. This is natural as faster cooling gas can be compressed to higher densities through gravity and shocks. For the same reason, the filaments are much better defined in S3 than they are in S1. A similar comparison is of course true for S2 and S4.

Simulation S5 is distinct from all the rest due to its small impact parameter, b, between the gas clouds (see 4.1). As a result of this fact and the relatively large value of  $\beta$ , a higher degree of mixing is achieved. Figure 4.8 shows a snapshot from this simulation, showing



Figure 4.6: Projected gas density for the inner two arcseconds, showing stars as well. Left: Simulation S1 at t = 1000; Right: Simulation S2 at t = 400.



Figure 4.7: Comparison of projected gas density for simulation S1 (left panel,  $\beta = 1$ ) and S3 (right panel,  $\beta = 0.3$ ) at t = 150. Note that filamentary structures are better defined in S3 than they are in S1.



Figure 4.8: Left: projected gas density for S5 at t = 2696, showing stars. Note that the stellar populations all possess similar eccentricities and thickness. Right: projected gas density for S6 at t = 664, at a slightly larger perspective than previous images. The overall gas dynamics are very similar to S2. At the time of writing the simulation has not yet progressed far enough for the filaments to collapse into stars. The two projections together are an indication of the range of behaviour that can be obtained from the cloud collision and the resulting flow with very little change to the parameters.

both gas and stars. Compared with all the other runs, the accretion disc extends out to as much as  $r \sim 20$  and represents a very coherent disc with little warping. Outer gas filaments do not form due to complete mixing of the original gas clouds.

Simulation S6 is again different in that it possesses the largest collision velocity of all the simulations, whilst retaining a cooling parameter of  $\beta = 1$ . One therefore expects the resulting thermal expansion of the clouds to significantly modify the velocities of the particles and encourage mixing. However, since the impact parameter is relatively large, this simulation actually has more in common with S2 that with S5, particularly in the inner parts where the gas forms into a double disc structure. The only major difference between these two simulations is that in S6 the secondary cloud has an orbit that is further out than its S2 counterpart, meaning the spatial scale of the secondary filament is shifted outwards. Figure 4.8 shows a snapshot from S6.

#### 4.4.1.2 Redistribution of angular momentum

It is useful to analyse the re-distribution of the gas as a result of the collision in terms of orbital parameters, namely, angular momentum and eccentricity. Gas with specific angular momentum  $l = |\mathbf{r} \times \mathbf{v}|$  circularises at the circularisation radius,  $r_c$ , given by the



Figure 4.9: The distribution of SPH particles over the circularisation radius of their orbits for S1. The black, magenta, blue and red curves corresponds to times t = 0, 25, 50, 500, respectively. As the result of angular momentum cancellation due to shocks, particles "leak" to small radii, establishing there a disc that eventually forms stars.

4.4. Results

where M(r) is the total enclosed mass within radius r. Figure 4.9 shows the distribution of gas particles over the circularisation radius of their orbits as a histogram for several snapshots. At time t = 0, the distribution is highly peaked around the clouds' initial positions. Due to the collision, however, a tail to small  $r_c$  appears in the distribution very quickly. This is not due to viscous angular momentum transport, as in thin accretion discs, but is rather due to cancellation of oppositely directed angular momentum components in shocks. Some of the gas particles aquire particularly small angular momenta and hence fall to small radial separations from Sgr A<sup>\*</sup> on nearly radial orbits. Particles with  $r_c < r_{acc}$  are in general accreted by Sgr A<sup>\*</sup> in our formalism (unless their orbit is changed by interactions before they make it inside  $r_{acc}$ ). Particles with  $r_c > r_{acc}$  settle into a thin rotationally supported disc, or into sheared out filaments (the remains of the clouds). Note that due to a complicated geometry, some SPH particles pass through shocks several times before reaching their final destination (i.e., the disc).

Figure 4.10 shows the profiles of orbital eccentricity defined on radial shells plotted for several snapshots. These profiles again show that soon after the collision the inner part of the computational domain is dominated by gas on plunging (high eccentricity) orbits. At later times finite net angular momentum of gas and shocks force the gas to circularise. It is notable that gas circularises faster at smaller radii, which is naturally expected as the orbital time is shortest there. Therefore, at the end of the simulation, nearly circular gaseous orbits are established in the innermost few arcseconds, whereas highly eccentric ones dominate at larger radii.

The structures that are formed in the simulation can also be identified by the orientation of the angular momentum vector of a particle's orbit, **L**. We define two angles,  $\theta$  and  $\phi$ , so that  $L_z = L \cos \theta$ , where  $L_z$  is the z-component of **L**, and  $\tan \phi = L_y/L_x$ . We then plot the values of these angles for individual SPH particles and stars in the Aitoff projection. As the number of SPH particles is too large to show each particle, we show only a fixed fraction of these, randomly chosen from the total number. This procedure keeps the resulting maps reasonably clear while also preserving the shape of the distribution.

The result is shown in Figure 4.11 for a selection of times. At t = 0 all the gas particles are concentrated in the two clouds, with the finite spread in the distribution arising as a result of all the particles within a cloud having the same velocity, whilst possessing a finite spread in positions about the centre. The difference in the orientations of the two clouds is almost 120° (see Table 4.1). By t = 100, the clouds have collided and the gas particles have experienced 'kicks' in different directions, filling the entire range in  $\theta$ . The  $\phi$  values are restricted to a range about +90° and -90°. Note that the dominant structure formed in these angular coordinates forms a plane  $L_x \approx 0$  in **L**-space. This can



Figure 4.10: Gas eccentricity defined on radial shells as a function of the shell's radius at different times. Black, blue and red curves corresponding to time t = 50, 100 and 250. Note that initially only gas on near-plunging high eccentricity orbits arrives in the innermost region, but with time gas circularises to  $e \ll 1$ .



Figure 4.11: Angular momentum direction in terms of the azimuthal,  $\theta$ , and the polar,  $\phi$ , angles. The frames are shown at t = 0,100,250,1000. Gas particles are colour-coded;  $r \leq 1''$  (red),  $r \leq 5''$  (orange), r > 5'' (blue).

be understood by noting that gas particles experience the collision close to the x-axis, i.e., where  $y/x \approx z/x \approx 0$ . Therefore, the x-component of the resulting angular momentum,  $L_x = yv_z - zv_y$  is very small compared to L.

The plots are colour-coded to show the behaviour of the gas in the inner arcsecond (red), the inner 5 arcseconds (orange), and outside the inner 5 arcseconds (blue). Whilst the distribution just after the collision is fairly spread out for all the three regions, we note a clear difference at later times. In particular, the innermost disc settles into a region which sits in between the two clouds' original positions. It is also the least spread out structure, apparently defining a thin disc which is only slightly warped by t = 1000. The innermost disc orientation does evolve with time, however, since matter infall on the disc continues throughout the simulation. The gas coloured in orange demonstrates a greater extent of warping, and the region outside cannot even be classified into a single structure, as Figure 4.4 shows. Nevertheless, it is clear that the inner and outer gas distributions are similarly oriented with respect to each other, with the majority of the outer gas angular momentum distributed in  $\theta$  and  $\phi$  between the initial values of the clouds.

Differences between S1 and S2 can be noted by looking at the respective histogram plots of circularisation radius (Figure 4.9 and Figure 4.12 respectively). A tail to small  $r_c$ appears faster in S2, by t = 25 rather than t = 50 in S1. The distribution of gas in S2 by t = 500 shows three individual peaks corresponding to the three populations; the inner disc ( $r \le 0.5$ ), the mid-range disc ( $0.5 \le r \le 2$ ), and streams ( $10 \le r \le 20$ ). This is in contrast to the more even distribution of S1.

#### 4.4.2 Star formation

#### 4.4.2.1 Star formation histories

Stars in S1 form in both the disc and the primary filament. The two populations are fairly co-eval, forming in the primary filament, at a radius of  $r \sim 25$ , by  $t \sim 900$  and in the disc by  $t \sim 950$ . Both populations initially consist entirely of low-mass stars ( $\sim 0.1 - 1 \,\mathrm{M}_{\odot}$ ), but high-mass stars ( $M \sim 100 \,\mathrm{M}_{\odot}$ ) soon appear in the disc, by t = 1100. As time goes on, both the inner (disc) and outer (filament) stellar populations increase in number. The outer population mass function remains dominated by low-mass stars whilst the inner population mass function becomes top-heavy. Of course, as noted in Section 4.2, the mass spectrum of stars formed depends on several physical parameters (e.g., see Nayakshin et al. (2007)) that are modelled only to a first approximation here. Furthermore, we have no star formation feedback implemented in our models. We would expect the mass spectrum of our models to change significantly if the proper cooling and feedback physics was included.

Stars form earlier in S2 than in S1, and they form in the disc first, by t = 400. The preferential sites of star formation in the disc appear to be dense concentric ring structures (see Figure 4.6) and we discuss these in more detail in Section 4.4.3. Owing to both the



Figure 4.12: The distribution of SPH particles over circularisation radius for simulation S2. The curves are for times t = 0, 25, 50, 500.



Figure 4.13: Angular momentum orientations for SPH particles (grey) and stars (colour) in simulations S1 (left) and S2 (right) at t = 1250.

primary and secondary clouds forming into filaments in S2, as well as those filaments containing a larger amount of mass than in S1, there is a far greater amount of star formation at large radii in S2 than in S1 at the same time. This can be seen in Figure 4.5. The stars formed in the filaments are all low mass  $(0.1 - 1 M_{\odot})$  and show a tendency to form into clusters; this effect we discuss in more detail in Section 4.4.2.2.

Star formation occurs earlier in S3 than it does in S1. In particular, first stars appear at t = 150 in S3 and only at  $t \approx 1000$  in S1. This is expected since faster cooling facilitates faster gravitational collapse. The S2 to S4 comparison is less extreme but still significant, with stars forming in S2 by t = 400 and in S4 by t = 100. In both  $\beta = 0.3$  runs the initial sites of star formation were the orbiting gas streams, with disc populations forming slightly later, t = 350 (S3) and t = 180 (S4).

The clearly defined concentric ring structures of the S1 and S2 discs were also seen in S3 and S4, although they did not provide the sites of star formation in this case, as fragmentation into sink particles occurred a long time before the rings have formed. In terms of the stellar mass spectra all the stars formed in S3 and S4 are low mass  $(\sim 0.1 - 1 M_{\odot})$ , both in the stream populations and in the disc. Again, this is naturally a consequence of a fast cooling time (Nayakshin, 2006; Shlosman et al., 1989a). Comparing S3 and S4 themselves yields some interesting results. Figure 4.14 contrasts S3 and S4 in terms of projected density and Figure 4.15 in terms of angular momentum of the stellar orbits. Once again, a larger impact parameter in S4 compared to that in S3 (see Table 4.1), allows part of the second (smaller) cloud to survive as a filament. There are therefore two filaments present rather than one. Star formation again occurs in both filaments which leads to a much more dynamically diverse stellar population than in test S3.

#### 4.4.2.2 Clustered star formation

Figure 4.15 shows the orientation of angular momentum of gas and star particles in S3 and S4. Amid obvious similarities with runs S1 and S2 (Figure 4.13), one notices extended linear structures, 'filaments' or 'arms' in Figure 4.15 for S3 and especially for S4. These structures, extended in velocity space, are in fact produced by the most compact features in coordinate space – the star clusters. These clusters are seen visually in Figure 4.14. The velocity dispersions of the stars in these very dense clusters is high; some are comparable to the orbital velocities of the clusters themselves. For example, in the right panel of Figure 4.15, the two longest continuous filaments, separated by 180° in  $\phi$ , are produced by the densest and most massive cluster located at  $z \approx 0$ ,  $x \approx -4$ ,  $y \approx -28$  in the right panel of Figure 4.14. Due to its location close to the y-axis, the y-component of the angular momentum of any star that belongs to the cluster satisfies  $l_y = -xv_z + zv_x \approx 0$ . Therefore, the cluster's stars form an incomplete plane  $l_y \approx 0$  in angular momentum space, producing the two (actually one) very long 'filament' in the right panel of Figure 4.15. This cluster



Figure 4.14: Projected gas density for simulations S3 (left) and S4 (right) at time t = 360, showing the locations of the stars. These plots are the equivalent visualisations of the Aitoff projection plots below.



Figure 4.15: Angular momentum distribution of individual gas particles (grey) and stars (colour) in simulation S3 (left) and S4 (right) at time t = 360. Extended stellar structures in these plots, such as concentrations and linear features, actually corresponds to extremely compact 'star clusters'. In the case of the most extended linear feature (right-hand plot), velocity dispersion inside the clusters is comparable to the orbital velocity of the cluster.

is the remnant of the primary cloud, and its mass is  $\sim 3000$  solar masses.

In fact most of the stars in simulations S3 and S4 belong to a cluster. This clustered mode of star formation in the fast-cooling runs is due to the gas collapsing promptly and thus forming very massive dense gas halos. The mass function that we find for the outer stellar population(s) is therefore likely to be a result of our particular star formation treatment. We note that in such a dense environment, any feedback from star formation activity, such as radiation or outflows, would not have been able to escape easily and would therefore have heated the surrounding gas, suppressing further fragmentation. Rather then than multiple low-mass stars in a cluster we would perhaps have seen fewer, higher mass stars formed. Indeed this effect of feedback has been documented a number of times in the literature (Nayakshin, 2006; Krumholz and Bonnell, 2007). We expect also that our results for star formation would change if the colliding clouds had net rotation (spin) before the collision, or if they had a turbulent structure, providing some stability against collapse.

#### 4.4.2.3 Radial distribution of stars

Interestingly, the most robust result from all of the six runs completed is the fact that distribution of stellar mass in an annulus,  $dM_* = \Sigma_*(R) 2\pi R dR$ , versus radius roughly follows the law

$$\Sigma_*(R) \propto \frac{1}{R^2} \,. \tag{4.15}$$

We show these distributions for runs S1, S2, S3 and S5 (all the simulations that have progressed far enough to transform a significant fraction of the gas into stars). To facilitate comparison with observational data, we use the *projected* radius instead of the proper 3D radius, which is not directly known in observations of young stars in the GC (e.g., Genzel et al., 2003). As the viewing angle of the stellar system modelled here is arbitrary, we chose to plot the distributions along the three axes of the simulations, so that  $R = \sqrt{x^2 + y^2}$  for the black symbols,  $R = \sqrt{y^2 + z^2}$  for the red symbols, and  $R = \sqrt{x^2 + z^2}$ for the blue symbols. The lines in the figure show  $R^{-2}$  (dotted) and  $R^{-3/5}$  (dashed) power laws. The latter corresponds to that predicted by *non self-gravitating* standard accretion disc theory (refer to Chapter 2, Section 2.3.4).

This result is remarkable given very large differences in the 3D arrangements of the stellar structures in our simulations. The  $1/R^2$  law is in good agreement with the observed distribution of 'disc' stars (Paumard et al., 2006). We have not been able to attribute a simple analytical explanation to this scaling, other than to suggest that cancellations of angular momenta in shocks appears to be very efficient in driving gas to smaller radii. We did however examine the distribution of the gas for a given N-body velocity profile emerging from the initial cloud collision, and attempted to fit the resulting circularisation radius  $r_c$  distribution to our result for  $\Sigma_*(R)$ . The thermal expansion of our clouds as they



Figure 4.16: Stellar surface density,  $\Sigma_*$ , versus projected radius, for runs S1 (top left, at t = 1955), S2 (top right, at t = 2243), S3 (bottom left, at t = 1396) and S5 (bottom right, at t = 2696). Different colours show the three different orientations, along z, x and y axes (black, red, and blue respectively). Power laws are overlaid for  $R^{-2}$  (dotted) and  $R^{-3/5}$  (dashed). All four plots correspond to the maximum time to which each simulation has been run.

collide can be modelled to a first approximation as an isotropic, delta-function velocity profile, which for small values of  $r_c$  yields a slope of  $\Sigma \propto R^{-1}$ . Introducing a gaussian velocity profile into the isotropic distribution, centered about low velocities, acts to steepen this slope closer to the desired  $R^{-2}$ . Beyond this simple N-body model, shocks and smallscale collisions elsewhere in the simulation will drive gas deeper into the potential well, increasing the slope even further. It would be interesting to explore a larger parameter space to quantify the robustness of these results.

#### 4.4.3 Inner disc

#### 4.4.3.1 Ring instability

Interestingly, in all of our simulations we discovered that the disc surface density exhibits eccentric rings in the inner parts. Star formation occurs preferentially in these rings (see Figure 4.6) in the slower cooling runs. The inner discs in runs S1-S6 are eccentric, warped, with a radial gradient in these properties, and evolving under the influence of a continuous but variable infall of material from larger radii. The rings therefore appear to form as a result of wave motion that propagates radially generating a series of regularly spaced overdensities in the form of rings that are enhanced through gas cooling. So far we have been unable to reproduce the effect in cleaner purpose-designed simulations. In particular, we ran simulations with eccentric and circular, flat and warped discs, but found no ring structures. Similarly, Nayakshin et al. (2007) did not encounter rings in their simulations of initially circular self-gravitating discs.

The most likely origin of this instability is reflection of waves at our inner boundary condition; namely, the accretion radius around Sgr A\*. By analogy with wavelike phenomena in other fields (such as optics), a wavelike disturbance moving through the disc will suffer partial reflection at the accretion boundary, as it marks a step change in the equivalent 'refractive index' of the medium through which the wave is travelling. Reflected waves will then set up an *interference pattern* with any incoming waves, creating a series of high- and low-density regions (Alexander, R., private communication). We have tested this hypothesis by altering the size of the accretion radius; with a smaller  $r_{\rm acc}$  the ring structures were far less marked, and did not propagate as far, while with a larger  $r_{\rm acc}$ the instability was even stronger. This suggests that it is indeed our inner boundary condition that is the cause of the problem. The strength of the instability, and our failure to reproduce it in purpose-designed tests, is the result of the strongly perturbing nature of the inflow that supplies the disc with material, and the lack of such behaviour in the cleaner, test discs. We note, somewhat fortunately, that when star formation does occur in these rings it does so, in general, with r < 1'', a region in which stars would not be expected to form in standard accretion disc theory (refer to Chapter 2, Section 2.3.4) as it lies within the self-gravity radius. We therefore do not take any notice of star formation

in this region.

#### 4.4.3.2 Midplane rotation

Simulation S4 is also interesting in that the inner disc undergoes complex evolution in angular momentum space. Due to the more grazing nature of the collision compared to S3, the inner disc feeding is more intermittent in its distribution of angular momentum. The newly arriving gas tweaks the disc orientation significantly. This is clearly seen in Figure 4.17, where the stellar disc remembers the 'old' orientation of the gaseous disc in which the stars were born, whereas the gaseous disc evolves to quite a different orientation. The possibility of this effect taking place was suggested, based on analytical arguments, by Nayakshin and Cuadra (2005). These authors found that stars will not follow disc midplane changes if these occur faster than the 'critical rotation time', which we briefly derive now. Firstly, we assume that the typical timescale for 'midplane rotation' of a disc is set by the mass condensation rate,  $\dot{M}_c$ , such that

$$t_{\rm rot} \sim \frac{M_{\rm disc}}{\dot{M}_c} \tag{4.16}$$

where  $M_{\text{disc}}$  is the mass of the disc. As the gas disc rotates perpendicular to the orbital rotation, and begins to separate from the stars, the characteristic relative velocity between the stars and the gas is  $\sim v_{\text{K}}/t_{\text{rot}}$ , where  $v_{\text{K}}$  is the Keplerian velocity. This defines an effective 'rotation acceleration', given by

$$a_{\rm rot} = \frac{v_{\rm K}}{t_{\rm rot}} \tag{4.17}$$

We must also consider the gravitational attraction between the disc and a star that has been left behind. Assuming the disc to be an infinite plane, this is given by

$$a_{\rm pl} = 2\pi G\Sigma \tag{4.18}$$

It is clear that we obtain the critical timescale for rotation when the 'rotation acceleration' of the turning disc midplane is equal to the gravitational acceleration between the star and the disc, namely  $a_{\rm pl} = a_{\rm rot}$ . The centrifugal acceleration of a star in a circular Keplerian orbit of cylindrical radius R around the central black hole is

$$a_{\rm c} = \frac{v_{\rm K}^2}{R} = a_{\rm pl} \frac{M_{\rm bh}}{2M_{\rm disc}} \tag{4.19}$$

since  $M_{\text{disc}} = \pi R^2 \Sigma$ . For critical rotation we therefore have

$$a_{\rm rot} = \frac{2M_{\rm disc}}{M_{\rm bh}} \frac{v_{\rm K}^2}{R} \tag{4.20}$$



Figure 4.17: The inner 1.5 arcseconds of the simulation S4 at t = 180 (left panel) and t = 260 (right panel). Note that stars born in the gaseous disc at the earlier time kept their orbital orientation whereas the gaseous disc has evolved due to deposition of new matter from larger radii.

yielding a timescale of

$$t_{\rm crit} = t_{\rm rot} = \frac{M_{\rm bh}}{2M_{\rm disc}} \frac{R}{v_{\rm K}} = \frac{M_{\rm bh}}{2M_{\rm disc}} t_{\rm dyn}$$
(4.21)

which is estimated to be about 500 in our code units for the disc in Figure 4.17. Our simulations are thus consistent with these predictions as the disc orientation changed on a timescale of just t = 80 in code units.

As a result of a several star forming events in gaseous discs of different orientations, the inner stellar disc in S4 is very much different from that found in the run S2, and even more so for S1. In the latter cases, when cooling is more gradual (i.e.  $\beta = 1$ ), the inner disc is fed in a steady-state manner. The S2 disc does show significant rotation out of the plane, but stars only form once it has settled into a stable orientation. Hence the resulting distribution of stars is that of a geometrically thin disc (see Figure 4.13). Such a distribution is broadly consistent with the observed orbits of young massive stars in the clockwise disc (Paumard et al., 2006). Due to gaseous disc midplane changes coupled with quicker fragmentation, the resulting stellar disc in S4 is much thicker, with  $H/R \sim 1$ . Simulation S4 thus fails to account for the most prominent feature of the observational data.



Figure 4.18: The time evolution of the accreted mass on to the central black hole (solid line) and the corresponding accretion rate (dotted line) compared to the Eddington limit for Sgr A\* (dotted red line), for simulations S1 (top left), S2 (top right), S3 (bottom left), and S4 (bottom right). Note that the variations in the accretion rate that lie below the Eddington limit should not be trusted in detail since for  $\dot{M} < \dot{M}_{\rm Edd}$  typically only  $\lesssim 10$  particles are accreted per orbit.

## 4.4.4 Feeding Sgr A\*

Figure 4.18 shows the mass accretion rate for the central black hole for simulations S1-S4. Formally, the Figure indicates that our simulations provide a sustained super-Eddington accretion rate (for Sgr A\* the Eddington rate is ~  $0.03 \,\mathrm{M_{\odot} yr^{-1}}$ ), in some cases for over  $10^4$  yrs. However, it must be remembered that we do not resolve the gas dynamics inside the accretion radius. We expect that material will form a disc there, and accretion will proceed viscously. The viscous time scale,  $t_{\rm visc}$ , depends on the temperature in the disc midplane and the viscosity parameter,  $\alpha$  (Shakura and Sunyaev, 1973). The midplane temperature in the inner arcsecond is ~  $10^3$  K in both the standard (non self-gravitating, e.g. Nayakshin and Cuadra (2005)) and the self-gravitating regimes (Nayakshin, 2006), yielding  $H/R \sim 0.01$ . Hence, the viscous time is

$$t_{\rm visc} = 6 \times 10^6 \text{ years } \alpha_{0.1}^{-1} \left[\frac{R}{100H}\right]^2 R^{3/2} ,$$
 (4.22)

where  $\alpha_{0.1} = \alpha/10$  and R is in our code units (one arcsecond). With these fiducial numbers, the viscous time scale coincides with the age of the young massive stars in the GC (e.g., Krabbe et al., 1995; Paumard et al., 2006). As  $\alpha$  is highly uncertain, we shall consider the two extreme cases. If  $t_{\rm visc} \ll$  a few million years (e.g., if  $\alpha = 0.1$ ,  $t_{\rm visc} \approx 10^6$  years at 0.3"), we expect that gas would have mainly accreted on to Sgr A\* by now. This accretion rate would be a significant fraction of the Eddington accretion rate. Standard disc accretion would then generate as much as  $2 \times 10^{56}$  erg  $M_{\rm acc,3}$  of radiative energy, where  $M_{\rm acc,3}$  is the gas mass accreted by Sgr A\* in units of  $10^3 \,\mathrm{M}_{\odot}$ . A similar amount of energy could have been released as energetic outflows. There is currently no evidence for such a bright and relatively recent accretion activity of Sgr A\*.

In the opposite limit, i.e., if  $\alpha \ll 0.1$ , and  $t_{\text{visc}} \gtrsim$  a few million years, the gaseous disc should still be there as self-consistent modelling predicts that stars should not be forming within R = 0.3'' - 1'' (Nayakshin, 2006; Levin, 2006). There are however very strong observational constraints on the absense of such a disc in the inner arcsecond of our Galaxy (e.g., Falcke and Melia, 1997; Narayan, 2002; Cuadra et al., 2003). Therefore, our simulations seem to over-predict the amount of gaseous material deposited in the inner  $\sim 0.3''$ . We take this as an indication that a better model would perhaps involve a gaseous cloud of a larger geometric size, thus shifting all spatial scales outwards (see further discussion of this in Section 4.5).

#### 4.4.5 Summary of results

In terms of the final distribution of gas and resulting stellar populations, our results can be divided into two main categories.

For those runs with a relatively small impact parameter, namely S1, S3 and S5, the initially small-scale disc around the black hole grows by steady accumulation of gas to extend out to r ~ 15". As a result, more than one stellar population is seen in the disc. In addition to the ubiquitous stars in the inner few arcseconds, a 'mid-range' population is seen, at r ~ 5 − 8". Populations in the filaments vary: in S1, only the primary filament forms a stellar population, as the secondary has accreted on to the disc by the time stars begin to form; in S3, although the gas dynamics are the same, the secondary filament forms stars before it can accrete, resulting in two filamentary populations; and in S5, due to the small impact parameter a single large disc is created, resulting in three disc stellar populations all at a similar orientation.

2. For the runs with a larger impact parameter, namely S2, S4 and S6, the small-scale disc grows very little over the course of the simulation, staying within a radius of a few arcseconds. The feeding of the disc is far from steady, generating enhanced midplane rotation, which, coupled with star formation (particularly in the case of faster cooling, e.g., S4) results in *geometrically thick* stellar populations. Both the primary and secondary filaments survive, and each form stellar populations inclined at a large angle to each other.

The other results of our simulations, presented roughly in the order of their commonality to the six runs, are the following. Firstly, the formation of a nearly circular, gaseous disc in the inner region of the computational domain is common to all runs, as is the ensuing formation of stars on similar near-circular orbits. This is natural as the dynamical time in the innermost disc is only  $\sim 60 r^{3/2}$  years. On the scale of the inner disc,  $r \sim 1$ , the disc makes tens to hundreds of revolutions during the simulation, allowing for near circularisation of the gaseous orbits. Conversely, the outer gaseous stream becomes selfgravitating much faster than it could circularise (or even become an eccentric disc), and hence orbits of stars in that region are more eccentric. This division on inner nearcircular and outer eccentric orbits is in broad agreement with the properties of the observed orbits (Paumard et al., 2006). One discrepancy however is the eccentricity of the counterclockwise feature of the observations compared to the outer stellar populations as seen in our model; the former is believed to possess a value of  $e \sim 0.6 - 0.8$  whilst the latter we find is only  $e \sim 0.2 - 0.4$ . We believe that simulations that use a strongly eccentric initial orbit for one of the clouds should be able to match the observations more closely.

Secondly, we found that the radial distribution of stellar mass closely follows the observed  $\Sigma_* \propto 1/R^2$  profile of the disc stellar populations (Paumard et al., 2006). This was observed for all the simulations, although we expect these results would change if the colliding clouds moved in similar directions, significantly reducing the angular momentum cancellation in the shock and the thermal 'kick' velocity due to the shock. Of course, this latter setup would prevent the formation of stellar populations at a large angle of inclination with respect to each other (one of the most prominent features of the observational data), so it is reasonable that we have not tested this possibility in our model.

For the same initial configurations, simulations with a comparatively long cooling time parameter,  $\beta = 1$ , lead to kinematically less dispersed stellar populations than those with faster cooling (i.e.,  $\beta = 0.3$ ). As a result, the longer the cooling time, the more closely the resulting stellar system can be fit by planar systems in velocity space. The innermost stellar disc is then reminiscent of the observed clockwise *thin* stellar system (Paumard et al., 2006) in terms of stellar orbits. Rapid cooling produces clumpier gas flows that lead to significant gaseous disc orientation changes during the simulations. Rapidly cooling gas flows naturally form less coherent discs; such geometrically thick systems are incompatible with the observations. Interestingly though, with the chosen initial conditions, faster cooling promotes survival of the gaseous streams corresponding to the orbits of the original clouds. These streams fragment and form stars mainly in a clustered mode, although this might be expected to depend on the details of the radiation feedback from young stars, which is not modelled in this set of simulations. Our slower cooling runs produce inner and outer stellar systems that are inclined to each other by only about 40°, which is much smaller than the observed angle of ~ 110° between the discs (Paumard et al., 2006). This is a natural consequence of angular momentum conservation in 'well-mixed' gas flow – the disc orientation in our model will always lie in-between the positions of the initial clouds in angular momentum space.

Finally, another robust result of our simulations is that all of the gaseous and ensuing stellar discs are significantly warped, by between  $30^{\circ}$  to  $60^{\circ}$  measured from the inner to the outer radii. Interestingly, the most recent analysis of the observations reveals strong warps in the clockwise disc system (Bartko et al., 2009).

# 4.5 Discussion

#### 4.5.1 Comparison to the Galactic centre

There are a number of aspects of our simulations that show promise in re-creating the observed properties of the GC in terms of bulk dynamics. The *in situ* accretion disc model suffers from two major difficulties (Paumard et al., 2006); namely, the presence of a massive stellar cluster (IRS13E) and a highly eccentric outer stellar population, and the presence of two co-eval discs inclined at a large angle to each other. In the latter case, the assumption that each disc was formed from a different cloud infall event requires that two such events occurred within  $\simeq 1$  Myr, which is quite unlikely. Our model, although simplistic, manages to account for each of these difficulties. We see the formation of a dense, massive stellar cluster, in the outer eccentric gaseous filament, and this filament forms stars that follow the eccentric orbits of the gas. We also form two distinct stellar populations from a single progenitor event, with these populations inclined at an angle to each other. It does appear, however, due to angular momentum conservation, that it is quite difficult to reproduce the large angle of inclination seen in the GC, as the inner population's orientation is formed from the *combination* of the initial angular momentum vectors of the clouds whilst the outer populations correspond to the individual clouds' angular momenta. The maximum angle that can be achieved between the inner disc and an outer filament in our model is therefore approximately half of the initial angle between the clouds' velocities for a collision of equal mass bodies.

Although the bulk dynamics of our simulations are reasonably matched to GC observations, we do not manage to reproduce the specific properties of the observed stars and their orbits in detail. However, there are important hints on a plausible scenario for the formation of these stars. The observed well defined, flat, geometrically thin and almost circular inner stellar system is best created via a gentle, slow accumulation of gas. Several independent major gas deposition events lead to a disc that is too warped, and/or 'mixed up' systems consisting of several stellar rings or discs co-existing at the similar radii. To avoid this happening, the inner disc must be created on time scale longer than the critical rotation time, which is estimated at  $t_{\rm cr} \sim \text{few} \times 10^4$  years. The deposition of gas in the inner disc takes place on the longest of two timescales: the cooling time  $t_{\rm cool}$  and the collision time,  $t_{\rm coll} \sim r_{\rm cl}/v_{\rm cl}$ , where  $r_{\rm cl}$  and  $v_{\rm cl}$  are the cloud's size and velocity magnitude.

#### 4.5.1.1 Estimate of the cooling time during the cloud collision

Let us first estimate realistically the clouds' cooling time as they collide. For this we must analyse the processes occuring during the collision.

1. Clouds moving at ~ a few hundred km s<sup>-1</sup> (namely, a highly supersonic velocity) would shock, reaching a temperature of up to a few million K. This value can be obtained from applying the Rankine-Hugoniot shock conditions (refer to Chapter 3, Section 3.2.5), with the assumption that the equation of state is adiabatic,  $\gamma = 5/3$ . This puts a limit on the ratio between the pre- and post-shock densities,  $\rho_2/\rho_1 \simeq 4$ . Using the 2nd shock condition (equation 3.76) we find

$$P_2 = P_1 + \frac{3}{4}\rho_1 v_1^2 \tag{4.23}$$

where subscripts 1 & 2 refer to pre- and post-shock regions respectively. Putting in the relevant values for the quantities (in particular the initial cloud density of  $\approx 6.7 \times 10^{17} \text{ g cm}^{-3}$ ), we find the post-shock temperature from

$$T_2 = \frac{\mu m_p P_2}{k_B \rho_2} \approx 1.4 \times 10^6 \text{K}$$
(4.24)

2. This shock temperature is above the typical ionisation threshold for H ( $\sim 10^4$  K), ensuring that the gas is ionized and optically thin. The gas will therefore cool through Bremsstrahlung (free-free) emission, the cooling function for which we derive in Appendix D. This results in a cooling time of

$$t_{\rm cool} = \frac{3nk_BT}{\Lambda n^2 T^{1/2}} \tag{4.25}$$

where n is the number density in the gas (for our clouds this is  $n \approx 4 \times 10^7 \text{ cm}^{-3}$ ) and

$$\Lambda = \left(\frac{q^6}{m_e^2 c^3}\right) \left(\frac{m_e k_B}{\hbar^2}\right)^{1/2} \tag{4.26}$$

where q is the charge on an electron,  $m_e$  the mass of an electron, and  $\hbar$  is Planck's constant. Putting the relevant quantities in yields  $t_{\rm cool} \simeq 5.8 \times 10^7$  s, 3-4 orders of magnitude shorter than the dynamical time at the position of the cloud collision,  $t_{\rm dyn} \simeq 2.6 \times 10^{11}$  s. This is in contrast to the simulations in our model, where  $t_{\rm cool} \sim t_{\rm dyn}$ .

3. The Bremsstrahlung emission will be in the form of soft X-rays ( $E_{\rm rad} \simeq 2.9 \times 10^{-10}$  erg s<sup>-1</sup>  $\simeq 200$  eV) which have energy far higher than the photoionisation energy for hydrogen (13.6 eV). The X-rays thus photoionise the cloud, making the dominant source of opacity that of *bound-free* absorption. The cross-section, above the ionisation threshold, scales with frequency as

$$\sigma(\nu) = \sigma_{\rm bf} \left(\frac{\nu}{\nu_{\rm thresh}}\right)^{-3} \tag{4.27}$$

where  $\nu_{\rm thresh}$  is the threshold frequency of  $3 \times 10^{15}$  Hz (corresponding to a wavelength of 91.2 nm), and for hydrogen  $\sigma_{\rm bf} \approx 10^{-17}$  cm<sup>2</sup> (Rybicki and Lightman, 1986). For soft X-rays with  $\nu \sim 10^{17}$  Hz, we find a cross-section of  $\sigma \approx 2.7 \times 10^{-22}$  cm<sup>2</sup>. For the clouds to be Compton-thick to the irradiating X-rays, therefore, we require a number density along the line of sight greater than or equal to  $1/\sigma$ . To an order of magnitude, this is  $\Sigma \sim \rho r_{\rm cloud} \simeq 2.4 \times 10^{25}$  cm<sup>-2</sup>. We can see therefore that the X-rays will be fully absorbed within the clouds.

4. Although X-rays are photoionising, the high density necessarily results in recombination of any free electrons, and the dominant effect will be *heating* of the cloud. The heated parts of the cloud, will, in turn, radiate as a blackbody. We can get the effective temperature of this radiation, assuming the cloud is optically thick, from the luminosity per surface area via the standard Stefan-Boltzmann law:

$$L = 4\pi r^2 \sigma T_{\text{eff}}^4 \tag{4.28}$$

The luminosity we assume to derive from the kinetic energy of the collision,  $E_{\rm kin} = 1/2 \ m(\Delta v)^2 \simeq 10^{52}$  erg, over a collision time of  $r_{\rm cloud}/\Delta v \simeq 10^{10}$  s. This comes out as  $\simeq 10^{42}$  erg s<sup>-1</sup>, giving  $T_{\rm eff} \simeq 500$  K. From Wien's law, the peak wavelength of the emission from a blackbody at this temperature is  $\simeq 6\mu$ m.

5. Finally, then, we must determine the clouds' optical depth to radiation with  $\lambda \simeq 6\mu$ m. We note that the energy is not high enough for photoionisation, and so we must consider bound-bound opacities, and for realistic molecular clouds, dust and metal opacities also. For this we turn to the Bell and Lin (1994) opacity graph, from which we read off the dust opacity (the dominant one at low temperatures since the

cloud will be mostly H<sub>2</sub>) as  $\kappa \simeq 2 \text{ cm}^2 \text{ g}^{-1}$ . Converting this into an optical depth,  $d\tau \sim \kappa \rho r_{\text{cloud}}$ , we find that  $\tau \simeq 80$ , i.e., very large. This means that the cloud will radiate as a blackbody, with a rate of flux per unit area according to  $F = \sigma T^4$ . This last estimate puts  $t_{\text{cool}}$  at least 2 orders of magnitude shorter than the dynamical time.

We conclude then that the cooling time *during the collision* is unlikely to exceed the required  $t_{\rm cr}$  unless we accept an unrealistically low mean gas density in the original cloud (say, 10 hydrogen atoms per  $cm^3$ , which is implausibly low for the inner parsecs of the Galaxy). We are thus left with the only option to require the collision itself be more prolonged than  $t_{\rm cr}$ . Estimating the velocity of the cloud at  $v_{\rm cl} \sim 150 \text{ km sec}^{-1}$ , which is of order of circular velocities in the inner Galaxy outside the inner parsec, we find  $t_{\rm coll} = r_{\rm cl}/v_{\rm cl} \sim (10^4 \, {\rm yrs}) \, r_{\rm cl,pc}$ , where  $r_{\rm cl,pc}$  is the size of the cloud in parsecs. We hence require the cloud to be larger than a few parsecs to satisfy  $t_{\rm coll} \gtrsim t_{\rm cr}$ . Note that this size is not necessarily the original size of the cloud if the cloud gets tidally disrupted before it makes the impact. In the latter case we can take  $r_{\rm cl}$  to be the radial distance to the centre of the Galaxy at which the tidal disruption took place. Finally, the location of the collision should not be too far from the central parsec, or else too much angular momentum would have to be lost to deposit a significant amount of gas at  $\sim 0.1$  pc region. Taking these constraints together, we believe that the most realistic scenario would be a giant molecular cloud (GMC) of the order of a few parsec in size striking the circumnuclear disc (CND) at the distance of a few parsec from Sgr A<sup>\*</sup>.

Another argument going in the same direction comes from a comparison of the radial distribution of gas and stars in our simulations with the observed stellar distribution (Paumard et al., 2006). The former is too compact, i.e., all of our simulations deposited too much mass within the inner arcsecond. In addition, if that was indeed the case 6 million years ago, then Sgr A<sup>\*</sup> would have received a significant amount of fuel, enough to become at least a bright AGN. Given the long viscous times in the inner arcsecond, Sgr A<sup>\*</sup> could actually continue to accrete this fuel now. However, it is well known that there is no geometrically thin and optically thick disc inside the inner arcsecond of Sgr A<sup>\*</sup> (Falcke and Melia, 1997; Narayan, 2002; Cuadra et al., 2003). Eliminating the gaseous disc by star formation is not an option as there are not enough massive young stars observed there.

#### 4.5.2 Alternative models

There are a number of authors who have considered similar models to the one that we use here. Bonnell and Rice (2008) have performed simulations of an infalling turbulent molecular cloud on a sub-parsec impact parameter with respect to the SMBH. Both found that the cloud became tidally disrupted and formed an eccentric disc around the black

hole. This disc later fragmented into stars with a top-heavy IMF, in agreement with Nayakshin et al. (2007). In addition, they found that the stellar orbits retain the eccentricity of the disc, which is more circular in its inner parts, with the eccentricity increasing with radius. This of course is in agreement with our results for the small-scale disc in our simulations.

Another paper, Wardle and Yusef-Zadeh (2008), used analytical arguments to suggest that a viable progenitor for the stellar discs is the radial passage of a large-scale molecular cloud, such as the 50 km s<sup>-1</sup> GMC (Armstrong and Barrett, 1985), which temporarily engulfs the SMBH, leaving behind some of its material to settle into a bound disc of gas. Their argument stems from the idea that the capture of gas by the black hole is aided by gravitational focusing of the passing cloud, such that regions of gas passing on opposite sides of Sgr A<sup>\*</sup> would have orbital angular momenta that would be largely equal and opposite. Any collision between them would therefore act to directly cancel their angular momentum. This model may be useful in explaining not only the observed stellar discs within the inner parsec but also the CND further out at ~ 10 pc.

Our approach uses similar principles to these authors, but is set apart by the fact that we model a collision of two clouds of gas, rather than one, and that the collision is offcentre. This allows us to achieve both an inner and an outer stellar population, where the inner is more circular and the outer more eccentric, in agreement with observations. The one cloud scenario is attractive both qualitatively and quantitatively as an explanation for the origin of the clockwise stellar disc, but we note that the young, massive stars that do not kinematically belong to the clockwise disc account for no less than  $\sim 50\%$  of the overall stellar population. Whether the majority of these stars form the second disc or not, their velocity vectors differ from the clockwise disc by up to a few hundred km  $s^{-1}$ (which is of the order of the circular velocity at these distances). The velocity dispersion even in a very compact GMC, with a mass of, say,  $3 \times 10^4 \,\mathrm{M_{\odot}}$ , and size of 1 parsec, is only  $\sim 10 \text{ km s}^{-1}$ . Thus the only way to create the observed kinematically distinct population of stars would be to postulate the existence of two or more streams (filaments) inside the cloud that pass on opposite sides of Sgr  $A^*$  and do not get completely mixed before forming stars. Given our numerical experiments in this paper, this may be a plausible scenario if the cooling time is short,  $\beta \lesssim 1$ . What is interesting in this picture is that the massive stars of the counter-clockwise population would then have to form very quickly, i.e., on a dynamical timescale, or else gaseous orbits would be mixed. The rotation period scales approximately as  $T_{\rm rot} = 3000 (R/5'')^{3/2}$  years, so this is quite fast indeed.

It should also be noted that recent N-body simulations conducted by Cuadra et al. (2008); Alexander et al. (2007) imply that it would have been very difficult for the high eccentricities and inclinations of the dynamically hotter counter-clockwise feature to have been formed from a flat, cold disc via scattering processes. A single disc progenitor for both

GC stellar features is therefore largely ruled out. In contrast, in the case of the collision of two clouds as considered here, it is almost too easy to obtain an inner near-circular disc and a kinematically diverse stellar population farther out. We therefore favour a model where a GMC collided with a pre-existing cloud or structure, such as a massive larger-scale disc, e.g., similar to the observed CND.

#### 4.5.3 Convergence tests

Finally, we must ensure that the results of our simulations are not dominated by numerical effects. The role of an artificial viscosity, while essential for the treatment of shocks, can cause unwanted dissipation and/or angular momentum transport that may modify the behaviour of the gas flow and invalidate our findings. To this end, we have performed *convergence* tests on both the artificial viscosity parameter,  $\alpha$ , and the resolution (the number of SPH particles) of the simulation. Both of these relate to the transport of the gas through the computational domain, which is of particular relevance to forming the required disc-like structures and feeding the central SMBH. A converged result is an indication that our conclusions are not motivated by numerical artifacts but are instead based on physical principles.

As can be seen in Figure 4.19, we find excellent convergence in both cases for the radial density profile, the radial mass profile, and the radial mass enclosed profile of the gas. Likewise, the visualisations (not shown here) of the various structures look very similar when changing these parameters. There are some minor differences in the very innermost part of the computational domain, namely r < 0.5 - 1'', but this region is largely unimportant for our conclusions.

# 4.6 Conclusion

In this Chapter we have presented several simulations of cloud-cloud collisions aimed at reproducing gas flows that form gaseous disc(s) and stellar populations in the central parsec of our Galaxy. We found the gas cooling time and the impact parameter of the collision to influence the outcome significantly. Comparison of the simulations with the observed data for the young stars in the GC suggests that the inner gaseous disc must have been assembled in a gentle way, i.e., on time scales longer than 10<sup>4</sup> years, which is much longer than the local dynamical time. These results agree with previous analytical arguments (Nayakshin and Cuadra, 2005) and arguments based on specific models for the top-heavy IMF in the GC (Nayakshin, 2006). This could occur if either the cooling time of the flow (after the collision) was long, or the duration of the collision that resulted in the deposition of the gas into the inner parsec was longer than 10<sup>4</sup> years. Analytical estimates suggest that the latter possibility is far more likely, and hence we suggest that



Figure 4.19: Convergence tests on the value of the artificial viscosity parameter,  $\alpha$  (left-hand side) and resolution (right-hand side) for simulation S1. The plots are: density defined on radial shells (top), mass defined on radial shells (middle) and mass enclosed defined on radial shells (bottom). For the viscosity parameter the lines correspond to  $\alpha = 0.67$  (blue),  $\alpha = 2$  (black) and  $\alpha = 6$  (red). In the resolution plots the lines correspond to  $N_{\rm sph} = 875,000$  (blue),  $N_{\rm sph} = 2,625,000$  (black) and  $N_{\rm sph} = 5,250,000$  (red). We note that although the latter variation is quite a  $N_{\rm sph}$ , it is a relatively small range in spatial resolution (which scales approximately as  $N_{\rm sph}^{1/3}$ ). Unfortunately, computational limitations prevent us from increasing this range, but we are nonetheless confident that the convergence is good enough to demonstrate the reliability of the results.

the initial cloud size or the location of the collision was at least a few parsecs, i.e., larger than that of our initial conditions. We arrive at the same conclusions also from looking at the deposition of material into the inner arcsecond (as discussed in Section 4.4.4).

We find that the transport of gas down the potential gradient of the gravitational potential (dominated by Sgr A<sup>\*</sup>) proceeds both dynamically, through cancellation of angular momentum as the clouds collide, and later viscously<sup>6</sup>, once the disc(s) form. However, even to later times the accretion rate on to Sgr A<sup>\*</sup>(actually the accretion radius) is at or near its Eddington limit. This appears to be a consequence of a highly disordered flow, with the hot gas left over from the cloud collision cooling, self-colliding, and generally continuing to infall on to the inner region in a non-axisymmetric, time-varying manner. It is notable that the accretion rate is significantly higher for a sustained period of time in those simulations where a small-scale, time-varying disc formed around the SMBH (e.g., S2, S4) rather than a large, planar disc (e.g., S1, S3, S5), as it is precisely this mode of 'stochastic' accretion that we proffer as a potential solution to the AGN feeding problem (refer to Chapter 1, Section 1.7.1.2 and references therein).

A significant factor in creating this complex flow is the thermal velocity 'kick' that the gas receives as the clouds collide, and more fundamentally the fact that the heat gained from the initial collision is retained in the gas for some time, allowing it to remain diffuse and occupy a large volume fraction of the computational domain for  $\sim 10^4$  yrs. It is at this point that we should be cautious. As we estimated in Section 4.5.1.1, the cooling time for the clouds as they collide and shock should realistically be a few orders of magnitude shorter than the dynamical time; this is counter to the numerical setup of our simulations, which necessarily cool on a dynamical time. From a purely hydrodynamical perspective, therefore, we might expect a far more structured flow than is seen, earlier star formation, and perhaps a lower accretion rate. On the other hand, our considerations of the cooling time completely neglect the action of magnetic fields, which would most likely be present in the pre-collision clouds (Levin, Y., private communication). If the radiative cooling time is short, then the shock is essentially isothermal, and we can expect the gas to be compressed by a few orders of magnitude for our typical conditions. A frozen-in magnetic field would then be amplified by similar factors, and magnetic pressure by four to six orders of magnitude. This might present another form of energy and strong pressure support against gravitational collapse, helping to keep the post-collision gas flow diffuse to late times.

In the next Chapter we move to the *general* environment of a galactic centre, focusing on the feeding of the central black hole from scales significantly further out than we have dealt with here. It is important to note that some aspects of our colliding clouds model are

 $<sup>^{6}</sup>$ it should be noted here that since there is no *physical* viscosity in the simulation this term really refers to *numerical* transport of angular momentum through the artificial viscosity

relevant to the model we consider next, as we stay with the theme of dynamical, disordered accretion that results from an initial 'kick' imparted to the gas distribution.

# 5

# Feeding SMBHs through supersonic turbulence and ballistic accretion

"You! Cake or death? Errr...cake please. Very well! Give him cake!"

Eddie Izzard

# 5.1 Introduction

The growth of supermassive black holes (SMBHs) at the centre of galaxy bulges is known to be correlated with observable properties of the host spheroid. The best established correlations are between the SMBH mass,  $M_{\rm bh}$ , and the velocity dispersion,  $\sigma$ , i.e., the  $M_{\rm bh}-\sigma$  relation (Gebhardt et al., 2000; Ferrarese and Merritt, 2000), and the  $M_{\rm bh}-M_{\rm bulge}$ relation (Magorrian et al., 1998; Häring and Rix, 2004), where  $M_{\rm bulge}$  is the stellar mass of the bulge. Understanding the formation and growth mechanisms of SMBHs is therefore believed to be important in determining the evolution of the larger host systems.

High luminosities of high redshift quasars at  $z \sim 6$  (when the age of the Universe was  $\sim 1 \text{ Gyr}$ ) suggest SMBH masses of up to  $\sim 10^9 \text{ M}_{\odot}$  (Kurk et al., 2007), implying accretion on to the black holes close to or even above the Eddington limit (King and Pringle, 2007) for up to  $\sim 1$  Gyr. Such a strong inflow is most likely the result of major mergers between galaxies, which have been shown via hydrodynamical simulations to drive gas inward through gravitational tidal fields (e.g., Barnes and Hernquist, 1991). Simulations have also shown that galaxy mergers within large dark matter halos at  $z \sim 14$  may be able to drive gas in to the centre of the galaxy at a sufficient rate to grow a seed BH of  $10^4 \text{ M}_{\odot}$  to an SMBH of  $10^9 \text{ M}_{\odot}$  in the required time i.e., by  $z \sim 6$  (Li et al., 2007). Recent cosmological simulations of dark matter halos that include gas physics and black hole feedback processes are able to build up the required mass by z = 6 via extended Eddington-limited accretion (e.g., Sijacki et al., 2009).

However, an ab-initio treatment of all the relevant physical processes (gravity, star formation, magneto-hydrodynamics, radiative feedback etc.) over the full range of scales required is still beyond current numerical capabilities. Therefore, little is known about the mechanism(s) that connect the large-scale flow (~ kpc) to the small-scale accretion flow in the immediate vicinity of the SMBH ( $\leq 1$  pc). Current cosmological simulations require sub-resolution prescriptions to encapsulate the accretion process (e.g., Springel et al., 2005; Booth and Schaye, 2009; Sijacki et al., 2009). These models commonly use Bondi-Hoyle accretion (Hoyle and Lyttleton, 1939; Bondi and Hoyle, 1944; Bondi, 1952), corrected upwards by orders of magnitude when the resolution limitations cause an under-prediction of the desired SMBH growth rate. Recently, DeBuhr et al. (2009) performed hydrodynamical simulations of major galaxy mergers, employing a sub-grid model for viscous angular momentum transport to dictate the accretion rate, the results of which suggest that the Bondi-Hoyle prescription actually *over-predicts* accretion on to the SMBH by up to two orders of magnitude. The applicability of the Bondi-Hoyle solution to the SMBH accretion rate is therefore not clear.

Looking at the problem from the other end, where the dynamical and viscous timescales are short compared with cosmological timescales, it is expected that geometrically thin and optically thick accretion discs (Shakura and Sunyaev, 1973) form due to a non-negligible amount of angular momentum in the gas. While it is possible to study such discs at a far higher resolution and with more physics, the approach is obviously limited by the fact that the disc is commonly assumed to just exist *a priori*. How these discs operate on  $\sim$  pc scales is also not clear as the discs are susceptible to self-gravity and hence may collapse into stars (Paczyński, 1978; Kolykhalov and Sunyaev, 1980; Collin and Zahn, 1999; Goodman, 2003; Nayakshin and Cuadra, 2005).

There is thus a genuine gap between these two scales and approaches. The purpose of this Chapter is to try to bridge this gap to some degree with the help of numerical hydrodynamical simulations of gas accreting on to a SMBH that is immersed into a static, spherical bulge potential. In particular, we study the 'intermediate-scale' flow, from the inner 100 parsecs of a galactic bulge (just below the resolution limit for some of the betterresolved cosmological simulations) to the inner parsec, where the gas would be expected to settle into a rotationally-supported feature. Instead of assuming the presence of the disc, we allow it to form self-consistently from the larger scale flow. For simplicity of analysis of the results, we start with a spherically-symmetric shell of gas of constant density, where the gas is isothermal throughout the simulation. The first non-trivial element in our study is the rotation of the shell, which should force the gas to settle into a rotating disc.

While this turns out to be an interesting system in itself, forming a narrow ring rather than a disc, we feel motivated to consider a more complex initial velocity field in the shell. First of all, cosmological simulations show the presence of cold streams (e.g., Dekel et al., 2009) immersed in a hotter, lower density medium. These complex gas features must propagate to smaller scales in the bulge as well. Secondly, the abundance of gas required to fuel the SMBH should also fuel star formation in the host. Massive stars deposit large amounts of energy and momentum in the surrounding gas (Leitherer et al., 1992), which is probably one of the dominant sources of *turbulence* in the interstellar medium (e.g., McKee and Ostriker, 1977; Mac Low and Klessen, 2004). In general, where we can observe it, it is evident that turbulence is ubiquitous in astrophysical flows over a wide range of scales and systems (for a review see e.g., Elmegreen and Scalo, 2004). Strong supersonic turbulence is a key ingredient in all modern simulations of star formation (e.g., Krumholz and Bonnell, 2007; Bate, 2009).

Therefore, we must expect that far from the SMBH, while the specific angular momentum of gas, l, is too large for it to be captured by the SMBH within the central parsec (or less), the *dispersion* in l is also large. It is not unreasonable to expect that there will be some gas counter-rotating with respect to the mean rotation of the flow. To test the importance of these ideas for accretion of gas on to SMBHs, we draw on the machinery developed for turbulent flows in the star formation field by adding a turbulent velocity spectrum to the gas in the initial shell. By varying the normalisation of the turbulent velocities we essentially control the dispersion in the initial gas velocities. We find that the role of the turbulence and in general any disordered supersonic velocity field is two-fold. First of all, it broadens the distribution of specific angular momentum, setting some gas on lower angular momentum orbits. Secondly, as is well known (e.g., McKee and Ostriker, 2007b), turbulence creates convergent flows that compresses gas to high densities. We find that the dynamics of such regions can be described reasonably well by a 'ballistic' approximation in which hydrodynamical drag and shocks are of minor importance for the high density 'bullets' moving through the lower density background fluid.

These two effects increase the accretion rate on to the SMBH by up to several orders of magnitude, largely alleviating the angular momentum barrier problem. Note that several authors have already suggested that star formation feedback and supersonic turbulence *inside* accretion discs can promote SMBH accretion by amplifying angular momentum transfer (Collin and Zahn, 2008; Kawakatu and Wada, 2008; Chen et al., 2009). An astrophysical conclusion from our simulations is that star formation in the host can actually promote AGN accretion.

The chapter is arranged as follows. In Section 5.2 we outline the computational method used, and in Section 5.3 we discuss our initial conditions. Section 5.4 presents an overview of the gas dynamics. Sections 5.5 & 5.6 present the results<sup>1</sup> for the no turbulence and turbulent cases, respectively, and Section 5.7 details the fate of the gas that makes it inside the accretion radius. Sections 5.8 & 5.9 comprise our discussion and conclusion respectively.

# 5.2 Computational method

For the simulations we use the three-dimensional SPH/N-body code GADGET-3, an updated version of the code presented in Springel (2005). Smoothing lengths in the gas are fully adaptive down to a minimum smoothing length of  $2.8 \times 10^{-2}$  pc, which is much smaller than the scales that we resolve in the simulation. The code employs the conservative formulation of SPH as outlined in Springel and Hernquist (2002), with the smoothing lengths defined to ensure a fixed mass within the smoothing kernel, for  $N_{\text{neigh}} = 40$ . Each simulation below starts with  $N_{\text{SPH}} \sim 4 \times 10^6$  particles, with an individual SPH particle possessing a fixed mass of  $m_{\text{sph}} \approx 12 \,\text{M}_{\odot}$ . The Monaghan-Balsara form of artificial viscosity is employed (Monaghan and Gingold, 1983; Balsara, 1995) with  $\alpha = 1$  and  $\beta = 2\alpha$ .

The calculations are performed in a static isothermal potential with a central constant density core to avoid divergence in the gravitational force at small radii. We also include the black hole as a static Keplerian potential. The mass enclosed within radius r in this

<sup>&</sup>lt;sup>1</sup>for movies of the simulations the reader is directed to http://www.astro.le.ac.uk/~aph11/movies.html

model is:

$$M(r) = M_{\rm bh} + \begin{cases} M_{\rm core} (r/r_{\rm core})^3 , & r < r_{\rm core} \\ M_a(r/a), & r \ge r_{\rm core} , \end{cases}$$
(5.1)

where  $M_{\rm bh} = 10^8 \,\mathrm{M_{\odot}}$ ,  $M_a = 10^{10} \,\mathrm{M_{\odot}}$ ,  $a = 1 \,\mathrm{kpc}$ , and  $r_{\rm core} = 20 \,\mathrm{pc}$  and  $M_{\rm core} = 2 \times 10^8 \,\mathrm{M_{\odot}}$ . The one dimensional velocity dispersion in the isothermal part of this potential is  $\sigma = (GM_a/2a)^{1/2} = 147 \,\mathrm{km \ s^{-1}}$ .

As we concentrate on the hydrodynamics of gas accreting on the SMBH, we make two simplifying assumptions which avoid further non-trivial physics (to be explored in our future work). First, the position of the SMBH is held fixed during the simulations here. This is a reasonable approximation as all of our initial conditions are either exactly spherically or azimuthally symmetric, or have these symmetries when averaged over the entire simulation volume (e.g., turbulence is assumed to be isotropic in this case). Allowing the black hole to move self-consistently would require relaxing the static potential assumption<sup>2</sup>. Secondly, gravitational forces between the particles are switched off to avoid complications that might arise from gas self-gravity. While this could be viewed as a shortcoming of our work, we believe that inclusion of self-gravity, ensuing star formation and stellar feedback would only strengthen our results and conclusions. The gravitational contraction of gas clouds would create even higher density contrasts and thus make ballistic trajectories even more likely. Star formation feedback would drive its own turbulence and hence amplify the effects of the turbulence that we seed.

One further simplification is that the gas is isothermal, at  $T = 10^3$  K, throughout the entirety of the simulations. This is a fair assumption as cooling times are short for the high densities we are considering. Accretion of gas on to the SMBH is modelled with the 'accretion radius' approach; particles that come within an accretion radius of  $r_{\rm acc} = 1$  pc are removed from the simulation and we track the total mass as well as each component of the net angular momentum vector of the accreted particles. This contrasts with the Bondi-Hoyle accretion formulations that are frequently used in cosmological simulations (e.g., Sijacki et al., 2009; Booth and Schaye, 2009). We believe the accretion radius approach, frequently used in the star formation field (e.g., Bate et al., 1995), is essential for the problem at hand as it prevents unphysical accretion of SPH particles with too large an angular momentum, whereas the Bondi-Hoyle formulation does not.

The units of length, mass, time and velocity used in the simulations are, respectively,  $r_{\rm u} = 3.08 \times 10^{21}$  cm (1 kpc),  $M_{\rm u} = 2 \times 10^{43}$  g (10<sup>10</sup> M<sub>☉</sub>),  $t_{\rm u} = t_{\rm dyn}(r_{\rm u}) \approx 5$  Myr, and  $v_{\rm u} = (GM_a/a)^{1/2} = 208$  km s<sup>-1</sup>.

 $<sup>^{2}</sup>$ we note of course that, similarly to Chapter 4, this 'nailing-down' of the potential means that momentum is no longer conserved exactly; nonetheless this effect is unlikely to be significant as the accretion is dominated by infall from the outer radii of the shell over a single dynamical time

# 5.3 Initial conditions

The starting condition for all our simulations is that of a uniform density, spherically symmetric gaseous shell centered on the black hole. The initial parameters for each run are given in Table 5.1. The inner and outer radii of the shell are  $r_{\rm in} = 0.03$  and  $r_{\rm out} = 0.1$  kpc, respectively, for most of the simulations. In principle, one can expect the outer radius of the shell to be much larger in a realistic bulge, with an effective radius of a few kiloparsecs, but we are forced to limit the dynamic range of the simulations due to computational resources. Furthermore, we believe we understand how our results scale with the outer radius of the shell (see Section 5.6.2.4), and hence the dynamic range limitation does not actually influence our conclusions. All simulations were performed with 4,075,686 SPH particles.

The total mass of the shell is  $M_{\rm sh} = 5.1 \times 10^7 \,\rm M_{\odot}$ . To minimise initial inhomogeneities we cut the shell from a relaxed, glass-like configuration. The initial velocity field is composed of two parts: net rotation and a seeded turbulent spectrum. The net rotation is described by the azimuthal velocity component,  $v_{\phi} = v_{\rm rot}$ , where  $v_{\rm rot}$  is a parameter of the simulation (see Table 5.1). In all our runs,  $v_{\rm rot}$  is below the circular velocity at the shell radii, meaning that our initial conditions are not in equilibrium. We stress that this is deliberate, as our investigation here is the formation of a disc from infalling gas, rather than from already rotationally-supported gas.

The turbulent spectrum introduced into the velocity field is best interpreted as a kinetic energy 'kick' to the gas that is assumed to arise from the action of star formation feedback processes (i.e., stellar winds, supernovae). The amount of energy injection is varied between the runs but is tailored to the energy input from typical supernovae rates in starburst galaxies. One can perform a quick order of magnitude calculation to confirm this; taking our fiducial value of  $v_{turb} = 200 \text{ km s}^{-1}$  in a gaseous shell of  $\sim 10^8 \text{ M}_{\odot}$  the kinetic energy of the seeded turbulence is  $\sim 10^{55}$  ergs, over a lifetime of  $\sim 1 \text{ Myr}$  (approximately the time for which each simulation is run). Assuming the energy of a single supernova to be  $\sim 10^{51}$  ergs (Padmanabhan, 2001), this yields a SNe rate of  $\sim 0.01 \text{ yr}^{-1}$ . This value is a few orders of magnitude above the observed SNe rate in the Milky Way (see, e.g., Cappellaro, 2003) but is towards the lower end of rates found in starburst galaxies (see, e.g., Mattila and Meikle, 2001).

Table 5.1: Initial conditions for each simulation. The ratio with respect to the gravitational potential energy gives an indication of how well supported the shell is against the external potential through virial motions:  $(E_{\rm turb} + E_{\rm therm})/|E_{\rm grav}| \gtrsim 1/2$  for a virialised shell.  $r_{\rm circ}$  here refers to the initial value of circularisation radius for each of the particles, calculated in the region of the potential that they start in, based solely on the rotation velocity.  $M_{\rm acc}$  is the accreted mass taken at t = 0.2 in code units.

ID	$r_{\rm in}$	$r_{\rm out}$	$v_{\rm rot}$	$v_{\rm turb}$	$\frac{E_{\text{turb}} + E_{\text{therm}}}{\langle  E_{\text{grav}}  \rangle}$	$< r_{\rm circ} >$	$r_{\rm acc}$	$M_{\rm acc}({ m M}_{\odot})$		
S00	0.03	0.1	0.0	0.0	$6 \times 10^{-5}$	0.0	0.001	$5.01 \times 10^7$		
S01	0.03	0.1	0.0	0.1	0.001	0.0	0.001	$5.01  imes 10^7$		
S02	0.03	0.1	0.0	0.2	0.004	0.0	0.001	$4.90\times 10^7$		
S03	0.03	0.1	0.0	0.3	0.009	0.0	0.001	$4.81\times 10^7$		
S04	0.03	0.1	0.0	0.5	0.026	0.0	0.001	$4.49\times 10^7$		
S05	0.03	0.1	0.0	1.0	0.103	0.0	0.001	$3.53  imes 10^7$		
S06	0.03	0.1	0.0	1.5	0.231	0.0	0.001	$1.72  imes 10^7$		
S07	0.03	0.1	0.0	2.0	0.410	0.0	0.001	$8.40\times10^6$		
							ſ			
S10	0.03	0.1	0.1	0.0	$6 \times 10^{-5}$	0.004	0.001	$5.86\times10^6$		
S11	0.03	0.1	0.1	0.1	0.001	0.004	0.001	$8.01  imes 10^6$		
S12	0.03	0.1	0.1	0.2	0.004	0.004	0.001	$9.38\times10^{6}$		
S13	0.03	0.1	0.1	0.3	0.009	0.004	0.001	$1.23\times 10^7$		
S14	0.03	0.1	0.1	0.5	0.026	0.004	0.001	$1.03\times 10^7$		
S15	0.03	0.1	0.1	1.0	0.103	0.004	0.001	$5.29\times10^{6}$		
S16	0.03	0.1	0.1	1.5	0.231	0.004	0.001	$4.36\times 10^6$		
S17	0.03	0.1	0.1	2.0	0.410	0.004	0.001	$3.40  imes 10^6$		
			'		•		'			
S20	0.03	0.1	0.2	0.0	$6 \times 10^{-5}$	0.011	0.001	$2.67\times 10^4$		
S21	0.03	0.1	0.2	0.1	0.001	0.011	0.001	$4.16\times 10^5$		
S22	0.03	0.1	0.2	0.2	0.004	0.011	0.001	$2.04\times10^{6}$		
S23	0.03	0.1	0.2	0.3	0.009	0.011	0.001	$2.90\times 10^6$		
S24	0.03	0.1	0.2	0.5	0.026	0.011	0.001	$2.87\times 10^6$		
S25	0.03	0.1	0.2	1.0	0.103	0.011	0.001	$2.37\times 10^6$		
S26	0.03	0.1	0.2	1.5	0.231	0.011	0.001	$2.04  imes 10^6$		
S27	0.03	0.1	0.2	2.0	0.410	0.011	0.001	$1.58\times 10^6$		
					•					
S30	0.03	0.1	0.3	0.0	$6 \times 10^{-5}$	0.019	0.001	$9.35\times10^2$		
S31	0.03	0.1	0.3	0.1	0.001	0.019	0.001	$9.69  imes 10^3$		
Continued on next page										

ID	$r_{\rm in}$	$r_{\rm out}$	$v_{\rm rot}$	$v_{\rm turb}$	$\frac{E_{\rm turb} + E_{\rm therm}}{\langle  E_{\rm grav}  \rangle}$	$< r_{\rm circ} >$	$r_{\rm acc}$	$M_{\rm acc}({\rm M_\odot})$
S32	0.03	0.1	0.3	0.2	0.004	0.019	0.001	$3.67 \times 10^5$
S33	0.03	0.1	0.3	0.3	0.009	0.019	0.001	$1.12  imes 10^6$
S34	0.03	0.1	0.3	0.5	0.026	0.019	0.001	$1.63  imes 10^6$
S35	0.03	0.1	0.3	1.0	0.103	0.019	0.001	$1.40 \times 10^6$
S36	0.03	0.1	0.3	1.5	0.231	0.019	0.001	$1.25  imes 10^6$
S37	0.03	0.1	0.3	2.0	0.410	0.019	0.001	$1.48 \times 10^6$
S40	0.03	0.1	0.4	0.0	$6 \times 10^{-5}$	0.026	0.001	$6.21 \times 10^2$
S41	0.03	0.1	0.4	0.1	0.001	0.026	0.001	$3.15  imes 10^3$
S42	0.03	0.1	0.4	0.2	0.004	0.026	0.001	$1.78 \times 10^4$
S43	0.03	0.1	0.4	0.3	0.009	0.026	0.001	$3.21 \times 10^5$
S44	0.03	0.1	0.4	0.5	0.026	0.026	0.001	$1.20 \times 10^6$
S45	0.03	0.1	0.4	1.0	0.103	0.026	0.001	$9.93  imes 10^5$
S46	0.03	0.1	0.4	1.5	0.231	0.026	0.001	$8.98  imes 10^5$
S47	0.03	0.1	0.4	2.0	0.410	0.026	0.001	$9.47  imes 10^5$
S50	0.03	0.1	0.5	0.0	$6 \times 10^{-5}$	0.034	0.001	$4.08 \times 10^2$
S51	0.03	0.1	0.5	0.1	0.001	0.034	0.001	$1.41 \times 10^3$
S52	0.03	0.1	0.5	0.2	0.004	0.034	0.001	$6.10 \times 10^3$
S53	0.03	0.1	0.5	0.3	0.009	0.034	0.001	$3.78 \times 10^4$
S54	0.03	0.1	0.5	0.5	0.026	0.034	0.001	$5.63  imes 10^5$
S55	0.03	0.1	0.5	1.0	0.103	0.034	0.001	$8.75 \times 10^5$
S56	0.03	0.1	0.5	1.5	0.231	0.034	0.001	$6.58 \times 10^5$
S57	0.03	0.1	0.5	2.0	0.410	0.034	0.001	$6.89  imes 10^5$
S60	0.015	0.05	0.3	1.0	0.091	0.008	0.001	$4.43 \times 10^{6}$
S61	0.06	0.2	0.3	1.0	0.119	0.041	0.001	$5.23  imes 10^5$
		-		-	-			
S70	0.03	0.1	0.3	1.0	0.103	0.019	0.002	$2.36  imes 10^6$
S71	0.03	0.1	0.3	1.0	0.103	0.019	0.003	$3.45 \times 10^6$
S72	0.03	0.1	0.3	1.0	0.103	0.019	0.004	$4.77 \times 10^6$
S73	0.03	0.1	0.3	1.0	0.103	0.019	0.005	$6.07  imes 10^6$
		-		-				

Table 5.1 - continued from previous page
# 5.3.1 Turbulent velocity field

Our approach to setting up the turbulent velocity field follows that of Dubinski, Narayan & Phillips (1995). We assume a Kolmogorov power spectrum (Kolmogorov, 1941) that is based on statistically isotropic, incompressible turbulence. Such a spectrum has a steep power law dependence for the energy E in the turbulent wavenumber  $k \equiv |\mathbf{k}|$ , namely

$$E_v(k) \sim k^{-5/3}$$
 (5.2)

where the dominant wavelength in the velocity field v is the largest, *stirring* scale on which the turbulence is driven. We can obtain a power spectrum by noting that the energy in a particular wavenumber is the power integrated over a spherical surface in k-space, such that

$$E_v(k) = 4\pi k^2 P_v(k)$$
 (5.3)

The associated Kolmogorov power spectrum  $P_v(k) \sim |\boldsymbol{v}|^2$  is therefore

$$P_v(k) \sim k^{-11/3},$$
 (5.4)

The key assumption here is that the velocity field is homogeneous and incompressible, and so we can define v in terms of a vector potential A such that  $v = \nabla \times A$ . Zero divergence is a defining property of a vector field constructed from a vector potential, as

$$\nabla \cdot (\nabla \times \boldsymbol{A}) = 0 \tag{5.5}$$

by the rules of vector calculus. We construct the vector potential in Fourier space as a Gaussian random field; in other words, independant, statistical realisations of a multi-variate Gaussian distribution<sup>3</sup>. The amplitudes and components of  $A_{\mathbf{k}}$  at each point in k-space i.e., at each value of  $(k_x, k_y, k_z)$  are drawn from a Rayleigh distribution<sup>4</sup> with a variance given by  $\langle |A_k|^2 \rangle$  and a phase that is uniformly distributed between 0 and  $2\pi$ .  $A_{\mathbf{k}}$  is related to the velocity field by  $\mathbf{v} = \nabla \times \mathbf{A}$ , which in Fourier space becomes

$$\boldsymbol{v}_{\boldsymbol{k}} = i\boldsymbol{k} \times \boldsymbol{A}_{\boldsymbol{k}} \tag{5.8}$$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$
(5.6)

where  $\mu$  is the mean and  $\sigma$  is the standard deviation <sup>4</sup>the Rayleigh PDF is given by

$$f(x) = \frac{x}{\sigma^2} e^{-x^2/2\sigma^2}$$
(5.7)

and relates to a multivariate Gaussian in an absolute value sense; we require both a real amplitude X and a complex amplitude Y to describe the distribution, where each are normally distributed with a mean of zero. In this case  $|Z| = \sqrt{X^2 + Y^2}$  will have the form of a Rayleigh distribution

<sup>&</sup>lt;sup>3</sup>the probability density function (PDF) of a Gaussian, or *normal* distribution for an independant variable x follows

such that

$$\boldsymbol{v}_{\boldsymbol{k}}|^2 = -k^2 |\boldsymbol{A}_{\boldsymbol{k}}|^2 \tag{5.9}$$

and we therefore find that, for  $|v|^2 \sim k^{-11/3}$  we have

$$P_A(k) \sim k^{-17/3}.$$
 (5.10)

This is a steep power-law, implying that the variance in  $|\mathbf{A}|$  will diverge sharply as k decreases, and so we introduce a small scale cut-off  $k_{\min}$ . Equation 5.10 can then be written as

$$<|\mathbf{A}_{k}|^{2}>=C(k^{2}+k_{\min}^{2})^{-17/6},$$
(5.11)

where C is a constant that sets the normalisation of the velocities. For our purposes we set it equal to unity and normalise the velocity field once the statistical realisation has been generated. Physically, the small scale cut-off  $k_{\min}$  can be interpreted as the scale  $R_{\max} \simeq k_{\min}^{-1}$  i.e., the likely *driving* scale of the turbulence.

We therefore obtain the Fourier components of the velocity field by taking the curl of  $A_k$  in k-space as per equation 5.8. Finally we take the Fourier transform of this to obtain the velocity field in real space. We use a periodic cubic grid of dimension 256<sup>3</sup> when generating the statistical realisation of the velocity field and employ an interpolation scheme based on a cubic spline (see, e.g., Chapter 3, Section 3.2.1) in three dimensions to estimate the components of the velocity field at the position of each SPH particle.

# 5.4 Overview of gas dynamics and main results

Before we embark on a quantitative study of the simulation results, we present several snapshots from the simulations that illustrate graphically the nature of the gas flow. In particular, here we discuss the overall gas dynamics for two simulations that typify the extremes of behavior that we find - S30 and S35 (refer to Table 5.1). Both simulations have an initial rotation velocity  $v_{\rm rot} = 0.3$  which results in a mean circularisation radius of  $r_{\rm circ} = 0.019$  (see Section 5.5.1). Simulation S30 has no imposed turbulence. Simulation S35 has turbulence characterised by  $v_{\rm turb} = 1$ , where  $v_{\rm turb}$  is the mean turbulent velocity in the imposed distribution, implying that gas turbulent motions are of the order of the velocity dispersion in the bulge.

# 5.4.1 A shell with no initial turbulence

Our no turbulence initial condition acts as a base comparison for the simulations with seeded turbulence. Figure 5.1 shows the gas column density and the velocity field in the angle-slice projection for the simulation S30 at t = 0.06. For projection along the z-axis,

the gas column density is calculated by

$$\Sigma(x,y) = \int_{-z(x,y)}^{z(x,y)} \rho(x,y) dz , \qquad (5.12)$$

where the limits of the integration are given by  $z(x, y) = R \tan \zeta$ , and  $R = (x^2 + y^2)^{1/2}$ . The angle  $\zeta$  is chosen to be  $\tan \zeta = 1/5$  for Figures 5.1 and 5.3. This projection method is convenient as it permits an unobscured view into the central regions where the black hole resides.

Figure 5.1 shows the gas flow at an early time, both an edge-on (the left hand side panel of the figure) as well as a top-down projection of the shell (the right panel). Due to nonzero angular momentum, gas in the shell is quickly pushed aside from the axis of symmetry, opening a cylindrical cavity. Initially gas falls closer to the centre of the potential than its circularisation radius, a consequence of highly eccentric orbits. As gas moves inside  $r_{\rm circ}$ , centrifugal force exceeds gravitational force. Interaction with neighboring gas streams results in 'radial' shocks in the xy-plane. A further set of shocks occurs due to gas initially moving supersonically in the vertical direction. As the gas from above the z = 0plane collides with the gas falling in from below the plane, the particles are shocked and accumulate at z = 0 due to symmetry, forming a disc. These two sets of shocks mix gas with different angular momentum. We shall discuss this interesting effect in greater detail below.

The overall dynamics of the simulation are thus relatively simple: angular momentum conservation and symmetry dictate the formation of a geometrically thin disc in the plane of symmetry of the shell. Accretion of gas on to the SMBH would be expected to proceed in an accretion disc mode, if at all – realistic models show that star formation timescales are much shorter than viscous times, depleting the disc of gas before it can feed the SMBH (e.g., Goodman, 2003; Nayakshin and Cuadra, 2005; Nayakshin et al., 2007).

We note that our particular choice of velocity field, namely a constant azimuthal velocity, when combined with a spherical shell, could be viewed as a somewhat peculiar initial condition. As we have mentioned, the constant  $v_{\phi}$  condition means that the polar regions of the shell are quickly evacuated, the gas here spiralling outward in cylindrical radius to encounter gas falling in with different angular momenta, and mixing with the latter. In fact what we are modelling here is the simplest case of a flow that undergoes an *angular momentum re-distribution shock*. We shall go into more detail on the consequences of this in Section 5.5, but for now we make the point that such a flow is likely to occur when gas is accreting in a stochastic fashion from large scales, perhaps as the result of a merger. The strong mixing of the gas with different angular momenta is exactly the situation that we wish to explore here, and so we have implemented what is essentially a symmetric case of this mixing process. We acknowledge that the spherical setup is



Figure 5.1: Edge-on angle-slice projection (see Section 5.4.1) of the gas flow at time t = 0.06 in the simulation S30 (left) and top-down projection (right). The gas falls in on eccentric orbits, giving rise to a radial shock that propagates outwards as the disc forms.

idealised, but it is also the best starting point from which to embark on a laboratory of tests where the turbulence and rotation is varied.

# 5.4.2 A shell with initial turbulence

The flow of gas in this case is far more complicated than in the case just considered. We first show in Figure 5.2 the *full* slice projection of the simulation volume, e.g., we use here  $x(z, y) = r_{\text{out}} = 0.1$  (refer to equation 5.12). The most outstanding features of the figure are the long dense filaments that form due to convergent turbulent velocity flows. The density contrast between the filaments and surrounding gas is over two orders of magnitude by this time.

Figure 5.3 shows both the y-axis and z-axis projections of this simulation at the same time as Figure 5.1 did for the no initial turbulence case, namely t = 0.06. Clearly, some of the filaments seen at the earlier time in Figure 5.2 survive and actually penetrate into the innermost region There are now density contrasts of as much as three orders of magnitude in regions that were completely uniform in simulation S30. We shall see that this distinction is a crucial one for the dynamics of the gas and SMBH feeding.



Figure 5.2: Projected gas column density and velocity field in simulation S35 at time t = 0.02, before gas has had a chance to accrete on the SMBH. Note the formation of multiple thin and dense filaments due to convergent, turbulent velocity flows.



Figure 5.3: Edge-on angle-slice projection of the gas flow at time t = 0.06 in the simulation S35 (left) and top-down projection (right). The velocity field appears far more isotropic that in Figure 5.1, although an imprint of the imposed net rotation can still be seen.

# 5.4.3 Turbulence and accretion

Figure 5.4 shows the mass accreted by the black hole versus time in the simulations S30–S37, e.g., the same rotation velocity (and thus angular momentum) but different levels of initial turbulence. This demonstrates the main result of our study. The accretion rate on to the SMBH strongly correlates with the strength of the imposed turbulence. The accretion rate increases rapidly with increasing  $v_{\text{turb}}$  while  $v_{\text{turb}} \ll v_{\text{rot}}$ , but then saturates at an approximately constant level for  $v_{\text{turb}} \gtrsim v_{\text{rot}}$ . The main qualitative explanation of the simulations is that turbulence decreases the degree to which gas with different angular momentum mixes, and creates gas streams with small angular momentum. In particular:

- At low  $v_{turb} \ll v_{rot}$ , an increase in the turbulent velocity leads to greater variations in the density fluctuations that are created by the turbulent velocity flows before gas circularises (refer to Figure 5.2). Greater density contrasts decrease the amount of angular momentum mixing, resulting in a disc rather than a narrow ring. The inner edge of the disc lies closer to the accretion radius of the simulations and hence feeds the SMBH more efficiently.
- At  $v_{turb} \sim v_{rot}$ , random initial velocity fields set some gas on orbits with a vanishingly small angular momentum compared with the mean in the shell. The turbulent 'kick



Figure 5.4: Accretion rate versus time for simulations S30–S37. Linestyles correspond to  $v_{turb} = 0$  (black solid),  $v_{turb} = 0.1$  (black dotted),  $v_{turb} = 0.2$  (black dashed),  $v_{turb} = 0.3$  (black dot-dashed),  $v_{turb} = 0.5$  (brown dot-dot-dash),  $v_{turb} = 1.0$  (red dashed),  $v_{turb} = 1.5$  (pink dotted) and  $v_{turb} = 2.0$  (blue dashed) in code units. All runs used  $v_{rot} = 0.3$ . This plot exemplifies the main result of the chapter: the accretion rate on the black hole strongly increases with increasing levels of turbulence when rotation is present (see Section 5.4.3).

velocity' in this case almost cancels the mean rotation for these regions. Since these regions move against the mean flow, they are also those that get strongly compressed. Reaching high densities, they continue to move on nearly ballistic trajectories, impacting the innermost region on randomly oriented orbits. Accretion in this regime is 'chaotic' (King and Pringle, 2007) rather than large-scale disc-dominated.

We shall spend the rest of the chapter investigating these results in more depth, suggesting and testing analytical explanations for the observed trends.

# 5.5 Without turbulence: why 'laminar' accretion is so low

# 5.5.1 Analytical estimates based on circularisation of gas

We shall now argue that the accretion rate of the simulation S30 (the one with rotation velocity  $v_{\rm rot} = 0.3$  and no initial turbulence), is surprisingly low compared with a straightforward and seemingly natural theoretical estimate. Let us start by estimating the fraction of gas that should be accreted by the SMBH in our simulations. The specific angular momentum of gas determines how close to the SMBH it circularises. For a circular orbit, the component of specific angular momentum out of the plane is given by  $l = [GM(r)r]^{1/2}$ . Using equation 5.1, one can obtain a general solution for the circularisation radius,  $r_{\rm circ}$ , for a given value this specific angular momentum component. We give it in two extremes. If  $r_{\rm circ} \ll r_{\rm core}$ , the point-mass Keplerian value applies:

$$r_{\rm circ} = \frac{l^2}{GM_{\rm bh}} . \tag{5.13}$$

In the opposite case,  $r_{\rm circ} > r_{\rm core}$ , we have

$$r_{\rm circ} = \left[\frac{al^2}{GM_a} + \frac{M_{\rm bh}^2 a^2}{4M_a^2}\right]^{1/2} - \frac{M_{\rm bh}a}{2M_a} , \qquad (5.14)$$

which simplifies to  $r_{\rm circ} \approx l[a/GM_a]^{1/2}$  for  $r_{\rm circ} \gg r_{\rm core}$ , when the second and the last terms on the right hand side of the equation are small.

We now make the simplest possible assumption here by suggesting that gas settles into a disc and that the distribution of gas in the disc follows the distribution of gas over the circularisation radius initially, at time t = 0. Essentially, we assume that the  $l_x$  and the  $l_y$  components of the initial angular momentum cancel out due to symmetry whereas the  $l_z$  component is conserved without any exchange with neighbouring cells.

To estimate the fraction of gas that will end up inside the accretion radius  $r_{\rm acc} \ll r_{\rm core}$ , we first note it is equation 5.13 that should be used for the circularisation radius of gas with a given z-projection of specific angular momentum,  $l_z$ . The requirement  $l_z \leq l_{\rm acc} = (GM_{\rm bh}r_{\rm acc})^{1/2}$  singles out a cylinder with cross sectional radius of  $R_1 = \sqrt{x^2 + y^2} < l_{\rm acc}/v_{\rm rot}$ . The intersection of this cylinder with the shell  $r_{\rm in} \leq r \leq r_{\rm out}$  has approximate volume  $2\pi R_1^2(r_{\rm out} - r_{\rm in})$ . For the shell the fraction of the volume that can be accreted is then given by the ratio of this volume to the total volume of the shell,  $(4\pi/3)(r_{\rm out}^3 - r_{\rm in}^3) \sim (4\pi/3)r_{\rm out}^3$ :

$$f_{\rm acc} \sim \frac{3GM_{\rm bh}r_{\rm acc}(r_{\rm out} - r_{\rm in})}{2v_{\rm rot}^2 r_{\rm out}^3} \,.$$
 (5.15)

For example, for  $r_{\text{out}} = 0.1$ ,  $r_{\text{in}} = 0.03$  and  $v_{\text{rot}} = 0.1$  this gives  $f_{\text{acc}} = 0.1$ , and at  $v_{\text{rot}} = 0.3$ 

we have  $f_{\rm acc} \approx 10^{-2}$  (note that G = 1 in code units).

The latter analytical estimate yields accreted mass  $f_{\rm acc}M_{\rm shell} \approx 5 \times 10^5 \,{\rm M}_{\odot}$ , whereas Figure 5.4 shows that the actually measured value to late times is  $\sim 10^3 \,{\rm M}_{\odot}$ . The analytical estimate thus significantly over predicts the amount of accretion.

# 5.5.2 Shock mixing of gas: ring formation and end to accretion

Figure 5.5 explains why our simple analytical theory did not work. Here we plot the distribution of gas in simulation S30 over the z-component of the angular momentum vector of particles,  $l_z$ , at three different times. The initial distribution is spread over a broad range of values, with a small but non neglibible fraction of gas having  $l_z < l_{\rm acc} \approx 0.003$ , e.g., the angular momentum of the circular orbit at  $r_{\rm in}$ . In our analytical estimate we assumed that this gas accretes on to the black hole. However, this is not what happens. The distribution of angular momentum at the later time, t = 0.2, shows strong radial mixing of gas with different angular momentum. The angular momentum distribution narrows due to shocks and eventually becomes a highly peaked Gaussian-like ring.

Note that the deficit of gas in the innermost  $r \lesssim 0.01$  region in the second curve in Figure 5.5 is caused not by the SMBH accretion but by the shocks described above. Low angular momentum gas shocks and mixes with high angular momentum material before it has a chance to travel into the SMBH capture region,  $r \leq r_{\rm in}$ . The mixing continues to late times and the peak in the angular momentum distribution actually moves outwards. This is to be expected as the gas that fell in earlier has a smaller angular momentum and is in the path of eccentric orbits of gas falling in from greater distances.

We believe this strong mixing of different angular momentum orbits is quite a general result of initially 'laminar' flows. Such flows thus initially form rings rather than discs. While our simulations deliberately omit gas self-gravity and hence star formation (refer to Section 6.4), previous theoretical work shows that the fate of the material is then decided by whether the viscous time of the ring is shorter than the star formation consumption time scale. Most authors agree that large-scale gas discs form stars more readily than they accrete (Goodman, 2003; Nayakshin et al., 2007). This would suggest that 'laminar' shells with a finite angular momentum would form stars more readily than feed the SMBH.

# 5.6 Accretion with seeded turbulence: why is it efficient?

# 5.6.1 Analytical expectations

#### 5.6.1.1 Weak turbulence

First we consider the case of the turbulent velocity fields with  $v_{\text{turb}} \ll v_{\text{rot}}$ . Figure 5.5 shows the distribution of gas over the angular momentum for three different times in



Figure 5.5: Top: the distribution of gas in the simulation S30 over the z-component of the angular momentum at three different times: t = 0.0, t = 0.1, t = 0.2 for solid, long dashed and dashed curves respectively. Note how the angular momentum distribution becomes narrower with time (through the action of shocks). Middle: same as top but for simulation S31, with an initial turbulent velocity  $v_{turb} = 0.1$ . There is a greater fraction of small angular material in this simulation initially, and more of it gets retained in the 'tail' to small  $l_z$ , e.g., the inner disc, at late times. Bottom: same as top and middle but for simulation S32. The features noted in the middle plot are even more pronounced here.

the simulations S31 and S32, respectively. As in Figure 5.5 for the no initial turbulence run S30, the initial angular momentum distribution is broader than those at later times. The small angular momentum tail of the initial angular momentum distribution is more pronounced for S31 and especially for S32 compared with S30, which helps to explain the higher accretion rates measured in these simulations.

The main effect, however, is the reduction in the shock mixing of gas. Consider a 'small' angular momentum part of the distribution; to be definitive, at  $l_z \sim 0.01$ . This part of the distribution disappears at later times in simulation S30, being completely assimilated into the peak region. In contrast, in simulations S31 and S32 this part of the distribution is only reduced by a factor of 2-3 compared with the initial curves. Apparently, the initial turbulent velocity flows create density contrasts that then propagate to smaller scales. The dynamics of a dense region are different from that of a low density region, leading to a larger spread in the angular momentum distribution at later times. This increases the amount of gas captured by the inner boundary condition.

#### 5.6.1.2 Strong turbulence: ballistic accretion

Now we consider the case when  $v_{turb} \gtrsim v_{rot}$ . As we have seen, the turbulent accretion rate is far larger in our simulations than that for the 'initially laminar' runs. We have also seen that differential and chaotic velocity flows form strong density enhancements consistent with well known results from star formation studies (e.g., McKee and Ostriker, 2007b). High density regions could in principle propagate through the mean density gas without much hydrodynamical drag, as long as their column densities are much higher than that of the surrounding gas. In this case we can approximate gas motion as ballistic and use the usual loss cone argument (e.g., page 406 in Shapiro and Teukolsky, 1983) for black hole accretion.

For a direct comparison to the simulation results, we consider a thin  $\Delta r \ll r$  shell of gas that has angular momentum  $l \sim rv_{\rm rot}$ , and calculate the fraction of gaseous mass that has angular momentum small enough to be captured within  $r_{\rm acc}$ . In doing so we assume that the energy and angular momentum of gas are both conserved as the orbits are approximately ballistic. The initial specific gas energy, E, is small compared with the gravitational binding energy at  $r \sim r_{\rm acc}$ , therefore we can set  $E \approx 0$ . An orbit that just reaches our inner boundary condition at  $r_{\rm acc}$  has radial velocity  $v_{\rm r}(r_{\rm acc}) = 0$ . This yields the maximum specific angular momentum of gas that could still be accreted:

$$l_{\rm max}^2 = 2GM_{\rm bh}r_{\rm acc} \ . \tag{5.16}$$

The case of a shell that rotates slowly compared with the local circular speed, i.e.,  $v_{\rm rot} \ll v_c = [GM_{\rm enc}(r)/r]^{1/2}$ , is most interesting, since the shell could not rotate faster than the

circular speed or else the centrifugal forces would exceed gravity; similarly if it rotates at the circular speed then it is rotationally supported and would most likely form a disc rather than a spherical shell.

## Assuming a monochromatic distribution

As a starting point for this estimate, we approximate the distribution of gas velocities in the shell with the turbulent velocity spectrum by a monochromatic distribution  $v = v_{turb}$ randomly and isotropically distributed in directions in the frame rotating with rotation velocity  $v_{rot}$ . We emphasise here that this is a simple assumption, but a sensible one, as on average one would expect the turbulent velocity field to have an isotropic character. We relate  $v_{turb}$  to the mean velocity amplitude in the initial turbulent velocity distribution.

In the simplest case  $v_{\rm rot} = 0$ , the fraction of solid angle that yields angular momentum smaller than  $l_{\rm max}$  is approximately  $l_{\rm max}^2/4(rv_{\rm turb})^2$ . Therefore, the fraction of mass that could end up inside the accretion radius is

$$\frac{\Delta M}{M_{\rm sh}} = \frac{GM_{\rm bh}r_{\rm acc}}{2v_{\rm turb}^2r^2} , \qquad (5.17)$$

where  $M_{\rm sh}$  is the mass of the shell. However, this approach neglects the rotation of the shell. For a non-zero rotation velocity the loss-cone approach is valid only for  $v_{\rm turb} \ge v_{\rm rot}$ , and equation 5.17 must be modified to take account of the orbital motion. We therefore proceed to do so. Our derivation here is initially similar to Loeb (2004), but taken to second order when applying the small-angle approximation.

We start by considering the axis of zero angular momentum material for an isotropic, monochromatic distribution at a point in orbit of the central black hole. This is defined by

$$v_{\rm rot} - v_{\rm turb} \cos \alpha = 0 \tag{5.18}$$

where  $\alpha$  is complementary to the angle between this axis and r, as shown in Figure 5.6. We consider a small perturbation in angle about this axis, given by  $\delta$ , which defines the opening angle of the loss-cone:

$$v_{\rm rot} - v_{\rm turb} \cos(\alpha \pm \delta) = \frac{l_{\rm max}}{r}$$
 (5.19)

We expand this to second-order in the limit that  $\delta \ll \alpha$ , giving us

$$v_{\rm turb} \frac{\delta^2}{2} \cos \alpha + v_{\rm turb} \,\delta \sin \alpha = \frac{l_{\rm max}}{r} \tag{5.20}$$



Figure 5.6: Schematic geometry of an isotropic (in the rest frame of the rotating shell) wind, an analogy for the effect of turbulence on the loss-cone. The shell rotates with velocity  $v_{\rm rot}$  and the central axis of the loss-cone is defined in terms of the angle  $\alpha$ . A perturbation,  $\delta$ , on either side of this central axis defines the opening angle of the loss-cone.

where  $\cos \alpha = v_{\rm rot}/v_{\rm turb}$ . We therefore find for the loss-cone opening angle the expression:

$$\delta = \left(\frac{2l_{\max}}{rv_{\text{rot}}} + \frac{v_{\text{turb}}^2 \sin^2 \alpha}{v_{\text{rot}}^2}\right)^{1/2} - \frac{v_{\text{turb}} \sin \alpha}{v_{\text{rot}}}$$
(5.21)

For an isotropic distribution in the rest frame of the orbiting body, the fraction of the solid angle that will be accreted can be calculated via

$$\Omega_{\rm lc} = \frac{\Delta\Omega}{\Omega} \approx \frac{\pi\delta^2}{4\pi} \tag{5.22}$$

In this case, at a given  $v_{\rm rot}$ , our second-order approximation for the loss-cone solid angle is

$$\Omega_{\rm lc} = \left[ \left( \frac{l_{\rm max}}{2rv_{\rm rot}} + \xi \right)^{1/2} - \xi^{1/2} \right]^2 \tag{5.23}$$

where  $\xi = (v_{\text{turb}}^2 - v_{\text{rot}}^2)/4v_{\text{rot}}^2$ . We can express this in the limit of two extremes, depending on the relative strength of the turbulence with respect to the rotation:

i) the case when  $v_{\text{turb}} \gg v_{\text{rot}}$ :

$$\Omega_{\rm lc} \approx \left[ \left( \frac{l_{\rm max}}{2rv_{\rm rot}} + \frac{v_{\rm turb}^2}{4v_{\rm rot}^2} \right)^{1/2} - \frac{v_{\rm turb}}{2v_{\rm rot}} \right]^2 \tag{5.24}$$

ii) when  $v_{\text{turb}} \approx v_{\text{rot}}$ :

$$\Omega_{\rm lc} \approx \frac{l_{\rm max}}{2rv_{\rm rot}} \tag{5.25}$$

## Assuming a Maxwellian distribution

The treatment we have presented so far is not the whole picture, as we have assumed a monochromatic distribution for the turbulence. In reality this is not the case. The turbulent velocity distribution over all three normally distributed components has the form of a Maxwellian<sup>5</sup> profile (this is true for the initial distribution as well as that of the high density regions that form). We find that the distribution can therefore be fitted (to within ~ 10% error) by the expression:

$$f(v) = \frac{4}{v_{\rm turb}^3 \pi^{1/2}} v^2 e^{-(v/v_{\rm turb})^2}$$
(5.26)

<sup>&</sup>lt;sup>5</sup>since for normally distributed variables  $k_x, k_y, k_z$  the form of  $|\mathbf{k}| = \sqrt{k_x^2 + k_y^2 + k_z^2}$  is a Maxwell-Boltzmann distribution

To obtain the fraction of particles in the loss-cone  $(N_{\rm lc} \equiv \Delta N/N_{\rm tot})$  we therefore need to integrate this expression over the relevant solid angle:

$$N_{\rm lc} = \int_0^\infty \Omega_{\rm lc} f(v) dv \tag{5.27}$$

were we should use equation 5.23 for the loss-cone solid angle, but replacing  $\xi$  with  $\xi_v = (v^2 - v_{\rm rot}^2)/4v_{\rm rot}^2$ . Unfortunately, performing the integration in this case is non-trivial and must be done numerically. Luckily, we can obtain a good analytical approximation by noting that the modified equation 5.23 is a maximum when  $v = v_{\rm rot}$ , and falls off sharply with increasing v. The majority of the accretion will therefore occur in the region of the distribution where  $v \approx v_{\rm rot}$  and so we can use equation 5.25 in the integral instead, giving us:

$$N_{\rm lc} = \int_{v_{\rm rot} - \Delta v}^{v_{\rm rot} + \Delta v} \frac{2l_{\rm max}}{r v_{\rm rot} v_{\rm turb}^3 \pi^{1/2}} v^2 e^{-(v/v_{\rm turb})^2} dv$$
(5.28)

where our limits are a perturbation either side of  $v_{\rm rot}$ , the size of which we choose to be  $\Delta v = v_{\rm rot}/4$ <sup>6</sup>. Performing the integral yields

$$N_{\rm lc} = \frac{l_{\rm max}}{2rv_{\rm rot}} \cdot \psi + \frac{3l_{\rm max}}{4rv_{\rm turb}\pi^{1/2}} \cdot \varphi$$
(5.29)

where

$$\psi \equiv \operatorname{erf}\left(\frac{5v_{\text{rot}}}{4v_{\text{turb}}}\right) - \operatorname{erf}\left(\frac{3v_{\text{rot}}}{4v_{\text{turb}}}\right)$$
$$\varphi \equiv e^{-9v_{\text{rot}}^2/16v_{\text{turb}}^2} - e^{-25v_{\text{rot}}^2/16v_{\text{turb}}^2}$$

For a thick shell we should generalise equation 5.29 by using the density weighted average of r over the shell,  $\langle r \rangle$ . For an initially constant density profile used in our simulations and  $r_{\rm out} \approx 3r_{\rm in}$ , we have  $\langle r \rangle \approx 2r_{\rm out}/3$ , and hence

$$\frac{\Delta M}{M_{\rm shell}} = \frac{3(GM_{\rm bh}r_{\rm acc})^{1/2}}{2\sqrt{2}r_{\rm out}v_{\rm rot}} \cdot \psi + \frac{9(GM_{\rm bh}r_{\rm acc})^{1/2}}{4\sqrt{2}rv_{\rm turb}\pi^{1/2}} \cdot \varphi$$
(5.30)

As before, we can analyse this equation in various limits:

i) when  $v_{\text{turb}} \gg v_{\text{rot}}$ :

$$\frac{\Delta M}{M_{\rm shell}} \approx \frac{(2GM_{\rm bh}r_{\rm acc})^{1/2}}{r_{\rm out}v_{\rm turb}} \left(1 + \frac{v_{\rm rot}^2}{v_{\rm turb}^2}\right)$$
(5.31)

ii) when  $v_{\text{turb}} \approx v_{\text{rot}}$ :

$$\frac{\Delta M}{M_{\rm shell}} \approx \frac{(GM_{\rm bh}r_{\rm acc})^{1/2}}{2r_{\rm out}v_{\rm rot}}$$
(5.32)

 $<sup>^{6}\</sup>mathrm{it}$  should be noted that this choice is somewhat arbitrary but our derivation here is of course approximate

iii) when  $v_{\text{turb}} \ll v_{\text{rot}}$ :

$$\frac{\Delta M}{M_{\rm shell}} \approx \frac{9(GM_{\rm bh}r_{\rm acc})^{1/2}}{4\sqrt{2}r_{\rm out}v_{\rm turb}\pi^{1/2}} \cdot \varphi \tag{5.33}$$

where in all of the above we have dropped numerical factors of approximately unity.

# 5.6.2 Detailed analysis of results

Our ballistic approximation to the accretion rate yielded a result that depends on several parameters of the simulations, e.g., the 'accretion' radius  $r_{\rm acc}$ , the turbulent velocity  $v_{\rm turb}$ , the rotation velocity  $v_{\rm rot}$  and the dimensions of the shell. We shall now look at simulations covering a range in the space of these parameters to discuss specific trends and determine whether the simulations support our simple analytical theory.

## 5.6.2.1 No net angular momentum case

Figure 5.7 shows the accreted mass (left panel) and the accretion rate (right panel) as functions of time for simulations S00 – S07 (see Table 1). The rotation velocity for these runs is set to zero, and the turbulent velocity parameter  $v_t$  is varied from 0 (solid curve) to 2 (blue short dashed).

The situation is rather simple for these runs. For turbulent velocity parameters much smaller than unity, the flow hardly deviates from that of a freely falling shell with a negligible pressure. Due to the constant density profile in our initial shell, this naturally leads to the accretion rate increasing as  $\dot{M} \propto t^2$  (see the dotted power law in the right panel of Figure 5.7) starting from the free fall time of the inner shell to that of the outer shell. This can be understood by noting that

$$\dot{M}_{\rm gas}(r) \sim \frac{M_{\rm gas}(r)}{t_{\rm ff}} \tag{5.34}$$

where  $t_{\rm ff} = (3\pi/32G\rho)^{1/2}$  is the free-fall time, and  $\rho = \rho(r)$  here is the density profile of the background potential. In the outer parts this is isothermal, with  $\rho(r) \propto r^{-2}$ . We therefore have  $t_{\rm ff} \propto r$ , and using  $M_{\rm gas}(r) \propto r^3$ , we find  $\dot{M}(r) \propto r^2$ , and thus  $\dot{M}(t) \propto t^2$ .

As turbulence increases to  $v_{turb} \gtrsim 1$ , however, two effects are noteworthy. Firstly, the accretion episode starts earlier and also finishes later. This is naturally due to the spread in the initial radial velocities of the turbulent gas, with some regions starting to fall in with negative velocities, and hence arriving at the SMBH earlier, and others doing the opposite. There is also a significant reduction in the accretion rate at the highest values of turbulence. We believe this is due to the same effects already discussed in Section 5.4.3. Convergent turbulent flows create high density regions with a large range of angular momentum. These regions do not mix as readily as mean density regions and thus retain their angular momentum for longer. This reduces the amount of gas accreted. However



Figure 5.7: Accreted mass vs. time (left) and accretion rate vs. time (right) for a velocity field that contains no net rotation ( $v_{\rm rot} = 0$ ). The different lines represent different strengths of the mean turbulent velocity,  $v_{\rm turb}$ : 0 (black solid), 0.1 (black dotted), 0.2 (black dashed), 0.3 (black dot-dashed), 0.5 (brown dot-dot-dash), 1.0 (red dashed), 1.5 (pink dotted), and 2.0 (blue dashed). Analytical line (thick red dashed) displays a  $\dot{M} \propto t^2$  slope.

we note that in this rather unrealistic set up (zero net angular momentum) gas accretion is still very efficient, with  $\sim 20 - 100\%$  of the shell being accreted by time t = 0.2.

#### 5.6.2.2 The highest rotation velocity case

In the other extreme, Figure 5.8 shows the accreted mass (left panel) and the accretion rate (right panel) for the simulations S50 – S57, where the rotation velocity is set to the maximum explored in the paper, e.g.,  $v_{\rm rot} = 0.5$ . The range of the turbulent velocity parameter and the respective curve coding are the same as in Figure 5.7.

Comparing Figures 5.8 and 5.7, it is obvious that rotation significantly decreases the amount of gas accreted, and the accretion rate. The reduction is most severe in the case of  $v_{turb} = 0$ , by about 5 orders of magnitude. The role of turbulence in simulations S50 – S57, as suggested in Section 5.4.3, is to increase the accretion rate. This occurs by enhancement of the angular momentum distribution into the low angular momentum end, and also by weakening the efficiency of shock mixing of the gas.

# 5.6.2.3 Rotation and turbulence parameter space

We combine the results of all our numerical experiments from S00 to S57 in Table 5.1 by considering a single characteristic – the accreted gas mass at time t = 0.2 in code units. All of these runs have the same initial geometrical arrangement of gas and a fixed accretion radius  $r_{\rm acc} = 10^{-3}$  but differ in the strengths of the initial turbulent and rotation velocities.



Figure 5.8: Accreted mass vs. time (left) and accretion rate vs. time (right) for a velocity field that contains high net rotation ( $v_{rot} = 0.5$ ). Linestyles are as per Figure 5.7.

The results are displayed in two ways. Figure 5.9 shows the accreted mass versus rotation velocity,  $v_{\rm rot}$ . The simulations with the same level of mean turbulent velocity,  $v_{\rm turb}$ , are connected by the different lines. In particular, the solid black curve shows the results for  $v_{\rm turb} = 0$ , and the dashed blue curve shows the results for  $v_{\rm turb} = 2$ . It is clear that increased rotation velocity, and thus net angular momentum, always reduces the accreted mass, for any value of the turbulence parameter. This is in a general agreement with analytical expectations explored in Section 5.6.1.2.

Together with the simulation results, we plot the predicted maximum accreted mass (equation 5.32), which shows excellent agreement with the maximum delineated by the levels of turbulence at saturation. The analytical fit here is described specifically by

$$\frac{\Delta M}{M_{\rm shell}} = \min\left[\frac{(GM_{\rm bh}r_{\rm acc})^{1/2}}{2r_{\rm out}v_{\rm rot}}, 1\right]$$
(5.35)

since the maximum fraction that can be accreted is unity and for  $v_{\rm rot} = 0$  we would expect the entire shell to have accreted by late times.

The increased level of turbulence acts to decrease the accretion rate in the case of very small rotation velocity  $v_{\rm rot} \approx 0$  but increase the accretion rate at higher rotation velocities (albeit up to saturation). This is not contradictory at all, however - the effect of turbulence in all cases is to spread the initial angular momentum distribution and prevent it from rapidly mixing into a single peak. In the case of no initial rotation this reduces the accretion rate by moving gas from zero to finite angular momentum orbits, whereas in the case of 'large' initial net angular momentum the accretion rate is increased as some gas is moved to low angular momentum orbits.



Figure 5.9: The accreted mass trend with rotation velocity, plotted for different levels of turbulence at t = 0.2 (left, linestyles as per Figure 5.7) and the accreted mass trend with mean turbulent velocity, plotted for different levels of rotation -  $v_{\rm rot} = 0.0$  (solid), 0.1 (dotted), 0.2 (dashed), 0.3 (dot-dash), 0.4 (dot-dot-dash) and 0.5 (long dashed) - at t = 0.2 (right).

#### 5.6.2.4 Accretion and the size of the shell

Equation 5.30 predicts that the fraction of accreted mass should be inversely proportional to the shell's size, which we characterise by  $r_{\rm out}$ . To test this prediction we repeated the simulation S35 with a scaled-up version,  $r_{\rm out} = 0.2$ ,  $r_{\rm in} = 0.06$  and a scaled-down,  $r_{\rm out} = 0.05$ ,  $r_{\rm in} = 0.015$  shell (see runs S60 and S61 in Table 1). To be definitive, we compare the accreted gas mass at the dynamical time at the outer radius of the shell, which corresponds to just after the initial peak in the accretion rate. This is approximately the time at which we would expect the ballistic mode to come to an end and a disc-dominated accretion mode to begin. The results are plotted in Figure 5.10 (left panel) as a function of  $r_{\rm out}$ .

# 5.6.2.5 Trend with accretion radius

The accretion radius is a 'nuisance' parameter of our study in the sense that it is introduced only to allow the simulations to run in a reasonable time. For physical reasons it would be desirable to make  $r_{\rm acc}$  as small as possible. Therefore it is important to test whether the suggested analytical scaling of the results indeed holds. With this in mind, we repeated same simulation S35 for 5 different values of  $r_{\rm acc}$ , spanning a range in  $r_{\rm acc}$  from 0.001 to 0.005. The accreted mass from these simulations is shown in Figure 5.10 (right panel). The red dotted power-law gives the analytical prediction. Although similar, it is somewhat less steep than the simulation results.



Figure 5.10: Scaling of accreted mass with  $r_{out}$  (left) and  $r_{acc}$  (right) for simulations with  $v_{rot} = 0.3$ and  $v_{turb} = 1.0$ . Symbols correspond to the dynamical time at the outer radius of the shell i.e., t = 0.05, t = 0.1, t = 0.2 (left) and t = 0.1 (right). Error bars are somewhat arbitrary, as it is not clear when the main ballistic mode of accretion ends - here they denote  $\pm t/3$ . Normalised analytical fits are equation 5.30 (red), the actual fit that would be expected for the simulation parameters, and equation 5.32 (green), the maximum accretion expected at saturation.

#### 5.6.2.6 Visualisation of the circumnuclear disc

We now take a look at the gaseous disc that forms to later times in our simulations. Figure 5.11 compares the surface density projected along the z-axis for simulations S30 (left) and S35 (right) in the inner r = 0.04 region. Figure 5.12 shows the same for these two snapshots but projected along the y-axis. As was already clear from Figure 5.5, the gas forms a narrow ring in simulation S30. This occurs due to efficient mixing of material with different angular momentum. This mixing is both radial (the ring is narrower) and vertical (the ring is vertically less extended).

In contrast, in the simulation S35, turbulence creates a broad distribution of angular momentum which does not mix so well. Mixing in the vertical direction is actually reasonably efficient, as the disc is thin and lies close to the xy-plane (see the right panel in Figure 5.12). However, the radial mixing is not so strong and the disc radial structure is very extended in S35 compared with S30.

## 5.6.2.7 Radial structure of the disc

Our analytical theory also makes a prediction for the surface density of the disc in the innermost part. Indeed, the mass captured by the inner boundary of radius  $r_{\rm acc}$  scales as  $M_{\rm capt} \propto r_{\rm acc}^{1/2}$ . This mass would likely sit in a rotationally supported disc of roughly  $R \sim r_{\rm acc}$  size. The surface density in the disc should then behave as  $\Sigma \propto M_{\rm capt}(r)/R^2 \propto R^{-3/2}$ .

This analytical prediction (red power-law) is compared to the simulation results in



Figure 5.11: Face-on projection of the disc that forms by t = 0.418 with no turbulence (left) and mean  $v_{\text{turb}} = 1.0$  (right). The initial rotation velocity for both was  $v_{\text{rot}} = 0.3$ .



Figure 5.12: Slices of the CND (using  $\tan \zeta = 2/5$ ) formed with no turbulence and with mean  $v_{\text{turb}} = 1.0$ , both for  $v_{\text{rot}} = 0.3$  at t=0.418

Figure 5.13 which shows the surface density defined on radial shells for all the different levels of turbulence for  $v_{\rm rot} = 0.5$ . While the region beyound  $r \sim 0.01$  shows separate rings and highly variable surface density profiles for different simulations, the disc inside the region  $r \sim 0.01$  has a similar radial surface density profile for all simulations with  $v_{\rm turb} \gtrsim 0.5$ . The agreement between simulations and theory is quite reasonable in that region except for runs with a small turbulent velocity  $v_{\rm turb} < v_{\rm rot}$ . For these simulations the ballistic approximation is not appropriate and the gas does not penetrate into the r < 0.01 region. At radii larger than 0.01 the surface density distribution is dominated by the interactions with the bulk of the initial shell as it continues to fall in. The ballistic approximation is of a limited use here because gas settling in the disc to late times (i) went through many collisions/interactions and hence the orbits are not ballistic; and (ii) has a relatively large angular momentum, so the small angle approximation we made in Section 5.5.1 is violated.

The observed  $\Sigma \propto R^{-3/2}$  dependence in the inner disc is another piece of simple but convincing evidence that the ballistic accretion approximation that we made in Section 5.6.1.2 is a reasonable one for the innermost region of the flow.

#### 5.6.2.8 Vertical structure

Figure 5.12 shows edge-on projections of the disc for S30 (no turbulence) and S35 (initial mean turbulent velocity  $v_{turb} = 1$ ). It is clear that the vertical stucture of the disc is actually rather complex, especially beyond the  $r \sim 0.02$  region. This is intuitively to be expected as the dynamical time is longest at larger radii, and therefore circularisation and decay of vertical velocity dispersion in the disc should take longer here. We can further quantify this by assuming that the vertical velocity dispersion is pumped at the rate at which the matter is brought in. Generation of momentum due to turbulence is therefore be described by  $\dot{p}_{turb} \sim \dot{\Sigma} v_{\rm ff}(r_{\rm out})$ , where  $v_{\rm ff}$  is the free-fall velocity and we have assumed that the generation of turbulence is dominated by infall from the radius where the majority of the mass resides i.e.,  $r_{\rm out}$ . These motions decay on the local dynamical time. Assuming a steady-state we have

$$\dot{\Sigma}v_{\rm ff}(r_{\rm out}) = \frac{\Sigma\sigma_{\rm vert}}{t_{\rm dyn}} \tag{5.36}$$

where  $\sigma_{\text{vert}}$  is the vertical velocity dispersion in the disc, generated through turbulence. Re-arranging we have

$$\frac{\Sigma t_{\rm dyn}}{\Sigma} = \frac{\sigma_{\rm vert}}{v_{\rm ff}(r_{\rm out})} \tag{5.37}$$

We use again the fact that the change in the surface density of the disc is dominated by infall from  $r_{\text{out}}$ , putting

$$\dot{\Sigma} \sim \frac{\Sigma}{t_{\rm ff}(r_{\rm out})} \tag{5.38}$$



Figure 5.13: Azimuthally-averaged surface density versus radius for runs S50-S57 at time t = 0.418. Coloured curves correspond to simulations with different levels of mean turbulent velocity, as in Figure 5.7. The red line above the curves shows the theoretically expected (un-normalised) power law  $\Sigma \propto r^{-3/2}$ .

where  $t_{\rm ff}$  is the free-fall velocity. This gives

$$\sigma_{\rm vert} \sim v_{\rm ff}(r_{\rm out}) \frac{t_{\rm dyn}}{t_{\rm ff}(r_{\rm out})}$$
(5.39)

In the outer parts of the disc, the potential is isothermal, so that  $M(r) \propto r$ . Thus we have  $t_{\rm dyn} \propto r$ , and  $t_{\rm ff} \propto r$ . Writing  $v_{\rm ff}(r_{\rm out}) = v_{\rm circ}(r_{\rm out}) \equiv v_{\rm out}$  we see that  $\sigma_{\rm vert} \propto v_{\rm out} R/r_{\rm out}$ . Therefore, the expected disc aspect ratio in the outer parts is

$$\frac{H}{R} = \frac{\sigma_{\text{vert}}}{v_{\text{circ}}} \propto \frac{R}{r_{\text{out}}} \propto R \tag{5.40}$$

since in an isothermal potential,  $v_{\rm circ} = (GM(r)/r)^{1/2} \propto \text{const.}$  We have thus defined a dynamical, azimuthally-averaged disc scaleheight for the circumnuclear disc that forms by using the projection of the velocity dispersion along the 'vertical' direction for each annulus i.e., the mean angular momentum vector in each case. In cylindrical radius R we set,

$$H^2 = \frac{R^3 \sigma_{\text{vert}}^2(R)}{GM_{\text{enc}}(r)} \tag{5.41}$$

where  $M_{\text{enc}}(r)$  is the enclosed mass at r,  $\sigma_{\text{vert}}(R)$  is the 'vertical' velocity dispersion at R. Figure 5.14 plots ratios of H/R as functions of radius for simulations S50-S57. The solid line is for simulation S50, for which no initial turbulence is present. It is notable that the disc formed by the flow in this case is geometrically much thinner than in all the cases S51-S57. Among the turbulent cases, the general trend in H/R is exactly as predicted above. The large scatter is explained by a strong time dependence: the disc is being stirred by strong waves that would eventually act to circularise the disc.

In the region inside r = 0.01, however, the observed H/R dependence is flat while the analytical expectation has  $M(r) \propto \text{const.}$  and therefore  $H/R \propto R^{3/2}$ . We note that the vertical disc scaleheight is much larger than the SPH smoothing length, implying that the result inside r = 0.01 is not due to under resolving that region. We believe that the larger than expected aspect ratio is due to dynamical heating of the inner disc by the waves already noted at larger radii. Indeed, the inner disc is far less massive than the outer one, and any waves present in the outer region have to either dissipate or get reflected in this region.

# 5.7 Accretion inside the capture radius

While we cannot resolve the hydrodynamics of gas within the accretion radius,  $r_{\rm acc}$  (1 parsec for most of the simulations presented here), we *can* track both the mass and the angular momentum of the particles that are captured. We can therefore calculate the mean angular momentum of the gas that settled inside  $r_{\rm acc}$ . If the disc viscous time is comparable



Figure 5.14: Dependence of vertical disc thickness, H/R, versus radius, for all levels of turbulence, with  $v_{\rm rot} = 0.5$  (S50-S57), at time t=0.418. Line styles of the curves are the same as in Figure 5.7. The red line above the curves shows the theoretically expected result, for the (un-normalised) slope  $H/R \propto R$ .



Figure 5.15: Plots of the accreted angular momentum direction in the Aitoff projection. Both plots are with  $v_{\rm rot} = 0.3$ , with no turbulence (left) and  $v_{\rm turb} = 2.0$  (right). Crosses mark every 0.001 in code time units. Note that the title of "SMBH spin direction" refers to the spin orientation that the SMBH would have if it had accreted all of the gas within the capture radius; in reality this is of course not the case.

or longer than the duration of our simulations, which is likely (e.g., Nayakshin and Cuadra, 2005), then the direction of the mean angular momentum of the gas captured within  $r_{\rm acc}$  region is related to that of the sub-parsec disc. Note that in principle the disc can be warped and then its structure is more complex (Lodato et al 2009).

Figure 5.15 shows the evolution of the direction of the mean specific angular momentum for the gas particles accreted inside  $r_{\rm acc}$  in the simulation S30 (left panel, no initial turbulence) and S37 (right panel,  $v_{\rm turb} = 2$ ). The latter simulation is initialised with the highest levels of turbulence we considered in this paper, and therefore the effects of disc plane rotation are the strongest. The 'north pole' in the diagram corresponds to a specific angular momentum vector oriented along the z-axis. Clearly, there is negligible deviation from that direction for the 'laminar' simulation S30, as expected. For the simulation S37, however, there is a complicated time evolution of the innermost disc direction. At one point the disc appears to be completely counter-aligned to the mean angular momentum of the shell.

The accretion disc 'plane wandering' at  $r \lesssim r_{\rm acc}$  should also be seen in the larger scale flow of gas. Figure 5.16 shows the innermost r = 0.005 region for the simulation S37 at time t = 0.072. We can clearly see here that the disc in the immediate vicinity of  $r_{\rm acc}$  is also strongly displaced in terms of direction from the nominally expected one. One should also note the flows of gas along directions roughly orthogonal to the disc plane. These flows must have very different orientations of angular momentum, and it is these that cause the plane of the 'sub-grid disc' to execute the non-trivial path seen in Figure 5.15.



Figure 5.16: Projected gas density and velocity field for simulation S37, the innermost r = 0.005 region, for t = 0.072. Note how strongly the disc is misaligned with the xy-plane. At later times the misalignment decreases as more material accumulates into the disc.

# 5.8 Discussion

# 5.8.1 Summary of the results

We have studied the hydrodynamics of an initially uniform rotating gaseous shell, seeded with an initial turbulent velocity spectrum as it falls into the inner part of a spherically symmetric static potential. Particular attention was paid to the amount of gas that made it inside the accretion radius,  $r_{\rm acc}$ .

In the control runs without initial turbulence, gas settles into a nearly circular rotationally supported ring (see Figures 5.5, 5.11, 5.12). This was somewhat of a surprise. Intuitively we expected a wide  $(\Delta R \gtrsim R)$  disc, as the initial distribution of angular momentum is broad (Figure 5.5). Furthermore, only a few hundred solar masses of gas were accreted, compared with the  $\sim 5 \times 10^5 \,\mathrm{M_{\odot}}$  expected on the basis of a simple circularisation radius argument (see Section 5.5.1). In hindsight, both of these effects are caused by the effective mixing of gas by the shocks. Low angular momentum gas mixes with high angular momentum gas before the former reaches the capture radius  $r = r_{\rm acc}$ . The inner edge of the ring is therefore a significant distance from the accretion radius. In the context of AGN accretion, such a ring would likely be strongly gravitationally unstable and form stars rather than feed the SMBH (e.g., Goodman, 2003; Nayakshin et al., 2007). In contrast, runs in which the initial shell was seeded with turbulence show qualitatively and quantitatively different results. The most significant of these is that the accretion rates are higher by up to 3-4 orders of magnitude compared with the simulations that had no initial turbulence imposed. This is particularly striking since the net angular momentum of these simulations is the same.

It is obvious that turbulence broadens the angular momentum distribution, setting some gas on low angular momentum orbits. A less trivial conclusion from our numerical results is that turbulence also prevents the efficient gas mixing pointed out above. Physically, high density regions are created by supersonic turbulent convergent flows. These regions can then travel nearly ballistically through the average density gas. Some of this gas is able to reach the accretion radius, increasing the SMBH accretion rate compared with runs without turbulence. This parameter space trend lasts until the initial mean turbulent velocity becomes greater than the net rotation velocity, at which point the accretion trend saturates and begins to slowly decrease as turbulence is increased further.

The development of high density regions in supersonic turbulence is by now a classical result from a number of papers (for reviews see e.g., Elmegreen and Scalo, 2004; McKee and Ostriker, 2007b). There is also a body of work that highlights the complicated dynamics of shocks in the multi-phase interstellar medium. For example, Bonnell et al. (2006); Dobbs (2008) suggest that shocks in a multi-phase medium generate (or rather preserve) differential motions, allowing the high density regions to slip through the shocks.

These effects are similar in nature to the ballistic behavior of high density clumps found in our simulations. Schartmann et al. (2009) simulated young nuclear star clusters injecting the energy into the surrounding gas via supernovae and winds. As in our study it was found that high density streams form. Those with small angular momentum feed the small scale 'torus'. These results, and earlier results by Cuadra et al. (2006) for stellar winds in the central parsec of the Milky Way also support the viability of a 'ballistic' AGN feeding mode.

We have attempted to build a simple analytical theory that would predict the accretion rate for our capture inner boundary condition. We first calculated the accretion rate in the ballistic approximation assuming that gas in the shell receives monochromatic velocity kicks  $v = v_{turb}$  in random directions in the frame rotating with velocity  $v_{rot}$ . This 'monochromatic' theory predicts that gas capture rate (fraction) is the highest at  $v_{turb} \approx v_{rot}$ . We then took into account in an approximate way the corrections associated with the fact that the real turbulent velocity distribution function is not monochromatic. The fraction of gas accreted within  $r_{acc}$  calculated in this way turned out to describe our numerical results well. Furthermore, the theory predicted that the inner disc surface density should have a power-law scaling:  $\Sigma(R) \propto R^{-3/2}$ , which is indeed born out by the simulations.

The approximate agreement of theory and simulations in this paper implies that we can use our results for systems of astrophysical interest by an appropriate rescaling of parameters, which we attempt to do in the next section.

# 5.8.2 Can quasars and AGN be fed by ballistic accretion?

The most serious challenge to theories of quasar and AGN feeding is gravitational instability of these discs and the ensuing fragmentation of discs into stars (e.g., Paczyński, 1978; Kolykhalov and Sunyaev, 1980; Collin and Zahn, 1999; Goodman, 2003; Nayakshin et al., 2007). Observations (e.g., Paumard et al., 2006; Bartko et al., 2009) and numerical simulations (Bonnell and Rice, 2008; Hobbs and Nayakshin, 2009) suggests that this process operated in the central 0.5 pc of the Milky Way, creating young stars and probably leaving the resident SMBH – Sgr A<sup>\*</sup> – without gaseous fuel to shine now (e.g., Levin and Beloborodov, 2003; Nayakshin and Cuadra, 2005).

One obvious way in which the self-gravity 'catastrophe' of AGN discs can be avoided is to deposit the gas at small enough radii (e.g., King and Pringle, 2007). Namely, one can show that gaseous discs are hot enough to avoid gravitational fragmentation within the self-gravity radius,  $R_{sg}$ , which is of the order of 0.01 - 0.1 pc depending on the SMBH mass and some of the disc parameters (see Goodman, 2003). Inclusion of feedback from star formation allows the disc to survive and yet maintain an interesting accretion rate on the SMBH out to a few times larger radii (Pedes and Nayakshin, in preparation). We shall now try to estimate the plausibility of such a small radius gas deposition in the context of the 'ballistic' gas deposition model we have developed. To this end we shall use the approximate scalings for the fraction of gas accreted from a rotating turbulent shell derived in this paper in Section 5.6.1.2. The size of the shell,  $r_{out}$ , enters this estimate directly. Existence of the  $M_{bh} - M_{bulge}$  relation implies (e.g., Häring and Rix, 2004) that the SMBH must be fed by gas located roughly the size of the bulge away from the SMBH (King, 2003, 2005), or else it is difficult to see why the quoted feedback link must exist (Nayakshin and Power, 2009). We estimate that the effective radius of the bulge,  $r_b$ , is about 2.5 kpc for a velocity dispersion  $\sigma = 150$  km sec<sup>-1</sup> (see Nayakshin et al., 2009). If the rotation velocity at  $r_b$  is a small fraction of  $\sigma$ , say  $v_{rot} = 50$  km s<sup>-1</sup>, then the fraction of the gas that can be accreted from a turbulent shell is

$$\delta_{\max} \approx 0.001 \, GM_8^{1/2} \left(\frac{r_{\rm acc}}{0.2 \, {\rm pc}}\right)^{1/2} \left(\frac{2.5 \, {\rm kpc}}{r_{\rm out}}\right) \left(\frac{50 \, {\rm km \, s^{-1}}}{v_{\rm rot}}\right)$$
(5.42)

where  $M_8 = M_{\rm bh}/10^8 \,\mathrm{M_{\odot}}$ . Here we have used equation 5.32, assuming that the mean turbulent velocity is of the order of the rotation velocity, which is our maximum in the accretion trend (although it should be noted that going to stronger turbulence does not decrease the accreted mass by much - see Section 5.6.2). The  $M_{\rm bh} - M_{\rm bulge}$  relation here gives the stellar mass in the bulge as ~ 500 $M_{\rm bh}$ , and if we assume that ~ a half of the bulge gas was used up in star formation (McGaugh et al., 2009), we find that the action of supernova feedback in the context of our model yields an accreted mass of ~ 5 × 10<sup>7</sup> M<sub>☉</sub>. This accretion would take place over roughly the dynamical time at the outer edge of the bulge, giving us an accretion rate ~ a few M<sub>☉</sub> yr<sup>-1</sup>. In this simple estimate, then, the SMBH is indeed able to grow as massive as the observations require, assuming that our ballistic accretion mode can be sustained for the required time. Future work should test these ideas more directly, with 'turbulence' driven directly by feedback from star formation in the bulge, where the bulge gas distribution has been derived from initial conditions in the gas on larger scales, e.g., the virial radius of the host dark matter halo.

# 5.8.3 A positive feedback link of star formation to AGN activity?

Observations show that starburst and AGN activity are correlated in a number of ways (e.g., Cid Fernandes et al., 2001; González Delgado et al., 2001; Farrah et al., 2003). One might think that this is entirely natural, as both phenomena require a source of gas to be present. However, star formation on kiloparsec scales does not have to be casually connected with SMBH activity. As an example of this, consider a case when the angular momentum of the gas is large, so that a kiloparsec-scale disc is formed. If star formation in this disc simply consumes the gas then there is nothing left to feed the SMBH. This is what may have happened on smaller scales in the central parsec of our Galaxy  $\sim 6$  million years

ago (Nayakshin and Cuadra, 2005; Nayakshin et al., 2007). Naturally, if such a process can rob the SMBH of fuel at parsec scales then the situation becomes even worse once we reach a kiloparsec.

The opposite situation is also possible because the SMBH mass is a small fraction of the bulge mass. It is not implausible to have a sub-parsec scale, non self-gravitating accretion disc that could feed a moderately bright AGN for a long  $(10^6 - 10^7 \text{ years}, \text{ say})$  time. The amount of gas involved in this could be  $10^5 - 10^7 \text{ M}_{\odot}$ , i.e., miniscule by galaxy standards. Why such activity should always be connected with a powerful starburst is not clear.

However, the inclusion of star formation feedback could alter this picture. For example, within a kiloparsec-scale disc, feedback from stellar winds and supernovae gives velocity kicks to the surrounding gas. For a standard IMF, the total momentum in supernova shells and winds from massive stars, integrated over their lifetime, is about  $\epsilon_* M_* c$ , where  $\epsilon \approx 10^{-3}$ , and  $M_*$  is the total mass of the stellar population (Leitherer et al., 1992; Thompson et al., 2005). If this momentum is absorbed by gas with mass  $M_g$  then the average kick velocity for the gas is  $v_{\text{kick}} = (M_*/M_g)\epsilon c$ , or  $\sim 300 \text{ km s}^{-1}$  for  $M_* \sim M_g$ . There is thus certainly enough momentum input to make the gas highly turbulent. The cancellation of oppositely directed momentum results in a reduction in this estimate. However, we have discussed here only the momentum input; including supernova kinetic energy could act to increase the above value (e.g., Dekel and Silk, 1986).

As we have shown in this Chapter, the effect of such feedback is likely to result in some small angular momentum clouds or streams that may feed the SMBH. A important feature of our model is therefore a correlation of starburst activity with SMBH activity, with the latter being somewhat offset in time more often than not (it is also possible to activate an AGN earlier if there is some low angular momentum gas in the initial shell). Further detailed physics-based modeling of AGN feeding, feedback, and star formation, as well as comparisons to observations should shed light on the mode of quasar feeding, and also on the mode of bulge formation.

# 5.8.4 Cosmological cold stream SMBH feeding?

We have employed turbulent initial conditions as a simple and mathematically convenient way to introduce strong disordered differential motions in the gas for controlled numerical experiments in AGN feeding. We expect realistic galaxies to be always in the 'turbulent' regime in the above sense because the gas contracting on to a dark matter halo is unlikely to be of uniform density or possess a uniform velocity field. In fact, large-scale numerical simulations of galaxy formation show cold gas streams that penetrate the shock-heated media of massive dark matter haloes (Dekel et al., 2009). This is physically similar to the ballistic motion of gas we observed in our simulations. While it is very unlikely that these very large-scale streams ( $\sim$  hundreds of kiloparsecs) would be sufficiently well aimed to actually feed the SMBH, collisions of several of such streams in the centre of a galaxy would pump very strong differential motions and lead to some gas being set on small angular momentum orbits. We therefore speculate that cold large-scale gas streams may promote AGN feeding by providing energy input to turbulence/differential motions on galactic scales.

# 5.8.5 Convergence tests

A great deal of the behaviour that we find in our simulations is the result of turbulent motions creating dense, shocked gas. It is natural to conclude, therefore, the the artificial viscosity parameter,  $\alpha$ , is of particular importance here. In order to demonstrate that numerical effects associated with  $\alpha$  are negligible, we present convergence tests on this parameter. The result can be seen for the accreted mass in Figure 5.17 for fiducial values of mean turbulent velocity,  $v_{turb} = 0.0$  and  $v_{turb} = 1.0$ , and rotation velocity,  $v_{rot} = 0.3$ . It is clear that the accretion we find in our simulations is reliable and not subject to numerical viscous effects, particularly at high turbulence. At zero turbulence there are small deviations, but we note that at such low accretion rates it is almost possible to see the signature of individual particles on the plot, and so even minor differences will tend to appear significant.

# 5.9 Conclusions

The classical challenge to AGN feeding is the large angular momentum of gas expected to result in circularised discs of  $\sim$  kpc scales (e.g., Shlosman, Frank & Begelman 1989). Transporting gas from these scales is extremely difficult due to long viscous times and the rapid consumption of gas in star formation (e.g., Goodman, 2003; Nayakshin et al., 2007). Here we pointed out that turbulence in the bulge, e.g., driven by supernova explosions, may be a way to overcome the AGN feeding debacle by setting some gas on ballistic orbits. To test these ideas, we have performed simulations of hydrodynamical gas flow with angular momentum, initially distributed in a spherical shell, in the static spherical potential of a bulge with a central SMBH. The duration of our simulations is a few dynamical times of the shell, allowing us to study the formation of an accretion disc and the capture of gas within a small inner boundary region. For all values of the rotation velocity, the shell was not in equilibrium with the background potential. This was done deliberately, as the goal of this project was to study the formation of a disc from an infalling gas distribution from first principles.

We found that angular momentum mixing in the runs without initial turbulence is very effective, resulting in an initial ring rather than an extended disc. Further evolution of



Figure 5.17: Convergence tests on the accreted mass through a boundary at r = 0.003, for the value of the artificial viscosity parameter,  $\alpha$ . The accretion was tracked at 3× the SMBH accretion radius to avoid numerical artifacts in the estimated accretion due to the inner vaccuum boundary condition. Solid linestyles correspond to zero turbulence, while dashed linestyles refer to an initial mean turbulent velocity of  $v_{turb} = 1.0$  in code units. Values are:  $\alpha = 1.0$  (black),  $\alpha = 0.7$  (red),  $\alpha = 0.5$  (magenta),  $\alpha = 0.2$ (blue),  $\alpha = 0.1$  (green). The amount of gas accreted by the SMBH shows excellent convergence at high turbulence, and at zero turbulence shows good convergence by the time the fiducial value (used for the rest of the simulations) of  $\alpha = 1.0$  is reached.

the ring depends on whether it can transfer angular momentum quickly enough and avoid completely collapsing into stars. If star formation consumes the ring quickly then such rings are a dead end as far as SMBH feeding is concerned. Turbulent motions in the shell overcome the angular momentum mixing problem by creating high density regions that can travel nearly ballistically, retaining their initial angular momentum. Provided there is enough gas with small angular momentum the SMBH can be fed much more efficiently than in the 'laminar' regime. The main conclusion here is that the star formation in the bulge may not simply deplete the gas (hence depriving the SMBH of its fuel) but may actually promote SMBH accretion by creating turbulent flows.

We note that turbulence in this particular project is simply a mathematically convenient formalism to introduce strong disordered differential motions in the gas. We expect that realistic galaxies are always in the turbulent regime because the state of the gas on large scales is unlikely to be that of uniform density and angular momentum. Further work with self-consistently driven turbulence from star formation and realistic initial (large-scale) boundary conditions is needed to establish the relevance of the 'ballistic' accretion mode to feeding the SMBH, although our research suggests that it is indeed a promising model.

There is a wider context to the work presented in this Chapter, beyond the specifics of SMBH feeding; namely, the modelling of an 'intermediate scale' between the standard cosmological simulations and the small-scale sub-parsec disc simulations such as we presented in Chapter 4. The presence of the strong correlations between the bulge and SMBH properties (refer to Chapter 1, Section 1.8) makes it a priority that we study the connection between the formation of a galaxy from large scales and the evolution of the galaxy and growth of the SMBH. The intermediate scale is likely to be extremely important in regulating this evolution, as it will contain large amounts of star-forming gas that can drive energy and momentum to smaller (and larger) scales. Furthermore, it is the only way that one can hope to develop a physically self-consistent model of AGN feeding and growth that can be employed in a cosmological simulation. In the next Chapter we discuss why this is needed; what is currently the state-of-the-art is somewhat lacking in this area and it is vital that it be improved. Cosmological simulations are the driving force behind our understanding of structure formation, galaxy evolution, and the expansion of the Universe since its inception, as well as one of the best ways to constrain the nature of dark matter short of resorting to particle physics. Their reliability is therefore paramount to our field.

# 6 On the modelling of accretion on to SMBHs

"The trouble with having an open mind, of course, is that people will insist on coming along and trying to put things in it."

Terry Pratchett

# 6.1 Introduction

Over the last decade, compelling observational evidence has revealed that many galaxies in the local Universe harbour SMBHs with masses  $10^6 \lesssim M_{\rm bh}/M_{\odot} \lesssim 10^9$  in their centres. During the same period, surveys of the distant Universe have uncovered the existence of quasars at  $z \sim 6$ , when the Universe was less than a  $1/10^{\rm th}$  of its current age; this implies that the component SMBHs had already assembled their mass by this time (see, e.g., Fan et al., 2006).

Our understanding of the physics that dictates the growth of SMBHs is incomplete. Black holes grow by accreting low angular momentum material from their surroundings, yet the character of the accretion flow on to an SMBH is governed by physical processes as diverse as galaxy mergers (e.g., Hopkins and Quataert, 2009), turbulence induced by stellar feedback (e.g., Chapter 5) and black hole accretion-driven outflows (e.g., Nayakshin and Power, 2010). Black hole growth is routinely modelled in galaxy formation simulations (see, e.g., Springel et al., 2005) and the importance of SMBHs in shaping the properties of galaxies is now well established (see, e.g., Croton et al., 2006; Bower et al., 2006). The majority of galaxy formation simulations published in the literature incorporate what we term the 'Bondi-Hoyle model' for black hole growth (see, e.g., Springel et al., 2005), which derives from the work of Bondi and Hoyle (1944) and Bondi (1952). This model assumes the simplest possible accretion flow, where the gas is at rest at infinity and accretes steadily on to a black hole, subject only to the (Newtonian) gravity of the latter, which is modelled as a point mass (refer to Chapter 2, Section 2.3.2).

Simulations that model the idealised physical problem as it is set out in Bondi and Hoyle (1944) and Bondi (1952) produce results that are in good agreement with the analytical solution (Ruffert, 1994). In galaxy formation simulations, gas is expected to have non-zero angular momentum that provides a natural barrier to accretion on to the SMBH (Power et al., 2010). Gas settles into a disc whose dimensions are set by the angular momentum of the accretion flow, and only the material with the very lowest specific angular momentum can accrete, as the viscous timescale for gas to be transported through the disc may already be comparable to the age of the Universe at  $r \sim 1$  pc (refer to Chapter 2, Section 2.3.4.6).

In this Chapter we suspend, for the moment, our disbelief that gaseous infall can proceed entirely radially from large scales. We consider spherically symmetric accretion flows (i.e., zero angular momentum) on to an SMBH embedded in the potential of a massive dark matter halo. This is an example of a situation where we might *expect* the Bondi (1952) formula to provide a reasonable estimate of the accretion rate on to the SMBH. We argue that, in fact, the Bondi (1952) formula, designed for accretion on to isolated black holes, can only be applied to accretion on to astrophysical SMBHs, i.e., those embedded in massive dark matter halos, if the gas in the latter is in or near hydrostatic equilibrium.
It is only in this case that one of the key assumptions of Bondi (1952) – the gas being at rest at infinity – is satisfied ('infinity' in this case meaning outside the Bondi radius - see Section 6.2.1). This is not what is expected in gas-rich epochs when SMBHs can grow efficiently. High density of gas implies efficient gas cooling; when the temperature of the gas falls below the virial one for the halo the gas cannot be in hydrostatic balance, as the halo potential must impose an inflow of gas on distances much larger than the Bondi radius. The accretion of gas on to the SMBH in this case is influenced by the properties of the host.

To demonstrate the validity of this point, we design and perform idealised spherically symmetric simulations in which gas is assumed to be isothermal for simplicity. We find that the Bondi-Hoyle accretion rate estimate,  $\dot{M}_{\rm Bondi}$ , can be orders of magnitude off the true  $\dot{M}_{\rm bh}$ . Worse yet, the factor by which it is off is a strong function of the temperature. At the same time, a *free-fall* accretion rate estimate comes out as a fairly accurate alternative for these tests. Therefore, we suggest that a better approach would be to employ an interpolative formula between the Bondi (1952) solution and the free-fall estimate.

The layout of this Chapter is as follows. In Section 6.2 we show analytically that in large-scale simulations of cosmological volumes the Bondi-Hoyle approach is invalid, and in Section 6.3 we present some numerical tests of this hypothesis. Finally in Section 6.4 we discuss our conclusions.

### 6.2 Analytical arguments

### 6.2.1 Classical Bondi-Hoyle accretion

We first recap the main assumptions underpinning the classical Bondi and Hoyle (1944); Bondi (1952) papers. These are nicely summarised in the first sentence of Bondi (1952)'s abstract: "The special accretion problem is investigated in which the motion is steady and spherically symmetrical, the gas *being at rest at infinity*". We have italicised the part of the sentence that bears the most importance for us here. While the presence or absence of vorticity and/or angular momentum in the gas has been pointed out by a long list of authors as a key determinant of the accretion process (e.g., Krumholz et al., 2005), the importance of the 'being at rest at infinity' condition is the key focus of this Chapter, however, and for the time being we accept<sup>1</sup> the assumptions of zero angular momentum and spherically symmetric infall.

Physically, gas can be at rest at infinity only when it is not subject to any forces. The only external force acting on the gas in the restricted Bondi and Hoyle (1944); Bondi (1952) problem, i.e., the gravitational force, is due to the black hole. Self-gravity of the

 $<sup>^{1}</sup>$  somewhat warily, and recognising that this entirely idealised picture should be viewed as *instructive* rather than realistic

gas is neglected. The 'infinity' is a region at a large enough distance that the gravitational potential of the SMBH is negligible with the internal energy of the gas. This is quantified by defining the Bondi (or the *capture*) radius,

$$r_B = \frac{2GM_{\rm bh}}{c_\infty^2} \tag{6.1}$$

where  $M_{\rm bh}$  is the mass of the central object and  $c_{\infty}$  is the sound speed of the gas far from the hole. The Bondi radius divides the flow into two distinct regions (Frank et al., 2002). Far from  $r_B$ , gas is hardly aware of the existence of the black hole, and the flow is very subsonic. The pressure and density of a subsonic flow are approximately constant, and so we can set  $\rho(r) \approx \rho_{\infty}$  at  $r \gg r_B$ .

Inside the capture radius, on the other hand,  $\rho(r)$  begins to increase above the ambient value, and the flow eventually reaches a sonic point where  $|v_r| = c_{\infty}$ , within which it plunges essentially at free-fall. The sonic point is found from  $r_s = GM_{\rm bh}/2c_s^2(r_s)$ , where  $c_s(r_s)$  is the sound speed at  $r_s$ . This local quantity is related to the sound speed at infinity via  $c_s(r_s) = c_{\infty}(2/(5-3\Gamma))^{1/2}$  where  $\Gamma$  is the polytropic index of the gas that relates the gas pressure and density by  $P = K\rho^{\Gamma}$ , with K a positive constant.

Applying the Bondi-Hoyle formalism to black hole growth assumes that the accretion rate on to the SMBH is set by the accretion rate through the Bondi radius, given by

$$\dot{M}_{\rm bh} = \pi \lambda(\Gamma) r_B^2 \rho_\infty c_\infty = \frac{4\pi \lambda(\Gamma) G^2 M_{\rm bh}^2 \rho_\infty}{c_\infty^3} \tag{6.2}$$

where  $\lambda(\Gamma)$  contains all the corrections arising due to the finite pressure gradient force in the problem. This function varies relatively weakly, i.e., between 1.12 for  $\Gamma = 1$  and 0.25 for  $\Gamma = 5/3$ . For the remainder of this paper our fiducial assumption is a soft equation of state i.e.,  $\Gamma \approx 1$ .

### 6.2.2 When is Bondi-Hoyle accretion applicable?

The Bondi-Hoyle formalism has been widely adopted for a a 'sub-grid' prescription for the accretion rate on to the SMBH in large-scale cosmological simulations. The argument seems to be that while one cannot usually resolve the scales of the Bondi radius for the black hole, one can at least use the smallest resolved scales to approximately determine the value of the gas density and the sound speed 'at infinity' for use in the Bondi formula. Indeed, the smallest resolvable scales are usually about a fraction of a kiloparsec, whereas the Bondi radius is of the order of a few to a few tens of parsecs.

Unfortunately, for the Bondi (1952) solution to be applicable, we still need to make sure that the gas being at rest at infinity assumption is satisfied. In cosmological simulations black holes are immersed in stellar bulges and dark matter haloes that are typically  $\sim 10^3$  to 10<sup>4</sup> times more massive than the SMBH. If gas in the halo (or bulge) is as hot as the halo virial temperature, it will be in hydrostatic balance. It seems that in this situation, which is common for low luminosity SMBHs in giant elliptical galaxies where the gas is rather tenuos and hot since the cooling time is long (Churazov et al., 2005), the Bondi-Hoyle solution should be a good approximation. However, in the epoch when SMBHs grow rapidly, their hosts are very gas rich. Higher density gas is likely to cool much faster and hence the gas is likely to be much cooler than the virial temperature. In this case the gas is not able to support its own weight, and must collapse to the centre, where it feeds the SMBH and forms stars. Therefore, we expect a radial inflow of gas to the centre rather than hydrostatic balance 'far' from the SMBH. As a first step, it is clear that the Bondi-Hoyle formalism must be modified in the presence of an external potential. Ricotti (2007) has demonstrated how this should be done for the specific case of the growth of primordial black holes in a dark matter halo with a power-law density profile.

More fundamentally, however, even the meaning of the Bondi radius loses its utility in this situation. As explained earlier, for the isolated SMBH problem,  $r_B$  delineates the region outside which the potential of the hole starts to become greater than the internal energy of the gas. For an SMBH plus host halo system, one should introduce a modified Bondi radius,

$$\tilde{r}_B = \frac{2GM_h}{c_\infty^2} \tag{6.3}$$

that takes into account the total mass of the halo,  $M_h$ . In order for gas to accrete efficiently on a dark matter halo from larger scales, its temperature must be at most comparable with the virial temperature of the halo (White and Frenk, 1991). Thus, we set  $c_{\infty} \lesssim GM_h/r_h$ , where  $r_h$  is the virial radius of the halo, giving us a modified Bondi radius of  $\tilde{r}_B \gtrsim 2r_h$ . We can see from this that the gravitational potential energy starts to dominate over the internal energy of the infalling gas before the latter has even reached the edge of the halo. The standard  $r_B$  is therefore meaningless in this case.

Figure 6.1 illustrates this point graphically, by comparing the potential energy of gas as a function of radius for a variety of halo profiles with  $c_s^2/2$ , assuming the virial gas temperature at  $r_h$ . These profiles are described in Section 6.3. The cosmology we have assumed for the halos is  $\lambda$ CDM with  $\Omega_m = 0.27$ ,  $\Omega_{\lambda} = 0.73$  and a virial overdensity parameter of  $\Delta = 200$  at a redshift of z = 2 (although we note that our conclusions are unchanged for a wide range in z, from the early Universe to the present day).

The  $\propto 1/r$  potential due to the SMBH of mass  $M_{\rm bh} = 10^8 \,\mathrm{M_{\odot}}$  in this Figure is shown with the dash-triple-dot power law. The traditional Bondi (or capture) radius is at r = 0.002, where the SMBH potential crosses the  $c_s^2/2$  line. For all the halos considered, the potential energy of the host makes a significant contribution at all radii except those within the very inner parts of the halo (Figure 6.1). The host potential can exceed the gas internal energy by a large factor everywhere outside the SMBH radius of influence. Therefore, if we assume the gas has a relatively soft equation of state and accretes on to the halo at or below  $T_{\text{virial}}(r_h)$ , where  $r_h$  is the halo truncation radius, the character of the inflow becomes supersonic inside the halo *instead of being stationary at infinity*. In Section 6.3 we quantify this by calculating the sonic point for isothermal gas flows at or below the virial temperature for all of the halo mass profiles considered. We show that for typical dark matter halos the sonic point is reached while still at very large distances from the central black hole.

From Figure 6.1 it can also be seen that for hotter gas, the classical  $r_B$  estimate becomes more accurate, as the potential energy of the (halo + SMBH) system starts to asymptote to the SMBH solution and no longer dominates over the value of  $c_s^2/2$ . If the gas thermal energy is comparable to that of the potential energy we naturally expect near hydrostatic equilibrium to be maintained and in this case the Bondi-Hoyle formalism is applicable.

### 6.3 The sonic point

In order to demonstrate quantitatively our findings so far, namely, the character of the spherically symmetric infall at large radii on to an SMBH embedded in a background potential, we consider the *sonic point* for gas infalling on to a dark matter halo. For an isothermal gas flow, determining the sonic point for an extended mass distribution reduces to the problem of finding a solution to the equality  $r = GM(r)/2c_s^2(r)$ , with M(r) as the mass enclosed within radius r (see, e.g., Frank et al., 2002). Assuming an isothermal equation of state for the moment and putting  $c_{\infty} = c_{\text{virial}}$  we find that for gas accreting from outside  $r_h$  the sonic radius should be  $r_h/2$ . Since this is within the halo, we must also take account of how the mass profile varies once inside  $r_h$ . For a given profile we therefore proceed to solve (numerically, with a root finding algorithm) for the sonic point, and plot its location versus gas temperature in Figure 6.2. We also consider the effect of varying the adiabatic index on the solution.

### 6.3.0.1 Power law profile

We consider first the simple case of an extended halo with a power-law density profile given by

$$\rho(r) = \rho_{\rm h} \left(\frac{r}{r_{\rm h}}\right)^{-q} \tag{6.4}$$

where  $\rho_{\rm h}$  is the dark matter overdensity at the truncation radius. For the sonic point we require the enclosed mass, which we obtain by integrating equation 6.4 over the spherical



Figure 6.1: Plot of potential energies for a selection of dark matter halo profiles (some realistic, some instructive) modified by a central SMBH, compared to the thermal energy of the gas at the virial temperature at the halo virial radius. The potential due to the central SMBH by itself is also plotted ('BH' in the figure key). The classical Bondi radius for three different values of temperature are indicated as crosses to demonstrate how the standard Bondi-Hoyle approach becomes more accurate at higher T. The modified Bondi radius for the fiducial  $T_{\rm vir}$  is denoted by a rhombus, demonstrating that at this temperature the potential energies are dominant for the entirety of the halo.

volume within radius r, giving us,

$$M(r) = M_{\rm h} \left(\frac{r}{r_{\rm h}}\right)^{3-q} \tag{6.5}$$

so that  $r_{\rm s} = 2^{1/(2-q)} r_{\rm h}$  for gas at  $T_{\rm vir}(r_{\rm h})$  and  $\Gamma = 1$ . We note that for the isothermal sphere,  $r_{\rm s}$  cannot be defined once inside the halo, as the mass enclosed falls  $\propto r$ .

Figure 6.2 shows how the sonic point varies for q = 3 (solid black line) as T is decreased below our fiducial (but upper limit) assumption of  $T_{\rm vir}(r_{\rm h})$ . As we can see,  $r_{\rm s}$  quickly reaches  $r_{\rm h}$  in this case. A dotted black line indicates the solution for q = 4, demonstrating how the sonic point is reached faster for a steeper mass profile once inside the halo.

In the context of cosmological modelling, we now consider some more realistic profiles for dark matter halos, such as would be found in a large-scale simulation. In each case we solve for the sonic point of gas accreting on to the halo and plot the solution for a range of temperatures in Figure 6.2.

### 6.3.0.2 NFW profile

The Navarro, Frenk and White (NFW) profile (Navarro et al., 1996) has been found to provide a good description of the average mass profiles of dark matter haloes in dynamical equilibrium in cosmological N-body simulations. The mass enclosed within the NFW profile can be conveniently expressed in the form

$$M(r) = 4\pi\rho_0 a^3 \left[ \ln\left(1+\frac{r}{a}\right) - \frac{r}{a(r+a)} \right]$$
(6.6)

where a is the scale radius of the halo and  $\rho_0$  is a characteristic density. Note that the scale radius a depends on the concentration  $c = r_{\rm h}/a$  of the halo, where  $5 \le c \le 15$  for halos of mass  $10^{14} - 10^{10} \,\mathrm{M_{\odot}}$  (see, e.g., Bullock et al., 2001).

### 6.3.0.3 Hernquist profile

The Hernquist profile (Hernquist, 1990) is a convenient analytical model for the mass distribution in elliptical galaxies and bulges that follow the  $R^{1/4}$  law of de Vaucouleurs (1948). It behaves similarly to the NFW profile at small radii ( $\rho \propto r^{-1}$ ) but drops off more rapidly at large radii ( $\rho \propto r^{-4}$ ); this has led to its use generating initial conditions for simulations of galaxy mergers (e.g. Springel et al., 2005). The Hernquist mass profile can be expressed as

$$M(r) = 4\pi\rho_0 a \left[\frac{r^2}{2(1+r/a)^2}\right]$$
(6.7)

### 6.3.0.4 Jaffe profile

The Jaffe cusp (Jaffe, 1983) is one of a family of spherical density profiles (which include the Hernquist profile, above) that are characterised by an exponent  $\gamma$  (Dehnen, 1993). For the Jaffe profile we have  $\gamma = 2$ . The mass enclosed goes as

$$M(r) = 4\pi\rho_0 a^3 \left(\frac{r}{r+a}\right) \tag{6.8}$$

This profile is motivated by observed luminosity profiles of galaxy bulges and is a reasonable model for the inner regions of galaxies and those that have recently undergone a major merger (Milosavljevic and Merritt, 2001).

Referring now to Figure 6.2, we can see that the isothermal solution for the sonic point is in all cases a significant fraction of the halo radius at  $T = T_{\rm vir}(r_{\rm h})$ . Indeed, one only has to decrease the temperature of the gas by ~ 30% from this fiducial value for the sonic radius to become greater than the halo radius itself. This means that for gas accreting (in a spherically-symmetric fashion) on to a dark matter halo, at or near the virial temperature for the virial radius, the infall must quickly tend to a free-fall solution once inside the halo.

Of course, we acknowledge that gas accreting on to a virial overdensity in a sphericallysymmetric fashion is unlikely to possess a purely isothermal equation of state. Indeed, accretion in the 'hot phase' may follow a nearly adiabatic trajectory (Kereš et al., 2005). In this case  $r_s$  cannot be defined, with the gas accreting subsonically down to the black hole. This of course is an extreme case and for a flow that has a mixture of cold and hot phases the adiabatic index will lie somewhere in between the two limits. We therefore plot the full range of  $\Gamma$  against the solution for the sonic point in Figure 6.3. We do this for what is arguably the most realistic halo profile of those we have discussed so far, namely the NFW profile. We display curves for differing fractions of the virial temperature, noting that although a harder equation of state might reduce the effectiveness of our conclusions, slightly decreasing the initial temperature below that of  $T_{\rm virial}(r_{\rm h})$  greatly assists the gas in reaching  $r_{\rm s}$  while still at large distances from the galaxy at the centre of the halo.

In the interest of completeness we now demonstrate *numerically* the form of the radial infall in a realistic background potential within a galaxy. We choose a dynamic range that lies inside the sonic point (as would always be the case for efficiently cooled gas at these scales, as we have shown) in order to highlight why the standard Bondi estimate is inaccurate in this case.



Figure 6.2: Plots of the sonic radius, scaled in units of  $r_{\rm h}$ , for a variety of halo profiles, assuming a halo of  $10^{11} \,\mathrm{M_{\odot}}$ . All of the profiles intersect the  $T = T_{\rm vir}(r_{\rm h})$  axis at a radius that lies within the halo but still at large distances from the black hole, and certainly well outside the resident galaxy. This demonstrates that for the majority of the infall within the halo, a spherically symmetric, radially infalling accretion flow with a soft equation of state will be travelling supersonically unless some form of external heating process takes place (or perhaps heating from AGN feedback or star formation, although we note that the latter processes would generally only operate well within the galaxy itself.)



Figure 6.3: Plots of how the sonic radius varies with the adiabatic index, for a halo with an NFW profile at different initial temperatures as fractions of  $T_{\rm vir}$ . It is clear that although moving to a harder equation of state can have a strong effect on the sonic radius for  $T \simeq T_{\rm vir}$ , reducing it significantly and allowing the flow to remain subsonic for far longer, even a small decrease in the temperature of the flow below  $T_{\rm vir}$  as it accretes on to the halo can completely counteract this effect. Naturally as  $\Gamma \to 5/3$ ,  $r_{\rm sonic} \to 0$ .

### 6.4 Numerical tests

To perform the simulations we employ the three-dimensional SPH code GADGET-3, as per Chapter 4 and as described in Chapter 3, Section 3.5. The gas is evolved in a static external potential which includes a point mass black hole at the centre. The computational domain extends from a kiloparsec down to an 'accretion radius' around the black hole at  $r_{\rm acc} = 1$  pc, and we remove the particles that come within this distance of the SMBH. We use adaptive SPH smoothing lengths down to a minimum of  $2.8 \times 10^{-2}$  pc, and employ the Monaghan-Balsara form of artificial viscosity (Gingold and Monaghan, 1983; Balsara, 1995) with  $\alpha = 1$  and  $\beta = 2\alpha$ .

For the external potential in our model we use a Jaffe cusp as per equation 6.8 but with a core at the centre of our computational domain to prevent divergence in the gravitational force to very small radii. The radius of the core,  $r_c$ , corresponds to approximately the dynamical influence radius of the SMBH. With this (modified) potential the mass enclosed within radius r is given by:

$$M(r) = M_{\rm bh} + \begin{cases} M_{\rm c} \left(\frac{r}{r_{\rm c}}\right)^3, & r < r_{\rm c} \\ M_{\rm c} + M_a \left(\frac{1}{r_{\rm c} + a} - \frac{1}{r + a}\right), & r \ge r_{\rm c} \end{cases}$$
(6.9)

where  $M_c = 2 \times 10^8 \,\mathrm{M_{\odot}}$ ,  $M_a = 10^{11} \,\mathrm{M_{\odot}}$ ,  $r_c = 20 \,\mathrm{pc}$ , and  $a = 10 \,\mathrm{kpc}$ . The mass of the SMBH is set to  $M_{\mathrm{bh}} = 10^8 \,\mathrm{M_{\odot}}$ . For simplicity, the gas is kept isothermal throughout the entirety of the simulation and self-gravity is turned off.

### 6.4.1 Initial conditions

The starting condition for our simulations is that of a uniform density, spherically symmetric thick gaseous shell centered on the black hole. The inner and outer radii of the shell are  $r_{\rm in} = 0.1$  kpc and  $r_{\rm out} = 1$  kpc, respectively, with  $M_{\rm shell} = 10^8 \,\mathrm{M_{\odot}}$ . The temperature of the gas is varied between tests, ranging from  $10^3 \,\mathrm{K}$  to  $10^5 \,\mathrm{K}$ . To minimise initial inhomogeneities we cut the shell from a relaxed, glass-like configuration. The gas is initially at rest, and allowed to infall within our static potential. After a time of the order of the dynamical time at the outer edge of the shell, a steady-state is reached and it is at this time that we make our comparisons between the various accretion rates in the next section.

### 6.4.2 Results

We define a radius-dependent, 'measured' accretion rate on radial shells as  $\dot{M}(r) = 4\pi r^2 \rho v_r$ , where  $v_r$  is the radial velocity of the gas on the shell r. Since the majority of the flow has been allowed to settle into an approximate steady state, this function is

largely constant with radius, and is thus the same as the time-averaged accretion rate measured at the black hole. We denote the average value of this function with the long dashed red curve in Figure 6.4. The solid, the dotted and the dashed curves in this Figure show the Bondi-Hoyle estimate for the accretion rate as a function of radius for three different values of gas temperature. The interpretation of the curves is as follows: in a cosmological isothermal simulation, where the smallest resolved scale is distance r from the black hole, one would would probably find the Bondi-Hoyle estimate to be similar to that shown by the curves in the figure. Clearly, therefore, the Bondi-Hoyle estimate is very inaccurate for these isothermal simulations. At intermediate temperatures, e.g.,  $10^5$  K, the formula in the inner parts results in a significant overestimate of the accretion rate and an overestimate at large radii.

### 6.4.2.1 Free-fall rate

A simple but physically well motivated alternative to the Bondi-Hoyle formula for cold spherically symmetric flows is a free-fall rate estimate,

$$\dot{M}_{\rm ff}(r) = \frac{M_{\rm gas,enc}(r)}{t_{\rm ff}(r)} ,$$
 (6.10)

where  $M_{\text{gas,enc}}$  is the enclosed gas mass within radius r, and  $t_{\text{ff}} = (r^3/2GM(r))^{1/2}$  is the free fall time. As the enclosed gas mass may be expensive to evaluate in a numerical simulation, one can try to approximate the enclosed gas mass in the above equation as  $(4\pi/3)r^3\rho_{\text{gas}}(r)$ , so that

$$\dot{M}_{\rm ff}(r) \sim \frac{4\pi r^3 \rho_{\rm gas}(r)}{3t_{\rm ff}(r)}$$
 (6.11)

Both of the above estimates provide a much better match to the accretion rate in our numerical tests than the Bondi-Hoyle approach, as Figure 6.4 shows. The estimate that includes the enclosed gas mass appears to be the better of the two. It should be noted too that since the free-fall estimates have no dependence on  $c_s$ , the profiles are converged regardless of the temperature.

### 6.5 Discussion and future work

We have argued that the Bondi-Hoyle formalism is directly applicable to SMBH accretion only if the accreting gas is fully adiabatic, which is the case when the cooling time is long. One expects that in this case the gas is in an approximate hydrostatic balance within the halo, thus satisfying the important explicit assumption of Bondi (1952) that gas is stationary at infinity. This may possibly be the case in giant, gas-poor elliptical galaxies, where the gas is rather tenuos and hot (Churazov et al., 2005). However, in the most interesting



Figure 6.4: The mass flux (average steady-state value) from the simulations through each radius (red) i.e.,  $\dot{M}(r)$ , together with the Bondi-Hoyle estimate as per Equation 6.2 and the free-fall estimate as per Equations 6.10 & 6.11.

phase of SMBH and galaxy buildup, when the halo is likely to be awash with gas to feed both the SMBH and star formation, the cooling time is expected to be short, and gas may cool much below the virial temperature. Hydrostatic balance is then impossible for gas in the halo; it is necessarily in free fall. We have run a series of simple numerical tests to explore this limit, allowing a thick, spherical shell of gas to accrete on to an SMBH at the centre of a background halo potential. These tests showed that a free fall accretion rate estimate is indeed considerably more accurate than the Bondi-Hoyle formalism in this case. What is most concerning is that the error of the latter strongly depends on the gas temperature (and therefore the cooling function) and cannot be 'predicted'. The Bondi-Hoyle accretion formalism may thus be wrong by significant (and unknown) factors in either direction. The error may well depend on the galaxy mass and type systematically. We must conclude that predictions based on the Bondi-Hoyle approach are unlikely to be robust.

The pertinent question then is how to rectify this situation. A potential solution is to use the 'accretion disc particle' method proposed by Power et al. (2010) and related to the 'sink particle' method of Bate et al. (1995). One introduces a reasonably small accretion radius as a region around the SMBH within which gas is accreted on the black hole. The advantage of this is that gas supported by its angular momentum at distances greater than the accretion radius forms a disc rather than being forced to accrete on the SMBH unphysically. One then introduces a sub-grid prescription for an accretion (AGN) disc at radii smaller than the accretion radius to treat processes such as viscous transport and star formation in AGN discs (see, e.g., Nayakshin et al., 2007).

The latter method works best at reasonably high particle numbers within the simulated galaxies. At poorer resolution one may still require a yet simpler method. What we desire is a Bondi-Hoyle *style* approach, that can yield a single number (or perhaps a set of numbers) for the accretion rate down to scales of the order of the SMBH capture radius, but after which point the accretion disc particle method can take over. Such an approach could be achieved by developing an *interpolation* formula between the near-hydrostatic balance/Bondi-Hoyle rate and the efficiently cooled/free-fall rate. The input for the accretion disc particle would therefore no longer be a physical boundary but rather the value(s) obtained from this formula.

## Discussion and conclusion

"If we knew what it was we were doing, it would not be called research, would it?"

Albert Einstein

This thesis has examined the feeding of supermassive black holes (SMBHs) in galactic centres. We have summarised the relevant theoretical topics for this work (Chapter 1) and derived the appropriate mathematical constructs (Chapter 2). We have presented in some detail the numerical approach that we have used to run the simulations and construct our models in Chapter 3. The science goals have been described in Chapters 4, 5, & 6, where we consider first the formation of the young stellar populations in our own Galactic centre and the character of the accretion flow from a cloud-cloud collision; second, a novel mode of AGN feeding driven by turbulence at scales of hundreds of parsecs; and third, an investigation into the validity of the accretion rate prescription for large scale, cosmological simulations. In the remainder of the current chapter we summarise each of the science chapters in more detail, discuss the bigger picture, and finally offer suggestions for future work.

### 7.1 Summary

In Chapter 4 we have studied the environment of our own Galactic centre using numerical simulations. Young, massive stars in the central parsec of our Galaxy are best explained by star formation within at least one, and possibly two, massive self-gravitating gaseous discs. We have explored the possibility that the observed population of young stars could have originated from a large angle collision of two massive gaseous clouds at  $r \simeq 1$  pc from Sgr A<sup>\*</sup>. In all of the simulations that we performed for this model, the post-collision gas flow forms an inner, nearly circular gaseous disc and one or two eccentric outer filaments, consistent with the geometry of the observed stellar features. Furthermore, the radial stellar mass distribution that we obtain is very steep,  $\Sigma_* \propto r^{-2}$ , again consistent with the observations. All of our simulations produce discs that are warped by between  $30^{\circ}$  and  $60^{\circ}$ . in accordance with the most recent observations. The three-dimensional velocity structure of the stellar distribution is sensitive to initial conditions (e.g., the impact parameter of the clouds) and the value of the cooling parameter. For example, the runs in which the inner disc is fed intermittently with material possessing fluctuating angular momentum result in multiple stellar discs with different orbital orientations, contradicting the observed data. In all cases the amount of gas accreted at our inner boundary is large, enough to allow Sgr A<sup>\*</sup>to radiate near its Eddington limit for over  $\sim 10^5$  yrs. This suggests that a refined model would consist of physically larger clouds (or a cloud and a disc such as the circumnuclear disc) colliding at a distance of a few parsecs rather than 1 parsec as we have assumed in our simulations.

In Chapter 5 we have generalised our picture to that of the inner 100 parsecs of a typical (gas-rich) galaxy bulge. We have examined specifically the gas accretion rate on to the SMBH at the centre of a galaxy from these scales. We point out that, while the

mean angular momentum of gas in the bulge is very likely to be large, the deviations from the mean can also be significant. Indeed, cosmological simulations show that velocity and angular momentum fields of gas flows on to galaxies are very complex. Furthermore, inside bulges the gas velocity distribution can be further randomised by the perturbations due to feedback from star formation. We perform hydrodynamical simulations of gaseous rotating shells infalling on to an SMBH, attempting to quantify the importance of velocity dispersion in the gas at relatively large distances from the black hole. We implement this dispersion by means of a supersonic turbulent velocity spectrum. We find that, while in the purely rotating case the circularisation process leads to efficient mixing of gas with different angular momentum, resulting in a low accretion rate, the inclusion of turbulence increases this accretion rate by up to several orders of magnitude. We show that this can be understood based on the notion of 'ballistic' accretion, whereby dense filaments, created by convergent turbulent flows, travel through the ambient gas largely unaffected by hydrodynamical drag. This prevents the efficient gas mixing that was found in the simulations without turbulence, and allows a fraction of gas to impact the innermost boundary of the simulations directly. Using the ballistic approximation, we derive a simple analytical formula that captures the numerical results to within a factor of a few. Rescaling our results to astrophysical bulges, we argue that this 'ballistic' mode of accretion could provide the SMBHs with a sufficient supply of fuel without the need to channel the gas via large-scale discs or bars. We therefore argue that star formation in bulges can be a strong catalyst for SMBH accretion.

Finally, in Chapter 6 we critically evaluate the modelling of accretion on to SMBHs in cosmological simulations. This is frequently dealt with, sub-grid, by a Bondi-Hoyle prescription. We examine this approach analytically and numerically, and argue that it is adequate only for hot, virialised gas with zero angular momentum. Gas with a subvirial temperature, on the other hand, would likely be in a state of free-fall long before it reaches the black hole. Our numerical tests demonstrate that in this regime the Bondi-Hoyle formalism can be erroneous by orders of magnitude in either direction. Not only it is inaccurate, but it may also impose unphysical trends in the results by being wrong by different factors for different halo masses. We propose that future numerical simulations of SMBH growth should either modify the prescription to include the free-fall regime and/or employ the 'accretion disc particle' method whereby only gas with low angular momentum is accreted (Power et al., 2010).

The work that we have presented in this thesis has been tailored to numerical and analytical modelling of galactic centres; specifically, to the feeding of the central SMBH via infall of gas from larger scales. A core theme of the models discussed in Chapters 4 & 5 has been feeding through a disordered flow that arises due to an initial 'kick' from an event that deposits energy and momentum in the surrounding gas. In Chapter 4 this took the form of a collision between two gas clouds, where the resulting kick was a *thermal* one; as a result the gas puffed up and remained diffuse to late times, ensuring an extended, disordered flow on to the central forming accretion disc as the diffuse gas self-collided and cooled. This gave rise to an effect noted by King and Pringle (2007), where the SMBH is fed through a small-scale disc that has no preferred orientation with respect to the host galaxy. In Chapter 5 the kick to the gas was *momentum*-driven, due to our isothermal assumption, and created dense clumps with low angular momenta that were able to travel to the centre of the computational domain without being affected by hydrodynamical drag. In this case too, the disc stayed small, and although it had a dominant angular momentum vector, the inner parts were significantly warped and subject to strong waves propagating through the disc as matter continued to fall on to it. In both instances we therefore avoid the creation of a large disc at a single orientation, which as we discussed in Chapter 1, Section 1.7.1.2, is highly undesirable if we are to maintain a large (~  $1 \,\mathrm{M}_{\odot} \,\mathrm{yr}^{-1}$ ) accretion rate on to the SMBH. Indeed, in our numerical models the amount of gas accreted by the black hole was sufficient for bright AGN activity for the duration of the simulations. We therefore favour this 'stochastic' mode of feeding as the means by which to grow SMBHs to the sizes that are observed  $(M_{\rm bh} \sim 10^9 \,\mathrm{M_{\odot}} \text{ by } z \sim 6)$ .

We find, in all of our models, that the inclusion of gas and stellar processes in galactic centres is of direct relevance to the feeding and growth of the central SMBH. Currently, standard paradigms for accretion do not allow rates fast enough to grow seed black holes in the early Universe to the masses of SMBHs that we observe in AGN. Central to this issue is the oft-quoted need for 'sensible', planar geometries and orientations; such assumptions necessarily lead to a large, thin, parsec scale disc through which transport of gas takes longer than the current age of the Universe, or a black hole that has been 'spun-up' to maximal with a radiative efficiency so high that its Eddington limit is reached at disappointingly low accretion rates. Fundamentally, the problem lies in the fact that the gas supply from large (galactic) scales must be transported some 10 orders of magnitude in scale to reach the actual accretion radius (not the numerical ones that we have used in our simulations, but the *physical*, last stable orbit one) of the SMBH. We really are therefore faced with (to use a popular analogy from the AGN community) figuring out how to feed a 'baby' with a 'giant spoon<sup>1</sup>.

We argue that this feeding must be processed by events occurring on the intermediate scale of a galaxy; say within a few hundreds of parsecs or less. These processes, of which momentum feedback from star formation is the particular example we have explored in this work, break up the gas into 'morsels' that are put on to low angular momentum orbits by collisions and shocks, sending them close enough to the black hole to be captured and accreted in a reasonable timeframe. In this way, the fuel does not have to wait its turn at

<sup>&</sup>lt;sup>1</sup>giant is an *understatement* 

large radii for all of the material inside of its radius to be accreted before it can join in; rather it can be 'kicked' right into the centre of the galaxy, arriving there on a dynamical time rather than a viscous one. The central black hole, in turn, does not continually swallow gas at the same orientation and orbital sense, resisting further accretion through efficient feedback, but instead devours each morsel that is sent its way.

To a certain degree, the idea that star formation can aid accretion on to the SMBH in this way runs counter to accepted wisdom, which tends to view the formation of stars as more of a death sentence for the progenitor gas. Furthermore, in order to fully explain the means by which SMBH growth to the observed masses occurred, we require this reprocessing the gaseous fuel to occur repeatedly, as a one-off event will merely produce a period of bright AGN activity for a short time, most likely of the order of the free-fall time from the appropriate radius. Our model therefore favours gas-rich surroundings and 'eager' star formation, so that feedback from supernovae have a chance to create a large dispersion in the velocity distribution before the majority of the gas is used up, and a way to replenish the gas supply from larger scales; in other words, a *major merger* of gas-rich galaxies. This may make it difficult for our model to be verified observationally, as it is hardly a surprising conclusion that gas-rich mergers should give rise to a period of strong AGN activity and SMBH growth. Further self-consistent modelling and the inclusion of *feedback* into our simulations may help to constrain this picture and the possible supporting trends that should be looked for in large observational AGN surveys.

### 7.2 Future work

There are a number of areas for the projects outlined in this thesis which warrant follow-up work. Firstly, it would be interesting to verify the prediction that we made in Chapter 4, namely that a better model for the formation of the stellar discs in the Galactic centre is one where the cloud collision occurs further out, for example within the circumnuclear discs a few parsecs away from the SMBH. Secondly, the project presented in Chapter 5 on turbulent, ballistic feeding of AGN is really only a first step in modelling the intermediate scale flow within a galaxy, in order to develop a physically self-consistent sub-grid prescription for AGN feeding that can be embedded into a cosmological simulation (where resolution is poor at galactic scales). This work is ongoing. The next step will be to take input from a cosmological simulation that can resolve down to kiloparsec scales as the outer boundary condition for our velocity field; currently the initial setup is very idealised. Finally, the critical evaluation of the Bondi-Hoyle sub-grid model for large-scale simulations (Chapter 6) would be improved considerably if we were to determine an appropriate interpolative formula that reproduced both the near hydrostatic equilibrium and free-fall regimes consistently. All of these extensions are planned, along with the coupling of the accretion model that we develop to an AGN feedback model that can further shed light on the symbiotic relationship between the SMBH and the host galaxy.

There is also room for improvement and extension in our numerical method. There has recently been substantial development of the SPH method (Cullen and Dehnen, 2010; Read et al., 2010) that would likely be very relevant to the work presented in this thesis. In particular, Cullen and Dehnen (2010) have devised an artificial viscosity scheme that only switches on in shocks, rather than being fixed for all particles for the entirety of the simulation. This brings SPH simulations into a new era; one of a fully physically consistent treatment of hydrodynamics (and, in theory, one that would perform far better in convergence tests that determine the contribution from numerical effects through the viscosity). The effect that this improved method would have on our simulations, and our results, is not clear. Almost certainly, it would lower significantly the numerical, viscous accretion rates through disc geometries that we find, although as we have mentioned in Chapters 4 & 5, this is generally not the dominant mode of angular momentum transport in our models. Collisions (such as between the two clouds in Chapter 4) and shocked regions due to turbulence (such as in Chapter 5) tend to put gas on plunging orbits and the accretion therefore proceeds dynamically. Nonetheless, it is an important priority for our future work on these models that the 'Cullen-Dehnen' switch be implemented in our code and the simulations re-run. Likewise, Read et al. (2010) has developed a method for SPH (termed 'optimised smoothed particle hydrodynamics', or OSPH) that is able to handle *mixing* in multiphase flow, something which traditional SPH falls down on (see, e.g., Agertz et al., 2007) due to entropies being discretely advected with the flow. The upshot here is that fluid instabilities, such as the Kelvin-Helmholtz instability (see, e.g., Gerwin, 1968) and the Rayleigh-Taylor instability (see, e.g., Chandrasekhar, 1961) can now be modelled self-consistently in an SPH simulation. This tends to have the effect of mixing fluid of different phases (i.e., densities) when they encounter each other, and in particular, shearing out dense clumps that would otherwise have formed. On the face of it this sounds somewhat disastrous for the work we present in Chapter 5, but in fact preliminary investigations (the fiducial turbulent feeding simulations run with OSPH) show little difference (aside from some aesthetics) in both the accretion rate on to the SMBH and the character of the flow. This is almost certainly because we have imposed an isothermal assumption on these simulations, as when the temperatures (and thus sound speeds) are identical this effect is reduced. It may also be because any enhanced mixing causes the extended turbulent filaments to break up more readily into dense clumps, counteracting the enhanced tendency for the dense clumps to be sheared out, or because the overdensities we find in these simulations are sufficiently above the ambient for surface instabilities to be negligible. Nonetheless, further tests on our models using this optimised method need to be done, and the results analysed and compared to standard SPH.

In conclusion of this thesis, we suggest that while the feeding of SMBHs remains an open problem (in terms of the required growth rate to match observations of black hole masses and AGN luminosities) the model that has been presented here is a promising one. The stochastic mode of feeding that we have developed constitutes an important step forward, and the concept that we have introduced - that of a 'reprocessing' of the gas inflow by feedback from star formation - will hopefully form the basis for an answer to the question of how AGN at the centres of galaxies are fed.



### Stability of particle orbits in Kerr spacetime

### Note: this derivation is taken from Hobson et al. (2006), Chapter 13, Sections 13-14.

The effective potential for a particle with rest mass m moving in the equatorial plane of a Kerr black hole with spin parameter a is given by

$$V(r) = \frac{J^2 - a^2 c^2 (k^2 - 1)}{2r^2} - \frac{GM(J^2 - ack)^2}{c^2 r^3} - \frac{GM}{r}$$
(A.1)

For the extrema we require dV(r)/dr = 0. It is easiest to work in terms of a variable u = 1/r, which when substituted into equation A.1 returns

$$V(u) = \frac{u^2}{2} \left[ J^2 - a^2 c^2 (k^2 - 1) \right] - \mu u^3 (J - ack)^2 - \mu u c^2$$
(A.2)

where we have put  $\mu \equiv GM/c^2$ . We equate this to the LHS of equation 2.10, so that

$$\frac{u^2}{2} \left[ J^2 - a^2 c^2 (k^2 - 1) \right] - \mu u^3 (J - ack)^2 - \mu u c^2 = \frac{c^2}{2} \left( k^2 - 1 \right)$$
(A.3)

and differentiate with respect to u, giving us

$$u\left[J^2 - a^2c^2(k^2 - 1)\right] - 3\mu u^2(J - ack)^2 - \mu c^2 = 0$$
(A.4)

We now introduce another variable to simplify the algebra, putting x = J - ack. The two equations above become

$$\frac{u^2}{2} \left[ x^2 + 2ackx + a^2c^2 \right] - \mu u^3 x^2 - \mu uc^2 = \frac{c^2}{2} \left( k^2 - 1 \right)$$
(A.5)

$$u\left[x^{2} + 2ackx + a^{2}c^{2}\right] - 3\mu u^{2}x^{2} - \mu c^{2} = 0$$
(A.6)

We take  $u \times (A.6) - (A.5)$ , giving us

$$x^{2}u^{2} + 2ackxu^{2} + a^{2}c^{2}u^{2} - 4\mu x^{2}u^{3} = c^{2}(1 - k^{2})$$
(A.7)

and we also re-arrange equation A.6 to get

$$2ackxu = x^2 u(3\mu u - 1) - c^2 (a^2 u - \mu)$$
(A.8)

which we substitute into equation A.7 to return

$$c^2k^2 = c^2(1-\mu u) + \mu x^2 u^3 \tag{A.9}$$

If we now eliminate k between equations A.9 and A.8 we obtain a quadratic in  $x^2$ :

$$u^{2} \left[ (3\mu u - 1)^{2} - 4a^{2}\mu u^{3} \right] x^{4} - 2c^{2}u \left[ (3\mu u - 1)(a^{2}u - \mu) - 2\mu a^{2}(\mu u - 1) \right] x^{2} + c^{4}(a^{2}u - \mu)^{2} = 0$$
(A.10)

for which the roots are given by

$$x^{2} = \frac{c^{2}(a\sqrt{u} \pm \sqrt{\mu})^{2}}{u(1 - 3\mu u \mp 2a\sqrt{\mu u^{3}}}$$
(A.11)

where to solve for x we must consider both signs of the subsequent square root. These determine whether the orbit is stable (-ve square root) or unstable (+ve square root). In the case of the former, we have

$$x = -\frac{c(a\sqrt{u} \pm \sqrt{\mu})}{[u(1 - 3\mu u \mp 2a\sqrt{\mu u^3})]^{1/2}}$$
(A.12)

which we substitute into equation A.9 to obtain

$$k = \frac{1 - 2\mu u \mp a \sqrt{\mu u^3}}{(1 - 3\mu u \mp 2a \sqrt{\mu u^3})^{1/2}}$$
(A.13)

and noting that J = x + ack we have

$$J = \mp \frac{c\sqrt{\mu}(1 + a^2u^2 \pm 2a\sqrt{\mu u^3})}{\sqrt{u}(1 - 3\mu u \mp 2a\sqrt{\mu u^3})^{1/2}}$$
(A.14)

We now consider a circular orbit of marginal stability, as similar to the Schwarzschild case the value of the ISCO will be at the corresponding radius. For this we require  $d^2V(r)/dr^2 = 0$ , and therefore with our substituted variables that  $d^2V(u)/du^2 = 0$ , since

$$\frac{\mathrm{d}^2 V}{\mathrm{d}r^2} = \frac{\mathrm{d}^2 V}{\mathrm{d}u^2} \left(\frac{\mathrm{d}u}{\mathrm{d}r}\right)^2 + \frac{\mathrm{d}V}{\mathrm{d}u}\frac{\mathrm{d}^2 u}{\mathrm{d}r^2} = u^3 \left(u\frac{\mathrm{d}^2 V}{\mathrm{d}u^2} + 2\frac{\mathrm{d}V}{\mathrm{d}u}\right) = 0 \tag{A.15}$$

Differenting equation A.6, which as we recall was the first derivative of V with respect to u, and using x = J - ack, we have

$$J^{2} - a^{2}c^{2}(k^{2} - 1) - 6\mu x^{2}u = 0$$
(A.16)

where we substitute for expressions for x, k, and J from equations A.12, A.13, A.14 respectively, and finally put back in that u = 1/r, so that

$$r^2 - 6\mu r - 3a^2 \mp 8a\sqrt{\mu r} = 0 \tag{A.17}$$

the solution to which is the radius of the ISCO.

### B

Derivation of the gravitational potential for Chapter 4: cloud collisions in the GC The gravitational potential of an arbitrary (spherically symmetric) mass distribution  $\rho(r)$  can be calculated from the enclosed mass by integrating the force contributions from each shell of radius dr between the limits of the distribution (Binney and Tremaine, 2008). For the stellar cusp + BH distribution used in Chapter 4, Section 4.2.2, the potential inside  $r_c \equiv r_{\text{cusp}}$  is given by

$$\Phi(r < r_{\rm c}) = \int_{r}^{r_{\rm c}} \frac{GM(r < r_{\rm c})}{r^2} \,\mathrm{d}r + \int_{r_{\rm c}}^{\infty} \frac{GM(r \ge r_{\rm c})}{r^2} \,\mathrm{d}r \tag{B.1}$$

where for the stellar cusp the density goes as

$$\frac{\rho_*(r)}{\rho_{*0}} = \begin{cases} \left(\frac{r}{r_c}\right)^{-1.4}, & r < r_c \\ \left(\frac{r}{r_c}\right)^{-2}, & r \ge r_c \end{cases}$$
(B.2)

The mass enclosed for each region is therefore

$$M(r < r_c) = \int_0^r 4\pi r^2 \rho_0 \left(\frac{r}{r_c}\right)^{-1.4} dr + M_{\rm bh}$$
(B.3)

$$M(r \ge r_c) = \int_{r_c}^{r} 4\pi r^2 \rho_0 \left(\frac{r}{r_c}\right)^{-2} dr + M_c + M_{\rm bh}$$
(B.4)

where  $M_c$  is the mass contained with  $r_c$ . We substitute the enclosed masses into equation B.1 to obtain

$$\Phi\left(r < r_{\rm c}\right) = \int_{r}^{r_{\rm c}} \frac{5\pi G\rho_0 r_{\rm c}^{1.4} r^{1.6}}{2r^2} + \frac{GM_{bh}}{r^2} \,\mathrm{d}r + \int_{r_{\rm c}}^{\infty} \frac{4\pi G\rho_0 r_{\rm c}^2}{r} - \frac{3\pi G\rho_0 r_{\rm c}^3}{2r^2} + \frac{GM_{bh}}{r^2} \,\mathrm{d}r$$

Performing the integration yields

$$\Phi (r < r_{\rm c}) = \left[\frac{25}{6}\pi G\rho_0 r_{\rm c}^{1.4} r^{0.6}\right]_r^{r_{\rm c}} - \left[\frac{GM_{\rm bh}}{r}\right]_r^{r_{\rm c}} + \left[4\pi G\rho_0 r_{\rm c}^2 \ln r\right]_{r_{\rm c}}^{\infty} + \left[\frac{3\pi G\rho_0 r_{\rm c}^3}{2r}\right]_{r_{\rm c}}^{\infty} - \left[GM_{\rm bh}r\right]_{r_{\rm c}}^{\infty}$$

giving us the full expression for  $\Phi$  at small r as

$$\Phi\left(r < r_{\rm c}\right) = \frac{25}{6}\pi G\rho_0 r_{\rm c}^{1.4} r^{0.6} - \frac{25}{6}\pi G\rho_0 r_{\rm c}^2 + 4\pi G\rho_0 r_{\rm c}^2 \ln r_{\rm c} + \frac{3}{2}\pi G\rho_0 r_{\rm c}^2 - \frac{GM_{bh}}{r} \quad (B.5)$$

where we have reversed the signs in order to consider the specific potential *energy*, and have dropped arbitrary constants that  $\rightarrow \infty$ .

For the potential *outside*  $r_c$  we require

$$\Phi(r \ge r_c) = \int_r^\infty \frac{GM(r \ge r_c)}{r^2} \,\mathrm{d}r = \int_r^\infty \frac{4\pi G\rho_0 r_c^2}{r} - \frac{3\pi G\rho_0 r_c^3}{2r} + \frac{GM_{\rm bh}}{r^2} \,\mathrm{d}r$$

which on integrating gives

$$\Phi\left(r \ge r_c\right) = \left[4\pi G\rho_0 r_c^2 \ln r\right]_r^\infty + \left[\frac{3\pi G\rho_0 r_c^3}{2r}\right]_r^\infty - \left[\frac{GM_{\rm bh}}{r}\right]_r^\infty$$

such that the specific potential energy at large r is

$$\Phi(r \ge r_c) = 4\pi G \rho_0 r_c^2 \ln r + \frac{3\pi G \rho_0 r_c^3}{2r} - \frac{GM_{\rm bh}}{r}$$
(B.6)

where once again we have dropped arbitrary constants.

# Derivation of the tidal shear density in the Newtonian potential of a black hole



Figure C.1: Supporting diagram (not to scale) for the derivation of the tidal shear density limit for a gas cloud in the gravitational potential of a Keplerian point mass.

We consider a cloud of gas, of mass  $M_{\text{cloud}}$ , moving in the (Newtonian) gravitational potential of a black hole with mass  $M_{\text{bh}}$ . As can be seen in Figure C.1, mass elements mon the nearest and farthest sides of the cloud, with respect to the black hole (BH), are located at a distance  $R_0 \pm r_0$  respectively. The gravitational force on these elements from the BH is

$$F_{\rm g} = \frac{GM_{\rm bh}m}{(R_0 \pm r_o)^2} \tag{C.1}$$

which can be written in a different way as

$$F_{\rm g} = \frac{GM_{\rm bh}m}{R_0^2} \frac{1}{(1 \pm r_0/R_0)^2} \tag{C.2}$$

Expanding this in a Taylor series we find

$$F_{\rm g} = \frac{GM_{\rm bh}m}{R_0^2} \left[ 1 \mp \frac{2r_0}{R_0} \pm \frac{6r_0^2}{R_0^2} + \cdots \right]$$
(C.3)

where we make the assumption that the size of the cloud,  $r_0$  is small in comparison to the distance from the BH, namely  $r_0 \ll R_0$ . This allows us to ignore terms quadratic in  $r_0/R_0$  and above. The gravitational force is therefore given by

$$F_{\rm g} = \frac{GM_{\rm bh}m}{R_0^2} \mp \frac{2GM_{\rm bh}mr_0}{R_0^3} \tag{C.4}$$

where the first term on the RHS is simply the standard gravitational force exerted by the BH on a mass m, while the second term represents the first-order tidal force correction.

This is

$$F_{\rm tidal} = \frac{2GM_{\rm bh}mr_0}{R_0^3} \tag{C.5}$$

To obtain the density at which the cloud begins to be sheared by the BH's gravity, we equate this to the gravitational self-binding energy of the cloud, namely

$$\frac{2GM_{\rm bh}mr_0}{R_0^3} = \frac{GM_{\rm cloud}m}{r_0^2}$$
(C.6)

so that we find

$$M_{\rm cloud} = \frac{2M_{\rm bh}r_0^3}{R_0^3} \tag{C.7}$$

Relating this to density via  $M_{\rm cloud} = 4/3 \, \pi r_0^3 \, \rho_{\rm tidal}$ , we see that

$$\rho_{\rm tidal} = \frac{3M_{\rm bh}}{\pi R_0^3} \tag{C.8}$$

## D

### Derivation of the Bremsstrahlung cooling function

We consider a scattering event between and electron of mass  $m_e$  and a proton of mass  $m_p$ , with an impact parameter b and relative velocity v. The electron will be accelerated by the proton by an amount  $a = q^2/m_e b$  for a time t = b/v, where q is the charge on each, assumed to be equal. The energy radiated during the encounter is given by Larmor's formula:

$$E \simeq \left(\frac{q^2 a^2}{c^3}\right) \left(\frac{b}{v}\right) \tag{D.1}$$

where c is the speed of light. On average, the impact parameter in terms of the number density of ions in the plasma can be approximated by  $b \approx n_i^{-1/3}$ , and to find the total energy radiated per unit volume we multiply by the number density of electrons,  $n_e$ , giving us

$$E_{\rm tot} \simeq \frac{q^6 n_i n_e}{c^3 m_e^2 v} \tag{D.2}$$

For a plasma in thermal equilibrium we can assume a Maxwell-Boltzmann distribution with speeds given by

$$v = \left(\frac{k_B T}{m_e}\right)^{1/2} \tag{D.3}$$

Therefore we have

$$E_{\rm tot} = \left(\frac{q^6}{m_e^2 c^3}\right) \left(\frac{m_e}{k_B T}\right)^{1/2} n_e n_i \tag{D.4}$$

Of course, not all electrons have the same energy, and the photons that are radiated will therefore also possess an energy range. The typical energy of an electron in our plasma is  $k_BT$ , and each collision lasts for a time  $\Delta t \sim b/v$ , so there will be very little radiation at frequencies greater than  $1/\Delta t$ , i.e.,  $\omega \lesssim (v/b)$ .

As a result the Bremssrahlung spectrum is flat for  $0 < \omega \lesssim k_B T/\hbar$  (since  $E = \hbar \omega \lesssim k_B T$ ) and falls rapidly for  $\omega \gtrsim k_B T/\hbar$ . An electron with typical energy  $k_B T$  cannot emit photons with frequency significantly higher than  $k_B T/\hbar$ . The total energy radiated must then be found by integrating over the frequency range  $0 < \omega \lesssim k_B T/\hbar$ :

$$E_{\rm rad} = \int_0^{\frac{k_B T}{\hbar}} \left(\frac{q^6}{m_e^2 c^3}\right) \left(\frac{m_e}{k_B T}\right)^{1/2} n_e n_i \,\mathrm{d}\omega \tag{D.5}$$

$$= \left(\frac{q^6}{m_e^2 c^3}\right) \left(\frac{m_e k_B T}{\hbar^2}\right)^{1/2} n_e n_i \tag{D.6}$$

per unit volume, per unit time. We therefore arrive at the conclusion that for Bremsstrahlung radiation from a plasma with number density n and temperature T, the energy emitted goes as

$$E_{\rm brem.} \propto n^2 T^{1/2} \tag{D.7}$$

### References

- Agertz, O., Moore, B., Stadel, J., Potter, D., Miniati, F., Read, J., Mayer, L., Gawryszczak, A., Kravtsov, A., Nordlund, Å., Pearce, F., Quilis, V., Rudd, D., Springel, V., Stone, J., Tasker, E., Teyssier, R., Wadsley, J., and Walder, R.: 2007, MNRAS 380, 963
- Albrecht, R., Albrecht, M., Dolensky, M., Micol, A., Pirenne, B., and Wicenec, A. J.: 1998, in U. Grothkopf, H. Andernach, S. Stevens-Rayburn, & M. Gomez (ed.), Library and Information Services in Astronomy III, Vol. 153 of Astronomical Society of the Pacific Conference Series, pp 261-+
- Alexander, R. D., Begelman, M. C., and Armitage, P. J.: 2006, ArXiv Astrophysics e-prints
- Alexander, R. D., Begelman, M. C., and Armitage, P. J.: 2007, ApJ 654, 907
- Arav, N.: 1996, ApJ 465, 617
- Arav, N., Korista, K. T., Barlow, T. A., and Begelman: 1995, Nature 376, 576
- Armstrong, J. T. and Barrett, A. H.: 1985, ApJS 57, 535
- Attwood, R. E., Goodwin, S. P., and Whitworth, A. P.: 2007, A&A 464, 447
- Baganoff, F. K., Maeda, Y., Morris, M., Bautz, M. W., Brandt, W. N., Cui, W., Doty, J. P., Feigelson, E. D., Garmire, G. P., Pravdo, S. H., Ricker, G. R., and Townsley, L. K.: 2003, ApJ 591, 891
- Bahcall, J. N. and Wolf, R. A.: 1976, ApJ 209, 214
- Balbus, S. A. and Hawley, J. F.: 1991, ApJ 376, 214
- Balsara, D. S.: 1995, Journal of Computational Physics 121, 357
- Bardeen, J. M. and Petterson, J. A.: 1975, *ApJL* 195, L65+
- Barnes, J. and Hut, P.: 1986, Nature 324, 446
- Barnes, J. E. and Hernquist, L. E.: 1991, *ApJL* 370, L65
- Barnes, J. E. and Hut, P.: 1989, ApJS 70, 389

Bartko, H.: 2010, ArXiv e-prints

- Bartko, H., Eisenhauer, F., Fritz, T., Genzel, R., Gillessen, S., Martins, F., Ott, T., Paumard, T., Pfuhl, O., and Trippe, S.: 2008, *Journal of Physics Conference Series* 131(1), 012010
- Bartko, H., Martins, F., Fritz, T. K., Genzel, R., Levin, Y., Perets, H. B., Paumard, T., Nayakshin, S., Gerhard, O., Alexander, T., Dodds-Eden, K., Eisenhauer, F., Gillessen, S., Mascetti, L., Ott, T., Perrin, G., Pfuhl, O., Reid, M. J., Rouan, D., Sternberg, A., and Trippe, S.: 2009, ApJ 697, 1741
- Bate, M. R.: 2009, MNRAS 397, 232
- Bate, M. R., Bonnell, I. A., and Price, N. M.: 1995, MNRAS 277, 362
- Bate, M. R. and Burkert, A.: 1997, MNRAS 288, 1060
- Baumgardt, H., Côté, P., Hilker, M., Rejkuba, M., Mieske, S., Djorgovski, S. G., and Stetson, P.: 2009, MNRAS 396, 2051
- Bell, K. R. and Lin, D. N. C.: 1994, ApJ 427, 987
- Bender, R., Kormendy, J., Bower, G., Green, R., Thomas, J., Danks, A. C., Gull, T., Hutchings, J. B., Joseph, C. L., Kaiser, M. E., Lauer, T. R., Nelson, C. H., Richstone, D., Weistrop, D., and Woodgate, B.: 2005, *ApJ* 631, 280
- Bertin, G.: 2000, Dynamics of Galaxies
- Binney, J. and Merrifield, M.: 1998, Galactic astronomy
- Binney, J. and Tremaine, S.: 1987, Galactic dynamics, Princeton University Press
- Binney, J. and Tremaine, S.: 2008, *Galactic Dynamics: Second Edition*, Princeton University Press
- Boley, A. C., Durisen, R. H., Nordlund, Å., and Lord, J.: 2007, ApJ 665, 1254
- Bondi, H.: 1952, MNRAS 112, 195
- Bondi, H. and Hoyle, F.: 1944, MNRAS 104, 273
- Bonnell, I. A., Dobbs, C. L., Robitaille, T. P., and Pringle, J. E.: 2006, MNRAS 365, 37
- Bonnell, I. A. and Rice, W. K. M.: 2008, Science 321, 1060
- Booth, C. M. and 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., and Lacey, C. G.: 2006, MNRAS 370, 645
- Bullock, J. S., Kolatt, T. S., Sigad, Y., Somerville, R. S., Kravtsov, A. V., Klypin, A. A., Primack, J. R., and Dekel, A.: 2001, MNRAS 321, 559
- Bushouse, H. A., Borne, K. D., Colina, L., Lucas, R. A., Rowan-Robinson, M., Baker, A. C., Clements, D. L., Lawrence, A., and Oliver, S.: 2002, ApJS 138, 1
- Cameron, A. G.: 1959, ApJ 130, 884
- Campbell, R. D., Le Mignant, D., van Dam, M. A., Conrad, A., Grace, K., Kassis, M., Lyke, J. E., Tran, H., Chin, J. C. Y., Chock, E., Doyle, S., Goodrich, R. W., Johansson, E. M., Krasuski, T., Melcher, C., Summers, D. M., Sumner, R. C., and Wizinowich, P. L.: 2008, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7016 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series
- Cappellaro, E.: 2003, in K. Weiler (ed.), Supernovae and Gamma-Ray Bursters, Vol. 598 of Lecture Notes in Physics, Berlin Springer Verlag, pp 37–46
- Chandrasekhar, S.: 1931, ApJ 74, 81

Chandrasekhar, S.: 1961, Hydrodynamic and hydromagnetic stability

- Chang, P., Murray-Clay, R., Chiang, E., and Quataert, E.: 2007, ApJ 668, 236
- Chen, Y., Wang, J., Yan, C., Hu, C., and Zhang, S.: 2009, ApJL 695, L130
- Churazov, E., Sazonov, S., Sunyaev, R., Forman, W., Jones, C., and Böhringer, H.: 2005, MNRAS 363, L91
- Cid Fernandes, R., Heckman, T., Schmitt, H., González Delgado, R. M., and Storchi-Bergmann, T.: 2001, ApJ 558, 81
- Clarke, C. J. and Carswell, R. F.: 2003, Principles of Astrophysical Fluid Dynamics
- Colina, L. and Wada, K.: 2000, ApJ 529, 845
- Collin, S. and Zahn, J.: 1999, A&A 344, 433
- Collin, S. and Zahn, J.: 2008, A&A 477, 419
- Collin-Souffrin, S. and Dumont, A. M.: 1990, A&A 229, 292

- Coskunoglu, B., Ak, S., Bilir, S., Karaali, S., Yaz, E., Gilmore, G., Seabroke, G. M., Bienayme, O., Bland-Hawthorn, J., Campbell, R., Freeman, K. C., Gibson, B., Grebel, E. K., Munari, U., Navarro, J. F., Parker, Q. A., Siebert, A., Siviero, A., Steinmetz, M., Watson, F. G., Wyse, R. F. G., and Zwitter, T.: 2010, ArXiv e-prints
- Cossins, P., Lodato, G., and Clarke, C.: 2010, MNRAS 401, 2587
- Cossins, P., Lodato, G., and Clarke, C. J.: 2009, MNRAS 393, 1157
- Cossins, P. J.: 2010a, ArXiv e-prints
- Cossins, P. J.: 2010b, pp 26–32
- Croton, D. J., Springel, V., White, S. D. M., De Lucia, G., Frenk, C. S., Gao, L., Jenkins, A., Kauffmann, G., Navarro, J. F., and Yoshida, N.: 2006, MNRAS 365, 11
- Cuadra, J., Armitage, P. J., and Alexander, R. D.: 2008, MNRAS 388, L64
- Cuadra, J., Nayakshin, S., Springel, V., and Di Matteo, T.: 2006, MNRAS 366, 358
- Cuadra, J., Nayakshin, S., and Sunyaev, R.: 2003, A&A 411, 405
- Cullen, L. and Dehnen, W.: 2010, MNRAS pp 1126-+
- Daly, R. A.: 1990, ApJ **355**, 416
- de Vaucouleurs, G.: 1948, Annales d'Astrophysique 11, 247
- DeBuhr, J., Quataert, E., Ma, C., and Hopkins, P.: 2009, ArXiv e-prints
- Deegan, P. and Nayakshin, S.: 2007, MNRAS 377, 897
- Dehnen, W.: 1993, MNRAS 265, 250
- Dehnen, W.: 2002, Journal of Computational Physics 179, 27
- Dekel, A., Birnboim, Y., Engel, G., Freundlich, J., Goerdt, T., Mumcuoglu, M., Neistein, E., Pichon, C., Teyssier, R., and Zinger, E.: 2009, *Nature* 457, 451
- Dekel, A. and Silk, J.: 1986, *ApJ* **303**, 39
- Demircan, O. and Kahraman, G.: 1991, Astrophysics and Space Science 181, 313
- Dobbs, C. L.: 2008, MNRAS 391, 844
- Eisenhauer, F., Genzel, R., Alexander, T., Abuter, R., Paumard, T., Ott, T., Gilbert, A., Gillessen, S., Horrobin, M., Trippe, S., Bonnet, H., Dumas, C., Hubin, N., Kaufer, A., Kissler-Patig, M., Monnet, G., Ströbele, S., Szeifert, T., Eckart, A., Schödel, R., and Zucker, S.: 2005, ApJ 628, 246

- Eisenhauer, F., Schödel, R., Genzel, R., Ott, T., Tecza, M., Abuter, R., Eckart, A., and Alexander, T.: 2003, *ApJL* 597, L121
- Elmegreen, B. G. and Scalo, J.: 2004, *AAR&A* 42, 211
- Everett, J. E.: 2007, Ap&SS 311, 269
- Fabian, A. C., Nandra, K., Reynolds, C. S., Brandt, W. N., Otani, C., Tanaka, Y., Inoue, H., and Iwasawa, K.: 1995, MNRAS 277, L11
- Falcke, H. and Melia, F.: 1997, ApJ 479, 740
- Fan, X., Strauss, M. A., Becker, R. H., White, R. L., Gunn, J. E., Knapp, G. R., Richards, G. T., Schneider, D. P., Brinkmann, J., and Fukugita, M.: 2006, AJ 132, 117
- Fan, X., Strauss, M. A., Schneider, D. P., Becker, R. H., White, R. L., Haiman, Z., Gregg, M., Pentericci, L., Grebel, E. K., Narayanan, V. K., Loh, Y., Richards, G. T., Gunn, J. E., Lupton, R. H., Knapp, G. R., Ivezić, Ž., Brandt, W. N., Collinge, M., Hao, L., Harbeck, D., Prada, F., Schaye, J., Strateva, I., Zakamska, N., Anderson, S., Brinkmann, J., Bahcall, N. A., Lamb, D. Q., Okamura, S., Szalay, A., and York, D. G.: 2003, AJ 125, 1649
- Farrah, D., Afonso, J., Efstathiou, A., Rowan-Robinson, M., Fox, M., and Clements, D.: 2003, MNRAS 343, 585
- Ferrarese, L. and Merritt, D.: 2000, ApJL 539, L9
- Figer, D. F.: 2004, in ASP Conf. Ser. 322: The Formation and Evolution of Massive Young Star Clusters, pp 49-+
- Forgan, D. and Rice, K.: 2009, MNRAS 400, 2022
- Frank, J., King, A., and Raine, D. J.: 2002, Accretion Power in Astrophysics: Third Edition, Accretion Power in Astrophysics, by Juhan Frank and Andrew King and Derek Raine, pp. 398. ISBN 0521620538. Cambridge, UK: Cambridge University Press, February 2002.
- Freitag, M., Amaro-Seoane, P., and Kalogera, V.: 2006, ApJ 649, 91
- Frenk, C. S., White, S. D. M., Bode, P., Bond, J. R., Bryan, G. L., Cen, R., Couchman, H. M. P., Evrard, A. E., Gnedin, N., Jenkins, A., Khokhlov, A. M., Klypin, A., Navarro, J. F., Norman, M. L., Ostriker, J. P., Owen, J. M., Pearce, F. R., Pen, U., Steinmetz, M., and et al.: 1999, ApJ 525, 554

Gammie, C. F.: 2001, ApJ 553, 174
- Gebhardt, K., Bender, R., Bower, G., Dressler, A., Faber, S. M., Filippenko, A. V., Green, R., Grillmair, C., Ho, L. C., Kormendy, J., Lauer, T. R., Magorrian, J., Pinkney, J., Richstone, D., and Tremaine, S.: 2000, *ApJL* 539, L13
- Genzel, R.: 2010, in Dynamics from the Galactic Center to the Milky Way Halo
- Genzel, R., Schödel, R., Ott, T., Eisenhauer, F., Hofmann, R., Lehnert, M., Eckart, A., Alexander, T., Sternberg, A., Lenzen, R., Clénet, Y., Lacombe, F., Rouan, D., Renzini, A., and Tacconi-Garman, L. E.: 2003, ApJ 594, 812
- Genzel, R., Thatte, N., Krabbe, A., Kroker, H., and Tacconi-Garman, L. E.: 1996, ApJ 472, 153
- Gerhard, O.: 2001, ApJ 546, L39
- Gerwin, R. A.: 1968, Reviews of Modern Physics 40, 652
- Ghez, A. M., Becklin, E., Duchêne, G., Hornstein, S., Morris, M., Salim, S., and Tanner, A.: 2003a, Astron. Nachr., Vol. 324, No. S1 (2003), Special Supplement "The central 300 parsecs of the Milky Way", Eds. A. Cotera, H. Falcke, T. R. Geballe, S. Markoff (astro-ph/0303151)
- Ghez, A. M., Duchêne, G., Matthews, K., Hornstein, S. D., Tanner, A., Larkin, J., Morris,
  M., Becklin, E. E., Salim, S., Kremenek, T., Thompson, D., Soifer, B. T., Neugebauer,
  G., and McLean, I.: 2003b, ApJ 586, L127
- Ghez, A. M., Klein, B. L., Morris, M., and Becklin, E. E.: 1998, ApJ 509, 678
- Ghez, A. M., Salim, S., Hornstein, S. D., Tanner, A., Lu, J. R., Morris, M., Becklin, E. E., and Duchêne, G.: 2005a, ApJ 620, 744
- Ghez, A. M., Salim, S., Hornstein, S. D., Tanner, A., Lu, J. R., Morris, M., Becklin, E. E., and Duchêne, G.: 2005b, ApJ 620, 744
- Gillessen, S., Eisenhauer, F., Fritz, T. K., Bartko, H., Dodds-Eden, K., Pfuhl, O., Ott, T., and Genzel, R.: 2009a, ApJL 707, L114
- Gillessen, S., Eisenhauer, F., Trippe, S., Alexander, T., Genzel, R., Martins, F., and Ott, T.: 2009b, ApJ 692, 1075
- Gingold, R. A. and Monaghan, J. J.: 1977, MNRAS 181, 375
- Gingold, R. A. and Monaghan, J. J.: 1983, MNRAS 204, 715

González Delgado, R. M., Heckman, T., and Leitherer, C.: 2001, ApJ 546, 845

Goodman, J.: 2003, MNRAS 339, 937

Gorda, S. Y. and Svechnikov, M. A.: 1998, Astronomy Reports 42, 793

- Greenhill, L. J.: 2005, in J. Romney & M. Reid (ed.), Future Directions in High Resolution Astronomy, Vol. 340 of Astronomical Society of the Pacific Conference Series, pp 203-+
- Gültekin, K., Richstone, D. O., Gebhardt, K., Lauer, T. R., Tremaine, S., Aller, M. C., Bender, R., Dressler, A., Faber, S. M., Filippenko, A. V., Green, R., Ho, L. C., Kormendy, J., Magorrian, J., Pinkney, J., and Siopis, C.: 2009, *ApJ* 698, 198
- Gürkan, M. A. and Rasio, F. A.: 2005, ApJ 628, 236
- Häring, N. and Rix, H.-W.: 2004, ApJL 604, L89
- Hasegawa, T., Sato, F., Whiteoak, J. B., and Miyawaki, R.: 1994, ApJL 429, L77
- Heckman, T. M.: 2004, Coevolution of Black Holes and Galaxies pp 358-+
- Hernquist, L.: 1990, ApJ 356, 359
- Hernquist, L. and Katz, N.: 1989, ApJS 70, 419
- Herrnstein, J. R., Moran, J. M., Greenhill, L. J., and Trotter, A. S.: 2005, ApJ 629, 719
- Hobbs, A. and Nayakshin, S.: 2009, MNRAS 394, 191
- Hobson, M. P., Efstathiou, G. P., and Lasenby, A. N.: 2006, General Relativity
- Hockney, R. W. and Eastwood, J. W.: 1981, Computer Simulation Using Particles
- Hoffmeister, C.: 1929, Astronomische Nachrichten 236, 233
- Hopkins, P. F. and Hernquist, L.: 2006, *ApJS* **166**, 1
- Hopkins, P. F. and Quataert, E.: 2009, ArXiv e-prints
- Hopkins, P. F. and Quataert, E.: 2010, MNRAS 405, L41
- Hoyle, F. and Lyttleton, R. A.: 1939, in Proceedings of the Cambridge Philosophical Society, Vol. 34 of Proceedings of the Cambridge Philosophical Society, pp 405-+
- Hubble, E.: 1929, Proceedings of the National Academy of Science 15, 168
- Hupfer, C., Käpylä, P. J., and Stix, M.: 2006, A&A 459, 935
- Hut, P. and Makino, J.: 1999, Science 283, 501

- Ishizuki, S., Kawabe, R., Ishiguro, M., Okumura, S. K., and Morita, K.: 1990, *Nature* **344**, 224
- Jaffe, W.: 1983, MNRAS 202, 995
- Jeans, J. H.: 1902, Royal Society of London Philosophical Transactions Series A 199, 1
- Kauffmann, G., Heckman, T. M., Tremonti, C., Brinchmann, J., Charlot, S., White, S. D. M., Ridgway, S. E., Brinkmann, J., Fukugita, M., Hall, P. B., Ivezić, Ž., Richards, G. T., and Schneider, D. P.: 2003, MNRAS 346, 1055
- Kawabe, R., Ishiguro, M., Ishizuki, S., Okumura, S. K., Kasuga, T., Morita, K., and Tosaki, T.: 1991, in A. D. Haschick & P. T. P. Ho (ed.), Atoms, Ions and Molecules: New Results in Spectral Line Astrophysics, Vol. 16 of Astronomical Society of the Pacific Conference Series, pp 19–+
- Kawakatu, N. and Wada, K.: 2008, ApJ 681, 73
- Kereš, D., Katz, N., Weinberg, D. H., and Davé, R.: 2005, MNRAS 363, 2
- Kerr, R. P.: 1963, Phys. Rev. Lett. 11(5), 237
- Kiessling, M.: 1999, ArXiv Astrophysics e-prints
- Kim, S. S., Figer, D. F., and Morris, M.: 2004, ApJ 607, L123
- Kim, S. S. and Morris, M.: 2003, ApJ 597, 312
- King, A.: 2003, ApJL 596, L27
- King, A.: 2005, ApJL 635, L121
- King, A. R., Lubow, S. H., Ogilvie, G. I., and Pringle, J. E.: 2005, MNRAS 363, 49
- King, A. R. and Pounds, K. A.: 2003, MNRAS 345, 657
- King, A. R. and Pringle, J. E.: 2007, MNRAS 377, L25
- King, A. R., Pringle, J. E., and Hofmann, J. A.: 2008, MNRAS 385, 1621
- King, I. R.: 1966, AJ 71, 64
- Kolmogorov, A.: 1941, Akademiia Nauk SSSR Doklady 30, 301
- Kolykhalov, P. I. and Sunyaev, R. A.: 1980, Soviet Astron. Lett. 6, 357
- Komissarov, S. S.: 2010, ArXiv e-prints

Kormendy, J.: 2004, Coevolution of Black Holes and Galaxies pp 1-+

- Krabbe, A., Genzel, R., Eckart, A., Najarro, F., Lutz, D., Cameron, M., Kroker, H., Tacconi-Garman, L. E., Thatte, N., Weitzel, L., Drapatz, S., Geballe, T., Sternberg, A., and Kudritzki, R.: 1995, ApJ 447, L95
- Krolik, J. H.: 1999, Active galactic nuclei : from the central black hole to the galactic environment, Active galactic nuclei : from the central black hole to the galactic environment /Julian H. Krolik. Princeton, N. J. : Princeton University Press, c1999.
- Krumholz, M. R. and Bonnell, I. A.: 2007, ArXiv e-prints 712
- Krumholz, M. R., McKee, C. F., and Klein, R. I.: 2005, ApJ 618, 757
- Kurk, J. D., Walter, F., Fan, X., Jiang, L., Riechers, D. A., Rix, H.-W., Pentericci, L., Strauss, M. A., Carilli, C., and Wagner, S.: 2007, ApJ 669, 32
- Kylafis, N. D. and Pavlakis, K. G.: 1997, Communications of the Konkoly Observatory Hungary 100, 231
- Lamers, H. J. G. L. M. and Cassinelli, J. P.: 1999, Introduction to Stellar Winds
- Landau, L. D. and Lifshitz, E. M.: 1959, Fluid mechanics
- Larson, R. B.: 1969, MNRAS 145, 271
- Lattanzio, J. C., Monaghan, J. J., Pongracic, H., and Schwarz, M. P.: 1985, MNRAS 215, 125
- Leitherer, C., Robert, C., and Drissen, L.: 1992, ApJ 401, 596
- Levin, Y.: 2006, ArXiv Astrophysics e-prints, astro-ph/0603583
- Levin, Y. and Beloborodov, A. M.: 2003, ApJ 590, L33
- Levin, Y., Wu, A., and Thommes, E.: 2005, ApJ 635, 341
- Levine, R., Gnedin, N. Y., and Hamilton, A. J. S.: 2010, ApJ 716, 1386
- Li, Y., Hernquist, L., Robertson, B., Cox, T. J., Hopkins, P. F., Springel, V., Gao, L., Di Matteo, T., Zentner, A. R., Jenkins, A., and Yoshida, N.: 2007, ApJ 665, 187
- Lightman, A. P. and Shapiro, S. L.: 1977, *ApJ* **211**, 244
- Lin, C. C. and Shu, F. H.: 1966, Proceedings of the National Academy of Science 55, 229
- Lin, D. N. C., Pringle, J. E., and Rees, M. J.: 1988, ApJ 328, 103
- Lo, K. Y., Cheung, K. W., Masson, C. R., Phillips, T. G., Scott, S. L., and Woody, D. P.: 1987, *ApJ* **312**, 574

Lodato, G.: 2007, Nuovo Cimento Rivista Serie 30, 293

- Lodato, G. and Bertin, G.: 2003, A&A 398, 517
- Lodato, G. and Rice, W. K. M.: 2004, MNRAS 351, 630
- Loeb, A.: 2004, MNRAS 350, 725
- Lu, J., Ghez, A., Hornstein, S., Morris, M., Matthews, K., Thompson, D., and Becklin, E.: 2006, JPhCS 54, 279
- Lucy, L. B.: 1977, AJ 82, 1013
- Lynden-Bell, D.: 1969, Nature 223, 690
- Lynden-Bell, D. and Pringle, J. E.: 1974, MNRAS 168, 603
- Lynden-Bell, D. and Rees, M. J.: 1971, MNRAS 152, 461
- Lynden-Bell, D. and Wood, R.: 1968, MNRAS 138, 495
- Mac Low, M. and Klessen, R. S.: 2004, Reviews of Modern Physics 76, 125
- Madau, P. and Rees, M. J.: 2001, ApJL 551, L27
- Magorrian, J. and Ballantyne, D.: 2001, MNRAS 322, 702
- Magorrian, J., Tremaine, S., Richstone, D., Bender, R., Bower, G., Dressler, A., Faber, S. M., Gebhardt, K., Green, R., Grillmair, C., Kormendy, J., and Lauer, T.: 1998, AJ 115, 2285
- Maiolino, R., Alonso-Herrero, A., Anders, S., Quillen, A., Rieke, M. J., Rieke, G. H., and Tacconi-Garman, L. E.: 2000, ApJ 531, 219
- Maoz, E.: 1998, ApJL **494**, L181+
- Masunaga, H., Miyama, S. M., and Inutsuka, S.-I.: 1998, ApJ 495, 346
- Mattila, S. and Meikle, W. P. S.: 2001, MNRAS 324, 325
- Mayer, L., Kazantzidis, S., Escala, A., and Callegari, S.: 2010, Nature 466, 1082
- McGaugh, S. S., Schombert, J. M., de Blok, W. J. G., and Zagursky, M. J.: 2009, ArXiv *e-prints*
- McKee, C. F. and Ostriker, E. C.: 2007a, *AAR&A* 45, 565
- McKee, C. F. and Ostriker, E. C.: 2007b, AAR&A 45, 565

- McKee, C. F. and Ostriker, J. P.: 1977, ApJ 218, 148
- Meglicki, Z., Wickramasinghe, D., and Bicknell, G. V.: 1993, MNRAS 264, 691
- Melia, F.: 2003, The black hole at the center of our galaxy
- Melia, F.: 2007, ArXiv e-prints
- Melia, F. and Falcke, H.: 2001, ARA&A 39, 309
- Mihos, J. C. and Hernquist, L.: 1994, *ApJL* **425**, L13
- Milosavljevic, M. and Merritt, D.: 2001, in Bulletin of the American Astronomical Society, Vol. 33 of Bulletin of the American Astronomical Society, pp 913-+
- Miralda-Escudé, J. and Gould, A.: 2000, ApJ 545, 847
- Miyoshi, M., Moran, J., Herrnstein, J., Greenhill, L., Nakai, N., Diamond, P., and Inoue, M.: 1995, Nature 373, 127
- Monaghan, J.: 1985, Computer Physics Reports 3, 71
- Monaghan, J. J.: 1989, Journal of Computational Physics 82, 1
- Monaghan, J. J.: 1992, AAR&A 30, 543
- Monaghan, J. J.: 1997, Journal of Computational Physics 136, 298
- Monaghan, J. J. and Gingold, R. A.: 1983, Journal of Computational Physics 52, 374
- Monaghan, J. J. and Lattanzio, J. C.: 1985, A&A 149, 135
- Morris, M.: 1993, ApJ 408, 496
- Muxlow, T., Beswick, R. J., Richards, A. M. S., and Thrall, H. J.: 2006, in *Proceedings* of the 8th European VLBI Network Symposium
- Narayan, R.: 2002, in Lighthouses of the universe: the most luminous celestial objects and their use for cosmology, ed. M.Gilfanov, R. Sunyaev & E. Churazov (Berlin: Springer), p. 405
- Navarro, J. F., Frenk, C. S., and White, S. D. M.: 1996, ApJ 462, 563
- Navarro, J. F. and White, S. D. M.: 1993, MNRAS 265, 271
- Nayakshin, S.: 2006, MNRAS 372, 143
- Nayakshin, S. and Cuadra, J.: 2005, A&A 437, 437

- Nayakshin, S., Cuadra, J., and Springel, V.: 2007, MNRAS 379, 21
- Nayakshin, S., Dehnen, W., Cuadra, J., and Genzel, R.: 2006, MNRAS 366, 1410
- Nayakshin, S. and King, A.: 2007, ArXiv e-prints
- Nayakshin, S. and Power, C.: 2009, ArXiv e-prints
- Nayakshin, S. and Power, C.: 2010, MNRAS 402, 789
- Nayakshin, S. and Sunyaev, R.: 2005, MNRAS 364, L23
- Nayakshin, S., Wilkinson, M. I., and King, A.: 2009, MNRAS 398, L54
- Nelson, R. P. and Papaloizou, J. C. B.: 1994, MNRAS 270, 1
- Newton, I.: 1999, The Principia : mathematical principles of natural philosophy
- O'Brien, P. T., Reeves, J. N., Simpson, C., and Ward, M. J.: 2005, MNRAS 360, L25
- Oke, J. B.: 1963, Nature 197, 1040
- Oppenheimer, J. R. and Volkoff, G. M.: 1939, Physical Review 55, 374
- O'Shea, B. W., Bryan, G., Bordner, J., Norman, M. L., Abel, T., Harkness, R., and Kritsuk, A.: 2004, ArXiv Astrophysics e-prints
- Paczyński, B.: 1978, Acta Astron. 28, 91
- Padmanabhan, T.: 2000, Theoretical Astrophysics Volume 1, Astrophysical Processes
- Padmanabhan, T.: 2001, Theoretical Astrophysics Volume 2, Stars and Stellar Systems
- Padmanabhan, T.: 2002, Theoretical Astrophysics, Volume III: Galaxies and Cosmology
- Parker, E. N.: 1960, ApJ 132, 821
- Pastorini, G., Marconi, A., Capetti, A., Axon, D. J., Alonso-Herrero, A., Atkinson, J., Batcheldor, D., Carollo, C. M., Collett, J., Dressel, L., Hughes, M. A., Macchetto, D., Maciejewski, W., Sparks, W., and van der Marel, R.: 2007, A&A 469, 405
- Paumard, T., Genzel, R., Martins, F., Nayakshin, S., Beloborodov, A. M., Levin, Y., Trippe, S., Eisenhauer, F., Ott, T., Gillessen, S., Abuter, R., Cuadra, J., Alexander, T., and Sternberg, A.: 2006, ApJ 643, 1011
- Paumard, T., Maillard, J.-P., and Morris, M.: 2004, A&A 426, 81
- Peebles, P. J. E.: 1972, ApJ 178, 371

Peebles, P. J. E.: 1993, Principles of physical cosmology

- Penrose, R.: 1973, in D. J. Hegyi (ed.), Sixth Texas Symposium on Relativistic Astrophysics, Vol. 224 of Annals of the New York Academy of Sciences, pp 125-+
- Peterson, B. M.: 1997, An Introduction to Active Galactic Nuclei
- Peterson, B. M. and Horne, K.: 2004, Astronomische Nachrichten 325, 248
- Plummer, H. C.: 1911, MNRAS 71, 460
- Pounds, K. A., Reeves, J. N., King, A. R., Page, K. L., O'Brien, P. T., and Turner, M. J. L.: 2003, MNRAS 345, 705
- Power, C., Nayakshin, S., and King, A.: 2010, ArXiv e-prints
- Price, D.: 2005, ArXiv Astrophysics e-prints
- Pringle, J. E.: 1981, AAR&A 19, 137
- Pringle, J. E. and King, A.: 2003, Astrophysical Flows
- Rafikov, R. R.: 2009, ApJ 704, 281
- Read, J. I., Hayfield, T., and Agertz, O.: 2010, MNRAS 405, 1513
- Rees, M. J.: 1989, MNRAS 239, 1P
- Reid, M. J., Readhead, A. C. S., Vermeulen, R. C., and Treuhaft, R. N.: 1999, ApJ 524, 816
- Rice, W. K. M., Lodato, G., and Armitage, P. J.: 2005, MNRAS 364, L56
- Ricotti, M.: 2007, ApJ 662, 53
- Robert, C., Leitherer, C., and Heckman, T. M.: 1993, ApJ 418, 749
- Rosswog, S.: 2009, New Ast. Rev. 53, 78
- Rosswog, S. and Bruggen, M.: 2003, Introduction to High-Energy Astrophysics
- Rovelli, C.: 2000, ArXiv General Relativity and Quantum Cosmology e-prints
- Ruffert, M.: 1994, ApJ 427, 342
- Rybicki, G. B. and Lightman, A. P.: 1986, Radiative Processes in Astrophysics

Sakamoto, K., Okumura, S. K., Ishizuki, S., and Scoville, N. Z.: 1999, ApJS 124, 403

Salpeter, E. E.: 1964, ApJ 140, 796

- Sanders, D. B., Soifer, B. T., Elias, J. H., Madore, B. F., Matthews, K., Neugebauer, G., and Scoville, N. Z.: 1988, ApJ 325, 74
- Sazonov, S. Y., Ostriker, J. P., Ciotti, L., and Sunyaev, R. A.: 2005, MNRAS 358, 168
- Schödel, R., Ott, T., Genzel, R., Hofmann, R., Lehnert, M., Eckart, A., Mouawad, N., Alexander, T., Reid, M. J., Lenzen, R., Hartung, M., Lacombe, F., Rouan, D., Gendron, E., Rousset, G., Lagrange, A.-M., Brandner, W., Ageorges, N., Lidman, C., Moorwood, A. F. M., Spyromilio, J., Hubin, N., and Menten, K. M.: 2002, *Nature* **419**, 694
- Schartmann, M., Meisenheimer, K., Klahr, H., Camenzind, M., Wolf, S., and Henning, T.: 2009, MNRAS 393, 759
- Schawinski, K.: 2010, ArXiv e-prints
- Schmidt, M.: 1963, Nature 197, 1040
- Schmidt, M.: 1969, AAR&A 7, 527
- Schödel, R., Eckart, A., Alexander, T., Merritt, D., Genzel, R., Sternberg, A., Meyer, L., Kul, F., Moultaka, J., Ott, T., and Straubmeier, C.: 2007, A&A 469, 125
- Seyfert, C. K.: 1943, ApJ 97, 28
- Shakura, N. I. and Sunyaev, R. A.: 1973, A&A 24, 337
- Shapiro, S. L. and Teukolsky, S. A.: 1983, Black holes, white dwarfs, and neutron stars: The physics of compact objects
- Shields, J. C.: 1992, *ApJL* **399**, L27
- Shlosman, I. and Begelman, M. C.: 1987, Nature 329, 810
- Shlosman, I., Begelman, M. C., and Frank, J.: 1990, Nature 345, 679
- Shlosman, I., Frank, J., and Begelman, M. C.: 1989a, in D. E. Osterbrock & J. S. Miller (ed.), Active Galactic Nuclei, Vol. 134 of IAU Symposium, pp 462-+
- Shlosman, I., Frank, J., and Begelman, M. C.: 1989b, Nature 338, 45
- Sijacki, D., Springel, V., and Haehnelt, M. G.: 2009, MNRAS 400, 100
- Silk, J.: 1977, ApJ 214, 152
- Silk, J. and Rees, M. J.: 1998, A&A **331**, L1

- Soltan, A.: 1982, MNRAS 200, 115
- Springel, V.: 2005, MNRAS 364, 1105
- Springel, V., Di Matteo, T., and Hernquist, L.: 2005, MNRAS 361, 776
- Springel, V. and Hernquist, L.: 2002, MNRAS 333, 649
- Springel, V., Yoshida, N., and White, S. D. M.: 2001, New Astronomy 6, 79
- Stern, B. E., Poutanen, J., Svensson, R., Sikora, M., and Begelman, M. C.: 1995, ApJL 449, L13+
- Stolte, A., Ghez, A. M., Morris, M., Lu, J. R., Brandner, W., and Matthews, K.: 2008, *ApJ* 675, 1278
- Sundman, K.: 1913, Acta Mathematica 36, 105, 10.1007/BF02422379
- Tasker, E. J., Brunino, R., Mitchell, N. L., Michielsen, D., Hopton, S., Pearce, F. R., Bryan, G. L., and Theuns, T.: 2008, MNRAS 390, 1267
- Thompson, R. I.: 2005, in R. de Grijs & R. M. González Delgado (ed.), Starbursts: From 30 Doradus to Lyman Break Galaxies, Vol. 329 of Astrophysics and Space Science Library, pp 289–+
- Thompson, T. A., Quataert, E., and Murray, N.: 2005, ApJ 630, 167
- Tonry, J. L., Dressler, A., Blakeslee, J. P., Ajhar, E. A., Fletcher, A. B., Luppino, G. A., Metzger, M. R., and Moore, C. B.: 2001, ApJ 546, 681
- Toomre, A.: 1964, ApJ 139, 1217
- Tremaine, S.: 1995, AJ 110, 628
- Tremaine, S., Gebhardt, K., Bender, R., Bower, G., Dressler, A., Faber, S. M., Filippenko, A. V., Green, R., Grillmair, C., Ho, L. C., Kormendy, J., Lauer, T. R., Magorrian, J., Pinkney, J., and Richstone, D.: 2002, ApJ 574, 740
- Urry, C. M. and Padovani, P.: 1995, PASP 107, 803
- Veilleux, S.: 2008, New Ast. Rev. 52, 289
- Volonteri, M.: 2010, A&A Rev. 18, 279
- Volonteri, M., Lodato, G., and Natarajan, P.: 2008, MNRAS 383, 1079
- Volonteri, M., Madau, P., Quataert, E., and Rees, M. J.: 2005, ApJ 620, 69

von Neumann, J. and Richtmyer, R. D.: 1950, Journal of Applied Physics 21, 232

Wachter, K. W., Strauss, M. A., and Filippenko, A. V.: 1988, ApJ 330, 91

Wada, K.: 1994, PASJ 46, 165

Wada, K.: 2004, Coevolution of Black Holes and Galaxies pp 186-+

- Wada, K. and Fukuda, H.: 2001, in M. Umemura & H. Susa (ed.), The Physics of Galaxy Formation, Vol. 222 of Astronomical Society of the Pacific Conference Series, pp 261-+
- Wada, K. and Habe, A.: 1992, MNRAS 258, 82
- Wada, K. and Habe, A.: 1993, in H. Dejonghe & H. J. Habing (ed.), Galactic Bulges, Vol. 153 of IAU Symposium, pp 285-+
- Wardle, M. and Yusef-Zadeh, F.: 2008, ApJL
- White, S. D. M. and Frenk, C. S.: 1991, ApJ 379, 52
- Williams, J. P., Blitz, L., and McKee, C. F.: 2000, Protostars and Planets IV pp 97-+
- Yu, Q. and Tremaine, S.: 2002, MNRAS 335, 965
- Zel'Dovich, Y. B.: 1964, Soviet Physics Doklady 9, 195