## Carnegie Mellon University MELLON COLLEGE OF SCIENCE

## THESIS

### SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

## DOCTOR OF PHILOSOPHY IN THE FIELD OF PHYSICS

TITLE: "Petascale Cosmological Hydrodynamic Simulation of Quasars"

PRESENTED BY: Yu Feng

#### ACCEPTED BY THE DEPARTMENT OF PHYSICS

Rupert Croft	7/30/14
RUPERT CROFT, CHAIR PROFESSOR	DATE

Stephen Garoff	7/30/14
STEPHEN GAROFF, DEPT HEAD	DATE

APPROVED BY THE COLLEGE COUNCIL

FRED GILMAN, DEAN

# Petascale Cosmological Hydrodynamic Simulation of Quasars

by

Yu Feng

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Carnegie Mellon University Department of Physics Pittsburgh, Pennsylvania

Advised by Professor Rupert Croft Advised by Professor Tiziana DiMatteo

July 17, 2014

#### Abstract

This thesis presents a theoretical investigation of supermassive black holes and the quasars they power through cosmological hydrodynamic simulations on petascale supercomputers. As the size of simulations increase, visualization and interaction with the data become difficult. We developed interactive visualization software on top of existing image deep-zoom and tagging infrastructure libraries and platforms, specifically in the context of cosmological hydrodynamic simulations. The visualization tools were instrumental in establishing the cold flow fueling model of high redshift quasar growth. We then proceed to further study this growth model with high resolution zoom-in re-simulations, and report that the cold flows feeding supermassive blackholes are numerically stable. We studied the morphology of HII regions with a parallel implementation of the ray tracing scheme SPHRAY, and found that the HII regions due to high redshift quasars are smoother and more extended that those due to galaxies. Finally, we developed a set of tools to produce and verify mocked correlated quasars and Lyman- $\alpha$  forest. We fit the linear theory correlation function of the Quasar-Quasar and Forest-Forest auto-correlation and Quasar-Forest cross-correlation of the mocked transmission fraction and quasar locations up to 160 Mpc/h.

## Acknowledgments

First and foremost I would like to thank my wonderful girlfriend and wife Jiani Niu for her incredible support throughout my years in Pittsburgh. Her continual support was invaluable, and I couldn't have done it without her.

I thank my advisors, Rupert Croft and Tiziana Di Matteo for their guidance and assistant on every level of the process. Thank you to all my colleagues and collaborators at Carnegie Mellon and abroad, for their ideas, discussions, and enrichment of my work. An extra thank you to my thesis committee, Hy Trac and Andrew Zentner for their questions, comments and suggestions.

Finally, I would like to thank my family for their love and encouragement during this process, and my life leading up to it.

This work has been funded by the National Science Foundation (NSF) PetaApps programme, OCI-0749212 and by NSF AST-1009781, and by the Moore foundation. The following computer resources were used in this research: BlueWaters (National Center for Supercomputing Applications), Kraken (National Institute for Computational Sciences), Blacklight (Pittsburgh Super-computing Center), Coma (Carnegie Mellon University), Warp (Carnegie Mellon University). Development of this work has made extensive use of the Bruce and Astrid McWilliams eScience Video Facility at Carnegie Mellon University.

# Contents

1	Intr	oducti	ion	1
	1.1	The S	tandard Model of Cosmology	1
		1.1.1	Components of the Universe	1
		1.1.2	The Hubble Parameter	3
		1.1.3	Cosmic Initial Condition	3
	1.2	Simula	ating the Universe with Supercomputers	4
		1.2.1	Motivation and History	4
		1.2.2	Contemporary Distributed Supercomputer Systems	5
		1.2.3	Domain Decomposition on Supercomputers	7
	1.3	Model	ling Physics Processes with Cosmological Simulation Algorithms	9
		1.3.1	Initial Condition of Simulations	9
		1.3.2	Gravity	11
		1.3.3	Hydrodynamics	13
		1.3.4	Atomic Cooling	14
		1.3.5	Feedback from Star Formation and Galaxies	16
		1.3.6	Feedback from Supermassive Black Holes	17
		1.3.7	Ultra-violet Radiation and Radiative Transfer	18
	1.4	Cosmo	ological Simulation Data Analysis Techniques	19
		1.4.1	Simulations as <i>Big Data</i>	19
		1.4.2	Visualization	19
	1.5	Astrop	physics	20
		1.5.1	Cold Flows and Quasars	20
		1.5.2	Re-ionization	21
		1.5.3	Lyman- $\alpha$ Forest	22
	1.6	Thesis	Overview	24
		1.6.1	Motivation	24
		1.6.2	Thesis plan	24
<b>2</b>	Visi	ualizat	ion and Interaction with Petascale Simulations	25
	2.1	Introd	uction	25
	2.2	Simula	ation	25
		2.2.1	MassiveBlack	26

		2.2.2 E5
	2.3	Spatial Domain Remapping 27
	2.4	Rasterization
	2.5	Image Rendering and Layer Compositing
	2.6	Parallelism and Performance
		2.6.1 Rasterization in Parallel
		2.6.2 Performance
	2.7	Image and Animation Viewing
		2.7.1 Gigapan
		2.7.2 GigaPan Time Machine
	2.8	Interactive Simulation Browser
	2.9	Conclusions
2	Cos	mological Hydrodynamic Simulation: Cold Flow and Quasars 44
J	2 1	Initial Conditions
	0.1	3.1.1 Selection of Halos 44
		3.1.2 Generating the Initial Conditions 46
	32	Simulations 48
	0.2	3.2.1 Blackhole Accretion Model 49
		3.2.2 Merger and Accretion history 50
		3.2.3 Deposition of AGN Feedback Energy 50
		3.2.4 Resolution 53
		3.2.5 SPH formulation: Pressure-Entropy 55
		3.2.6 Seeding Mass
	3.3	Feeding Blackholes via Cold Flows 58
	3.4	Conclusion         67
4	Par E	allel Radiative Transfer: Growth and Anisotropy of Ionization
	<b>FFO</b>	ILS (U Hudrodynamic Simulation and Density Field 70
	4.1	Selection of Oussars 72
	4.2 1 3	Analytic Crowth of Strongron Spheros
	4.0	Redictive Transfer Simulation 76
	4.4	$\begin{array}{c} \text{Adiative frame of Scheme} \\ 1 \\ 4 \\ 4 \\ 1 \\ 1 \\ 2 \\ 2 \\ 3 \\ 3 \\ 3 \\ 3 \\ 3 \\ 3 \\ 3 \\ 3$
		4.4.2 Parametrization of the Ionized Bubble 76
		4.4.2 I arametrization of the folized Dubble
	15	4.4.5       Shot-hoise and Convergence       78         Pogulta and Discussion       70
	4.0	1     1
		4.5.2 MassiveBlack: Visualization 70
		4.5.2 Massive Black, visualization
		4.5.4 Massive Black: Apisotropy of Jonized Regions
	16	4.5.4 MassiveDiack. Anisotropy of follized Regions
	4.0	Source Lummosity Models

		4.6.1 Band luminosity of Quasars	91
		4.6.2 Flux of UV stellar photons	91
	4.7	Conclusions	92
<b>5</b>	Qua	asar Correlated Lyman-Alpha Forest	93
	5.1	Introduction	93
	5.2	The Multi-level Fluctuation Approximation	94
	5.3	From Density Field to Quasar Catalogue	95
	5.4	From Density Field to Lyman- $\alpha$ Forest	96
	5.5	Redshift Distortions	98
	5.6	Measuring Linear Theory Redshift Distortion Parameters	99
		5.6.1 Estimating the Covariance Matrix	99
		5.6.2 Best Fit to Linear Theory Correlation Functions	101
	5.7	Software Implementation and Data Format	102
		5.7.1 Software Implementation	102
		5.7.2 Data Format: Quasars	102
		5.7.3 Data Format: Forest Spectra	102
	5.8	Conclusions	103
6	Cor	clusion and Future Work	105
	6.1	Conclusions	105
	6.2	Future Directions	106

# List of Tables

1.1	$\Lambda CDM$ model parameters	1
2.1	Transformations	30
2.2	Time take to rasterize the MassiveBlack Simulation	36
3.1	Cosmology Parameters	46
3.2	Simulation matrix	49
3.3	Correlation between heating time and entering time	61
4.1	UV Flux of Sub-volumes	73
4.2	Correlation coefficients in convergence test	78
4.3	Band definitions	91
5.1	Break down of the Covariance Matrix	100
5.2	Best Fit of Redshift Distortion Parameters	101
5.3	List of Free parameters in the Mock Algorithm	104

# List of Figures

1.1	Components of the Universe	2
1.2	Fluctuations in the Cosmic Microwave Background	3
1.3	Quasar Luminosity Function at $z = 5.0$	4
1.4	Architecture of Supercomputer	6
1.5	Software Stack on a Supercomputer	8
1.6	Recursive Domain Decomposition	9
1.7	Power spectrum measured from observation	10
1.8	Long-range Gravity Force and Short-range Gravity Force	13
1.9	The Mixing Problem Demonstated with Kelvin-Helmholtz Instability	15
1.10	Cool Rate Function	16
1.11	Quasar Spectrum and Lyman- $\alpha$ Forest	23
2.1	Transformation of the Primitive Cell	27
2.2	Boundary effects and smoothed particles	29
2.3	Fiducial Color-map for the Gas Density Field	34
2.4	MassiveBlack Simulatoin Rasterizatoin Rate	36
2.5	GigaPan View of the MassiveBlack Simulatoin at $z = 4.75$	38
2.6	GigaPan Time Machine Animation of the E5 simulation	39
2.7	Interface of the Interactive Browser	41
2.8	Visualization of MassiveBlack II simulation	43
3.1	Mass of Black Holes and Mass of their Hosting Halos	45
3.2	Environment of Resimulated Halos	45
3.3	Multi-level Initial Condition	47
3.4	Merger Trees of Three Most Massive Black Holes	51
3.5	Accretion History and Prescription of Feedback Energy Deposition	52
3.6	Resolution dependence of accretion history	54
3.7	Effects of SPH formulation on accretion history	56
3.8	Accretion history with varying seeding mass	57
3.9	Visualization of Gas Density at Varying Resolutions	59
3.10	Time a Particle Enters a Halo and Time the Particle Heats Up	62
3.11	Properties of HALO and BH particles	64
3.12	The Trajectories of BH and HALO Particles (RA phase)	65

3.13	The Trajectories of BH and HALO Particles (EA phase)	66
4.1	Central and Total Ionizing Photon Flux	74
4.2	CIR Growth in a Uniform Density Field	80
4.3	Ionized Bubbles in MassiveBlack	81
4.4	Evolution of the Ionization Front	82
4.5	HII and HeIII CIR Front Radius	83
4.6	Correlation between HeII CIR and HII CIR	84
4.7	Evolution of Three Quasar driven sub-volumes	85
4.8	Anisotropy .vs. Radius	87
4.9	Angular Distribution of the HII Radius	89
4.10	Time Evolution of $x_{\rm HI}$	90
5.1	The Parameter $A$ and $B$ used in FGPA	98
5.2	Fitting to the transmission model. $\langle F \rangle$ and $\sigma_F(z)$	98
5.3	Estimated Covariance Matrix	00
5.4	Best fit Linear Theory averaged over 50 mocks	.01

# Chapter 1 Introduction

## 1.1 The Standard Model of Cosmology

The model of the Universe adopted in this thesis is the ACDM model. The model describes the components of the Universe, its initial state immediately after the Big Bang, and how the components evolve. A minimal set of parameters corresponding to this model is listed in Table 1.1. Measuring these parameters is a core task of modern day physical cosmology. [For in-depth discussion of modern cosmology, see Dodelson, 2003, Loeb and Furlanetto, 2013, Padmanabhan, 2006]

ACDM model become the standard model in cosmology following the discovery of accelerate expansion of the Universe in 1998[Perlmutter et al., 1999, Riess et al., 1998] with supernovae data, and consistent with a wide list of observations: WMAP[Hinshaw et al., 2013], Planck[Planck Collaboration et al., 2013] 2dFGRS[Cole et al., 2005], SDSS/BOSS [Reid et al., 2010, Eisenstein et al., 2005, Percival et al., 2010], and Lyman- $\alpha$  forest [Croft et al., 1999, Seljak et al., 2006, McDonald et al., 2006].

#### 1.1.1 Components of the Universe

The three major components of present day universe in the ACDM model are dark energy, cold dark mater, and baryonic matter. The amounts of these quantities are expressed as fractions of the critical density of the Universe, the density of matter

$\Omega_{\Lambda}$	Dark energy density
$\Omega_{\rm DM}$	Dark matter energy density
$\Omega_{\rm B}$	Matter energy density
h	Hubble Parameter
$\sigma_8$	Power spectrum measured with a $8{\rm Mpc/h}$ top hat window

Table 1.1: ACDM model parameters.



Figure 1.1: Components of the Universe as Measured by the WMAP experiment. [Hinshaw et al., 2013]

and energy which makes space Euclidean ("flat") at a given redshift. The current measurements of the fractions of the three components are: dark energy:  $\Omega_{\Lambda} \approx 0.72$ , cold dark matter:  $\Omega_M \approx 0.24$  and baryonic matter:  $\Omega_B \approx 0.05$ . (Figure 1.1, reproduced from [Hinshaw et al., 2013])

The dark energy component is the component that is understood to be currently driving the expansion of the Universe. This anti-gravity effect of dark energy can be understood as a negative pressure

$$p = w\rho$$
,

where w is the equation of state parameter, and is believed to be close to -1 for dark energy. At present very little is known about dark energy. It can only be measured after carefully studying the expansion of the Universe and accounting for the effect of other components.

The dark matter component is collisionless, hence has no pressure (w = 0). In ACDM model, the velocity of that random motion of microscopic particles that constitute the dark matter is small, hence the name "Cold". Dark matter does not interact through electro-magnetic interactions, although it has been hypothesised that it does interact through weak-interactions. [Bertone et al., 2005] Dark matter can be probed through its gravitational effects, for example, most notably, from lensing and galaxy rotation curves. [see, e.g. Kent, 1987, Rubin et al., 1985, Begeman et al., 1991] Several experiments aim for direct detection of dark matter through its decay into baryons, although with no conclusive results have been reported. [see, e.g. Bernabei et al., 2003, Weniger, 2012, Angloher et al., 2012]



Figure 1.2: Fluctuations in the Cosmic Microwave Background as measured by the WMAP experiment. [Hinshaw et al., 2013]

The baryonic matter component is made of the ordinary Standard Model particles. Two most important aspect of this matter in cosmology are that it is hydrodynamical (has thermal and pressure aspects), and that it interacts with radiation, hence providing a channel to release the gravitational energy as feedback that regulates the growth of structure. The formation of quasars and galaxies is directly related to the interaction of matter. [Dekel and Silk, 1986] On very large scale where gravity is the only dominant interaction, baryonic matter in general follows the motion of dark matter. However in recent years it has been realized that the pull from matters have a much stronger effect on dark matter structures than previously thought. [see, e.g. Khandai et al., 2014, Governato et al., 2012]

#### 1.1.2 The Hubble Parameter

The additional parameter h is the (normalized) Hubble parameter. The Hubble Parameter describes the present day expansion rate of the Universe:  $H_0 = 100h \text{ km/s/Mpc}$ . Conventionally h has been absorbed into the units of distance and time, especially in the expressions of distances that are fixed with the expansion of universe (co-moving distance). The current measurement suggests that  $h \approx 0.70$ . [Freedman and Madore [2010], and recently Planck Collaboration et al. [2013], Hinshaw et al. [2013]]

#### 1.1.3 Cosmic Initial Condition

The initial conditions in the Universe are set by the remnant density fluctuations of the inflation field, and hence not part of the  $\Lambda$ CDM model. [Guth and Kaiser, 2005] The earliest observable effect of the initial fluctuations is the cosmic microwave background (CMB). The temperature of CMB is highly isotropic, with fluctuations on the order of one part of 10<sup>5</sup>. Figure 1.2 shows the temperature fluctuation as measured by the 9 year data release of the WMAP experiment. The overall amplitude of the initial



Figure 1.3: Quasar Luminosity Function at z = 5.0, comparing simulation with measurements. [McGreer et al., 2013]

power-spectrum is the encoded in the parameter  $\sigma_8$ , which measures the standard deviation of matter fluctuations in an 8 MPC/h window.  $\sigma_8$  can be measured from analysis of Lyman- $\alpha$  forest, galaxy clustering and the CMB. [Seljak et al., 2005] The initial fluctuations eventually grow into galaxies and quasars, which are the objects that we are studying with simulations in this thesis.

## **1.2** Simulating the Universe with Supercomputers

#### 1.2.1 Motivation and History

The first challenge in simulation of the Universe with computers is the infiniteness of the Universe and the finiteness of computers. Compromises (or necessary approximations) must be made to bridge the two together.

First, a periodic box (usually cubical) of a fixed finite volume is taken as a representative piece of the Universe. It is therefore debatable how representative any volume actually is. The finite size of the box prevents us from including effects due to scales greater than the size of the box. The effect on the simulated universe due to the truncation of power can be significant, for example [Power and Knebe, 2006, Bagla and Ray, 2005, Gelb and Bertschinger, 1994b,a]

- 1. The perturbation from large scale forces directly affects the morphology of small scale structures, for example, large scale tides are believed to be important in the formation of disk galaxies.
- 2. If the box is not sufficiently large, it becomes improbable that the selected region would host rare objects, such as first quasars. For example, in Figure 1.3, the simulation MassiveBlack II, does not give predictions for the luminosity function beyond M < -26, due to the rarity of such objects (less than 1 per  $\text{Gpc}^3$ ).

Second, a simulation can only resolve to a finite scale, expressed both in terms of mass and spatial resolution. It turns out though, that the lack of small scale power initial power due to finite resolution does not have to be a severe issue. It has been shown using numerical experiments that that the truncation of power spectrum at small scales does not affect large scale clustering in any significant way. [Peebles, 1974, 1985, Little et al., 1991, Bagla and Padmanabhan, 1997] The most important implication of the finite resolution of simulation is therefore to require development of physically motivated sub-grid physical models. The sub-grid models are designed to capture physical processes below or near the resolved scale. As the resolution of simulations increase, some models will inevitably be replaced with first principles calculations.

#### **1.2.2** Contemporary Distributed Supercomputer Systems

The cosmological simulations in this thesis have been largely carried out on contemporary supercomputers. These are characterized by their massively parallel architecture. [see Culler et al., 1998, for a comprehensive introduction] A massively parallel supercomputer usually consists of distributed central processing units (CPU), distributed memory, distributed filesystems, and a complicated inter-connect system that binds the distributed components together. Figure 1.4 illustrates the architecture of Blue-Waters, a Cray XE/XK system at National Center for Supercomputing Applications (NCSA). On BlueWaters, 30 thousand computing nodes, each equipped with its own memory are connected through a high speed, low latency network. The system also emphasizes the balance between the computing power and interaction between peripheral devices, the near-line and scratch file systems.

Because the computers are distributed, programming on such massively parallel systems requires the use of a communication layer. Message Passing Interface (MPI) a language-independent communications protocol used to program parallel computer, is the de-facto industry standard. The ongoing development of MPI and its implementation involves heavy interaction between academia and the industry : several open



Figure 1.4: Architecture of a Massively Parallel Supercomputer - NCSA BlueWaters. Image obtained from the BlueWaters users guide. [for Computational Applications, 2014]

implementation of MPI were developed by universities, MPICH, LAM, and most recently OpenMPI, and most private implementations of MPI provided by hardware vendors are based on these open implementations. MPI has a very primitive semantics, providing functions for synchronization and communication (Peer to Peer, and multi-cast) of structured data. As a result of this simplicity, vendor provided MPI implementations usually directly map to the hardware operations, and thus utilize the full performance of the underlying hardware.

The primitive nature of MPI also means software development with MPI is verbose. Several standards now compete with MPI. Parallel Virtual Machine (PVM) aims to provide a fault tolerant environment for heterogeneous clusters / grids. Parallel Objects (Charm++) is similar to PVM in spirit, but represents computation task with objects. Other approaches include Unified parallel C (UPC) and Fortran Co-Array which each provide a unified address space for accessing the distributed memory on different nodes.

The communication standards are usually built upon vendor libraries that directly interact with the underlying hardware systems. Figure 1.5 describes the software layers that bridges the higher level libraries (MPI, UPC / Fortran) to the hardware.

#### **1.2.3** Domain Decomposition on Supercomputers

Simulation in cosmology is usually carried out as the time evolution of a 3 dimensional spatial volume. The distributed architecture of supercomputers requires that the full domain of a simulation be decomposed into subdomains, a process usually called Domain Decomposition.

The ultimate design goals of a Domain Decomposition scheme are flexibility and orthogonality: a scheme must be sufficient flexible that it can divide the problem evenly into any number of domains (which is usually set by the number of available processing units on the supercomputers); a scheme must also be orthogonal, in the sense that the domains are independent from each other. In practice, a domain decomposition scheme can never perfectly reach these goals, due to the complexity of the problems to be solved. [Quinn, 2003]

Two flavors of domain decomposition schemes are usually used in cosmological simulations: the straight-forward scheme and the recursive scheme. In the straight-forward scheme, the volume of the simulation is split into cubical volumes of identical (or near identical sizes) [see, e.g. Khandai and Bagla, 2009, Habib et al., 2013] The recursive scheme involves indexing the 3 dimensional volume with a recursive space-filling curve, then chopping the curve into segments that are spatially localized (but with a irregular shape). [see, e.g. Springel, 2005] Figure 1.6 illustrates the recursive domain decomposition scheme used in the simulation code RAMSES. [Teyssier, 2002] Two popular space-filling curves used are the Peano-Hilbert curves and the Morton-Z-order curves. The former has more locality, while the latter is directly related to the construction of oct-trees, a hierarchical spatial data structure often used in



Figure 1.5: Layers of Software Libraries on a Supercomputer. [Cray, 2011] (From The Cray System Documentation: Using the GNI and DMAPP APIs.)



Figure 1.6: Recursive Domain Decomposition, illustrated in 2D obtained from RAM-SES website. [Teyssier, 2002]

simulations.

## 1.3 Modelling Physics Processes with Cosmological Simulation Algorithms

In this section, we examine the algorithms used in cosmological simulations to simulate physical processes. Since the simulation is decomposed into domains that correspond to remote processing units on a supercomputer, it is important to analyze if an algorithm is sub-grid, local or long range: a sub-grid algorithm requires no communication across domains; a local algorithm requires communications only across the boundaries of neighbouring domains; and a long range algorithm requires communication between all of the domains. The long range algorithms require expensive global all-to-all communications, and are the most complicated to implement. As we will see, most of the algorithms used in cosmological simulations are local or subgrid, with the only exception the 'long range' component of gravity and radiative transfer.

#### **1.3.1** Initial Condition of Simulations

The initial conditions for simulations consist of the perturbations evolved from the cosmic initial state in  $\Lambda$ CDM model. The earliest direct constraint on these perturbations is in the form of the fluctuations (on the order of 1 part in 10<sup>5</sup>) seen in the cosmic wave background radiation (CMB). The power spectrum of fluctuations (the amplitude of the fluctuations as a function of scale in Fourier space) is usually



Figure 1.7: Present day power spectrum, image obtained from Tegmark et al. [2004]

normalized to the present day according to linear theory,  $P_0(k)$ , as seen in Figure 1.7.

At a very early stage of the Universe, the density fluctuations grow linearly on all scales. The initial conditions of a cosmological simulation are therefore usually constructed directly (on large-scales, together with an extrapolation to small-scales) from the linearly evolved power spectrum of the CMB,

$$P_{\rm IC} = \left(\frac{D_{\rm IC}}{D_0}\right)^2 P_0$$

where  $DIC/D_0$  is the amount of growth in the density fluctuation at the present day compared to the time of the initial conditions.

 $P_{\rm IC}(k)$  describes the correlations present in the overdensity field of the initial condition. A uncorrelated Gaussian field  $g(\mathbf{k})$  in Fourier space is then produced. The power spectrum  $P_{\rm IC}(k)$  in the Fourier space is used to introduce correlations, to form

a correlated over-density field:

$$\delta(\mathbf{k}) = (P_{\rm IC})^{\frac{1}{2}} g(\mathbf{k}).$$

The field  $\delta(\mathbf{k})$  is the overdensity of the initial conditions. The initial density field  $\rho(\mathbf{x})$  and initial displacement field  $\Phi(\mathbf{x})$  then immediately follow

$$\rho(\mathbf{x}) = (1 + \delta(\mathbf{x}))\,\bar{\rho},$$

and

$$abla^2 \mathbf{\Phi}(\mathbf{x}) = 
abla \delta(\mathbf{x})$$

The initial displacement field is proportional to the initial velocity field

$$\mathbf{v}(\mathbf{x}) = \frac{F_{\Omega}}{H} \mathbf{\Phi}(\mathbf{x}).$$

In numerical implementations, the differentiation is usually carried out in Fourier space,

$$\mathbf{\Phi}(\mathbf{k}) = \frac{i\mathbf{k}}{k^2}\delta(k)$$

If the density field is represented by particles, the particles are usually displaced according to the Zel'Dovich approximation using  $\Phi(\mathbf{q})$ , where  $\mathbf{q}$  is the initial position of the particle. [see, e.g. White, 2014]

#### 1.3.2 Gravity

In the Friedman-Lemaitre model, a gravitationally interacting fluid is described by the collisionless Boltzmann equation coupled to the Poisson equation in an expanding background universe. The system can be tracked by quantizing the phase space elements with particles.

The many-particle system is governed by the gravitational potential  $\phi(x)$ , which is determined from the Poisson equation

$$\nabla^2 \phi(x) = 4\pi G \left[ \rho(x) - \bar{\rho} \right],$$

where  $\rho(x)$  is the point mass density field convolved with a gravitational softening kernel of comoving scale  $\varepsilon$ .

The gravitational force on a particle is

$$g_i = -\nabla \phi(x_i).$$

A brute force approach to solving the gravitational force requires  $O(N^2)$  computations, as the force of a particle receives contribution from every other particle, due to the non-locality of gravity. Because gravity converges as  $r^{-2}$ , several faster solution of time complexity  $O(N \log N)$  have been developed. The first category of algorithm (Tree) is the based on multipole expansions. [see, e.g. Hernquist and Katz, 1989] The particles are clustered into a spatial hierarchical structure (usually an OctTree or a KD-Tree)), with the particles represented by the leaves. The data structure is augmented with the total mass and higher order mass moments of each internal node. The potential and force at any given point x is calculated by transversing the tree from top down: a node is opened only if it is so close to x that the stored multiple moments can not accurately reproduce its contribution to the potential and force at point x. It can be shown that the method has a time complexity of  $O(N \log N)$ . Nevertheless, the tree methods suffers from practical difficulties: the multiple moments introduce a somewhat large memory overhead, and more importantly, a non-local tree transverse of a distributed tree data structure on a distributed supercomputer is inefficient.

A second category of algorithms (Particle-Mesh, PM) involve the use of Fast Fourier Transform (FFT) algorithms. [see, e.g. Klypin and Shandarin, 1983, White et al., 1983] FFT is a method to compute the discrete Fourier transform in  $O(M \log M)$ time where M is the total number of mesh points. [Cooley and Tukey, 1965]. The method makes use of the fact that the Poisson equation determining the potential  $\phi(x)$  can be rewritten in Fourier space, as

$$-k^2\hat{\phi}(k) = 4\pi G\hat{\rho}(k),$$

and therefore solved with a forward Fourier Transform, multiplying by the transfer function of gravity and forces, and then applying an inverse Fourier Transform.

On computers, the Fourier Transforms have to be discretized onto a regular mesh, and implemented as Discrete Fourier Transforms. The density field needs to be quantized to a regular mesh, with a finite scale. Unfortunately a full mesh covering both the simulation volume and resolve the smoothing scale  $\varepsilon$  is impractical. As a consequence, PM methods do not resolve the small scale gravitational force, giving unphysical suppression of small scale structure growth.

The third category of algorithms (P3M and TreePM) separates gravity into a nonlocal long range component, and a local short range component. [see, e.g. Khandai and Bagla, 2009, Yoshikawa and Fukushige, 2005, Bagla, 2002] Here we briefly describe the treatment of the potential, as the treatment of the gravitational force are similar. Figure 1.8 illustrates the split scale of the long range component,

$$\hat{\phi}_l(k) = \hat{\phi}(k) \exp\left(-k^2 r_s^2\right),\,$$

and the short range component,

$$\phi_s(x) = -G \sum_i \frac{m_i}{r_i} \operatorname{erfc}\left(\frac{r_i}{2r_s}\right).$$

(the smoothing by comoving scale  $\varepsilon$  in the short range term is omitted for clarity). The short range component drops off very quickly, and can be solved with either a



Figure 1.8: Separation of long range and short range force. From Springel [2005]. Solid line represents the long range force, dashed line represents the short range force.

brute force algorithm  $[O(N_{\text{short}}^2)]$  or a tree expansion  $[O(N_{\text{short}} \log N_{\text{short}})]$  algorithm. The error of the force calculation introduced by the split can be less than 3 percent. [Springel, 2005]

The mesh used in the range force is usually chosen such that the total number of elements  $N_{\text{cells}} \sim 1.0 \times N$ . Because  $N_{\text{short}} \ll N$ , the overall computation time is still  $O(N \log N)$ .

The TreePM and P3M family of algorithms are a particularly good fit for supercomputers for the following reasons. First, there are several well developed FFT libraries on distributed computers [Frigo and Johnson, 2005, Pekurovsky, 2012, Pippig and Potts, 2013]. Secondly, the data structure for the local force calculation is simple: the tree in TreePM is local to each distributed domain; the pair summing in P3M does not need any special data structure, making it a suitable candidate for hardware accelerator units. [Habib et al., 2012].

#### 1.3.3 Hydrodynamics

Hydrodynamics is usually expressed in either Eulerian or Lagrangian formulations. The difference is in the treatment of the time derivative of density. A partial derivative is used in the Eulerian formulation, while a full derivative is used in the Lagrangian formulation. Intuitively, in the Eulerian formulation, one looks at the density field as a function of space which evolves with time; in the Lagrangian formulation, one follows the motion of conserved mass elements.

The numerical methods of hydrodynamics derived from the Eulerian formulation are mesh based: the density field in a mesh based method is a function of mesh index i,

rather than the continuous spatial coordinates. The construction of meshes vary with methods. Adaptive mesh refinement (AMR, Berger and Colella [1989]), for example, adaptively uses higher resolution meshes in high density regions. Conservation of mass is not implicit, and usually the introduction of a mesh breaks any spherical symmetry present.

Smoothed Particle Hydrodynamics, (SPH, Lucy [1977], Gingold and Monaghan [1977]) is instead based on the Lagrangian formulation. In SPH, the motions of conserved mass elements (particles) are traced in a simulation, with the free variable and its spatial derivatives at particle locations estimated using an adaptive smoothing kernel.

SPH explicitly conserves energy. If density is used as an independent variable, then the formulation will suffer from the "mixing" problem: an unphysical surface tension at the boundary between high density regions and low density regions prohibits high density particles from dissipating into low density regions. The problem can be seen in cosmological simulations in situations when the hydrodynamical force is dominating over gravity. It was thought that mixing and energy conservation could not co-exist, until [Hopkins, 2013] pointed out that the choice of density as a free variable in the SPH formulation is quite arbitrary and that giving up on density – replacing it with an "entropy variable" that is directly related to pressure, mixing and energy conservation can be simultaneously achieved in SPH (pressure-entropy SPH). Figure 1.9 illustrates the mixing problem in density-entropy and pressure-entropy SPH with a standard Kelvin-Helmholtz instability simulation. We see that with pressure-entropy (on the right), the mixing between the high and low density fluid occurs naturally.

Because hydrodynamics is a local interaction, it is naturally a suitable problem for implementation on distributed computer systems. The interaction happens between a particle and its nearest neighbours, incurring communication only at the domain boundaries.

#### 1.3.4 Atomic Cooling

Hydrodynamics conserves energy. It is the baryon chemistry that removes energy from the system through atomic cooling. The cooling rate is a function of local density, temperature, metal abundance, and radiation field [see, e.g. Vogelsberger et al., 2013]. For more detailed discussion about the cooling process, refer to Gnedin and Hollon [2012], Wiersma et al. [2009], Katz et al. [1996], Sutherland and Dopita [1993], Katz et al. [1992], Ikeuchi and Ostriker [1986].

$$\Lambda = \Lambda(\rho, T, Z, \Gamma).$$

Figure 1.10 from Vogelsberger et al. [2013] shows the dependence of cooling rate on these variables. High radiation suppresses cooling, while metals enhances cooling. At high temperatures, the cooling is dominated by the free-free process.



Figure 1.9: Kelvin-Helmholtz instability in Density-Entropy SPH and Pressure-Entropy SPH [Hopkins, 2013].



Figure 1.10: Cooling Rate as a function of local density and radiation field. This figure is taken from Vogelsberger et al. [2013].

In SPH simulations, the atomic cooling process is implemented using heat sinks (sources, when cooling rate is negative) in individual particles. The implementation of cooling is completely sub-grid, requiring no interaction between neighbours, and thus no communication across domains.

#### **1.3.5** Feedback from Star Formation and Galaxies

#### Multi-phase Star Formation Model

Once the gas has cooled and collapsed to high densities, the star formation process occurs. A widely used star-formation model was developed by Springel and Hernquist [2003] (SH03). The SH03 model is a self-regulated multi-phase subgrid model, where the feedback of supernovas regulates the fraction (x) of the cold phase component relative to the total mass of a mass element – the remaining mass is assumed to be in a low density hot ambient phase.

The star-formation rate  $\phi$  is

$$\phi = \begin{cases} x(\rho, T)m/t_{\star} & \rho > \rho_{\star}, \\ 0 & \rho <= \rho_{\star}, \end{cases}$$

where  $t_{\star}$  is the star formation scale (an input parameter), x being the fraction of cold phase and  $\rho_{\star}$  is the density threshold for star formation. The cold fraction x is a function of the density and temperature of the mass element.

#### Wind Feedback

In addition to the sub-grid thermal self-regulation between hot and cold phases, star formation can also be suppressed though a momentum-like "wind" mechanism. In the model prescribed by SH03, star forming gas has a probability of being turned into non star-forming wind elements with a large velocity, until is is ejected from the dense star forming region. The velocity of the wind and the mass unloading ratio are related by the energy conservation:

$$\frac{1}{2}\eta_w v_w^2 = \chi \varepsilon_{\rm SN},$$

where  $\varepsilon_{\rm SN}$  is the supernova energy released per unit star mass, and  $\eta_w$  is the wind efficiency.

#### Modifications to the Multi-phase Model

Many variations to SH03 have been proposed in recent years. Here we describe three of them that are most relevant to the formation of high redshift quasars:

- 1. Gnedin et al. [2009] argued that the star formation is closely related to the formation of molecular hydrogen, thus the fraction of star forming gas gains an extra factor of  $x_{\rm HII}$ , which is the fraction of molecular hydrogen in cold gas.
- 2. Hopkins et al. [2013] argued that dense gas does not guarantee star formation: star formation occurs in gas that is undergoing gravitational collapse; hence an additional 'self-gravity' criterion is added to the prescription.
- 3. Okamoto et al. [2010] argued that the efficiency of winds depends on the depth of the gravitational well of the star forming region, hence it is a property of the dark matter halo.

These modified models improve the star-formation model of high resolution simulations of small mass halos. The multi-phase star formation model is implemented as a sub-grid model. The wind feedback as described in Okamoto et al. [2010] is local, as each newly formed star mass element drives neighbouring gas elements into wind, something which involves a neighbour walk.

#### **1.3.6** Feedback from Supermassive Black Holes

Supermassive black holes are believed to power the radiation from quasars at the centers of galaxies. Cosmological simulations lack the resolution to directly resolve the Bondi accretion scale of black holes (< pc), thus can only model the process with a sub-grid model. Di Matteo et al. [2005] introduced a simple model used to incorporate supermassive black holes into cosmological simulations. We describe the model below.

The first consideration is that each massive halo shall host a black hole. Based on this assumption, as a simulation evolves, each newly formed halo with  $M_{\rm HALO} > 5 \times 10^{10} M_{\odot}$  will be associated with a black hole seed in its center.

The accretion rate of a black hole is taken as a function of the local density, sound speed, and the mass of the black hole

$$\frac{dM}{dt} = \begin{cases} \dot{M}_{\rm BONDI} = \alpha \frac{4\pi G^2 M^2 \rho}{c_s^3} & \dot{M}_{\rm BONDI} < \dot{M}_{\rm Eddington} \\ \dot{M}_{\rm Eddington} = \frac{\eta 4\pi G c m_p M}{\sigma_T} & \dot{M}_{\rm BONDI} > = \dot{M}_{\rm Eddington} \end{cases}$$

On the resolutions of current cosmological simulation (100s of parsec), the quasar mode thermal feedback from super-massive black holes is the most important form of feedback. The feedback energy is proportional to the accretion rate,

$$\frac{dT}{dt} = \eta c^2 \frac{dM}{dt}$$

where  $\eta$  is the efficiency of converting mass to thermal feedback energy.

#### **1.3.7** Ultra-violet Radiation and Radiative Transfer

Radiative transfer is one of the most numerical challenging problems for cosmological simulations. After reionization, the universe is mostly transparent. Photons, especially Ultra-violet(UV) photons, travel almost freely in the universe unless they encounter dense clumps. The mean free path of UV photons is on the order of 100 Mpc/h, the typical size of a cosmological simulation box. The long mean free path and high speed of light means that once the radiative transfer is included, a simulation becomes non-local. The alternative is to follow the finite (albeit large) speed of light, which results smaller time steps – a problem also encountered in direct simulation of neutrinos in cosmological simulations. Because cosmological hydrodynamic simulations are already very computationally expensive, radiative transfer is usually carried out as a post processing step with low resolution resample.

Two types of methods exist in the computer modelling of radiative transfer: one can either solve the multi-moments of the radiation field, or directly model the radiation field with photon rays (Ray Tracing, hereafter RT). In RT, photon rays are cast from sources and then traced through the simulation volume. As photons interact with mass elements representing baryons, the total photon deposit in the ray is decreased, and optionally new rays are cast to represent the recombination photons. [See, eg Altay et al., 2008, Maselli et al., 2009, for RT decoupled with hydrodynamical simulation]; [see, e.g. Trac and Cen, 2007, for coupled RT].

## 1.4 Cosmological Simulation Data Analysis Techniques

#### 1.4.1 Simulations as *Big Data*

In the 40 years that N-body simulations have been used in Cosmology research, visualization has been the most indispensable tool. Physical processes have often been identified first and studied via images of simulations. A few examples are: formation of filamentary structures in the large-scale distribution of matter [Jenkins et al., 1998, Colberg et al., 2005, Springel et al., 2005b, Dolag et al., 2006], growth of feedback bubbles around quasars [Sijacki et al., 2007, Di Matteo et al., 2005]; cold flows of gas forming galaxies [Dekel et al., 2009, Kereš et al., 2005], and the evolution of ionization fronts during the re-ionization epoch [Shin et al., 2008, Zahn et al., 2007]. The size of current and upcoming petascale simulation datasets can make such visual exploration to discover new physics technically challenging. Here we present techniques that can be used to display images at full resolution for datasets of hundreds of billions of particles in size.

#### 1.4.2 Visualization

Several implementations of visualization software for cosmological simulations already exist. IRFIT [Gnedin, 2011] is a general purpose visualization suite that can deal with mesh based scalar, vector and tensor data, as well as particle based datasets as points. YT [Turk et al., 2011] is an analysis toolkit for mesh based simulations that also supports imaging. Splash [Price, 2007] is a visualization suite specialized for simulations that use smoothed particle hydrodynamics(SPH) techniques. Aside from the CPU based approaches mentioned above, Szalay et al. [2008] implemented a GPU based interactive visualization tool for SPH simulations.

The Millennium I & II simulations [Springel et al., 2005b, Boylan-Kolchin et al., 2009] have been used to test an interactive scalable rendering system developed by Roland Fraedrich and Westermann [2009]. Both Splash and the Millennium visualizer support high quality visualization of SPH data sets, while IRFIT treats SPH data as discrete points.

Continuing improvements in computing technology and algorithms are allowing SPH cosmological simulations to be run with ever increasing numbers of particles. Runs are now possible on scales which allow rare objects, such as quasars to form in a large simulation volume with uniform high resolution (see Section 2.1; and Di Matteo et al. [2012], Degraf et al. [2011], Khandai et al. [2012]). Being able to scan through a vast volume and seek out the tiny regions of space where most of the activity is occurring, while still keeping the large-scale structure in context necessitates special visualization capabilities. These should be able to show the largest scale information but at the same time be interactively zoomable. However, as the size of the datasets quickly exceeds the capability of moderately sized in-house computer clusters, it becomes difficult to perform any interactive visualizations. For example, a single snapshot of the MassiveBlack simulation (Section 2.1) consists of 8192 files and is over 3 TB in size.

Even when a required large scale high resolution image has been rendered, actually exploring the data requires special tools. The GigaPan collaboration <sup>1</sup> has essentially solved this problem in the context of viewing large images, with the GigaPan viewer enabling anyone connected to the Internet to zoom into and explore in real time images which would take hours to transfer in totality. The viewing technology has been primarily used to access large photographic panoramas, but is easily applicable to simulated datasets. A recent enhancement to deal with the time dimension, in the form of gigapixel frame interactive movies (GigaPan Time Machine<sup>2</sup>) turns out to give particularly novel and exciting results when applied to simulation visualization.

## **1.5** Astrophysics

#### 1.5.1 Cold Flows and Quasars

Deep sky surveys have revealed populations of distant quasars at redshifts z > 6, [Fan et al., 2001, 2003, 2004, 2006b, Jiang et al., 2009, Abazajian et al., 2009, Mortlock et al., 2011, Morganson et al., 2012, Willott et al., 2013]. The masses of the central black holes in high redshift quasars are estimated to be  $M_{\rm BH} \sim 10^9 M_{\odot}$ ; the feasibility of growing such black holes on a timescale of less than 1 billion years poses tight constraints on astrophysical mechanisms for doing this. Two aspects of growing such super-massive black holes are seeding and the subsequent accretion of mass. On the seeding side, seed masses of from  $100 - 10^5 M_{\odot}$  at redshifts z > 10 have been proposed by various authors. On the lower end, they could form from the remnants of PopIII stars, [Bromm et al., 2005, Tanaka and Haiman, 2009, Tanaka et al., 2013], or for the most massive seeds from direct gravitational collapse [Koushiappas et al., 2004, Begelman et al., 2006, Mayer et al., 2010, Bellovary et al., 2011, Choi et al., 2013, Latif et al., 2013].

Given a seed, the black holes must endure a sustained period of Eddington limited growth [Volonteri and Rees, 2005, Lodato and Natarajan, 2006, Pelupessy et al., 2007] to grow to masses of  $10^9 M_{\odot}$ . An important factor in studying the growth of these objects is understanding whether sufficiently strong gas inflows are present in high redshift halos and how self-regulation and feedback from supermassive blackholes may affect them [Di Matteo et al., 2005, Cattaneo and Teyssier, 2007, Johansson et al., 2008, Booth and Schaye, 2009, Mayer et al., 2010, Debuhr et al., 2011, Gaspari et al.,

<sup>&</sup>lt;sup>1</sup>http://www.gigapan.org

<sup>&</sup>lt;sup>2</sup>http://timemachine.gigapan.org

2011, Hobbs et al., 2012, Dubois et al., 2012].

Numerical simulation of the growth of supermassive black holes in a cosmological volume is a challenging problem. This is especially the case for the first quasars. Constraints from the quasar luminosity function [e.g. Fan et al., 2001] imply that such quasars are rare objects, with a density of  $\sim 1 - 10 \,\mathrm{Gpc}^{-3}$ . As a consequence, to simulate the first quasars, a large volume on the order of  $\mathrm{Gpc}^3$  with sufficient mass resolution to resolve their host galaxies and the gas inflow within them is required.

To work around this difficulty, many authors have adopted a method of "resimulation", where high resolution, "zoomed-in" initial conditions are generated for high redshift massive halos selected from large volume, lower resolution simulations [see, e.g., Dubois et al., 2013, Anglés-Alcázar et al., 2013, Bellovary et al., 2013, Costa et al., 2013, Romano-Diaz et al., 2011, Bournaud et al., 2011, Hopkins and Quataert, 2010, Sijacki et al., 2009, Li et al., 2007]. In this commonly adopted method however the growth of the most massive black holes is then typically assumed to be associated with the most massive dark matter halos. This may not be the case at all redshifts: it is still controversial whether halo mass is a good indicator of the black hole properties of the rare first quasars [see, e.g., Fanidakis et al., 2013, Husband et al., 2013, Kim et al., 2009].

Another approach is to directly run a large volume simulation with hydrodynamics and a model for supermassive black hole formation and growth see, e.g., Booth and Schaye, 2013, Wurster and Thacker, 2013, Booth and Schaye, 2010, 2009, Bonoli et al., 2009, Schaye and Dalla Vecchia, 2008, Di Matteo et al., 2008, Priddey et al., 2008]. The largest simulation in this approach is the MassiveBlack (hereafter MB) simulation by Di Matteo et al. [2012], designed to study the first quasars. With a 0.75 Gpc box side length and  $2 \times 3200^3$  particles, the Smoothed Particle Hydrodynamics (hereafter SPH) simulation run from uniform cosmological initial conditions produced around ten supermassive black holes with  $M_{\rm BH} \sim 10^9 M_{\odot}$  at  $z \sim 6$ . The growth of these black holes was found to be mainly due to cold flows. Gas in the vicinity of the black holes was shown to originate from cold dense filaments that survive well within the virial radius of the halo [see also Bournaud et al., 2011, Dekel et al., 2009]. Even though the mass resolution of MB is sufficient to resolve the host halos of the first quasars, the spatial resolution was not sufficient to follow gas inflows on sub-kpc scales. Another drawback with large uniform simulations such as MB is that it becomes prohibitively expensive to experiment with the numerical schemes for hydrodynamics and feedback.

#### 1.5.2 Re-ionization

The current consensus is that the contribution to the global ionizing budget from quasars during the Epoch of Reionization (EoR) is small compared to that from early stars and galaxies [see, e.g., Loeb, 2009, Giroux and Shapiro, 1996, Trac and Gnedin, 2011]. The EoR began as the population III stars and primordial galaxies ionized their most immediate vicinity, as studied by Thomas and Zaroubi [2008], Chen

and Miralda-Escudé [2008], Venkatesan and Benson [2011]. On the other hand, the characteristic proper radius of ionizing bubbles at the end of EoR is constrained to be on the order of 10 Mpc/h [Maselli et al., 2007, Alvarez and Abel, 2007, Wyithe and Loeb, 2004], and the quasar contribution is limited to be less than 14% [Srbinovsky and Wyithe, 2007] of the total. Even though the contribution to global reionization by quasars is constrained in this way, bright quasars may still leave a signature on the growth of individual ionized regions. Several authors have investigated this signature in mock 21cm emission spectra taken from simulations of isolated quasars [Datta et al., 2012, Majumdar et al., 2011, Alvarez et al., 2010, Datta et al., 2008, Geil and Wyithe, 2008] at redshift  $z \sim 8$ . In observations, an object recently reported by Mortlock et al. [2011], ULAS J1120+0641, has a luminosity of  $6.3 \times 10^{13} L_{\odot}$  at  $z \sim 7$  and a proper near-zone radius of less than 2 Mpc/h. The near-zone radius is consistent with the possibility of bright quasar driven growth in a near neutral intergalactic medium background [Bolton et al., 2011].

#### **1.5.3** Lyman- $\alpha$ Forest

Quasars converts gravitational energy into radiation; the spectrum of radiation generated during the infalling of matter is expected to be a continuous function of wavelength over a large range. The continuous spectra from quasars are then modulated by the ISM and IGM around quasars, leaving absorption and emission lines on the spectra. [see, e.g. Meiksin, 2009]

One particular interesting family of absorption lines is the Lyman- $\alpha$  forest. The Lyman- $\alpha$  forest is the absorption arising at wavelength  $\lambda_{\alpha} = 121.6$  nm of neutral hydrogen in the IGM. Figure 1.11 is the spectrum of a z = 3.7 quasar taken from McDonald et al. [2006]. We can see that the Lyman- $\alpha$  forest starts from an observed wavelength of  $(1+3.7) \times 121.6$  nm, and ends at an observed wavelength of  $(1+3.7) \times 121.6$  nm, and ends at an observed wavelength of  $(1+3.7) \times 102.6$  nm, where the Lyman- $\alpha$  absorption features are superimposed with Lyamn- $\beta$  absorption features.

The physics of the Lyman- $\alpha$  forest is simple. After re-ionization, the IGM is nearly completely ionized. However, the Lyman- $\alpha$  cross section is sufficiently large that even a tiny fraction ( $x_{\rm HI} < 10^{-3}$ ) of neutral hydrogen is able to significantly absorb the redshifted photons emitted from quasars. On large scales, the neutral hydrogen absorption traces the IGM density.

$$\tau \propto \rho^{\beta},$$

where  $\tau$  is the optical depth, and  $\beta = 2 - 0.7\gamma \approx 1.6$  is a parameter related to the equation of state of the IGM. This approximation is known as the Fluctuating Gunn Peterson Approximation [Croft et al., 1998].

Because the Lyman- $\alpha$  forest is correlated with the hydrogen density, observations of the Lyman- $\alpha$  forest provide a complementary way to study the large scale structure of in the universe, in addition to CMB and galaxy surveys.



Figure 1.11: Example of a Quasar Spectrum and its Lyman- $\alpha$  forest at z=3.7. [McDonald et al., 2006]

## 1.6 Thesis Overview

#### 1.6.1 Motivation

As such extreme objects, quasars provide great ways to test cosmological theories. Remarkably strong correlations have been found between black holes and properties of their host galaxies, suggesting significant interaction between the evolution of a galaxy and the supermassive black hole it contains. As a result it has recently been realized that black holes play a vital role in the evolution of all galaxies. Observational cosmological surveys are producing large catalogs of quasars and the Ly- $\alpha$  forest; to fully derive and understand the constraints on cosmological parameters from these surveys, large simulations and mock catalogs are needed.

The growth of the capability of supercomputers has followed the predictions of Moore's law for the last four decades, arriving most recently at the era of petascale computing: supercomputers nowadays can perform more than 1 peta floating operations per second. We have reached the point where we can actually simulate huge cosmological volumes to form these rare objects. This thesis is therefore motivated by the recent developments in two fields, bringing cosmological simulations to petascale supercomputers.

#### 1.6.2 Thesis plan

In this thesis we first develop a new method to visualize and interact with large data sets produced by cosmological simulations in Chapter 2. In Chapter 3, we develop a zoom-in resimulation scheme, and apply it to study the gas supply of high redshift quasars found in large simulations. In Chapter 4 we develop a parallel version of the radiative transfer scheme SPHRAY, and apply it to study the HII region of high redshift quasars. The work in Chapters 1, 2, and 3 have been published in peer reviewed journals: Feng et al. [2014, 2013, 2011].

Some of the work featured in the thesis are taken from the work of the author in following publications In addition some work by the author taken from the following publications features in this thesis: Khandai et al. [2014], Stevens et al. [2014], Tenneti et al. [2014], Wilkins et al. [2013b,a,c], Khandai et al. [2012], Di Matteo et al. [2012]. Chapter 5 describes some recent work producing physically motivated quasar and Lyman- $\alpha$  forest mock datasets.

# Chapter 2

# Visualization and Interaction with Petascale Simulations

## 2.1 Introduction

In this chapter we combine an off-line imaging technique together with GigaPan technology to implement an interactively accessible visual probe of large cosmological simulations. While GigaPan is an independent project (uploading and access to the GigaPan website is publicly available), we release our toolkit for the off-line visualization as Gaepsi<sup>1</sup>, a software package aimed specifically at Gadget [Springel et al., 2001, Springel, 2005] SPH simulations.

The layout of this chapter is as follows. In Section 2 we give a brief overview of the physical processes modeled in Gadget, as well as describing two P-GADGET simulations which we have visualized. In Section 3 we give details of the spatial domain remapping we employ to convert cubical simulation volumes into image slices. In Section 4, we describe the process of rasterizing an SPH density field, and in Section 5 the image rendering and layer compositing. In Section 6 we address the parallelism of our techniques and give measures of performance. In Section 7 we briefly describe the GigaPan and GigaPan Time Machine viewers and present examples screenshots from two visualizations (which are both accessible on the GigaPan websites). Finally we introduce the interactive browers we developed for MassiveBlack II simulation in the final chapter.

## 2.2 Simulation

Adaptive Mesh Refinement(AMR, e.g., Bryan and Norman [1997]) and Smoothed Particle Hydrodynamics (SPH, Monaghan [1992]) are the two most used schemes for

 $<sup>^{1}</sup> http://web.phys.cmu.edu/~yfeng1/gaepsi$
carrying out cosmological simulations. In this work we focus on the visualization of the baryonic matter in SPH simulations run with P-Gadget [Springel, 2005].

Gadget is an SPH implementation, and P-Gadget is a version which has been developed specifically for petascale computing resources. It simultaneously follows the self-gravitation of a collision-less N-body system (dark matter) and gas dynamics (baryonic matter), as well as the formation of stars and super-massive black holes. Dark matter particles and gas particle positions and initial characteristics are set up in a comoving cube, and black hole and star particles are created according to sub-grid modeling [Springel and Hernquist, 2003, Di Matteo et al., 2008] Gas particles carry hydrodynamical properties, such as temperature, star formation rate, and neutral fraction.

Although our attention in this chapter is limited to imaging properties the of gas, stars and black holes in Gadget simulations, similar techniques could be used to visualize the dark matter content. Also, the software we provide should be easily adaptable to the data formats of other SPH codes (e.g. Gasoline, [Wadsley et al., 2004])

### 2.2.1 MassiveBlack

The MassiveBlack simulation is the state-of-art SPH simulation of a  $\Lambda$ CDM universe [Di Matteo et al., 2012]. P-Gadget was used to evolve  $2 \times 3200^3$  particles in a volume of side length  $533 h^{-1}$ Mpc with a gravitational force resolution of  $5 h^{-1}$ Kpc. One snapshot of the simulation occupies 3 tera-bytes of disk space, and the simulation has been run so far to redshift z = 4.75, creating a dataset of order 120 TB. The fine resolution and large volume of the simulation permits one to usefully create extremely large images. The simulation was run on the high performance computing facility, Kraken, at the National Institute for Computational Sciences in full capability mode with 98,000 CPUs.

### 2.2.2 E5

To make a smooth animation of the evolution of the universe typically requires hundreds frames directly taken as snapshots of the simulation. The scale of the MassiveBlack run is too large for this purpose, so we ran a much smaller simulation (E5) with  $2 \times 336^3$  particles in a 50 h<sup>-1</sup>Mpc comoving box. The model was again a  $\Lambda$ CDM cosmology, and one snapshot was output per 10 million years, resulting in 1367 snapshots. This simulation ran on 256 cores of the Warp cluster in the McWilliams Center for Cosmology at CMU.



Figure 2.1: Transformation of the Primitive Cell. The cubical unit cell is shown using solid lines. The new primitive cell, generated by  $\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$  is shown with dash-dotted lines. The transformed domain is shown in gray.

## 2.3 Spatial Domain Remapping

Spatial domain remapping can be used to transform the periodic cubic domain of a cosmological simulation to a patch whose shape is similar to the domain of a sky survey, while making sure that the local structures in the simulation are preserved [Carlson and White, 2010, Hilbert et al., 2009]. Another application is making a thin slice that includes the entire volume of the simulation. Our example will focus on the latter case.

A Gadget cosmological simulation is usually defined in the periodic domain of a cube. As a result, if we let f(X = (x, y, z)) be any position dependent property of the simulation, then

$$f(X) = f((x + \mu L, y + \nu L, z + \sigma L)),$$

where  $\mu, \nu, \sigma$  are integers. The structure corresponds to a simple cubic lattice with lattice constant a = L, the simulation box side-length. A bijective mapping from the cubic unit cell to a remapped domain corresponds to a choice of the primitive cell. Figure 2.1 illustrates the situation in 2 dimensions.

Whilst the original remapping algorithm by Carlson and White [2010] results in the correct transformations being applied, it has two drawbacks: (i) the orthogonalization is invoked explicitly and (ii) the hit-testing for calculation of the shifting (see below) is against non-aligned cuboids. The second problem especially undermines the performance of the program. In this work we present a faster algorithm based on similar ideas, but which features a QR decomposition (which is widely available as a library routine), and hit-testing against an AABB (Axis Aligned Bounding Box).

First, the transformation of the primitive cell is given by a uni-modular integer matrix,

$$M = \begin{pmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{pmatrix},$$

where  $M_{ij}$  are integers and the determinant of the matrix |M| = 1. It is straightforward to obtain such matrices via enumeration. [Carlson and White, 2010] The QR decomposition of M is

$$M = QR,$$

where Q is an orthonormal matrix and R is an upper-right triangular matrix. It is immediately apparent that (i) application of Q yields rotation of the basis from the simulation domain to the transformed domain, the column vectors in matrix  $Q^T$ being the lattice vectors in the transformed domain; (ii) the diagonal elements of R are the dimensions of the remapped domain. For imaging it is desired that the thickness along the line of sight is significantly shorter than the extension in the other dimensions, thus we require  $0 < R_{33} \ll |R_{22}| < |R_{11}|$ . Note that if a domain that is much longer in the line of sight direction is desired, for example to calculate long range correlations or to make a sky map of a whole simulation in projection, the choice should be  $0 < |R_{33}| < |R_{22}| \ll |R_{11}|$ .

Next, for each sample position X, we solve the indefinite equation of integer cell number triads  $I = (I_1, I_2, I_3)^T$ ,

$$\tilde{X} = Q^T X + a Q^T I,$$

where a is the box size,  $\tilde{X}$  is the transformed sample position satisfying  $\tilde{X} \in [0, R_{11}) \times [0, R_{22}) \times [0, R_{33})$ . In practice, the domain of  $\tilde{X}$  is enlarged by a small number  $\epsilon$  to address numerical errors. Multiplying by Q on the left and re-organizing the terms, we find

$$I = \frac{Q\dot{X}}{a} - \frac{X}{a}.$$

Notice that  $Q\tilde{X}$  is the transformed sample position expressed in the original coordinate system, and is bounded by its AABB box. If we let  $(Q\tilde{X}/a)_i \in [B_i, T_i]$ , where  $B_i$  and  $T_i$  are integers, and notice  $\frac{X_i}{a} \in [0, 1)$ , the resulting bounds of I are given by

$$I_i \in [B_i, T_i].$$

We then enumerate the range to find X.

When the remapping method is applied to the SPH particle positions, the transformations of the particles that are close to the edges give inexact results. The situation is similar to the boundary error in the original domain when the periodic boundary



Figure 2.2: Boundary effects and Smoothed Particles. Four images of a particle intersecting the boundary are shown. The top-right image is contained in the transformed domain, but the other three are not. The contribution of the two bottom images is lost. By requiring the size of the transformed domain to be much larger than typical SPH smoothing lengths, most particles do not intersect a boundary of the domain and the error is contained near the edges.

Simulation	MassiveBlack	E5	
Matrix $M$	$\begin{bmatrix} 5 & 6 & 2 \\ 3 & -7 & 3 \\ 2 & 7 & 0 \end{bmatrix}$	$\begin{bmatrix} 3 & 1 & 0 \\ 2 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix}$	
$R_{11}(\mathrm{h}^{-1}\mathrm{Mpc})$	3300	180	
$R_{11}$ Pixels(Kilo)	810	36.57	
$R_{22}(\mathrm{h^{-1}Mpc})$	5800	101	
$R_{22}$ Pixels(Kilo)	1440	20.48	
$R_{33}(\mathrm{h^{-1}Mpc})$	7.9	6.7	
$Resolution(h^{-1}Kpc)$	4.2	4.9	

Table 2.1: Transformations

condition is not properly considered. Figure 2.2 illustrates the situation by showing all images of the particles that contribute to the imaging domain. We note that for the purpose of imaging, by choosing a  $R_{33}$  (the thickness in the thinner dimension) much larger than the typical smoothing length of the SPH particle, the errors are largely constrained to lie near the edge. These issues are part of general complications related to the use of a simulation slice for visualization. For example in an animation of the distribution of matter in a slice it is possible for objects to appear and disappear in the middle of the slice as they pass through it. These limitations should be borne in mind, and we leave 3D visualization techniques for future work.

The transformations used for the MassiveBlack and E5 simulations are listed in Table 2.1.

## 2.4 Rasterization

In a simulation, many field variables are of interest in visualization.

- scalar fields : density  $\rho$ , temperature T, neutral fraction  $x_{\rm HI}$ , star formation rate  $\phi$ ;
- vector fields: velocity, gravitational force.

In an SPH simulation, a field variable as a function of spatial position is given by the interpolation of the particle properties. Rasterization converts the interpolated continuous field into raster pixels on a uniform grid. The kernel function of a particle at position  $\mathbf{y}$  with smoothing length h is defined as

$$W(\mathbf{y},h) = \frac{8}{\pi h^3} \begin{cases} 1 - 6\left(\frac{y}{h}\right)^2 + 6 & \left(\frac{y}{h}\right)^3, 0 \le \frac{y}{h} \le \frac{1}{2} \\ 2\left(1 - \frac{y}{h}\right)^3, & \frac{1}{2} < \frac{y}{h} \le 1, \\ 0 & \frac{y}{h} > 1. \end{cases}$$

Following the usual prescriptions [e.g. Springel, 2005, Price, 2007], the interpolated density field is taken as

$$\rho(\mathbf{x}) = \sum_{i} m_i W(\mathbf{x} - \mathbf{x}_i, h_i),$$

where  $m_i$ ,  $\mathbf{x}_i$ ,  $h_i$  are the mass, position, and smoothing length of the *i*th particle, respectively. The interpolation of a field variable, denoted by A, is given by

$$A(\mathbf{x}) = \sum_{i} \frac{A_{i}m_{i}W(\mathbf{x} - \mathbf{x}_{i}, h_{i})}{\rho(\mathbf{x}_{i})} + O(h^{2}),$$

where  $A_i$  is the corresponding field property carried by the *i*th particle. Note that the density field can be seen as a special case of the general formula.

Two types of pixel-wise mean for a field are calculated, the

1. volume-weighted mean of the density field,

$$\bar{\rho}(P) = \frac{M(P)}{P} = \frac{1}{P} \int_{P} d^{3}\mathbf{x} \,\rho(\mathbf{x})$$
$$= \frac{1}{P} \sum_{i} m_{i} \int_{P} d^{3}\mathbf{x} \,W(\mathbf{x} - \mathbf{x}_{i}, h_{i})$$
$$= \frac{1}{P} \sum_{i} M_{i}(P),$$

where  $M_i(P) = m_i \int_P d^3 \mathbf{x} W(\mathbf{x} - \mathbf{x}_i, h_i)$  is the mass overlapping of the *i*th particle and the pixel, and the

2. mass-weighted mean of a field  $(A)^2$ ,

$$\bar{A}(P) = \frac{\int_P d^3 \mathbf{x} A(\mathbf{x}) \rho(\mathbf{x})}{\int_P d^3 \mathbf{x} \rho(\mathbf{x})}$$
$$= \frac{\sum_i A_i M_i(P)}{M(P)} + O(h^2).$$

To obtain a line of sight projection along the third axis, the pixels are chosen to extend along the third dimension, resulting a two dimensional final raster image.

<sup>2</sup>The second line is an approximation. For the numerator,

$$\begin{split} &\int_{P} d^{3}\mathbf{x} A(\mathbf{x})\rho(\mathbf{x}) = \\ &\int_{P} d^{3}\mathbf{x} \sum_{i} \frac{A_{i}m_{i}W(\mathbf{x} - \mathbf{x}_{i}, h_{i})}{\rho_{i}} \sum_{j} m_{j}W(\mathbf{x} - \mathbf{x}_{j}, h_{j}) \\ &= \sum_{i} \frac{A_{i}m_{i}}{\rho_{i}} \sum_{j} m_{j} \int_{P} d^{3}\mathbf{x} W(\mathbf{x} - \mathbf{x}_{i}, h_{i})W(\mathbf{x} - \mathbf{x}_{j}, h_{j}) \end{split}$$

The calculation of the overlapping  $M_i(P)$  in this circumstance is two dimensional. Both formulas require frequent calculation of the overlap between the kernel function and the pixels. An effective way to calculate the overlap is via a lookup table that is pre-calculated and hard coded in the program. Three levels of approximation are used in the calculation of the contribution of a particle to a pixel:

- 1. When a particle is much smaller than a pixel, the particle contributes to the pixel as a whole. No interpolation and lookup occurs.
- 2. When a particle and pixel are of similar size of (up to a few pixels in size), the contribution to each of the pixels are calculated by interpolating between the overlapping areas read from a lookup table.
- 3. When a particle is much larger than a pixel, the contribution to a pixel taken to be the center kernel value times the area of a pixel.

Note that Level 1 and 3 are significantly faster than Level 2 as they do not require interpolations.

The rasterization of the z = 4.75 snapshot of MassiveBlack was run on the SGI UV Blacklight supercomputer at the Pittsburgh Supercomputing Center. Blacklight is a shared memory machine equipped with a large memory for holding the image and a fairly large number of cores enabling parallelism, making it the most favorable

If we apply the mean value theorem to the integral and Taylor expand, we find that

$$\begin{split} &\int_{P} d^{3}\mathbf{x} W(\mathbf{x} - \mathbf{x}_{i}, h_{i}) W(\mathbf{x} - \mathbf{x}_{j}, h_{j}) \\ &= W(\xi_{ij} - \mathbf{x}_{j}, h_{j}) \int_{P} d^{3}\mathbf{x} W(\mathbf{x} - \mathbf{x}_{i}, h_{i}) \\ &= [W(\mathbf{x}_{i} - \mathbf{x}_{j}, h_{j}) + W'(\mathbf{x}_{i} - \mathbf{x}_{j}, h_{j})(\xi_{ij} - \mathbf{x}_{i}) \\ &+ O[(\xi_{ij} - \mathbf{x}_{i})^{2}]] \\ &\times \int_{P} d^{3}\mathbf{x} W(\mathbf{x} - \mathbf{x}_{i}, h_{i}) \\ &= W(\mathbf{x}_{i} - \mathbf{x}_{j}, h_{j}) \int_{P} d^{3}\mathbf{x} W(\mathbf{x} - \mathbf{x}_{i}, h_{i}) + O(h^{2}). \end{split}$$

Both W' and  $\xi_{ij} - \mathbf{x}_i$  are bound by terms of  $O(h_j)$ , so that the extra terms are all beyond  $O(h^2)$  (the last line). Noticing that  $\rho_i = \sum_j m_j W(\mathbf{x}_i - \mathbf{x}_j, h_j) + O(h^2)$ , the numerator

$$\begin{split} &\int_{P} d^{3}\mathbf{x} A(\mathbf{x})\rho(\mathbf{x}) \\ &= \sum_{i} A_{i}m_{i} \int_{P} d^{3}\mathbf{x} W(\mathbf{x} - \mathbf{x}_{i}, h_{i}) + O(h^{2}) \\ &= \sum_{i} A_{i}M_{i}(P) + O(h^{2}). \end{split}$$

machine for the rasterization. The rasterization of the E5 simulation was run on local CMU machine Warp. The pixel dimensions of the raster images are also listed in Table 2.1. The pixel scales have been chosen to be around the gravitational softening length of  $\sim 5 \, h^{-1}$ Kpc in these simulations in order to preserve as much information in the image as possible.

## 2.5 Image Rendering and Layer Compositing

The rasterized SPH images are color-mapped into RGBA (red, green, blue and opacity) layers. Two modes of color-mapping are implemented, the simple mode and the intensity mode.

In the simple mode, the color of a pixel is directly obtained by looking up the normalized pixel value in a given color table. To address the large (several orders of magnitude) variation of the fields, the logarithm of the pixel value is used in place of the pixel value itself.

In the intensity mode, the color of a pixel is determined in the same way as done in the simple mode. However, the opacity is reduced by a factor  $f_m$  that is determined by the logarithm of the total mass of the SPH fluid contained within the pixel. To be more specific,

$$f_m = \begin{cases} 0 & \log M < a, \\ 1 & \log M > b, \\ \left[\frac{\log M - a}{b - a}\right]^{\gamma} & \text{otherwise,} \end{cases}$$

where a and b are the underexposure and overexposure parameters: any pixel that has a mass below  $10^a$  is completely transparent, and any pixel that has a mass above  $10^b$  is completely opaque.

The RGBA layers are stacked one on top of another to composite the final image. The compositing assumes an opaque black background. The formula to composite an opaque bottom layer B with an overlay layer T into the composite layer C is [Porter and Duff, 1984]

$$C = \alpha F + (1 - \alpha)T,$$

where C, B and T stand for the RGB pixel color triplets of the corresponding layer and  $\alpha$  is the opacity value of the pixel in the overlay layer T. For example, if the background is red and the overlay color is green, with  $\alpha = 50\%$ , the composite color is a 50%-dimmed yellow.

Point-like (non SPH) particles are rendered differently. Star particles are rendered as colored points, while black hole particles are rendered using circles, with the radius proportional to the logarithm of the mass. In our example images, the MassiveBlack simulation visualization used a fast rasterizer that does not support anti-aliasing, whilst the frames of E5 are rendered using matplotlib [Hunter, 2007] that does antialiasing.



Figure 2.3: Fiducial color-map for the gas density field. The colors span a darkened red through yellow to blue.

The choice of the colors in the color map has to be made carefully to avoid confusing different quantities. We choose a color gradient which spans black, red, yellow and blue for the color map of the normalized gas density field. This color map is shown in Figure 2.3. Composited above the gas density field is the mass weighted average of the star formation rate field, shown in dark blue, and with completely transparency where the field vanishes. Additionally, we choose solid white pixels for the star particles. Blackholes are shown as green circles. In the E5 animation frames, the normalization of the gas density color map has been fixed so that the maximum and minimum values correspond to the extreme values of density in the last snapshot.

## 2.6 Parallelism and Performance

The large simulations we are interested in visualizing have been run on large supercomputer facilities. In order to image them with sufficient resolution to be truly useful, the creation of images from the raw simulation data also needs significant computing resources. In this section we outline our algorithms for doing this and give measures of performance.

### 2.6.1 Rasterization in Parallel

We have implemented two types of parallelism, which we shall refer to as "tiny" and "huge", to make best use of shared memory architectures and distributed memory architectures, respectively. The tiny parallelism is implemented with OpenMP and takes advantage of the case when the image can be held within the memory of a single computing node. The parallelism is achieved by distributing the particles in batches to the threads within one computing node. The raster pixels are then color-mapped in serial, as is the drawing of the point-like particles. The tiny mode is especially useful for interactively probing smaller simulations.

The huge version of parallelism is implemented using the Message Passing Interface (MPI) libraries and is used when the image is larger than a single computing node or the computing resources within one node are insufficient to finish the rasterization in

a timely manner. The imaging domain is divided into horizontal stripes, each of which is hosted by a computing node. When the snapshot is read into memory, only the particles that contribute to the pixels in a domain are scattered to the hosting node of the domain. Due to the growth of cosmic structure as we move to lower redshifts, some of the stripes inevitably have many more particles than others, introducing load imbalance. We define the load imbalance penalty  $\eta$  as the ratio between the maximum and the average of the number of particles in a stripe. The computing nodes with fewer particles tend to finish sooner than those with more. The color-mapping and the drawing of point-like particles are also performed in parallel in the huge version of parallelism.

### 2.6.2 Performance

The time spent in domain remapping scales linearly with the total number of particles N,

$$T_{\text{remap}} \sim O(N)$$

The time spent in color-mapping scales linearly with the total number of pixels P,

$$T_{\rm color} \sim O(P).$$

Both processes consume a very small fraction of the total computing cycles.

The rasterization consumes a much larger part of the computing resources and it is useful to analyze it in more detail. If we let  $\bar{n}$  be the number of pixels overlapping a particle, then  $\bar{n} = K^{-1}N^{-1}P$ , where  $K^{-1}$  is a constant related to the simulation. Now we let t(n) be the time it takes to rasterize one particle, as a function of the number of pixels overlapping the particle. From the 3 levels of detail in the rasterization algorithm (Section 4), we have

$$t(n) = \begin{cases} C_1, & n \ll 1\\ C_2 n & n \sim 1 \\ C_3 n, & n \gg 1 \end{cases}$$

with  $C_2 \gg C_1 \approx C_3$ . The effective pixel filling rate R is defined as the total number of image pixels rasterized per unit time,

$$R = P[t(\bar{n})N]^{-1} = \bar{n}K[t(\bar{n})]^{-1}$$
$$= \begin{cases} \bar{n}KC_1^{-1}, & \bar{n} \ll 1\\ KC_2^{-1} & \bar{n} \sim 1\\ KC_3^{-1}, & \bar{n} \gg 1 \end{cases}$$

The rasterization time to taken to create images from a single snapshot of the MassiveBlack simulation (at redshift 4.75) at various resolutions is presented in Table 2.2 and Figure 2.4.



Figure 2.4: MassiveBlack Simulation Rasterization Rate. We show the number of pixels fixed as a function of resolution (pixel scale). The rate peaks at  $KC_3^{-1}$  at the high resolution limit and approaches  $KC_2^{-1}$  as the resolution worsens. The  $\bar{n} \ll 1$  domain is not explored.

Pixels (Kpc/px) (hours) (K/sec)	$\begin{array}{c} \operatorname{Res} \\ \bar{n} \\ \eta \end{array}$	CPUs Rate	Wall-time			
5.6G	58.4	80	128	1.63	1.45	11.3
22.5G	29.2	330	256	3.17	1.66	13.4
90G	14.6	1300	512	3.35	1.57	24.0
90G	14.6	1300	512	3.06	1.39	23.3
360G	7.3	5300	512	7.65	1.39	37.2
1160G	4.2	16000	1344	10.1	1.56	37.2

Table 2.2: Time Taken to Rasterize the MassiveBlack Simulation. Here N is the number of SPH particles, Res is the resolution (pixel scale),  $\overline{n}$  is mean number of pixels that overlap each particle,  $\eta$  is the load unbalance penalty averaged over patches, and the final column, Rate, is the number of kilopixels rasterized per second.

The rasterization of the images were carried out on Blacklight at PSC. It is interesting to note that for the largest images, the disk I/O wall-time, limited by the I/O bandwidth of the machine, overwhelms the total computing wall-time. The performance of the I/O subsystem shall be an important factor in the selection of machines for data visualization at this scale.

## 2.7 Image and Animation Viewing

Once large images or animation frames have been created, viewing them presents a separate problem. We use the GigaPan technology for this, which enables someone with a web browser and Internet connection to access the simulation at high resolution. In this section we give a brief overview of the use of the GigaPan viewer for exploring large static images, as well as the recently developed GigaPan Time Machine viewer for gigapixel animations.

### 2.7.1 Gigapan

Individual gigapixel-scale images are generally too large to be shared in easy ways; they are too large to attach to emails, and may take minutes or longer to transfer in their entirety over typical Internet connections. GigaPan addresses the problem of sharing and interactive viewing of large single images by streaming in real-time the portions of images actually needed by the viewer of the image, based on the viewers current area of focus inside the image. To support this real-time streaming, the image is divided up and rendered into small tiles of multiple resolutions. The viewer pulls in only the tiles needed for a given view. Many mapping programs (e.g., Google Maps) use the same technique.

We have uploaded an example terapixel image<sup>3</sup> of the redshift z = 4.75 snapshot of the MassiveBlack simulation to the GigaPan website, which is run as a publicly accessible resource for sharing and viewing large images and movies. The dimension of the image is 1440000 × 810000, and the finished image uncompressed occupies 3.58 TB of storage space. The compressed hierarchical data storage in Gigapan format is about 15% of the size, or 0.5 TB. There is no fundamental limits to size, provided the data can be stored on the disk. It is possible to create directly the compressed tiles of a GigaPan, bypassing the uncompressed image as an intermediate step, and thus reducing the requirement on memory and disk storage. We leave this for future work.

On the viewer side, static GigaPan works well at different bandwidths; the interface remains responsive independent of bandwidth, but the imagery resolves more slowly as the bandwidth is reduced. 250 kilobits/sec is a recommended bandwidth

<sup>&</sup>lt;sup>3</sup>image at http://gigapan.org/gigapans/76215/



Figure 2.5: GigaPan View of the MassiveBlack Simulation at z = 4.75. The images are screen grab from the GigaPan viewer: we have left the magnification bar visible. The background is an overall view of the entire snapshot. We also show zooms into region around one of the most massive blackholes in the simulation, as shown in the right most zoom with the largest circle. The three zoom levels are  $80 h^{-1}$ Mpc,  $8 h^{-1}$ Mpc,  $800 h^{-1}$ Kpc, from left to right.



Figure 2.6: GigaPan Time Machine Animation of the E5 simulation. These are screen grabs from the GigaPan Time Machine viewer. In the left column we show 8 frames (out of 1367 in the full animation) which illustrate the evolution of the entire simulation volume (at time intervals of 2 Gyears. The middle panel zooms in to show formation of the largest halo through merger event. The right panel shows some of the history of a smaller halo.

for exploring with a  $1024 \times 768$  window, but the system works well even when the bandwidth is lower.

An illustration of the screen output is shown in Figure 2.5. The reader is encouraged to visit the website to explore the image.

### 2.7.2 GigaPan Time Machine

In order to make animations, one starts with the rendered images of each individual snapshot in time. These can be gigapixel in scale or more. In our example, using the E5 simulation (Section 2.2) we have 1367 images each with 0.75 gigapixels.

One approach to showing gigapixel imagery over time would be to modify the single-image GigaPan viewer to animate by switching between individual GigaPan tile-sets. However, this approach is expensive in bandwidth and CPU, leading to sluggish updates when moving through time.

To solve this problem, we created a gigapixel video streaming and viewing system called GigaPan Time Machine [Fin, 2010], which allows the user to fluidly explore gigapixel-scale videos across both space and time. We solve the bandwidth and CPU problems using an approach similar to that used for individual GigaPan images: we divide the gigapixel-scale video spatially into many smaller videos. Different video tiles contain different locations of the overall video stream, at different levels of detail. Only the area currently being viewed on a client computer need be streamed to the client and decoded. As the user translates and zooms through different areas in the video, the viewer scales and translates the currently streaming video, and over time the viewer requests from the server different video tiles which more closely match the current viewing area. The viewing system is implemented in Javascript+HTML, and takes advantage of recent browser's ability to display and control videos through the new HTML5 video tag.

The architecture of GigaPan Time Machine allows the content of all video tiles to be precomputed on the server; clients request these precomputed resources without additional CPU load on the server. This allows scaling to many simultaneously viewing clients, and allows standard caching protocols in the browser and within the network itself to improve the overall efficiency of the system. The minimum bandwidth requirement to view videos without stalling depends on the size of the viewer, the frame rate, and the compression ratios. The individual videos in "Evolution of the Universe" (the E5 simulation, see below for link) are currently encoded at 25 FPS with relatively low compression. The large video tiles require a continuous bandwidth of 1.2 megabits/sec, and a burst bandwidth of 2.5 megabits/sec.

We have uploaded an example animation<sup>4</sup> of the E5 simulation, showing its evolution over the interval between redshift z = 200 and z = 0 with 1367 frames equally spaced in time by 10 Myr. Again, the reader is encouraged to visit the website to explore the image.

<sup>&</sup>lt;sup>4</sup>http://timemachine.gigapan.org/wiki/Evolution\_of\_the\_Universe



Figure 2.7: Interface of the Interactive Browser.

## 2.8 Interactive Simulation Browser

To ease the exploration of the large dataset in MassiveBlack2 simulation, we developed an interactive simulation browser web-application. The browser allows real-time zooming, panning in the simulation, and enables searching and locating of halos and subhalos in the simulation. The application is built upon existing web technology. Two main libraries used are Gigapan<sup>5</sup>, and the Microsoft Seadragon library <sup>6</sup>.

Figure 2.7 shows the interface of the interactive browser. The browser can be accessed from URL http://mbii.phys.cmu.edu . It consists of a viewport and three floating control panels: the MAIN panel, located at the top-right corner of the interface; the INFOrmation panel, located at the left side of the interface; and the NAVigation panel, located at the bottom right corner of the interface.

The viewport is where the Gigapan image of the selected snapshot is displayed. On the viewport, sub-halos are marked with green crosses. In addition, central sub-halos  $(M_{\rm sub-halo} > 0.1M_{\rm group})$  are marked with an additional circle. Interactive zooming and panning in the viewport is implemented via mouse clicking and dragging.

The MAIN panel provides the following functionalities:

- 1. selecting an epoch from the snapshot number;
- 2. switching between the gas and stellar layer;
- 3. jumping among FOF groups;

<sup>&</sup>lt;sup>5</sup>http://www.gigapan.org

 $<sup>^{6}</sup> http://gallery.expression.microsoft.com/SeadragonAjax$ 

4. querying sub-halos in current view.

The INFO panel displays the properties of the currently selected sub-halo or group. In Figure 2.7, for example, the panels shows the property of the currently selected sub-halo (marked with a rectangle).

The NAV panel provides zoom-in and zoom-out controls, and a switch to toggle the visibility of other control panels.

Figure 2.8 shows a collage of images extracted from the browser. We selected three halos in the simulation at redshift z = 1.0: (I) at a major confluence of filaments and halo majors; (II) a moderate halo of three small majors; (III) a relatively isolated halo. For each of the halo we show the stellar component in their subhalos, embedded in their surrounding gas environment.

The gigapan images used in the browser are high resolution 2-D images of the full simulation rendered with the visualization software Gaepsi [Feng et al., 2011]. The gas images (panels O, I, II, III) are rendered with the divergent Cool-Warm color-map introduced by Moreland [2009]. The density information is encoded in the brightness of the pixels: brighter pixels have higher column density, and voids are represented with black (zero-brightness). The temperature of gas is encoded in the hue of the pixels, blue represents low temperature ( $T < 10^{3.5}$  K), and red represents high temperature ( $T > 10^{7.5}$  K). The stellar images (panels a, b, c, d, e, and f) are composed from the simulated i, r, and g band luminosity (band definition follows the convention used by the Sloan Digital Sky Survey). The procedure is similar to that described in Lupton et al. [2004].

## 2.9 Conclusions

We have presented a framework for generating and viewing large images and movies of the formation of structure in cosmological SPH simulations. This framework has been designed specifically to tackle the problems that occur with the largest datasets. In the generation of images, it includes parallel rasterization (for either shared and distributed memory) and adaptive pixel filling which leads to a well behaved filling rate at high resolution. For viewing images, the GigaPan viewers use hierarchical caching and cloud based storage to make even the largest of these datasets fully explorable at high resolution by anyone with an internet connection. We further extend the framework with an interactive simulation browser. We make our image making toolkit publicly available, and the GigaPan web resources are likewise publicly accessible.



Figure 2.8: Visualization of MassiveBlack II simulation. Panel O shows the snapshot (z = 1.0) of the full simulation box mapped into a  $8 h^{-1}$ Mpc thick slice. Panel I, II, III show the gas environment of three FoF groups. Panel a to f show the stellar component of the sub-halos. The central subhalo is marked by dots, and 10 of the brightest sub-halos are marked with stars. Please see text for a description of the color scheme. The interactive simulation browser is available at http://mbii.phys.cmu.edu .

## Chapter 3

# Cosmological Hydrodynamic Simulation: Cold Flow and Quasars

In this chapter we apply the resimulation method to a sample of three halos hosting  $10^9 M_{\odot}$  black holes in the hydrodynamical MassiveBlack simulation. Our main goal is to study the growth of these extreme objects at higher resolution. We will examine the effect of different numerical schemes on the accretion history of black holes, then proceed to investigate how gas arrives to and participates in the growth of supermassive black holes.

The chapter is organized as follows: in Section 2 we discuss the selection and construction of the initial conditions; in Section 3, we study how method of feedback energy deposition, resolution and SPH formulation affect the accretion history of the supermassive black holes; in Section 4 we investigate in detail the picture of cold flow feeding.

## **3.1** Initial Conditions

### 3.1.1 Selection of Halos

The MB simulation can be used to examine the relationship between black hole mass and halo mass for the entire population of hosting halos, and this is shown in Figure 3.1 at three redshifts z = 7, z = 6, and z = 4.75.

We can see that for the same host mass  $M_{\text{HALO}}$  of a few times  $10^{12} M_{\odot}$ , the black hole mass can vary significantly (by greater than an order of magnitude). Similar scatter is also reported by Fanidakis et al. [2013] from semi-analytic modelling of black hole growth in dark matter simulation. The scatter shows that picking target halos from a hydrodynamic simulation is necessary to ensure that we are sampling the distribution of supermassive black holes that we want.



Figure 3.1: The mass of black holes  $M_{\rm bh}$  compared to the total mass of their host dark matter halos,  $M_{\rm halo}$ . The color in the histogram represents the number density of halos per logarithm of black hole mass and per logarithm of halo mass. The three halos we have targeted for resimulation are marked with colored circles: red is halo 1, green is halo 2, and blue is halo 3. The black hole in halo 3 at z = 7 is not shown in the figure because its mass is less than  $10^8 M_{\odot}$ .



Figure 3.2: The environment of the halos targeted for resimulation as seen in the MB simulation at redshift z = 6. From left to right: Halo 1, Halo 2 and Halo 3. Shown is the projected gas density color coded by temperature (red =  $10^8$ K, blue =  $10^4$ K). The scale in each panel marks the virial radius (typically 100 Kpc at z = 6.0.

h	$\Omega_b$	$\Omega_{\Lambda}$	$\Omega_M$	$\sigma_8$	$z_{ m init}$	$z_{\rm end}$
0.72	0.044	0.74	0.26	0.8	159.0	4.75

 Table 3.1: Cosmology Parameters

We select three halos from the MB simulation based on their black hole mass at z = 6. We pick 3 of the most massive black holes (shown as circles in Figure 3.1). Their large scale gas environment is shown in figure 3.2. The three halos evolve differently even though they have a similar mass of a few times  $10^{12} M_{\odot}$  at z = 6. The ranking of halos by black hole mass can change significantly between redshifts. For example, at z = 7 the black hole mass in halo 2 does not even show up in Figure 3.1 ( $M_{\rm BH} < 10^8 M_{\odot}$ ), but at z = 6.0 it becomes one of the most massive black holes ( $M_{\rm BH} \sim 10^9 M_{\odot}$ ).

We now describe the environment and black holes in each halo in turn (see also Figure 3.4):

- Halo 1 lies along the most prominent filament in the MB simulation (first panel in Figure 3.2. The halo eventually amalgamates with nearby halos into one "super halo" that is several times more massive than any other halos in the simulation. The most massive progenitor black hole grew rapidly at high redshift z = 9 via a continuous supply of cold gas flows through the filament. At low redshift the black hole mass increased due to a major merger event. Two nearby black holes are also approaching the most massive black hole, though they have not merged at z = 4.75 (end of simulation).
- Halo 2 is located at the confluence of 3 filaments (second panel in Figure 3.2). From a general visualization of the MB simulation [Feng et al., 2011] we can infer that this is one of the typical configurations for massive halo growth by accretion from filaments. The black hole went though a rapid growth phase between z = 7 and z = 6, when the black hole travels though a region of high density gas  $(n_b \sim 100 \,\mathrm{cm}^{-3})$ .
- Halo 3 is hosted by a quiescent environment. The growth of the black hole has been quenched at z = 7 by feedback which has reduced the cold gas supply. As we can see from the third panel in Figure 3.2, the halo is surrounded by a large region of hot gas.

### 3.1.2 Generating the Initial Conditions

The cosmological parameters used for the initial conditions are identical to that used in MB (listed in Table 3.1).

Our algorithm for generating the initial conditions uses as a base the code from the initial condition generator that was used for the MB simulation (N-GenIC, author



Figure 3.3: Illustration of gradual degrading of resolution of initial conditions with 2-d projection of a slice along z-direction. Color represents the initial Zel'Dovich x-velocity; the points represent particles: larger points correspond to lower resolution; in this particular set up, there are 5 resolution levels. For illustration purpose, the particles in the figure are not displaced by their initial displacement. (The initial conditions are.) The large scale fluctuation blends smoothly across the boundary between zoom regions as seen for the x-velocity field. The total mass and center of mass of particles are conserved by ensuring that whenever a lower resolution particle is replaced by a higher resolution particle, it is fully replaced by 8 particles (shown as four in the 2-d projection).

Volker Springel). This is in order to preserve the same randomly sampled large scale Fourier modes used in MB. We modified N-GenIC to produce multiple levels of displacement at a selected region, and implemented a post-processing algorithm that assembles the particles and their displacement at different resolution levels while conserving the total mass and the center of mass of particles.

To define the high resolution zoom region we first find the dark matter particles in the friends-of-friends [FOF, see Davis et al., 1985] group corresponding to the selected halo at redshift z = 6. We then find the initial positions of these particles, and put down a bounding sphere. We enlarge the bounding sphere by a factor of 1.5, and the region inside this sphere is our highest resolution zoom region. We populate this region with high resolution gas and dark matter particles whose initial perturbation is the sum of the interpolated (4th order spline) large scale mode perturbations in the original MB initial conditions and additional small scale mode perturbations (with random Fourier phases) drawn from smaller boxes. Outside the spherical high resolution region, a series of spherical shells are populated with lower resolution matter particles (collisionless), each shell radius being a factor of 1.14 larger than that interior to it, until we have reached a spatial resolution 8 times worse than that of MB. The rest of the cubical volume is populated with particles of this mass. The setup is illustrated with a 2d-projection of the x velocity field in Figure 3.3.

During the simulation, we measure the minimal distance from the nearest low resolution particle to the central black hole. The closest distance is  $1h^{-1}$ Mpc in comoving units, which is greater than the virial radius of the halo. This provides enough evidence that particles from the low resolution regions do not contaminate the black hole accretion and halo growth.

## 3.2 Simulations

Our fiducial SPH simulation code is the same as that used to run the original MB model [Di Matteo et al., 2012], P-GADGET3 [see Springel, 2005, for details of the hydrodynamics and gravity computation]. The formulation of SPH used in MB is density-entropy SPH [Springel and Hernquist, 2002, Lucy, 1977, Gingold and Monaghan, 1977], with a cubic spline kernel, the black hole accretion model from Di Matteo et al. [2005] and the multiphase star-formation model from Springel and Hernquist [2003]. In order to assess the effects of choices for the numerical schemes, we incorporate additional features into the code:

- a fixed feedback deposition volume in the black hole accretion model;
- an alternative, pressure-entropy SPH formulation to alleviate the problem of unphysical SPH surface tension [Hopkins, 2013, Read et al., 2010];
- a quintic smoothing kernel to alleviate the problem of pairing instability; we also change the variable controlling the SPH resolution from  $N_{\text{NGB}}$  (number

Symbol	Halo	$\varepsilon_{\rm grav} \ h^{-1}{\rm Kpc}$	$m_{0{ m DM}}~{ m M}_{\odot}$	SPH	Kernel	Feedback
1HDCV	1	1.5	$4.3 \times 10^6$	DE	Cubic	Volume
1MDCV	1	3.0	$3.5 \times 10^7$	DE	Cubic	Volume
1MDCA	1	3.0	$3.5  imes 10^7$	DE	Cubic	Adaptive
1MDQV	1	3.0	$3.5 \times 10^7$	DE	Quintic	Volume
1MPQV	1	3.0	$3.5  imes 10^7$	$\mathbf{PE}$	Quintic	Volume
$1LDCA^{\dagger}$	1	5.5	$2.8 \times 10^8$	DE	Cubic	Adaptive
2HDCV	2	1.5	$4.3 \times 10^6$	DE	Cubic	Volume
2MDCV	2	3.0	$3.5 \times 10^7$	DE	Cubic	Volume
2MDCA	2	3.0	$3.5  imes 10^7$	DE	Cubic	Adaptive
2MDQV	2	3.0	$3.5 \times 10^7$	DE	Quintic	Volume
2 MPQV	2	3.0	$3.5  imes 10^7$	$\mathbf{PE}$	Quintic	Volume
$2LDCA^{\dagger}$	2	5.5	$2.8 \times 10^8$	DE	Cubic	Adaptive
3HDCV	3	1.5	$4.3 \times 10^6$	DE	Cubic	Volume
3MDCV	3	3.0	$3.5 \times 10^7$	DE	Cubic	Volume
3MDCA	3	3.0	$3.5  imes 10^7$	DE	Cubic	Adaptive
3MDQV	3	3.0	$3.5 \times 10^7$	DE	Quintic	Volume
3MPQV	3	3.0	$3.5 \times 10^7$	$\mathbf{PE}$	Quintic	Volume
$3LDCA^{\dagger}$	3	5.5	$2.8 \times 10^8$	DE	Cubic	Adaptive

Table 3.2: Simulation matrix. D: density-entropy formulation; P: pressure-entropy formulation; V: radius of feedback region fixed to 0.5 physical  $h^{-1}$ Kpc; A: feedback region is set to 64 nearest neighbour (cubic spline) or 224 nearest neighbour (quintic spline); H, M, L stands for the resolution (high, medium and low, see text); Q: quintic spline kernel; C: cubic spline kernel.

† the LDCA simulations use the same accretion and feedback model as MassiveBlack.

of nearest neighbours) to  $\eta$  (ratio between SPH resolution and mean particle separation) recommended by Price [2012] [see also Dehnen and Aly, 2012].

In this work we present 18 simulation with various combinations of these models and the resolutions. The matrix of the simulations and their abbreviated names are listed in Table 3.2. We note that our LDCA simulations use the same model and resolution as MassiveBlack; and the HDCV simulations are the simulations with the highest resolution (64 times finer mass resolution and 4 times finer spatial resolution than MB).

### 3.2.1 Blackhole Accretion Model

The black hole accretion model was introduced by Springel et al. [2005a] and Di Matteo et al. [2005]. Here we summarize some relevant features. The black holes

are seeded in the simulations on-the-fly in newly formed FOF halos with  $M_{\text{Halo}} \leq 5 \times 10^{10} h^{-1} M_{\odot}$ . The seed mass is  $M_{\text{bh,seed}} = 5 \times 10^5 h^{-1} M_{\odot}$ . The accretion rate of a black hole is calculated from

$$\frac{dM}{dt} = 4\pi \cdot \min \begin{cases} A\rho(x)c_{\text{eff}}(x)^{-3}(GM)^2 & \text{Bondi-Scaling} \\ Bm_{\text{p}}(\eta\sigma_T c)^{-1}(GM) & \text{Eddington Limit} \end{cases}$$

,

where the local gas density  $\rho(x)$  and local effective sound speed  $c_{\text{eff}}(x)$  are evaluated from the nearest neighbour SPH estimate at the black hole position x.  $\sigma_T$  is the Thompson cross section, G the gravitational constant and  $m_p$  the proton mass. We have absorbed other constants in A and B. The light to mass ratio  $\eta$  is taken as 10%. We also assume that 5% of the light emitted becomes thermal feedback energy.

Two black holes are merged when two nearby (within the SPH resolution) black holes satisfy the condition  $v_{\rm rel} < 1/2c_{\rm sound}$ , where  $v_{\rm rel}$  is their pairwise relative velocity and  $c_{\rm sound}$  is the sound speed of the surrounding gas.

### **3.2.2** Merger and Accretion history

When examining the accretion history of black holes in each zoom region in Sections 3.3 onward we will restrict ourselves to the most-massive progenitor of the most-massive black hole at z = 6. We find that halo 2 and halo 3 contain one central black hole each at the end of the simulation, while halo 1 contains three major black holes of similar mass. This can be seen in Figure 3.4 where we show the mass evolution of the 3 most massive (final redshift) black holes and their progenitors in each halo. The difference in the environments of the black holes seen in Figure 3.2 shows up strikingly here too, with for example halo 3 hosting only one early merger events between black holes at close to z = 8.

We note that where there are differences in the precise ranking of black holes for different versions of a simulation we define as the central black hole the most massive black hole in the highest resolution simulation.

In the following subsections we explore the effect of varying the accretion and feedback models on the growth of the central black hole.

### 3.2.3 Deposition of AGN Feedback Energy

The black holes deposit feedback energy into nearby gas environment. In our simulations, we model this process by depositing thermal feedback energy to the neighbouring gas particles, weighted by their mass. The exact details of the physical mechanism by which this occurs in real galaxies is likely dictated by radiative transfer through the medium surrounding the black hole [Alvarez et al., 2009, Ciotti and Ostriker, 2007, Hayes et al., 2006, Jeon et al., 2012, Novak et al., 2012, Milosavljević et al., 2009, Park and Ricotti, 2011, 2012]. As the length and mass scales over which this occurs



Figure 3.4: The merger tree of three most massive black holes in each simulated halo. The data has been taken from the simulations with the highest resolution (HDCV). From left to right we show halos 1, 2 and 3. The thick lines show the most massive progenitors in each case; black holes in the same merger tree are shown with the same color. The lower panel shows the accretion rate of the most massive progenitors; grey lines indicate the Eddington rate. The circles and squares mark the snapshot redshifts at which the particles used in the EA phase and RA phase analysis are selected in Section 3.3. The shaded area shows the duration of the EA phase.



Figure 3.5: The dependence of black hole accretion history on feedback energy deposition prescription. From left to right we plot halo 1, 2, 3. The blue curves show results for the fixed-volume (finite physical radius of  $0.5h^{-1}$ Kpc) feedback kernel. The green curves are for the fixed-mass kernel, which uses the 64 nearest neighbours. We also show the star formation rate of the halo in the bottom panels. Refer to the text in Section 3.2.3.

are not certain, it is important to test how our sub-grid prescription for depositing this feedback energy affects growth of the black holes.

In the original MB simulation the energy deposition was done using the SPH smoothing kernel of the black hole particle to distribute energy to the 64 nearest neighbors. We refer to the MB feedback method "adaptive", as the size of the feedback region is directly proportional to the mean separation of SPH particles close to the black hole (a ratio of  $\eta = 1.26$ ). In order to preserved the ratio, we use 224 neighbours in the quintic spline kernel simulations.

An alternative to the "adaptive" model is to fix the proper volume of the gas receiving the feedback energy. For the fixed-volume model, we use a region corresponding to a spline smoothing kernel with proper radius  $h = 0.5 h^{-1}$ Kpc. The total mass of gas receiving the feedback energy is therefore proportional to the gas density around the black hole. Note that we do not use the fixed-volume model at low resolution as the feedback region is smaller than the gravitational smoothing when even at very high redshift (z < 10).

In Figure 3.5, we compare the results of simulations with the adaptive model for distributing feedback energy and those with the fixed-volume model at the medium resolution (MDCA and MDCV simulations in Table 3.2). We can see that the black hole mass, accretion and the halo star formation in all three halos are very similar for both feedback models (within the same order of magnitude). The only exception is in halo 1, where there is some difference in the early growth: the accretion rate in the adaptive model remains 2 orders of magnitude lower than the fixed-volume model until redshift z = 8.5. The difference in accretion rate disappears after z < 8, and eventually the black hole mass between two models become very similar. The star formation in the halo also experiences a burst during the same period, indicating an accumulation of cold dense gas in the halo before the black hole accretion kicks in. After the accretion is started, the black hole mass quickly picks up. We therefore regard the similarity in the late time black hole masses and accretion rates as more definitive. The choice of feedback region therefore does not appear to significantly alter the accretion history.

#### 3.2.4 Resolution

The MB simulation has a gravitational force softening length of  $\varepsilon = 5.5 h^{-1}$ Kpc. The re-simulations are conducted with three different gravitational force softening lengths (in comoving coordinates):  $5.5 h^{-1}$ Kpc (low resolution),  $3.0 h^{-1}$ Kpc (med resolution), and  $1.5 h^{-1}$ Kpc (high resolution). Our highest resolution simulation therefore has a proper resolution of 300pc at z = 6. The simulations do not resolve the influence sphere of the supermassive black holes, which has a radius close to 50pc at  $z \sim 6$ . This is consistent with the model used in this work, which was calibrated for a subgrid treatment of the black hole vicinity. [see, e.g. Di Matteo et al., 2008]

We plot the accretion history for the 3 different black holes at different resolutions



Figure 3.6: Resolution dependence of the accretion history of the most massive black holes in our 3 selected target halos. From left to right we plot Halo 1, Halo 2, and Halo 3. The blue (LDCA) curves are the lowest resolution, green and red (MDCV, MDCA) medium resolution and yellow (HDCV) is high resolution (see Section 3.2.4).

in Figure 3.6, showing black hole mass and accretion rate and the star formation rate in the halo. We find that the resolution does not appear to be affecting the black hole growth and the star formation in the halo. The accretion history appear to have reached convergence with the medium resolution simulations (MDCV and MDCA), which shows no significant difference with the high resolution simulations (HDCV) that has 8 times better mass resolution. We also note that in the resimulations, the suppression in star formation rate is tightly coupled with the suppression of black hole accretion, in agreement with Khandai et al. [2012]. The level of suppression varies from case to case. We attribute the variance to the difference of amount of black hole feedback: the simulations with more massive black holes are in general associated with strong suppression in star formation. At high redshift when both black hole and star formation are determined by the cold gas supply, the star formation shows very good agreement.

The major difference again lies in the adaptive feedback model. We see that the initial growth (z > 8) of the adaptive model shows stronger resolution dependence than the fixed-volume model in both halo 1 and halo 3: the lower resolution simulation (LDCA) produces slower initial growth at z > 8 than the higher resolution (MDCA). The difference in halo 3 is less drastic than the difference in halo 1 and is not associated with a burst in star formation rate. We believe this is because the halo is in a relatively isolated region, and the cold gas supply is lower than halo 1.

We point out that eventually (z < 6) the lower resolution simulation (LDCA) does also reach the same final black hole mass as the medium resolution simulation (MDCA), with the dependence on resolution disappearing. This is because at z < 6 the black hole is sufficiently massive that the accretion becomes regulated by feedback, regardless of how the feedback is implemented.

### 3.2.5 SPH formulation: Pressure-Entropy

Recent developments in SPH include the introduction of the so-called pressure-entropy formulation [Hopkins, 2013, Read et al., 2010], and the quintic smoothing kernel [Price, 2008]. These new developments aimed to address several difficulties noticed in prior formulations of SPH, namely: an unphysical surface tension across surface boundaries that forbids particle exchange at a density discontinuity, the fluctuation in density estimation due to a small number of nearest neighbours (32), and nearby particles bonding into pairs causing a loss of resolution in high density regions [see e.g. Agertz et al., 2007, Morris, 1996, Schuessler and Schmitt, 1981, Monaghan, 2000, Wadsley et al., 2008].

We have run our zoomed simulations with pressure-entropy SPH and the traditional density-entropy SPH formulation. The black hole accretion histories and the halo star formation rates are shown in Figure 3.7. We can see that going from density-entropy to pressure-entropy formulations appears to consistently result in faster black hole accretion. The star-formation history is largely intact, although



Figure 3.7: The effect of changing SPH formulation on the accretion history of the black holes. From left to right we show halo 1, halo 2, and halo 3. The blue (MDCV) line is the density-entropy formulation with a cubic spline smoothing kernel; the green (MPCV) line is the pressure-entropy formulation with a cubic spline smoothing kernel and the red (MPQV) line is the pressure-entropy formulation with quintic spline smoothing kernel.



Figure 3.8: the dependence of black hole accretion history on seeding mass. The blue (M5), red (M4) and yellow (M3) represent the simulation at medium resolution with seeding mass of  $5 \times 10^3 h^{-1} M_{\odot}$ ,  $5 \times 10^4 h^{-1} M_{\odot}$  and  $5 \times 10^5 h^{-1} M_{\odot}$ . H4 (green dotted line) represents a high resolution simulation that is run to redshift 6.5.

with some variance in the level of suppression at low redshift. (See discussion in the previous section.) Comparing the results for quintic spline and cubic kernels (both are plotted) we can see that some of the contribution to the faster black hole accretion is due to the switch to a quintic spline kernel. The quintic spline kernel samples more particles (112) and reduces the noise in the quantities used the black hole accretion model. The black holes are up to a factor of 5 times more massive at the final time in the pressure-entropy case compared to density entropy case. We note that there is a free parameter in the black hole model, namely the feedback efficiency (set to 10 % here). This parameter was set [Di Matteo et al., 2005] to reproduce black hole masses observed in the local Universe and so with pressure-entropy SPH a recalibration of this parameter could be employed.

### 3.2.6 Seeding Mass

The origin and nature of the black hole seed population remain uncertain. Two distinct populations of black hole seeds, in the range of  $100 - 10^6 M_{\odot}$  have been

proposed, either from the remnants of the first generation of Population III stars at  $z \sim 20 - 30$  or from direct dynamical collapse of gas. In large volume cosmological simulations such as MassiveBlack it is not currently possible to resolve the collapse of the (small mass) halos at redshifts commensurate with black hole seed formation and hence resolve these processes directly. As a result, the black hole seed mass is a parameter of our models [see also Costa et al., 2013], and the early growth phase cannot be accurately followed.

In this section we explore the effect of changing our BH seed mass on the growth of the first quasars. We have performed a series of resimulations of Halo II, where we decrease the seed mass by up to a factor of 100. In particular, we use  $M_{\text{seed}} = 5 \times 10^3 h^{-1} M_{\odot}$  (M3),  $5 \times 10^4 h^{-1} M_{\odot}$  (M4),  $5 \times 10^5 h^{-1} M_{\odot}$  (M5) and seed them in halos of  $M_{\text{HALO,seed}} = 5 \times 10^8 h^{-1} M_{\odot}$ ,  $5 \times 10^9 h^{-1} M_{\odot}$ ,  $5 \times 10^{10} h^{-1} M_{\odot}$ , respectively. In these zoom runs, we are limited by the minimum mass of a resolved halo (which corresponds to  $10^8 M_{\odot}$  in the medium resolution simulations): reducing the BH seed mass further (e.g., closer to  $100 M_{\odot}$ ) would require resolving halos of the order of  $10^6 M_{\odot}$  commensurate with those forming first PopIII stars. We show the accretion history and star formation history in Figure 3.8.

The accretion history in the simulations with different seed mass follows a similar growth history. In general, by lowering the seeding mass,

- 1. the early (z > 8) growth of the supermassive black hole is somehat suppressed;
- 2. the early star formation history is neither enhanced or suppressed, indicating that black hole accretion and associated feedback are not affecting the gas supply for star-formation;
- 3. smaller black holes are seeded earlier due to the reduced seeding halo mass; the earlier seeding compensates for the smaller seeding mass, and the black hole undergoes a longer Eddington growth period;
- 4. eventually, regardless of the mass of the seed, the black holes grow to a final mass of about  $10^9 M_{\odot}$ .

In conclusion, we observe no significant change to the black hole accretion history in our simulations as long as a reasonable seeding mass is used.

## **3.3** Feeding Blackholes via Cold Flows

We now investigate the fuel supply onto black holes and study the thermal history of high redshift gas that participates in black hole accretion. We focus on the question of whether the gas that fuels the black hole is different from the typical gas in the halo. We want to further test the indication that at these high redshifts ( $z \sim 6$ ) fast black hole growth can be achieved via cold flows [Di Matteo et al., 2012]. Further



Figure 3.9: Visualization of Gas Density at Varying Resolutions. Top 3 rows: Cold gas filaments surrounding black holes in the simulations, run at different resolution. The color represents the projected density (in a cube) of cold ( $T < 10^5$ K) gas at z = 6.5. The columns show Halos 1, 2, and 3, from left to right. The first 3 rows are, from top to bottom, low resolution, medium resolution and high resolution. The bottom row shows the density distribution of the hot gas in the high resolution simulation.

we aim to investigate how black hole feedback may play a role in disrupting the cold streams close to the black hole.

The virial temperature of the hosting halo of  $10^{12} M_{\odot}$  is close to  $10^{6}$ K, while in the simulation the cold IGM temperature is close to  $10^{4}$ K. We set the split between hot and cold gas at  $10^{5}$ K. In Figure 3.9, we compare the morphology of the cold gas ( $T < 10^{5}$  K) and that of the hot gas ( $T > 10^{5}$  K) around the black holes in the three halos. We first notice that the cold gas forms dense compact filaments that are unlikely due to resolution effects: as we increase resolution, the cold gas filaments are identical in morphology except have increased sharpness. Secondly, we notice that the hot gas is distributed in a diffuse manner, forming thicker structures containing large blobs that extend to the entire halo.

The striking morphological difference between hot and cold gas motivates us to consider two species of gas particles:

- 1. the first gas species we call "HALO" particles<sup>1</sup> and are randomly selected gas particles from the halo, weighted by volume; because hot gas particles occupy most of the volume of the halo, HALO particles are most likely the hot environment gas in the halo shown in the bottom panel of Figure 3.9.
- 2. the second gas species are the "BH" particles<sup>2</sup>, and are the nearest neighbour particles contributing to the evaluation of gas properties at the black hole. The BH particles are the gas particles that are actively participating in accretion onto the black hole.

After a gas particle (if it has not been converted to a star particle) enters the halo, depending on its cooling efficiency, the particle will either be heated to the virial temperature of the halo, losing its initial inflow motion, or remain cold, arriving to the center of the halo to participate in star formation and black hole accretion. In this picture, the history of a gas particle can be quantified by two characteristic times:

- 1. the entering time  $t_{enter}$ , the time since the Big Bang at which a gas particle first enters the virial radius;
- 2. the heating time  $t_{\text{heated}}$ , the time since the Big Bang at which a gas particle is first heated above the virial temperature of the gas.

We also calculate the correlation coefficient (CC) of the two characteristic times

$$\mathsf{CC} = \rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

where X and Y represent  $t_{\text{enter}}$  and  $t_{\text{heated}}$ ,  $\sigma$  and  $\mu$  the standard-deviation and the mean of the data, and E the mean operator.

<sup>&</sup>lt;sup>1</sup>Not to be confused with the dark matter particles in the halo.

 $<sup>^{2}</sup>$ Not to be confused with particles that represents the black holes in the simulation.

Halo	1		2		3	
Species	BH	HALO	BH	HALO	BH	HALO
EA (high redshift)	0.069	0.574	-0.075	0.778	0.131	0.337
RA $(z = 5.5)$	0.389	0.893	0.356	0.745	0.557	0.827

Table 3.3: Correlation coefficient (CC) between  $t_{\text{heated}}$  and  $t_{\text{enter}}$ . Also see Figure 3.10.

CC measures the correlation between the time a gas particle enters the halo and the time a particle is heated to the virial temperature of the halo. If a particle is heated by the virial heating from the halo, we expect CC to be large, for particles entering the halo early shall be heated early, while those entering late shall be heated late. On the other hand, if a particle is heated by the central black hole, we expected CC to be small, for these particles are heated only after they arrive to the black hole, no matter when they enter the halo. As we will show later, these expectations agree with our simulations.

We calculate  $t_{enter}$ ,  $t_{heated}$ , and CC for BH and HALO particles at two epochs from the high resolution simulations (HDCV): (1) the first epoch is when the black hole is undergoing Eddington accretion (EA phase,  $\dot{M}_{BH} \sim \dot{M}_{Edd}$ ), and (2) the second epoch is when the black hole accretion has been regulated (RA phase,  $\dot{M}_{BH} \ll \dot{M}_{Edd}$ ). The EA phases of three halos are marked with shaded areas in Figure 3.4. We choose the particles in redshifts selected a priori (without the knowledge of the behavior of particles) at z = 9, 6.5, and 7.5 for each halo respectively to perform the analysis. Similarly, in the RA phase analysis, we use the particles at redshift of z = 5.5 for all three halos.

The results are shown in Figure 3.10 and Table 3.3. We first look in the EA phase. During the EA phase, the correlation coefficient (CC) between the heating time and entering time of the BH particles is much smaller than that of the HALO particles as seen in the first row in Table 3.3. The difference can also be seen in Figure 3.10, where the BH particles are clustering around the same  $t_{\text{heated}}$  that corresponds to the time the particle becomes the nearest neighbour of the black hole, while the HALO particles follows a line with a positive slope. The lower CC in BH particles is consistent with heating by black hole feedback: the BH particles remain cold after they enter the halo, and are then heated up shortly after they arrive at the black hole. The larger CC with HALO particles in EA phase is consistent with the virial heating picture: the hot gas halo is formed by virial heating from the halo (which may indirectly contain the feedback energy from the central black hole and the feedback energy from satellites); if a HALO particle enters the halo early, it is also heated early.

This difference suggests that the gas that participates the black hole accretion in an Eddington accretion phase is unlikely to be from the hot halo environment. The gas must have come directly from outside of the halo; and its thermal history is consistent with that of the cold flows.


Figure 3.10: Time since the Big Bang that a particle enters a halo  $(t_{enter})$  plotted against time since the Big Big for a particle to heat up to the virial temperature  $(t_{heat})$ HALO particles and BH particles (see text) are shown in red and blue respectively. The panels show halos 1, 2 and 3 from left to right. The upper panels are at the EA phase when the black holes are undergoing Eddington accretion. The lower panels are in the RA phase when the black hole growth is regulated by feedback. The red horizontal dashed line indicates the time when most of the BH particles are heated, acting as a visual guide to the timescale of black hole heating. The blue dashed line is a linear fit to the HALO particles, visually guide the time scale of halo heating.

We then move on to the RA phase, during which black hole accretion has been regulated. We notice that the CC of the BH particles have significantly increased. In the lower panels of Figure 3.10, we find this increase is due to a larger population of the BH particles that behaves similarly to the HALO particles. The existence of a HALO-like population within the BH particles shows that the cold flows (though they still exist) are being somewhat disrupted. Because the HALO-like particles are heated before they arrive to the black hole, the temperature around the black hole rises, and the black hole accretion is regulated.

We investigate further the motion of BH particles and HALO particles in the RA phase with Figure 3.11 and Figure 3.12 & 3.13. In Figure 3.11, we plot the history of properties of BH and HALO particles in simulation 1MDCV, where we saved one snapshot every 3 Myear of simulation time. In this case we have traced 32 particles of each population originating from three snapshots close to z = 5.5 (three columns

are three snapshots; two rows are BH and HALO from top to bottom, note that the time increases from right to left). The considered properties are:

- T: the temperature of the particle; we compare the temperature with the virial temperature of the halo.
- r: the distance between the particle and the black hole (proper); we compare the distance with the virial radius of the halo.
- $v_r$ : the radial velocity of the particle; negative value ( $v_r < 0$ ) is infalling and positive ( $v_r > 0$ ) is out-flowing.
- *n*: the density of the particle.

We visualize the history of each property with a 2-d histogram binned by the quantity and time in Figure 3.11. The histogram demonstrates the behavior of the entire population as a function of time: the particles with similar property at a given time will contribute to the same bin, resulting a darker pixel. For example, the dark horizontal line at  $T = 10^4$ K in the top panel T plots of Figure 3.11 indicates a large fraction of gas remains cold for a very long time. We also mark the probability distribution of  $t_{\text{heated}}$  and  $t_{\text{enter}}$  in the figure. We are more interested in the behavior of the gas between  $t_{\text{enter}}$  and  $t_{\text{heated}}$ .

In the three top panels for BH particles, we see that most of the BH particles have significantly increased their density while remaining cold for as long as 0.5 Gyear after they enter the halo, as suggested by the dark lines in T,  $v_r$  and n. We do notice that for some of the particles, the infalling motion starts to become disrupted much sooner (0.2 Gyear after entering), while their temperature rise to the virial temperature. As we will see after analyzing the HALO particles, this early heating population within BH gas are very similar to the HALO particles: they correspond to the HALO-like population we identified in the lower panels of Figure 3.10, which are responsible for the increased correlation between  $t_{enter}$  and  $t_{heated}$ . The population indicates that the cold flows in a regulated growth phase are being disrupted, and can no longer provide the Eddington accretion of black holes. All of the BH particles are heated after they get close to the black hole ( $r < 1 \,\mathrm{Kpc}$ ), but the temperature reaches up to  $10^8 \,\mathrm{K}$ , much higher than the virial temperature of the halo  $(10^7 \,\mathrm{K})$  at this redshift. Some of the heated particles can receive an out-flowing velocity up to  $1000 \,\mathrm{km/s}$ , and can reach a substantial distance away from the black hole, though none of the heated particles have reached a distance further than the virial radius of the halo. We do note that our method cannot exclude the possibility that a fraction of heated particles escape from the halo. We plan to investigate whether any of the BH particles are unbound from the halo due to feedback from the black hole in a further study.

In the three low panels for HALO particles, we see that the heating happens much sooner (within 0.2 Gyear) after they enter the halo. The infalling motion is disrupted as the particles are heated, and most of the particles remains at a low density of



Figure 3.11: The properties of HALO and BH particles selected from three snapshots close to z = 5.5, as function of time. Top row : BH particles (nearest neighbours of the black hole). Bottom row: HALO particles (randomly picked from inside the halo). The sub-panels with in a figure are: temperature (T), distance to black hole (r), radial velocity  $(v_r)$  and density (n). The 2-d histogram visualizes the trajectories of particles, with the mean value shown in blue solid lines. (See text) The black dashed lines mark the virial temperature, virial radius, and zero radial velocity. The black vertical lines mark the average of the time a particle enters the halo  $(t_{enter}, see text)$ .



Figure 3.12: The trajectories of BH and HALO gas particles, projected into the x - y plane, during the RA phase. Top panels: BH particles that participate in accretion onto the black hole; Bottom panels: HALO particles. The trajectories are color-coded by temperature: red is  $10^8$ K, blue is  $10^3$ K. The wedges shows the direction of motion.



Figure 3.13: The trajectories of BH and HALO gas particles, projected into the x - y plane, during the EA phase. Top panel: BH particles that participate in accretion onto the black hole; Bottom panel: HALO particles. The trajectories are color-coded by temperature: red is  $10^8$ K, blue is  $10^3$ K. The wedges shows the direction of motion.

 $10^{-6}$  cm<sup>-3</sup>. After being heated, the temperature of HALO particles tightly tracks the virial temperature of the halo, as the halo grows.

We proceed to visualize the motion of the in-flowing gas with Figure 3.12 for the particles selected from the three snapshots near z = 5.5. In Figure 3.12, we show the trajectories of the BH and HALO particles projected in the x - y plane, colored by their temperature. The BH particles, being initially cold  $(T \sim 10^4 \text{K})$ , flow into the black hole (located at the center of the plots), apparently following several preferred directions. The majority of the BH particles remain cold until they arrive to the black hole. However, we can also see some of the particles are heated to  $T > 10^6 \text{K}$  before arriving the the black hole; these particles correspond to the HALO-like particles. Afterwards, the particles are ejected towards directions that are unaligned to the directions they come in from. The ejected gas can reach 100 Kpc away from the black hole, though still within the virial radius of the halo. The HALO particles show a very different pattern. Although the HALO particles are also initially cold  $(T \sim 10^4 \text{K})$ , the gas is quickly heated to the virial temperature of the halo  $(T \sim 10^6 \text{ K})$ . Afterwards, the motion of particles visually resembles a random-walk within the virial radius of the halo (a few times 100 Kpc).

The fate of BH particles in the EA phase is drastically different. As shown in the top panels of Figure 3.13, we see that instead of being ejected away, most of the BH particles remain cold after arriving at the black hole and they are eventually consumed by black hole accretion and star formation at the very center of the halo. The fate of HALO particles in the EA phase is rather similar to that in the RA phase: particles are heated to the virial tempearture as they fall into the halo.

#### 3.4 Conclusion

We have studied the fueling and gas supply onto high-redshift supermassive black holes in high resolution re-simulations of halos selected from the MassiveBlack hydrodynamic cosmological simulation (covering a volume close to 1 Gpc<sup>3</sup>). Using Massive-Black, Di Matteo et al. (2012) showed that steady high density cold gas flows were able to produce the high gas densities that can lead to sustained critical (Eddington) accretion rates and hence rapid growth commensurate with the existence of  $10^9 M_{\odot}$ black holes as early as  $z \sim 7$ . We have tested this scenario further for a subsample of three of the halos hosting  $10^9$  solar mass black holes at z = 6.0 in MassiveBlack. We have used zoom-in techniques to investigate the nature of the gas inflows at scales that could not be resolved from the large volume MassiveBlack simulation.

We have shown that for the three re-simulated halos the growth history of their central super massive black holes is consistent with a sustained Eddington limited fueling provided by cold streams of gas that penetrate all the way to the innermost region of the galaxies undisrupted (i.e. consistent with our earlier picture). Our conclusions remain unchanged even though the numerical scheme we use is varied in several significant aspects:

- a change in resolution does not change the fuelling history;
- AGN feedback energy is spread over different size regions of constant mass or volume,
- an improved formulation of SPH (pressure-entropy and quintic kernel) is used (in this case, the growth of the black holes is systematically faster).
- the black hole growth history is insensitive to changes of the seed black hole mass (in the range of  $M_{\text{seed}}5 \times 10^3 5 \times 10^5$ , commensurate with the simulation resolution).

There were several recent studies on the supermassive black holes based on the resimulation method. For example, Costa et al. [2013] applied a similar resimulation method to 18 halos selected from the Millennium simulation (dark matter only) and reported that halos in large scale over dense regions are able to experience critical growth commensurate with  $10^9 M_{\odot}$  by z = 6. Our results are in consistency with Costa et al. Bellovary et al. [2013] also studied the growth history of black holes in high redshift halos. Their zoomed regions were extracted from an initial cosmological box of only 50 Mpc per side, unlikely to contain high enough over dense regions. We note that their final black hole masses are consistent with what is expected from more moderately over-dense regions, such as those studied by Costa et al. [2013].

We have investigated the nature of the gas fueling the central black holes by computing the correlation between the time required for gas particles (in the halo) to reach the virial temperature and the time since they first enter the halo. We also investigated the correlation of gas participating in accretion and the correlation of the hot gas in the halo environment. We have found that while gas that fuels the black hole preserves its low temperature until heated by the feedback from the black hole (low correlation), the gas ending up in the halo environment typically heats quickly upon entering the halo (high correlation). With this method, we have shown that during the Eddington growth phase, the accretion gas appears to have directly arrived to the black hole through cold flows without disruption from the hot environment; while during the regulated growth phases (z < 6) a larger population of the accretion gas behave very similar to the hot environment, indicating that the cold flows are becoming disrupted by the feedback.

After being heated by the black hole the gas particles are ejected with a high velocity up to 1000km/s from the black hole, though none of the particles we have investigated have left the halo. Several authors [see, e.g. Dubois et al., 2013, McCarthy et al., 2011] reported that the strong outflows during the regulated phase at high redshift can reduce the baryon fraction in the halo up to 30%, increasing the entropy of halos at lower redshift. In our halos, the disruption appears to be not as strong, as most of the gas surrounding the black hole is cold even after the accretion is regulated. One possible source of the discrepancy is that their halos are less massive, and the supply of cold gas is less efficient: for example, our smallest halo (halo 3)

has the largest fraction of HALO-like gas population which is heated before they arrive to the black hole. We note that the faster black hole accretion in pressureentropy formulation suggests a higher feedback efficiency, which may contribute to the disruption of cold flow and produce a stronger hot outflow. We plan to investigate the hot outflows with further studies.

We have not evaluated the effects of resolving into the black hole zone of influence for our black holes. It is unclear whether the accretion model we used requires a recalibration. Modelling the zone of influence may require more detailed treatment of the physics processes, such as radiative transfer. Our simulations were only able to resolve halos more massive than  $\sim 10^8 M_{\odot}$ , which are too massive for seeding stellar originated black hole progenitors as the halos are formed. We shall leave these topics for further study.

## Chapter 4

# Parallel Radiative Transfer: Growth and Anisotropy of Ionization Fronts

In this chapter we study the growth of the ionization front of bright quasars in an almost neutral cosmological context. The quasars and their surrounding medium are selected from a large hydrodynamic simulation [the MassiveBlack simulation, introduced in Di Matteo et al., 2012, and then post-processed with a radiative transfer code. This allows us to simulate 8 rare quasars using reasonable computing resources. Our focus is on the evolution and properties of the largest individual ionized bubbles, the sources that produce them, and the relationship between the two. Because the simulation forms quasars and star forming galaxies ab initio, we are able to make use of the luminosity and positions of the radiation sources that the simulation produces, rather than setting them in by hand. However we do not deal with the full reionization of the volume of the simulation, which would require following the evolution of the entire density and ionization field from high redshifts down to at least z = 6. Instead we restrict ourselves to the growth of ionized regions around an early period in this process (at z = 8), where the photon path lengths are still smaller than the computational sub-volumes we analyze. We leave the study of the full reionization of the volume to future work.

## 4.1 Hydrodynamic Simulation and Density Field

The SPH output we use is from the MassiveBlack simulation [see Di Matteo et al., 2012, DeGraf et al., 2012, Khandai et al., 2012, for further details], which was run with a  $\Lambda$ CDM cosmology with parameters ( $\Omega_{\Lambda}, \Omega_{M}, \Omega_{b}, h, \sigma_{8}$ ) = (0.74, 0.26, 0.044, 0.72, 0.8). A total number of 2 × 3200<sup>3</sup> gas and dark matter particles were followed in a box of 0.75 Gpc<sup>3</sup> from redshift z = 159 to redshift z = 4.75. This simulation is by far the largest cosmological hydrodynamics simulation run with the P-GADGET program.

This run not only contains gravity and hydrodynamics but also the extra physics (sub-grid modeling) for star formation [Springel and Hernquist, 2003], black holes and associated feedback processes.

The basics aspects of the black hole accretion and feedback model [Di Matteo et al., 2008] consist of representing black holes by collisionless particles that grow in mass (from an initial seed black hole) by accretion of gas in their environments. The accretion rate is given by

$$M = \min(M_{\text{Bondi}}, M_{\text{Edd}}), \qquad (4.1)$$

and

$$\dot{M}_{\text{Bondi}} = \frac{4\pi G^2 M_{BH}^2 \rho}{(c_s^2 + v^2)^{3/2}},\tag{4.2}$$

where  $\rho$  and  $c_s$  are the density and sound speed of the ISM gas respectively, and v is the velocity of the black hole relative to the gas, and  $\dot{M}_{\rm Edd} = L_{\rm Edd}/(\eta c^2)$ , where  $L_{\rm Edd} = 1.26 \times 10^{38} \, {\rm erg s^{-1}}$ , is the Eddington Luminosity. At the high redshift and high halo masses we are carrying out the analysis here, black holes are growing at their Eddington rates [DeGraf et al., 2012] making any detail of the sub-grid models for black holes virtually irrelevant.

For the black hole feedback, a fraction of the radiative energy released by the accretion of material is assumed to couple thermally to nearby gas and influence its motion and thermodynamic state. The radiated luminosity,  $L_{\rm r}$ , from the black hole is related to accretion rate,  $\dot{M}_{\rm BH}$ , as

$$L_{\rm bol} = \eta \dot{M}_{\rm BH} c^2, \tag{4.3}$$

where we take the standard mean value  $\eta = 0.1$ . Some coupling between the liberated luminosity and the surrounding gas is expected: in the simulation 5% of the luminosity is isotropically deposited as thermal energy in the local black hole kernel, providing some form of feedback energy.

With our simulations the activity of quasars is directly derived from the accretion history of rapidly growing super-massive black holes, and their fueling is driven from the large scales and occurs through high density cold flows along the cosmic filaments [Di Matteo et al., 2012].

The luminosity of the stars and galaxies can be derived from the star formation rate history which is provided by the multiphase star formation model in the simulation that depends on a single free parameter,  $t_{\star}$  the global star formation timescale [Springel and Hernquist, 2003]. The multiphase star formation model also provides a mechanism to remove the self-shielded interstellar medium (ISM) from the matter density field, leaving only the intergalactic medium (IGM) density for the radiative transfer simulation.

In the multiphase star formation model, a dense gas particle is divided into a

non-star forming IGM component and a star forming ISM component:

$$m = \begin{cases} m_{\rm IGM} + m_{\rm ISM} = (1 - x)m + xm, & \text{if } \rho > \rho_{\rm th}, \\ m_{\rm IGM}, & \text{if } \rho \le \rho_{\rm th}, \end{cases}$$
(4.4)

where x is the mass fraction of the ISM component. The threshold density  $\rho_{\rm th}$  is determined from the global star formation time scale  $t_{\star}$ , as described in Springel and Hernquist [2003]. The IGM component of gas particles forms a hot ambient medium and occupies the entire volume of the particle. The ISM component of gas particles condenses into cold star-forming clouds which are self-shielded from cosmic radiative transfer, except that they may host stellar radiative sources. As a result we excise them from the matter density field when performing the ray tracing calculation. By doing this we also assume that their small cross-section would not have affected ray tracing for the rest of the gas. The ISM fraction, x, is an increasing function of the mean density of the gas particle, effectively removing the densest particles from the density field in a manner similar to the threshold method used to calculate the clumping factor by Pawlik et al. [2009], and, the removal of the cold high density gas in the X-ray emission calculation by Croft et al. [2001].

We also note that in MassiveBlack the mean baryon density (IGM + ISM) around quasar sources can be as high as  $60 \text{ cm}^{-3}$ . Were the ISM not excised from the ray tracing, the high mean density would shield off the radiation and prevent the growth of any cosmic scale ionized regions.

## 4.2 Selection of Quasars

We use the quasars and density field of the z = 8 snapshot of MassiveBlack in this study. There are two reasons for this choice:

- The EoR in MassiveBlack is modeled by a uniform UV background radiation field that is introduced near the end of the EoR (z = 6) in the optically thin approximation [see e.g. Bolton and Haehnelt, 2007]. By choosing an earlier redshift we do not contaminate the ionization fronts with this global radiation field.
- Extremely bright quasar sources at high redshifts are rare objects. Limited by the  $0.75 \,\mathrm{Gpc}^3$  volume of MassiveBlack, we cannot find many bright quasar systems at a much higher redshift than z = 8.

We select a quasar system based on halo mass, taking the 10 most massive halos, numbered from 0 to 9. We note that in general larger halos host brighter quasars unless (i) the quasar system is turned off by feedback, or (ii) the quasar system has not yet grown its black hole mass significantly. 8 unique hosting sub-volumes (50 Mpc/h per side) are identified, which we refer as sub-volume 0 to 7 throughout the rest of

#H	#V	$M_{\rm HALO}$	$M_{\rm BH}$	Total		Center	
				QSO	STR	QSO	STR
0	0	148.35	74.6	8.35	3.38	7.24	0.94
1	1	77.9	49.6	4.97	3.45	4.43	0.79
2	2	75.5	34.7	2.06	3.93	1.39	0.45
3	0	84.9	7.3	8.35	3.71	0.49	0.22
4	3	60.1	119.0	8.18	2.72	7.76	0.54
5	0	67.1	16.0	8.35	3.71	0.56	0.21
6	4	70.8	5.5	0.50	2.72	0.06	0.16
7	5	65.6	13.2	0.95	1.98	0.69	0.30
8	6	66.6	6.4	1.29	3.82	0.20	0.21
9	7	64.9	7.2	0.38	2.22	0.15	0.15

Table 4.1: UV Flux of Sub-volumes. The first column #H (0 - 9) identifies the ten most massive halos. The second column #V (0 - 7) identifies the eight unique subvolumes that contain the halos. The halo mass ( $M_{\text{HALO}}$ ) is in units of  $10^{10} \,\mathrm{M_{\odot}h^{-1}}$ . The blackhole mass ( $M_{\text{BH}}$ ) is in units of  $10^6 \,\mathrm{M_{\odot}h^{-1}}$ . The flux is in units of  $10^{55} \,\mathrm{sec^{-1}}$ . Halos 0, 3, and 5 are clustered and belong to the same sub-volume 0. See Appendix 4.6 for the definition of the spectra.

the chapter: Three of the halos (halo number 0, 3 and 5 in Table 4.1) are located at spatial locations within sub-volume 0.

Next, we compute the ionizing photon flux due to the quasars and star formation occurring in each of the sub-volumes. The details of the assignment are described in Appendix 4.6. In addition to the full volume flux, we also define a Central Source Flux, which is the flux of sources within the 1 Mpc/h radius of the center of the sub-volume. The sources in the center are associated with the ionized bubble of the central halo. The ionizing photon flux of the sub-volumes is summarized in Table 4.1 and Figure 4.1.

There is an expected correlation between the central quasar and central stellar flux. Based on the ratio of the central flux to total ionizing flux within the subvolume, as well as the composition (quasar or stellar) of the central flux, we divide the sub-volumes into three groups:

- Quasar (Q) : The central quasar dominates the entire sub-volume.
- Sub-dominant Quasar (QS) : The central quasar dominates the central flux, but is not a significant fraction of the total flux of sources in the sub-volume.



Figure 4.1: Central and total ionizing photon flux. The left panel compares the total UV photon flux of the stars and the quasars within the sub-volumes. The right panel compares the UV photon flux from the sources within 1 Mpc/h of the center. The lines represent the linear regression of the data. In the bottom panel the central photon flux of the quasar sources and the central photon flux of the stellar sources as a fraction of the total flux are shown. It is interesting to observe that the central stellar photon flux is always a small fraction of the total flux yet the central quasar can contribute a significant fraction of the total. See Appendix 4.6 for the definition of the spectra.

• Stellar (S): The quasar flux in the central sources is smaller than the sum of stellar flux, and the entire central flux is small compared to the ionizing flux of the entire box. We note there can be two possible distributions of sources within the box: one being the presence of another major source that is not near the center of the sub-volume; the other being ionizing sources that are rather more uniformly distributed throughout the sub-volume. In our sub-volumes, the latter is more often to the case.

## 4.3 Analytic Growth of Strongren Spheres

The growth of Strongren spheres in a clumpy cosmological environment can be analytically modeled using [Cen and Haiman, 2000],

$$\frac{dR_t^3}{dt} = 3H(z)R_t^3 + \frac{3N_{\rm ph}}{4\pi \langle n_{\rm H} \rangle} - C_{\rm H} \langle n_{\rm H} \rangle \,\alpha_{\rm B}R_t^3, \tag{4.5}$$

where the first term represents the Hubble expansion, and can be neglected in this context. The solution is straightforward [eg, Shapiro and Giroux, 1987]

$$R_{t} = R_{s} \left( 1 - \exp\left(-\frac{t}{t_{s}}\right) \right)^{-1/3} \approx R_{s} \left(\frac{t}{t_{s}}\right)^{1/3}$$

$$t_{s} = (C_{\rm H} \langle n_{\rm H} \rangle \alpha_{\rm B})^{-1}$$

$$R_{s} = \left(\frac{3\dot{N}_{\rm ph}}{4\pi C_{\rm H} \langle n_{\rm H} \rangle^{2} \alpha_{\rm B}}\right)^{1/3}.$$
(4.6)

We will use this analytic solution to compare to our numerical results. The IGM clumping factor  $C_{\rm H}$  is calculated using a method appropriate for an SPH simulation, described by Pawlik et al. [2009],

$$C_{\rm H} = \frac{\sum \rho_i^2 h_i^3}{\langle \rho \rangle^2 \sum h_i^3},\tag{4.7}$$

where  $h_i$  is the SPH smoothing length, and  $\rho_i$  is the IGM density. The clumping factor in the entire sub-volumes has a value of  $C_H \sim 4$ , giving a recombination time about  $t_S \sim 100$  Myear  $\gg t_Q$ . However we note the clumping factor can go up to  $\sim 20$ within 1 Mpc/h radius from the center of the sub-volume, reducing  $t_S$  to  $\sim 20$  Myear.

For the source flux, we make the simplest approximation, summing the flux of all sources with in the sub-volume to obtain one effective  $\dot{N}_{\rm ph}$ . We also assume that the volume consists of pure hydrogen (i.e. no helium,  $X_{\rm H} = 1$ ) in the calculation of the analytic model. The Case B recombination rate  $\alpha_{\rm B}$  is taken from Hui and Gnedin [1997], in consistency with the recombination rate table used in the radiative transfer simulation.

## 4.4 Radiative Transfer Simulation

#### 4.4.1 Ray-tracing Scheme

For this study we have rewritten the SPHRAY Monte Carlo radiative transfer code [Altay et al., 2008, Croft and Altay, 2008] in C with OpenMP(TM), so that it runs in parallel on shared memory systems (P-SPHRAY). We have also eliminated the on-the-spot approximation of recombination, so that instead recombination rays are emitted when the recombination photon deposit at a spot exceeds a threshold, as done in the code CRASH [Maselli et al., 2003]. The parallel version allows us to trace many rays at the same time on several CPUs within a single time step, achieving a near-linear speed up with a small number ( $10 \sim 30$ ) of CPUs. Monochromatic rays are sampled from frequency space according to the source spectrum. Hydrogen and Helium ionization and recombination are both traced. The tabulated atomic reaction rates of Hui and Gnedin [1997] taken from the serial version SPHRAY are used.

As a new addition to the code we model the secondary ionization of high energy photons using a fit to the Monte Carlo simulation results of Furlanetto and Stoever [2010], Shull and van Steenberg [1985]. These fast non-thermal-equilibrium electrons produced from high energy ionization photons (e.g., X-ray photons) may collide and secondarily ionize more neutral atoms, increasing the ionizing efficiency of the quasar sources by allowing one high energy photon to ionize tens of neutral atoms. On the other hand, our current post-processing approach is incapable of modeling the heating from the residual energy of the ionizing photons, for the thermal evolution (hydro) is decoupled from radiative transfer. We note that the local heating of ionizing photons can be thought of as part of the local thermal coupling between the liberated luminosity and the surrounding environment described in Section 2. The long range heating of the harder photons is not directly modeled.

#### 4.4.2 Parametrization of the Ionized Bubble

For this study, we are more interested in the ionized bubbles located at the centers of sub-volumes, because they are associated with the main halos in the sub-volumes. We will call these central ionized bubbles the central ionized regions (CIR).

When quantifying the size of a CIR we measure the spherically averaged fraction of a species, x using (we show the example of HII regions, similar definitions hold for other species):

$$x_{\rm HI}(r) = \frac{\int \delta(r' - r) x_{\rm HI} \rho dV}{\int \delta(r' - r) \rho dV} \doteq \frac{\sum_{\rm shell} x_{\rm HI,i} m_i}{\sum_{\rm shell} m_i}.$$
(4.8)

Here the sums are carried out by binning SPH particles (index i) in shells. Particles are assigned as a whole to shells by their center position rather than being integrated over the SPH kernel with in the shell. We define the averaged CIR radius at a given

threshold  $x^*$  to be the first crossing of x(r) with  $x^*$ , or

$$\hat{R}_S(x) = \inf\{r | x_{\rm HI}(r) = x^*\}.$$
(4.9)

The first crossing corresponds to the edge of the CIR, and later crossings indicate the edges of the nearby ionized regions. The motivation for this definition comes from its similarity to that used to quantify ionized regions seen in quasar absorption line spectra, the Ly $\alpha$  absorption Near-Zone radius  $R_{NZ}$  by Fan et al. [2006a]. In practice, we take as  $\hat{R}_S$  the average of the two nearest bins  $r_L$  and  $r_R$  which satisfy  $[x_{\rm HI}(r_L) - x][x_{\rm HI}(r_R) - x] < 0.$ 

The spherically averaged radius  $\hat{R}_S$  is a simple quantity to use in comparisons but the angular variations in the properties of the CIR are completely lost. To quantify the angular dependency, we bin the SPH particles into angular cones and calculate the averaged fraction within the cones,

$$x_{\rm HI}(r,\theta,\phi) \doteq \frac{\sum_{\rm cone, shell} x_{\rm HI,i} m_i}{\sum_{\rm cone, shell} m_i}.$$
(4.10)

The angular dependent CIR radius is

$$R_s(x,\theta,\phi) = \inf\{r | x_{\rm HI}(r,\theta,\phi) = x^*\}.$$
(4.11)

Note that in general  $\langle R_s(x,\theta,\phi) \rangle \neq \hat{R}_s$ , because averaging and binning do not commute. For measuring  $R_s(x,\theta,\phi)$ , we use  $12 \times 16^2$  cones. We note that increasing to  $12 \times 32^2$  cones does not significantly alter the results.

We measure the anisotropy of the ionized bubble from the variance of  $R_s$ 

$$A_s(x) = ST_k[R_s(x,\theta_k,\phi_k)], \qquad (4.12)$$

where  $ST_k$  stands for the standard derivation. We run the simulation with sufficient number of rays so that the typical shot-noise contribution towards  $A_s$  is negligible, as shown in the next section.

Three different levels of species fraction x are used to define three different levels of CIR fronts in this study:

- Inner front, x = 0.1 for the neutral fraction, or x = 0.9 for the ionized fraction;
- Middle front, x = 0.5 for the neutral fraction;
- Outer front, x = 0.9 for the neutral fraction, or x = 0.1 for the ionized fraction.

The Inner front corresponds to the near-ionized edge of the Stromgren sphere, and the Outer front corresponds to the near-neutral edge of the Stromgren sphere. We however note that these choices are for illustrative purposes and they do not have any direct correspondence with threshold values used to detect 21 cm or  $Ly\alpha$  observation signatures.

#### 4.4.3 Shot-noise and Convergence

One issue which must be addressed in Monte Carlo ray tracing (RT) schemes is the presence of shot-noise and its effect on convergence of results, something particularly important when sampling is also in frequency space [eg, Iliev et al., 2006, comments on CRASH]. Shot-noise artificially increases the angular anisotropy  $A_s$  measure, and so is of direct concern for this study.

We define a shot-noise parameter  $\gamma$  to be the ratio between the number of photons in a ray packet,  $n_r^0$  and the number of atoms in an SPH particle  $n_p^0$ ,

$$\gamma = \frac{n_r^0}{n_p^0}.\tag{4.13}$$

A small  $\gamma$  guarantees that the ionization front cannot advance by more than one particle in one time step. When  $\gamma \ll 1$ , the ionization front advances slowly and the shot-noise is controlled. One of course still needs to ensure the rays have a sufficient angular resolution to resolve the angular scale used in the calculation of the anisotropy.

Runs	HII Inner	HII Middle	HII Outer	
$\gamma = 0.003$ and $0.03$	0.96	0.98	0.86	
$\gamma = 0.03$ and $0.3$	0.74	0.94	0.68	

Table 4.2: Correlation coefficients in convergence test for number of rays. See Equation 4.13 for definition of  $\gamma$ . Shown in the table are the correlation coefficient of the center bubble radius  $R_s(\theta, \phi)$  between different runs.

In MassiveBlack, a typical SPH gas particle has a mass of  $5 \times 10^7 \,\mathrm{M_{\odot}/h}$ , equivalent to  $n_p^0 = 9 \times 10^{64}$ . For ray tracing with  $10^5$  steps and 128 rays per time step,  $(1.28 \times 10^7 \,\mathrm{rays})$ , and with the total luminosity listed in Table 4.1, an average packet contains  $n_r^0 = 3 \times 10^{63}$  photons. The shot-noise parameter is therefore  $\gamma = 0.03 \ll 1$  for our typical runs.

We can also empirically confirm that convergence is reached by increasing the number of rays used in the simulation and showing that the quantity of interest is insensitive to further increases. We perform such a convergence test with 3 runs on sub-volume 4 with (i)  $1.3 \times 10^8$  rays,  $\gamma = 0.003$ , (ii)  $1.3 \times 10^7$  rays,  $\gamma = 0.03$ , and (iii)  $1.3 \times 10^6$  rays,  $\gamma = 0.3$ . We quantify the similarity of the ionization front using the correlation coefficient of  $R_s(\theta, \phi)$  between runs, shown in Table 4.2. The higher correlation between the runs with more rays indicates that the simulation has effectively converged. Therefore, for the runs we use  $1.3 \times 10^7$  rays.

### 4.5 **Results and Discussion**

#### 4.5.1 Uniform Density Field

We first test P-SPHRAY with a source in a uniform density field at z = 8, with H number density  $n_{\rm H} = 2 \times 10^{-4} \,{\rm cm}^{-3}$ , and uniform temperature  $T = 10^4 \,{\rm K}$ . The hydrogen mass fraction is  $X_{\rm H} = 0.76$ , the cubical box has a side length of 50 Mpc/h, and we evolve the radiation and ionization fractions for a duration of  $2 \times 10^7 \,{\rm yrs}$ . We carry out 3 separate simulations, considering different central source properties for each one as follows: (i)  $1.2 \times 10^{56} \,{\rm sec}^{-1} \,{\rm UV}$ ; (ii)  $1.2 \times 10^{56} \,{\rm sec}^{-1} \,{\rm UV}$ ,  $0.3 \times 10^{55} \,{\rm sec}^{-1}$  soft X-Ray,  $0.5 \times 10^{55} \,{\rm sec}^{-1}$  hard X-Ray; (iii) same as (ii), with secondary ionization. The luminosity of the source is motivated by the luminosity of the center sources in sub-volume 0. The growth of the CIR fronts in the simulations are shown as a function of time in Figure 4.2. We show as separate panels the behaviors of the three parts of the ionization front: inner, middle and outer (as defined in Section 4.4.2).

There are three highlights from the uniform density field simulations: (i) The fronts in general agree well with the analytic model described in 4.6. (ii) There is a smooth transition from neutral to ionized state at the front, indicated by the 1 Mpc/h difference between the inner front and the outer front. The transition is mostly due to the penetration of harder UV photons. (iii) The effect of X-ray photons and secondary ionization is more evident on growth of the outer front than of the inner and middle fronts. This is because the secondary ionization affects the way the harder photons interact with the matter the most, and harder photons travel further into the neutral region. We note that by adding in the effect of secondary ionization increases the outer radius of ionized bubbles by approximately 10%.

#### 4.5.2 MassiveBlack: Visualization

We show the CIR bubble of the most extreme Q type and S type sub-volumes (subvolume number 3 and 4 in Table 4.1 respectively) in Figure 4.3. The halo mass of both are similar,  $60 \times 10^{10} \,\mathrm{M_{\odot}h^{-1}}$  and so are their total star formation rates. However sub-volume 3 has a large ionized HII region associated with the bright active quasar in the center of the sub-volume, whilst in sub-volume 4 the HII region formed merely from stellar sources are small. Note that, even though in a comparable mass halo, the central black hole is much smaller in sub-volume 3 and hence its activity has a negligible impact. This example aims to illustrate how, in the presence of an active early quasar, the ionized bubble is far more extensive than one driven by star formation alone.

The time evolution of the HII bubble in sub-volume 0, where three halos coexist, is shown graphically with a slice through the center of the simulation volume every  $2 \times 10^{6}$  yr in Figure 4.4. It is interesting to observe that a neighboring bubble on the right merges with the center bubble and this increases the size of the outer front in



Figure 4.2: CIR Growth in a Uniform Density Field. The symbols are: (i) cross[+]: UV only; (ii) cross[x]: UV and X-ray (iii) circle: UV and X-ray, with Secondary ionization. The analytic model (dashed lines) neglects Helium and assumes  $X_H = 1.0$ .



Figure 4.3: Q (quasar, left) type and S (stellar, right) type ionized bubbles in MassiveBlack. Both panels are of comoving length 15 Mpc/h per side. A camera is put at about 60 Mpc/h from the center of the sub-volume and a perspective transformation is applied to form a projection onto the image plane of the camera. Red color corresponds to fully neutral IGM and blue color corresponds to fully ionized IGM. Yellow is in between the two states. Crosses are the sources; both stellar and quasar sources in the entire sub-volume are shown. It is interesting to notice the lack of a major source (compared to the bright quasar in sub-volume 3) in sub-volume 4.



Figure 4.4: Evolution of the ionization front. The evolution of the ionization front in sub-volume #0 is shown. The ionization front  $(x_{\rm HI} \in [0.1, 0.9])$  is marked in black, and the fully ionized near zone  $(x_{\rm HI} < 0.1)$  is marked with gray.

that direction, contributing to the anisotropy of the front. We investigate this issue in the next section in a more quantitative way.

### 4.5.3 MassiveBlack: Spherically Averaged Radius of Ionized Regions

We now examine the results of the radiative transfer post-processing of the eight sub-volumes from MassiveBlack. The averaged CIR radius  $\bar{R}_s$  for HII and HeIII, as defined in Section 4.4.2, are shown in Figure 4.5 as functions of the central source flux. We fit the growth against the scaling relation described by Equation 4.6, and find that at the low luminosity end (mainly S type and QS type sub-volumes) there is a significant deviation from the fit for the inner and middle fronts. The outer front is more extended than the simple uniform density field simulation, albeit given the similar source spectra, hinting that the structure in the IGM is also contributing to the smoothing of the front. This smoothing is a phenomenon similar to that described by Wyithe and Loeb [2007]. We also attribute the effect to the clustering of sources; however unlike the smoothing due to secondary ionization, the clustering contribution is anisotropic, and we discuss it in Section 4.5.4.

In Figure 4.5, we can see that the HII and HeIII CIR radii appear to grow together. The correlation between the two can be seen directly by plotting one against the other, which we do in Figure 4.6, where we find there is a strong correlation between the HII radius and HeIII radius. The HeIII radius is smaller than the HII radius, agreeing with the finding of Friedrich et al. [2012] who used the ray tracing code C2RAY. We note however that the treatment of secondary ionization and Helium ionization in P-SPHRAY is similar to that of C2RAY, and an agreement is not surprising.

The time evolution history of the HII CIR radius for all sub-volumes is shown in Figure 4.7. We compare the evolution of the three parts of the fronts with the prediction of the analytic model (in equation 4.6), in which we have used the clumping factor and mean H density measured within the final CIR in the simulation to estimate a fiducial recombination time  $t_s$ .

The growth of the CIR in the Q type sub-volumes flattens off much earlier than



Figure 4.5: HII and HeIII CIR Front Radius. The left panel is for the HII CIRs, the right panel for the HeIII CIRs. The horizontal axis is the central source flux. The dotted line in the first two panels shows the scaling assuming the analytic model in Equation 4.6.



Figure 4.6: Correlation between HeIII CIR and HII CIR. HeIII CIRs is plotted against HII CIRs, and the linear regression for three fronts are shown with the dotted line.



Figure 4.7: Evolution of three quasar driven sub-volumes. The top panels show the CIR radius as function of time; the thick black line is the fiducial analytic model as described in the text (Equation 4.6). The vertical short lines show the recombination time in an analytic model as if it is fitted as a parameter to the simulation data points, which are marked with squares (Inner), triangles (Middle), and circles (Outer). The bottom panels show the ratio between the radius and the fiducial model.

one might expect from a free streaming law. In order to ascertain the relevant physical timescales, we fit the time evolution of the fronts to the full analytic model in 4.6, and extract the effective recombination time  $\hat{t}_S$  as a fit parameter, then mark the time in the plot. We can see that for the quasar driven (type Q) sub-volumes, all three fronts (inner, middle and outer) growth stop at about 10 Myear, on the same order of the life time of the quasar  $t_Q \sim 20$  Myear and the fiducial recombination time  $t_S \sim 20$  Myear. The deviation from free streaming indicates that by merely increasing the life span  $t_Q$  of the central quasar, we can not substantially increase the size of their CIR.

In S type type sub-volumes, there is a similar tailing off for the inner fronts. However, the outer fronts continue their free streaming growth, behaving differently from the analytic model, which has stopped due to reaching the recombination time. We attribute this apparent excessive growth of the averaged S type outer front to the anisotropic growth via merging with other small ionized bubbles that are close to the CIR, as described later in the chapter.

We note that our Monte-Carlo RT approach allows for superluminal growth of the CIR. This nonphysical situation happens well before the first snapshot time which is 2 million years after the sources are turned on. The first snapshot, in which the Strongren sphere typically has grown to about 10% of the full size, gives an upper bound to the contribution from superluminal growth. We conclude that the superluminal growth occurring at early time (< 2 Myear) and small radii (< 10%)

does not play an important role in the final shape of the ionized bubble. We refer the readers to Shapiro et al. [2006] for a more detailed discussion of the role of the nonphysical superluminal phase of the growth of an ionization front.

#### 4.5.4 MassiveBlack: Anisotropy of Ionized Regions

In Figure 4.8 we show the anisotropy measured using Equation 4.12 of all 8 subvolumes. The anisotropy of different ionization fronts as a function of radius are marked with different symbols. As a reminder, in Equation 4.12, the anisotropy  $A_s$ is the standard deviation of the ionization front radius, so that by comparing to the radius plotted on the x-axis it can be seen that the fronts are not very far from spherical symmetry (~ 10% variations). We find that larger ionized regions (corresponding to brighter quasar sources) are in general associated with more anisotropy.

In the simulation, the anisotropy of the inner and middle radii of the HII regions do not significantly depend on the radius (the curve is relatively flat), but the outer fronts have more anisotropy. The HeIII regions show a similar feature, except for the inner fronts of three type Q sub-volumes, which have significantly stronger anisotropy.

These phenomena lead us to the following explanation, incorporating two contributing mechanisms:

- 1. The anisotropic distribution of gas that attenuates the ionizing photons; when the density is high in a particular direction, the extra absorption decreases the radius of the ionized region in that direction.
- 2. The merging of nearby bubbles from clustered halos; when the density is sufficient to host bright sources lying in a certain direction, their extra photoionization increases the bubble radius in that direction. The outer front is more sensitive to merging than the inner and middle front.

For a small CIR (or a CIR in its early growing stage), no merging has occurred and only the density induced anisotropy is present. As the CIR grows, it overlaps with nearby ionized regions, resulting in the second type of anisotropy (due to overlapping). We note that this does not contradict the finding that clustering contributes to the smoothness of the extended ionization front [Wyithe and Loeb, 2007], because by definition, after spherical averaging, the existence of surrounding bubbles will result a smoother averaged front.

In order to better visualize the structures which cause the anisotropy, we plot the distance to the different parts of the ionization fronts in a Molleweide projection as seen from the point of view of the central source. These maps of the angulardependent HII bubble radius,  $R_s$  are shown in Figure 4.9, where we plot the quasar driven sub-volume 0 and stellar driven sub-volume 4.

Looking from the top panels downwards for sub-volume 0, we can first see clumps of mostly neutral gas close to the quasar which restrict the distance to the  $x_{\rm HI} = 0.1$ 



Figure 4.8: Anisotropy .vs. Radius. The left panel shows the anisotropy of three (inner, middle, outer) HII CIR fronts defined as in Equation 4.12. The right panel shows the anisotropy of the HeIII fronts. The dotted lines are fits against a square-root + linear offset model and are only meant to guide the eye. All 8 sub-volumes are displayed.

fronts to be very close by (< 1 Mpc/h), compared to the mean distance for this neutral fraction of 6 Mpc/h. In the  $x_{\rm HI} = 0.9$  plot, for sub-volume 0 a prominent red region can be seen at the top. This corresponds to a bubble which has merged with the central bubble. Next to each panel in Figure 4.9 we show a histogram of the  $R_s$  values. The secondary bubble is responsible for the small tail of high values region in  $x_{\rm HI} = 0.9$  histogram, as well as giving an extra contribution to  $A_s$ .

Moving on to the stellar driven sub-volume 4 (right hand panels in Figure 4.9), we can see that the ionized region radius is smaller by approximately a factor of 3. Because this radius is small compared to the distance to nearby major halos, there is no sign of merging with other large bubbles, and the ionized region remains more isotropic. In the bottom panels of Figure 4.9 we show the Molleweide-projected mass density within the inscribed sphere of the sub-volume. It is interesting to compare this clumpiness with the structure in the ionized bubble radius. We can see some correlation with some structures, but not as much as might be expected, indicating that the interaction of radiation with the clumpy medium surrounding the sources is a complex process.

In Figure 4.10, we show an orthogonal view to Figure 4.9, showing structure along rays moving out from the central source in some different directions. The outer HII front radii in three chosen directions are shown as a function of time. This allows us to see in detail how the clustering of halos affects the center bubble radius through merging as the simulation develops. The three directions are (i) one directed towards the second brightest source in the sub-volume, (ii) one directed towards the merging bubble seen on the right hand side of Figure 4.4, and finally (iii) one oriented in an arbitrarily choosed direction not crossing any nearby secondary bubbles. In order to show the stronger bubble merging effect which would happen with more luminous sources, on the right hand side of Figure 4.10, we plot the same 3 sight-lines, but after increasing the stellar luminosity by a factor of 10.

Merging happens as the central bubble touches the surrounding ones and meanwhile the radius along that direction significantly increases, compared to the radius without a nearby halo. This can be seen most clearly for sight-line (iii) on the right in Figure 4.10. The development of the anisotropy due to merging of nearby ionized bubbles is responsible for the apparent rapid growth of the averaged outer Stromgren radius after the growth of the inner radius is stabilized for S type and QS type sub-volumes. As the neutral fraction at the overlapping edge slowly drops below the threshold, two bubbles merge and the averaged radius increases. However, at the luminosity seen in the simulation (left hand panels of Figure 4.10), such a merging mechanism is mostly limited to the growth of the outer front. It is also worth noting that in the absence of the effect of smoothing due to secondary ionization the steep bubble edges will make the merging even more unlikely.



Figure 4.9: Angular Distribution of the HII Radius. The left column is sub-volume 0 (Q type) and the right column is sub-volume 4 (S type). From top to the bottom the inner  $(x_{\rm HI=0.1})$ , middle  $(x_{\rm HI=0.5})$ , and outer  $(x_{\rm HI=0.9})$  fronts are shown. The histograms measure the variance of  $R_s$  in different angles, characterizing the anisotropy. The red region in the  $x_{\rm HI} = 0.9$  plot of sub-volume 0 is due to merging with a further away HII bubble. The less significant red in the center corresponds to the merging shown in Figure 4.4.



Figure 4.10: Time evolution of  $x_{\rm HI}$ . Time evolution of  $x_{\rm HI}$  in sub-volume 0, along three selected sight-lines: (i) the blue line is directed towards the second brightest halo in the sub-volume; (ii) the green line is directed towards the merging bubble on the right shown in the slice plots; (iii) the red line is directed towards an arbitrary direction with no nearby bubbles. The left panel is the luminosity model used throughout the chapter, while on the right the stellar luminosity is boosted by a factor of 10 to demonstrate a stronger merging effect.

Band	Range
UV	$13.6\mathrm{eV}$ to $250\mathrm{eV}$
Soft X-Ray	$250\mathrm{eV}$ to $2\mathrm{KeV}$
Hard X-Ray	$2\mathrm{eV}$ to $10\mathrm{KeV}$

Table 4.3: Band definitions.

#### 4.6 Source Luminosity Models

#### 4.6.1 Band luminosity of Quasars

We calculate the bolometric luminosity  $L_{\rm bol}$  of each quasar from the accretion rate M of the super-massive black hole with Equation 4.3. We choose to use as  $\dot{M}$  a single value throughout the calculation, which is averaged accretion rate measured from the simulation over a timescale  $t_Q = 2 \times 10^7$  yrs, where  $t_Q$  is also the length of time over which we follow the evolution of the quasar ionization front. We estimate the flux in different bands from the bolometric luminosity according to the fitting formula of Hopkins et al. [2007]. The band definitions are listed in Table 4.3. The soft and hard X-Ray band luminosity are directly calculated with the program distributed by Hopkins et al. [2007]. We obtain the UV band luminosity from the reported B band luminosity following the broken power law described in the chapter ( $\alpha_{\rm UV} = 1.76$  and  $\alpha_{\rm B} = 0.44$ ).

$$L_{\rm UV} = \frac{L_B}{\nu_B} \left(\frac{\nu_X}{\nu_B}\right)^{-\alpha_B} \left(\frac{\nu_I}{\nu_X}\right)^{-\alpha_{\rm UV}} \frac{\left(\frac{250}{13.6}\right)^{1-\alpha_{\rm UV}} - 1}{1 - \alpha_{\rm UV}},\tag{4.14}$$

where  $\nu_X = c/120 \text{ nm}$  is the pivot from optical to UV,  $\nu_B = c/445 \text{ nm}$  and  $\nu_I = c/91.1 \text{ nm}$ . We calculate the photon flux of the quasar sources according to these power law spectra applied to the various bands.

#### 4.6.2 Flux of UV stellar photons

The stellar photon flux of a halo is given by

$$\dot{N}_{\rm SFR} = f_{\rm esc} N_{\gamma/\rm H} X_{\rm H} \Phi/m_p, \qquad (4.15)$$

where  $\Phi$  is the star formation rate of the halo,  $X_{\rm H} = 0.76$ , and  $m_p$  is the proton mass. We take the escape fraction to be  $f_{\rm esc} = 0.1$ , and the photon to hydrogen ratio to be  $N_{\gamma/\rm H} = 4000$ , a value found to produce a reionization history consistent with observations [Furlanetto et al., 2004, Sokasian et al., 2003]. For simplicity, the stellar sources are assigned to the center of the halo; in future work we plan to investigate the effect of the positioning of the stellar sources by either following the star forming gas particles in the simulation directly or the gravitationally bound sub-halos or galaxies. To ease the modeling, we only include and assume the same UV band spectrum for the stellar radiation to that of the quasar spectrum.

## 4.7 Conclusions

We have presented results from radiative transfer simulations in the vicinity of high redshift quasars in the MassiveBlack simulation. We find that the rare brightest quasars drive a much more significant growth of ionized regions than in the purely stellar driven case. The ionized regions associated with active quasars are characterized by (i) a smooth ionized fraction transition from the middle to the outer front, and (ii) an increased anisotropy in the front when it starts to overlap the nearby ionized regions. The nature of such growth is significantly more complex than a simple analytic growth of a single center bubble with clumping correction.

The largest HII bubble obtained in this simulation has a comoving radius of 10 Mpc/h, which is smaller than the general expectation that can fulfill the reionization of the universe [Trac and Gnedin, 2011, Wyithe and Loeb, 2004, Majumdar et al., 2011]. The quasar near zones we have presented in this chapter are the primordial ancestors of the later much larger Stromgren spheres which will form near the end of the EoR. They are however relatively isolated regions that could be interesting objects for study in future 21cm surveys. After the z = 8 epoch we have modeled in this chapter, we expect that the global star formation in the MassiveBlack simulation increases significantly. This will eventually lead to the global reionization of the universe. We plan to study this process and role of quasars in future work.

## Chapter 5

# Quasar Correlated Lyman-Alpha Forest

## 5.1 Introduction

The Lyman- $\alpha$  forest is a unique probe of the intergalactic medium. [Rauch, 1998, Gnedin and Hui, 1998, McDonald et al., 2000, Petitjean et al., 1993, Viel et al., 2004, Croft et al., 1998, Rauch et al., 1997, Bi and Davidsen, 1997, Davé et al., 1999]. The physics of Lyman- $\alpha$  forest is relatively clean. [see, e.g. Meiksin, 2009]. It can be used to probe the initial density perturbations across a large range of scales (from Kpc to Gpc) at a range of redshifts range between z = 2 and z = 4. The Sloan Digital Sky Survey (SDSS) and the upcoming Dark Energy Spectroscopy Instrument (DESI) experiment have measured and will measure the spectra of thousands and then millions of quasars. [Pâris et al., 2012, Levi et al., 2013] The abundance of data in Lyman- $\alpha$  forest allows high precision measurements of the Baryon Acoustic Oscillation (BAO) from the auto-correlation of the forest. [Slosar et al., 2011, Seljak et al., 2006, Croft et al., 1998] Recently interest of Lyman- $\alpha$  forest has been in the cross correlation between quasars and the forest. The correlation is a complementary to the auto-correlation because of different composition of the systematics. [Font-Ribera et al., 2014]

A large number mock catalogs are needed to understand systematic effects in the extraction of cosmological parameters from the data. Producing mocks cheaply poses a worthwhile computational challenge. Font-Ribera et al. [2012] has developed a fast method to generate mocks of the Lyman- $\alpha$  forest with identical statistical properties to the observations. However, as the authors noted, the method requires quasar positions (the starting point of each forest) to be given a priori, and does not allow them to be picked from the density field which also generates the forest. This method is thus incapable of modelling the cross-correlation between quasars and the forest. In order to model the quasar and forest in a coherent way, a method with more complete physical treatment of the density field is required.

The most straight-forward way is to construct a consistent density field from which both the quasars and the forest are sampled. Direct construction of a density field using a fluctuation power spectrum covering the full dynamical range needed for Lyman- $\alpha$  forest surveys can be very costly. The brute force approach (generating a single uniform grid with the large volume and small cell size required) would be impossible, even with the currently largest supercomputers. For example, Le Goff et al. [2011] used this method to produce mocks of uncorrelated quasar and Lyman- $\alpha$  forest with an anisotropic density mesh that is thin along the line of sight direction. The method can not account for the spatial distribution of quasars, thus needs extension for correlated quasar and forest mocks.

In this chapter we develop a affordable algorithm that generates correlated quasar and forest mocks which are equivalent for our purposes to those that would be made with the brute force approach, but at enormously lower computational cost. The method can be run on home computer hardware, or equivalently single computing nodes of super-computers. For any  $\Lambda$ CDM cosmology and any survey profile (quasar number density per redshift and sky mask), it produces a catalog of quasars (redshift, RA and DEC), and the optical depth  $\tau$ , transmission fraction F of Lyman- $\alpha$  forest pixels in uniform log- $\lambda$  pixels. Redshift distortions are also included.

#### 5.2 The Multi-level Fluctuation Approximation

We begin with the linear theory matter power spectrum at z = 0,  $P_0(k)$ . The power spectrum is produced from CAMB [Lewis et al., 2000].

The linear theory matter density field  $\delta_0(\mathbf{x})$  is the Fourier transform of the correlated Gaussian field generated by the power spectrum  $P_0(k)$ ,

$$\delta_0 = \int \frac{d^3k}{(2\pi)^3} P_0(k)^{\frac{1}{2}} g(\mathbf{k}),$$

where  $g(\mathbf{k})$  is a Hermitian random Gaussian field.

Numerically, this procedure is implemented as a Discrete Fourier Transform on a regular mesh, just like the generation of initial conditions for a cosmological simulation.

The challenge is the size of the box and the required resolution to resolve the forest. The comoving distance to z = 4 is  $d_c(z = 4) \approx 5000 \,\mathrm{h^{-1}Mpc}$  (to make the situation worse, to modelling the correct geometry which involves both halves of the sky, a 10  $\mathrm{h^{-1}Gpc}$  box is needed), while the Fluctuating Gunn Peterson approximation calls for a forest resolution on the order of  $R_{\rm LN} \sim 100 \,\mathrm{h^{-1}Kpc}$ . A naive approach calls for a cubical mesh of 100,000 points per side, beyond the capability of contemporary supercomputers.

However, given that we are most interested in the BAO scale at  $100 \,h^{-1}$ Mpc, the size of the problem can be significantly reduced with two approximations.

The first approximation is to recognize that correlations on small scale are important to produce the correct bias (ratio of forest to matter fluctuations) of the forest, but unimportant in the shape of the BAO signal. In fact, in the analysis of BAO from observational data the Lyman- $\alpha$  pixels are typically combined into  $4 \,\mathrm{h^{-1}Mpc}$ sized *bins* in any case. Hence, we can set a split scale  $R_{\alpha}$ , below which we only model the variance of the density field:

$$V_{\alpha} = \int^{2\pi R_{\alpha}} \frac{d^3k}{(2\pi)^3} P_0(k).$$

Note that in practice we measure  $V_{\alpha}$  in real space with several boxes of size  $R_{\alpha}$ and resolving to  $R_{\rm LN}$ , to account for the sampling resolution effect. The small scale fluctuations  $\delta_{\alpha}(\mathbf{x})$  are sampled from a Gaussian distribution with variance  $V_{\alpha}$ , to ensure the correct small scale power in the FGPA. A factor of  $\sim 10^3$  reduction in the total size of the mesh is achieved with the first approximation, and the problem is now within the reach of supercomputers.

The second approximation is that correlations on large scales only provide a fluctuating background on some scale  $R_B$  much larger than  $100 \,\mathrm{h^{-1}Mpc}$ . We divide the full volume according to  $R_B$ , and independently generate the density fluctuations using the power spectrum in each sub-volume  $\iota(\mathbf{x})$ . On top of the independent sub-volume fluctuation, we add the correlated large scale fluctuation from the full volume, but smoothed at very low resolution  $\eta(\mathbf{x})$ , such that

$$\delta_1(\mathbf{x}) = \eta(\mathbf{x}) + \iota(\mathbf{x}).$$

The sub-volumes are sufficiently small that each of the Fourier transforms fits into the typical 2 gigabyte memory available per processing unit on contemporary consumer computers and as well as single nodes of supercomputers. With these independent sub-volumes even a laptop can drive these simulations – however there is a trade off between total computation time and computation power: running less processing power takes more processing time.

We note that the overlapping Gaussian modes in k space for  $\iota$  and  $\eta$  fields must be consistent (the low k modes of  $\iota$  and high k modes of  $\eta$ . We fill the k space Gaussian variables by moving along an inside-out space-filling curve.

The full matter density field is given by the sum of the three pieces

$$\delta_0 = \delta_\alpha + \iota + \eta.$$

The velocity field is obtained by differentiating the matter density field in Fourier space, following the Zel'Dovich approximation.

### 5.3 From Density Field to Quasar Catalogue

The bias of quasars in SDSS/BOSS has been measured to be close to  $b_Q = 3.0$  by various authors. [White et al., 2012, Shen et al., 2013, see, e.g.] Given a matter density

field, the problem of building a quasar catalog is translated into sampling biased point-like objects from a matter density field. We use a method motivated by the peak-background split introduced in Kaiser [1984] to generate the quasar catalogue, even thought strictly speaking the conditions of the split as stated in Kaiser [1984] are not met: the correlation length of quasars is on the order of Mpcs, and use the matter density field  $\delta_1(\mathbf{x})$  is resolved to a similar scale. This breaks the assumption that the number counts of objects be well resolved on the background scale.

We require the expected number of quasars in a given  $R_{\alpha}$  cell, at redshift z to be

$$E[N_Q(x;z)] = R_\alpha^3 \langle n_Q(x,z) \rangle \exp\left\{ \left( \beta(z)\delta_1(x) - \frac{1}{2}\beta^2(z) \left\langle \delta(x)^2 \right\rangle \right) \right\}.$$

The parameter  $\beta(z)$  controls the bias of quasars, and the relation between  $\beta$  and  $b_Q$  can be measured from mocks. Between  $\beta = 1$  and  $\beta = 4$ , we have roughly  $b_Q \approx \beta$ .

The actual number of quasars in a cell is a random variable. We assume it is Poisson distributed with mean  $E[N_Q(x; z)]$ . It is straight-forward to show that the distribution does not contaminate the correlation function, as long as the numbers in different cells are kept independent.

The position of a quasar with in a cell  $(r_{ab})$  is then drawn from a uniform distribution, and the final position of b-th quasar in cell a is

$$y_{a,b} = x_a + r_{ab}.$$

It can be shown that the positioning within a cell has no effect on the correlation function as long as the positions are independent.

As a final step, we apply a sky mask: quasars not in the sky coverage are removed from the catalogue. The comoving position of quasars are converted to RA, DEC and redshift z, where z only contains the Hubble recession component for now.

## 5.4 From Density Field to Lyman- $\alpha$ Forest

After the position of quasars is determined, we draw sightlines from the positions of quasars  $y_i$  to the observer at  $x_0$ .

Unlike the discrete point-wise quasars, the "over"-transmission fraction of forest  $(\delta_F)$  is a continuous field: the pixels on the sightlines sample this continuous field at some given scale resolved by the instrument used to measure the spectra.

Lyman- $\alpha$  forest arises from [Davé et al., 1999] In order to calculate this field from the density fluctuations, we largely follow the approximation used in Le Goff et al. [2011], where a lognormal transformation is first applied to the density fluctuations  $\delta(\mathbf{x})$ , then Fluctuating Gunn Peterson Approximation (FGPA) is introduced to convert the correlated Gaussian density field  $\delta(x)$  to a dimensionless "column density" field. [Croft et al., 1998]

$$N_{\rm LN}(x,z) = \exp\left\{D(z)\delta(x) - \frac{1}{2}D(z)^2\left\langle\delta(x)^2\right\rangle\right\}.$$

The lognormal transformation is applicable on a comoving scale  $R_{\rm LN}$  of around 200 h<sup>-1</sup>Kpc [Bi and Davidsen, 1997]. In Bi & Davidsen, the authors choose a scale where the variance of the smoothed density field is around 1.0 – we follow the same guideline in selecting the comoving log-normal transformation scale . Our density field has redshift evolution:  $\delta(x; z) = D(z)\delta(x; z = 0)$ . We choose a smoothing scale where the variance between redshift z = 2.0 and z = 3.0 is close to 1.0; the scale is  $R_{\rm LN} = 250 \text{ h}^{-1}\text{Kpc}$ .

For the transformation from the dimensionless column density field to the neutral hydrogen column density, we use the FGPA formula,

$$T(x, z; R_{\rm LN}) = R_{\rm LN} A(z) (N_{\rm LN}(x, z))^{B(z)},$$

where the Lyman- $\alpha$  cross section is absorbed to a free parameter A(z). A(z) mean controls the overall transmission fraction at a given redshift  $\langle F \rangle(z)$ , while the other parameter B(z) controls the variance of the transmission fraction  $\sigma_F(z)$ . These parameters can be fitted for by requiring the mocks to match up with the models. [See Lee et al., 2013, for a collection of fitting formulas]

We cannot perform the fitting at this stage, because the optical depth field along the line of sight  $T(x, z; R_{\rm LN})$  is smoothed to  $R_{\rm LN} = 250 \ {\rm h^{-1}Kpc}$ , a scale very different from the length scale of a pixel in the survey.

To match the length scales, we resample  $T(x, z; R_{\rm LN})$  onto the uniform log- $\lambda$  grid of the SDSS pixels, following the method used in Le-Goff et al. The integral for the *j*-th pixels on the *i*-th quasar is

$$\tau_i(j) = \int_{r_{j1}}^{r_{j2}} R_{\rm LN}^{-1} T(x_i(r), z_i(r); R_{\rm LN}) dr,$$

where  $r_{j1}$  is the comoving distance at the beginning of the *j*-th pixel,  $r_{j2}$  being the end.  $x_i(r)$  is the position of along the line of sight of *i*-th quasar, relative to the observer at  $x_0$ .

After the resampling, we solve for the FPGA parameters A(z) and B(z) as functions of redshift, such that the transmission fraction of  $F = \exp -\tau_i(j)$  matches the constraints on  $\langle F \rangle(z)$  and  $\sigma_F(z)$ .

In FGPA, B(z) directly corresponds to the equation of state parameter of the IGM  $\gamma \ b = 2 - 0.7(\gamma - 1)$ , Allowing a running B(z) is equivalent to allowing a running equation of state for the IGM. B(z) varies slowly with redshift from 1.2 to 2.0 from z = 2.0 to z = 4.0. A canonical value used in previous literature is  $\gamma = 1.6$ , b = 1.58. [see, e.g. Le Goff et al., 2011, Croft et al., 1998, Lee et al., 2014]

Interestingly, A(z) can be fit by an exponential function of 1/(1 + z) and B(z) can be fit by a second order polynomial in 1/(1 + z). The origin of these fits may be related to the analytical form of the fits in the mean and variance of the transmission fraction, but we leave this relation to further study. See Figure 5.1. The fitted mean and intrinsic variance of forest pixels are shown in Figure 5.2.


Figure 5.1: The Parameter A and B used Figure 5.2: Fitting to the transmission in FGPA. model.  $\langle F \rangle$  and  $\sigma_F(z)$ .

### 5.5 Redshift Distortions

Redshift distortion is the apparent displacement in redshift for objects due to peculiar velocity along light of sight. It was first studied by Kaiser [1987], and more recently [e.g. Szalay et al., 1998, Percival and White, 2009, Vlah et al., 2013]

In this section we describe our method of introducing the redshift distortion in real space. Alternatively, redshift distortion can be introduced to the density field in Fourier space through a modified power spectrum. [see, e.g. Font-Ribera et al., 2012]

Along with the initial Gaussian field at  $R_S$ , we also generate the linear theory displacement field  $\Phi(x, z)$ . Three components x, y, z are generated with a simple differential kernel in K space. The redshift evolution is linear

$$\Phi(x,z) = D(z)\Phi(x,z=0).$$

The linear theory displacement  $\Phi(x, z)$  is converted to RSD displacement via the Kaiser (1987) formula Kaiser [1987].

$$\Psi(x,z) = F_{\Omega}(a)\Phi(x,z),$$

where

$$F_{\Omega}(a) = \left[\frac{\Omega_M}{a^3 E^2(a)}\right]^{0.6}$$

The redshift distortion displacement  $\Psi(x, z)$  is projected along the sight line. The position of the  $R_{\rm LN}$  scale sampled points is shifted by  $\Psi(\mathbf{x})$  before the resampling to spectra pixels; the operation is equivalent to a convolution by the redshift distortion field. The location of quasars is shifted by the same RSD displacement field.

# 5.6 Measuring Linear Theory Redshift Distortion Parameters

The redshift distortion (RSD) in the mocks can be measured by fitting the correlation function to the parametrized linear theory prediction. In the linear redshift distortion theory developed by Kaiser [1987], the distortion can be described by two parameters, b, and  $\beta$ , where b is the bias of the object, and  $\beta$  is the RSD parameter. Redshift distortion introduces a quadrupole contribution to the correlation function, and may significantly change the magnitude of the monopole.

We have implemented a fitter that is similar to the fitter used by the SDSS survey. [Xu et al., 2013] The fitter minimizes

$$\chi = (D - M)^T \mathcal{C}(D - M),$$

where D is the simulated correlation function, M the model correlation function (for now, linear theory is used), and C is the covariance matrix.

The linear theory model is calculated from the linear theory power spectrum following following the method described in Slosar et al. [2011], but extended by O'Connell et. al. to include cross correlation as well

$$\xi_F(r,\mu) = \sum b_1 b_2 C_l(\beta_1,\beta_2) \xi_{Fl}(r) P_l(\mu),$$

where  $P_l(\mu)$  are the legendre polynomials,  $\beta_i$  and  $b_i$  are the RSD-parameter and bias of two types of correlated objects, and

$$C_{0} = 1 + \frac{1}{3}(\beta_{1} + \beta_{2}) + \frac{1}{5}\beta_{1}\beta_{2}$$
$$C_{2} = \frac{2}{3}(\beta_{1} + \beta_{2}) + \frac{4}{7}\beta_{1}\beta_{2}$$
$$C_{4} = \frac{8}{35}\beta_{1}\beta_{2},$$

. The eigenstates used for fitting,  $\xi_{Fl}$  are spherical Bessel expansions of the power spectrum,

$$\xi_{Fl} = (2\pi)^{-3} \int P(k) j_l(kr) d^3k.$$

#### 5.6.1 Estimating the Covariance Matrix

The covariance matrix C is measured from a bootstrap sample of 100,000 realizations from 50 independent mocks. We divided each mock into 192 cone-shaped chunks originating from the observer with Healpix [Górski et al., 2005], and perform pair counting with each chunk against the full data set for all three correlation functions (quasar-quasar, quasar-forest, and forest-forest). Pair counting is slow, thus we save the results for faster calculation in the assembly of the random realizations.



Figure 5.3: Estimated Covariance Matrix with 100,000 realizations, shown as the correlation matrix (normalized by the diagonal terms). The break down of the layout of the matrix is described in Table 5.1

Row / Column range	0 - 960	960 - 2880	2880 - 3840
Correlation function	Quasar-Quasar	Quasar-Forest	Forest-Forest
$R~{ m h^{-1}Mpc}$	0 to $160, 40$ bins	0 to $160, 40$ bins	0  to  160, 40  bins
$\mu$	0 to $1, 24$ bins	-1 to 1, 48 bins	0 to $1, 24$ bins

Table 5.1: Break down of the Covariance Matrix

To assemble a random realization, 192 uncorrelated random numbers between 0 and 50 (the total number of mocks, excluded) are generated, each corresponding to the mock whose corresponding chunk is selected. This guarantees that the geometry of the realization is consistent with the mock. The pair counts in the selected chunks are then added up to form the correlation functions of the random realization.

The estimated correlation matrix is shown in Figure 5.3. A breakdown of the



Figure 5.4: Correlation functions on 50 Mocks with the fiducial cosmology. Power spectrum cut at 8  $h^{-1}$ Mpc.

$b_F$	$b_Q$	$\beta_F$	$eta_{m{Q}}$
$-0.264 \pm 0.013$	$2.74\pm0.42$	$1.65\pm0.16$	$1.35\pm0.50$

Table 5.2: Best Fit of Redshift Distortion Parameters

layout of the matrix by degrees of freedom is listed in Table 5.1.

#### 5.6.2 Best Fit to Linear Theory Correlation Functions

Figure 5.4 shows the averaged best fit cross 50 mocks, and the averaged correlation functions of 50 mocks. The best fit value of the bias and redshift distortion parameter is listed in Table 5.2

# 5.7 Software Implementation and Data Format

#### 5.7.1 Software Implementation

We provide Python implementations of the mock generation algorithm, the correlation function calculator, and the linear theory fitter.

The mock generation algorithm uses pyFFTW [Gomersall, 2014], PyCAMB, [Zuntz, 2013], and a modified version of the ndimage module in scipy [Jones et al., 2001–]. The correlation function calculator is based on kdcount [Feng, 2013a]. The linear theory fitter is based on the fitting routines in scipy [Jones et al., 2001–].

The parallel infrastructure of the software is provided by sharedmem [Feng, 2013b].

#### 5.7.2 Data Format: Quasars

The quasars in the simulation are saved in one file per mock. Usually saved in datadir/QSOcatelogue.raw. RA and DEC angle of the quasar is saved in radians. Z\_RED is the redshift distorted (RSD) redshift of a quasar. Z\_REAL is the non-RSD redshift of a quasar.

The file can be accessed with numpy, as illustrated in the following example.

#### 5.7.3 Data Format: Forest Spectra

The forest is more convoluted. There are several involved files:

- QSONpixel.raw The number of forest pixels in each spectra line. It contains one 32 bit integer per QSO.
- SpectraOutputTauRed.raw The RSD optical depth of pixels. One single precision floating number per pixel. All QSO spectra lines are concatenated into one file.
- SpectraOutputTauReal.raw The RSD optical depth of pixels. One single precision floating number per pixel. All QSO spectra lines are concatenated into one file.
- SpectraOutputLogLam.raw. The restframe wavelength of pixels, in  $\log_{10}$  /A. One single precision floating number per pixel. All QSO spectra lines are concatenated into one file.

The file can be accessed with numpy, as illustrated in the following example.

```
Npixels = numpy.fromfile('QSONpixel.raw', 'i4')
End = Npixels.cumsum()
Start = numpy.concatenate([[0], END])
tau = numpy.fromfile('SpectraOutputTauRed.Raw', 'f4')
loglam = numpy.fromfile('SpectraOutputLogLam.Raw', 'f4')
for i in range(len(Npixels)):
    s = slice(Start[i], End[i])
    print loglam[s], tau[s]
```

# 5.8 Conclusions

We developed a highly affordable method to generate correlated quasar and Lyman- $\alpha$  forest mocks. The software we developed scales down to computers as small as laptops. The log-normal transformation we use to generate the density fields for quasars and the forest wa inspired by the peak-background split. The Fluctuating Gunn-Peterson Approximation is the fundamental approximation used in the mocks. In addition, a split power spectrum at different scales is used to ease the numerical feasibility of the generation of the matter density field. We allow a floating *B* parameter in FGPA approximation in addition to the floating *A* parameter, so that the mocks can simultaneously fit to the first and second moments (mean and variance) of the transmission fraction of forest. We show that the parameter *A* and *B* agree with simple analytical forms.

The correlation function of the mocks shows good agreement with linear theory models at the large scale as expected. We fit the model with linear theory in a manner similar to observations, and extract the vanilla redshift distortion parameters from the transmission fraction directly produced from the mocks. We note that there is tension with the recently reported numbers from the SDSS survey, which can be resolved by tuning the input parameters of the mocks.

The input parameters used in the mocks are summarized in Table 5.3. In the future, we plan to incorporate a quasar continuum model, and the instrument noise model to build a full mock pipe-line. The mock catalog generator will be used in work with the SDSSIV/eBOSS and DESI survey, both of which will observe many more quasars even than SDSS/BOSS. The mock generator can easily scale to these sizes and more, using distributed computer architectures.

Cosmology Cosmology Parameters Controls the power spectrum output from CAMB. ( $\Omega_{\Lambda}$ , etc.)

- $\beta_Q$  Quasar Bias Parameter: Controls the bias of quasar
- $\langle F \rangle (z)$  Mean Transmission Fraction: The mean transmission fraction of Lyman- $\alpha$  forest
- $\sigma_F(z)$  Intrinsic Variance in Transmission Fraction: The intrinsic variance in Lyman- $\alpha$  forest pixels; this parameter controls the bias of the forest.
- $R_\alpha$  Decorrelation Scale: The short range scale where correlation in density fluctuation is replaced with uncorrelated fluctuation
- $R_B$  Large Box Scale : The long range scale where the long range power spectrum split is performed.

Table 5.3: List of Free parameters in the Mock Algorithm

# Chapter 6 Conclusion and Future Work

# 6.1 Conclusions

The increasing size of cosmological simulations has led to the need for new visualization techniques. We focus on Smoothed Particle Hydrodynamical (SPH) simulations run with the Gadget code and describe methods for visually accessing the entire simulation at full resolution. The simulation snapshots are rastered and processed on supercomputers into images that are ready to be accessed through a web interface (GigaPan). This allows any scientist with a web-browser to interactively explore simulation datasets in both in spatial and temporal dimensions, datasets which in their native format can be hundreds of terabytes in size or more. We present two examples, the first a static terapixel image of the MassiveBlack simulation, a P-Gadget SPH simulation with 65 billion particles, and the second an interactively zoomable animation of a different simulation with more than one thousand frames, each a gigapixel in size. Both are available for public access through the GigaPan web interface. We also make our imaging software publicly available.

We use zoom-in techniques to re-simulate three high-redshift ( $z \ge 5.5$ ) halos which host 10<sup>9</sup> solar mass black holes from the ~ Gpc volume, MassiveBlack cosmological hydrodynamic simulation. We examine a number of factors potentially affecting supermassive black hole growth at high redshift in cosmological simulations. We find insignificant differences in the black hole accretion history by (i) varying the region over which feedback energy is deposited directly (ii) changing mass resolution by factors of up to 64, (iii) changing the black hole seed mass by a factor of 100. Switching from the density-entropy formulation to the pressure-entropy formulation of smoothed particle hydrodynamics slightly increases the accretion rate. In general numerical details/model parameters appear to have small effects on the main fueling mechanism for black holes at these high redshifts. The insensitivity to simulation technique seems to be a hallmark of the cold flow feeding picture of these high-z supermassive black holes. We show that the gas that participates in critical accretion phases, in these massive objects at  $z > 6 \sim 7$  is in all cases colder, denser, and forms more coherent streams than the average gas in the halo. This is also mostly the case when the black hole accretion is feedback regulated (z < 6), however the distinction is less prominent. For our resimulated halos, cold flows appear to be a viable mechanism for forming the most massive black holes in the early universe, occurring naturally in  $\Lambda$ CDM models of structure formation, without requiring fine tuning of numerical parameters.

We use radiative transfer to study the growth of ionized regions around the brightest, z = 8 quasars in a large cosmological hydrodynamic simulation that includes black hole growth and feedback (the MassiveBlack simulation). We find that in the presence of the quasars the comoving HII bubble radii reach  $10 \,\mathrm{Mpc/h}$  after 20 Myear while with the stellar component alone the HII bubbles are smaller by at least an order of magnitude. Our calculations show that several features are not captured within an analytic growth model of Strongren spheres. The X-ray photons from hard quasar spectra drive a smooth transition from fully neutral to partially neutral in the ionization front. However the transition from partially neutral to fully ionized is significantly more complex. We measure the distance to the edge of bubbles as a function of angle and use the standard deviation of these distances as a diagnostic of the anisotropy of ionized regions. We find that the overlapping of nearby ionized regions from clustered halos not only increases the anisotropy, but also is the main mechanism which allows the outer radius to grow. We therefore predict that quasar ionized bubbles at this early stage in the reionization process should be both significantly larger and more irregularly shaped than bubbles around star forming galaxies. Before the star formation rate increases and the Universe fully reionizes, quasar bubbles will form the most striking and recognizable features in 21cm maps.

## 6.2 Future Directions

Beyond petascale, supercomputers are already advancing into a new era of co-processors and accelerators, which allows on chip massive parallellism. It has been demonstrated for simple algorithms, like Lattice QCD, fully embracing the co-processors can result in speed ups of up to factor of 30. However, for more non-local algorithms such as those used in cosmological simulations, no such advances have been seen yet. In particular, although cosmological simulation algorithms are in principle friendly towards massively parallelism, the implementations used are more frequently not. One particular barrier that we have been encountering is domain decomposition. Domain decomposition, due to its complexity, can quickly become unpractically slow on large supercomputers. One workaround for this issue is to increase the efficiency of in-domain parallelism with threads. However, a true solution is to increase the parallelism of the domain decomposition algorithms. Another issue is dealing with the longrange particle mesh gravity. Currently, solving the long range gravity in cosmological simulations requires carrying out some of the largest discrete Fourier transforms with brute force methods. The naive slab domain decomposition used in popular parallel FFT software libraries fails to use all of the processors and becomes

 $O(N^2)$  in the limit of large processor counts. Incorporating new FFT libraries will fully utilize all of the processing units. We are starting to carry out such work, which will lay out the foundation for the next generation of software to simulate quasars from cosmological initial conditions.

We have not evaluated the effects of resolving the black hole zone of influence in simulations. Directly modelling the zone of influence may require more detailed treatments of the physical processes likely to be most important, such as radiative transfer. We also plan to study the global reionization process and role of quasars in the process in future work. Both of these topics point to the development of petascale hybrid radiative transfer and hydrodynamics software.

With the mocks observations of quasars and the Lyman- $\alpha$  forest, although the most computationally expensive part of the algorithm has been developed, there are several major physical models that will be added to increase their usefulness for real scientific studies. Two features needed by observers are the inclusion of a quasar continuum and noise model, and addition of metal absorbers.

# Bibliography

- K. N. Abazajian, J. K. Adelman-McCarthy, M. A. Agüeros, S. S. Allam, C. Allende Prieto, D. An, K. S. J. Anderson, S. F. Anderson, J. Annis, N. A. Bahcall, and et al. The Seventh Data Release of the Sloan Digital Sky Survey. ApJS, 182:543, June 2009. doi: 10.1088/0067-0049/182/2/543.
- T. Abel, G. L. Bryan, and M. L. Norman. The Formation and Fragmentation of Primordial Molecular Clouds. ApJ, 540:39–44, September 2000. doi: 10.1086/309295.
- O. Agertz, B. Moore, J. Stadel, D. Potter, F. Miniati, J. Read, L. Mayer, A. Gawryszczak, A. Kravtsov, Å. Nordlund, F. Pearce, V. Quilis, D. Rudd, V. Springel, J. Stone, E. Tasker, R. Teyssier, J. Wadsley, and R. Walder. Fundamental differences between SPH and grid methods. MNRAS, 380:963–978, September 2007. doi: 10.1111/j.1365-2966.2007.12183.x.
- G. Altay, R. A. C. Croft, and I. Pelupessy. SPHRAY: a smoothed particle hydrodynamics ray tracer for radiative transfer. MNRAS, 386:1931–1946, June 2008. doi: 10.1111/j.1365-2966.2008.13212.x.
- M. A. Alvarez and T. Abel. Quasar HII regions during cosmic reionization. MNRAS, 380:L30–L34, September 2007. doi: 10.1111/j.1745-3933.2007.00342.x.
- M. A. Alvarez, J. H. Wise, and T. Abel. Accretion onto the First Stellar-Mass Black Holes. ApJ, 701:L133–L137, August 2009. doi: 10.1088/0004-637X/701/2/L133.
- M. A. Alvarez, U.-L. Pen, and T.-C. Chang. Enhanced Detectability of Prereionization 21 cm Structure. ApJ, 723:L17–L21, November 2010. doi: 10.1088/2041-8205/723/1/L17.
- D. Anglés-Alcázar, R. Davé, F. Ozel, and B. D. Oppenheimer. Cosmological Zoom Simulations of z = 2 Galaxies: The Impact of Galactic Outflows. ArXiv e-prints, March 2013.
- G. Angloher, M. Bauer, I. Bavykina, A. Bento, C. Bucci, C. Ciemniak, G. Deuter, F. von Feilitzsch, D. Hauff, P. Huff, C. Isaila, J. Jochum, M. Kiefer, M. Kimmerle, J.-C. Lanfranchi, F. Petricca, S. Pfister, W. Potzel, F. Pröbst, F. Reindl, S. Roth, K. Rottler, C. Sailer, K. Schäffner, J. Schmaler, S. Scholl, W. Seidel,

M. v. Sivers, L. Stodolsky, C. Strandhagen, R. Strauß, A. Tanzke, I. Usherov, S. Wawoczny, M. Willers, and A. Zöller. Results from 730 kg days of the CRESST-II Dark Matter search. *European Physical Journal C*, 72:1971, April 2012. doi: 10.1140/epjc/s10052-012-1971-8.

- J. S. Bagla. TreePM: A Code for Cosmological N-Body Simulations. Journal of Astrophysics and Astronomy, 23:185–196, December 2002. doi: 10.1007/BF02702282.
- J. S. Bagla and T. Padmanabhan. Cosmological N-body simulations. *Pramana*, 49: 161, August 1997. doi: 10.1007/BF02845853.
- J. S. Bagla and S. Ray. Comments on the size of the simulation box in cosmological N-body simulations. MNRAS, 358:1076–1082, April 2005. doi: 10.1111/j.1365-2966.2005.08858.x.
- M. C. Begelman, M. Volonteri, and M. J. Rees. Formation of supermassive black holes by direct collapse in pre-galactic haloes. MNRAS, 370:289–298, July 2006. doi: 10.1111/j.1365-2966.2006.10467.x.
- K. G. Begeman, A. H. Broeils, and R. H. Sanders. Extended rotation curves of spiral galaxies - Dark haloes and modified dynamics. MNRAS, 249:523–537, April 1991.
- J. Bellovary, M. Volonteri, F. Governato, S. Shen, T. Quinn, and J. Wadsley. The First Massive Black Hole Seeds and Their Hosts. ApJ, 742:13, November 2011. doi: 10.1088/0004-637X/742/1/13.
- J. Bellovary, A. Brooks, M. Volonteri, F. Governato, T. Quinn, and J. Wadsley. The Relative Role of Galaxy Mergers and Cosmic Flows in Feeding Black Holes. ArXiv e-prints, July 2013.
- M. J. Berger and P. Colella. Local adaptive mesh refinement for shock hydrodynamics. *Journal of Computational Physics*, 82:64–84, May 1989. doi: 10.1016/0021-9991(89)90035-1.
- R. Bernabei, P. Belli, F. Cappella, R. Cerulli, F. Montechia, F. Nozzoli, A. Incicchitti, D. Prosperi, C. J. Dai, H. H. Kuang, J. M. Ma, and Z. P. Ye. Dark Matter search. *Nuovo Cimento Rivista Serie*, 26(1):010000–74, January 2003.
- G. Bertone, D. Hooper, and J. Silk. Particle dark matter: evidence, candidates and constraints. Phys. Rep., 405:279–390, January 2005. doi: 10.1016/j.physrep.2004.08.031.
- H. Bi and A. F. Davidsen. Evolution of Structure in the Intergalactic Medium and the Nature of the Ly $\alpha$  Forest. ApJ, 479:523–542, April 1997.

- J. S. Bolton and M. G. Haehnelt. The observed ionization rate of the intergalactic medium and the ionizing emissivity at z ¿= 5: evidence for a photon-starved and extended epoch of reionization. MNRAS, 382:325–341, November 2007. doi: 10.1111/j.1365-2966.2007.12372.x.
- J. S. Bolton, M. G. Haehnelt, S. J. Warren, P. C. Hewett, D. J. Mortlock, B. P. Venemans, R. G. McMahon, and C. Simpson. How neutral is the intergalactic medium surrounding the redshift z = 7.085 quasar ULAS J1120+0641? MNRAS, 416:L70–L74, September 2011. doi: 10.1111/j.1745-3933.2011.01100.x.
- S. Bonoli, F. Marulli, V. Springel, S. D. M. White, E. Branchini, and L. Moscardini. Modelling the cosmological co-evolution of supermassive black holes and galaxies -II. The clustering of quasars and their dark environment. MNRAS, 396:423–438, June 2009. doi: 10.1111/j.1365-2966.2009.14701.x.
- C. M. Booth and J. Schaye. Cosmological simulations of the growth of supermassive black holes and feedback from active galactic nuclei: method and tests. MNRAS, 398:53–74, September 2009. doi: 10.1111/j.1365-2966.2009.15043.x.
- C. M. Booth and J. Schaye. Dark matter haloes determine the masses of supermassive black holes. MNRAS, 405:L1–L5, June 2010. doi: 10.1111/j.1745-3933.2010.00832.x.
- C. M. Booth and J. Schaye. The interaction between feedback from active galactic nuclei and supernovae. *Scientific Reports*, art. 1738, May 2013. doi: 10.1038/srep01738.
- F. Bournaud, A. Dekel, R. Teyssier, M. Cacciato, E. Daddi, S. Juneau, and F. Shankar. Black Hole Growth and Active Galactic Nuclei Obscuration by Instability-driven Inflows in High-redshift Disk Galaxies Fed by Cold Streams. ApJ, 741:L33, November 2011. doi: 10.1088/2041-8205/741/2/L33.
- M. Boylan-Kolchin, V. Springel, S. D. M. White, A. Jenkins, and G. Lemson. Resolving cosmic structure formation with the Millennium-II Simulation. MNRAS, 398:1150–1164, September 2009. doi: 10.1111/j.1365-2966.2009.15191.x.
- V. Bromm, P. S. Coppi, and R. B. Larson. Forming the First Stars in the Universe: The Fragmentation of Primordial Gas. ApJ, 527:L5–L8, December 1999. doi: 10.1086/312385.
- G. L. Bryan and M. L. Norman. A Hybrid AMR Application for Cosmology and Astrophysics. ArXiv Astrophysics e-prints, October 1997.
- J. Carlson and M. White. Embedding Realistic Surveys in Simulations Through Volume Remapping. ApJS, 190:311–314, October 2010. doi: 10.1088/0067-0049/190/2/311.

- A. Cattaneo and R. Teyssier. AGN self-regulation in cooling flow clusters. MNRAS, 376:1547–1556, April 2007. doi: 10.1111/j.1365-2966.2007.11512.x.
- R. Cen and Z. Haiman. Quasar Strömgren Spheres Before Cosmological Reionization. ApJ, 542:L75–L78, October 2000. doi: 10.1086/312937.
- X. Chen and J. Miralda-Escudé. The 21 cm Signature of the First Stars. ApJ, 684: 18–33, September 2008. doi: 10.1086/528941.
- J.-H. Choi, I. Shlosman, and M. C. Begelman. Supermassive Black Hole Formation at High Redshifts via Direct Collapse: Physical Processes in the Early Stage. ApJ, 774:149, September 2013. doi: 10.1088/0004-637X/774/2/149.
- L. Ciotti and J. P. Ostriker. Radiative Feedback from Massive Black Holes in Elliptical Galaxies: AGN Flaring and Central Starburst Fueled by Recycled Gas. ApJ, 665: 1038–1056, August 2007. doi: 10.1086/519833.
- J. M. Colberg, K. S. Krughoff, and A. J. Connolly. Intercluster filaments in a ACDM Universe. MNRAS, 359:272–282, May 2005. doi: 10.1111/j.1365-2966.2005.08897.x.
- S. Cole, W. J. Percival, J. A. Peacock, P. Norberg, C. M. Baugh, C. S. Frenk, I. Baldry, J. Bland-Hawthorn, T. Bridges, R. Cannon, M. Colless, C. Collins, W. Couch, N. J. G. Cross, G. Dalton, V. R. Eke, R. De Propris, S. P. Driver, G. Efstathiou, R. S. Ellis, K. Glazebrook, C. Jackson, A. Jenkins, O. Lahav, I. Lewis, S. Lumsden, S. Maddox, D. Madgwick, B. A. Peterson, W. Sutherland, and K. Taylor. The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications. MNRAS, 362:505–534, September 2005. doi: 10.1111/j.1365-2966.2005.09318.x.
- James W. Cooley and John W. Tukey. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 19(90):297–301, May 1965. ISSN 0025-5718. doi: 10.1090/S0025-5718-1965-0178586-1. URL http://dx.doi.org/10.1090/S0025-5718-1965-0178586-1.
- T. Costa, D. Sijacki, M. Trenti, and M. Haehnelt. The environment of bright QSOs at z ~ 6: Star forming galaxies and X-ray emission. ArXiv e-prints, July 2013.
- INC Cray. Using the GNI and DMAPP APIs. Cray, INC, 2011.
- R. A. C. Croft and G. Altay. Radiation-induced large-scale structure during the reionization epoch: the autocorrelation function. MNRAS, 388:1501–1520, August 2008. doi: 10.1111/j.1365-2966.2008.13513.x.
- R. A. C. Croft, D. H. Weinberg, N. Katz, and L. Hernquist. Recovery of the Power Spectrum of Mass Fluctuations from Observations of the Lyα Forest. ApJ, 495: 44–62, March 1998. doi: 10.1086/305289.

- R. A. C. Croft, D. H. Weinberg, M. Pettini, L. Hernquist, and N. Katz. The Power Spectrum of Mass Fluctuations Measured from the Ly $\alpha$  Forest at Redshift z = 2.5. ApJ, 520:1–23, July 1999. doi: 10.1086/307438.
- R. A. C. Croft, T. Di Matteo, R. Davé, L. Hernquist, N. Katz, M. A. Fardal, and D. H. Weinberg. Hydrodynamic Simulation of the Cosmological X-Ray Background. ApJ, 557:67–87, August 2001. doi: 10.1086/321632.
- David Culler, J.P. Singh, and Anoop Gupta. Parallel Com-Architecture: *Hardware/Software* Morputer A Approach.1998. URL Kaufmann, 1stedition, ISBN 1558603433. gan http://www.amazon.com/Parallel-Computer-Architecture-Hardware-Software/dp/155860 The Morgan Kaufmann Series in Computer Architecture and Design.
- K. K. Datta, S. Majumdar, S. Bharadwaj, and T. R. Choudhury. Simulating the impact of HI fluctuations on matched filter search for ionized bubbles in redshifted 21-cm maps. MNRAS, 391:1900–1912, December 2008. doi: 10.1111/j.1365-2966.2008.14008.x.
- K. K. Datta, M. M. Friedrich, G. Mellema, I. T. Iliev, and P. R. Shapiro. Prospects of observing a quasar H II region during the epoch of reionization with the redshifted 21-cm signal. MNRAS, 424:762–778, July 2012. doi: 10.1111/j.1365-2966.2012.21268.x.
- R. Davé, L. Hernquist, N. Katz, and D. H. Weinberg. The Low-Redshift Ly $\alpha$  Forest in Cold Dark Matter Cosmologies. ApJ, 511:521–545, February 1999. doi: 10.1086/306722.
- M. Davis, G. Efstathiou, C. S. Frenk, and S. D. M. White. The evolution of largescale structure in a universe dominated by cold dark matter. ApJ, 292:371–394, May 1985. doi: 10.1086/163168.
- J. Debuhr, E. Quataert, and C.-P. Ma. The growth of massive black holes in galaxy merger simulations with feedback by radiation pressure. MNRAS, 412:1341–1360, April 2011. doi: 10.1111/j.1365-2966.2010.17992.x.
- C. Degraf, T. Di Matteo, and V. Springel. Black hole clustering in cosmological hydrodynamic simulations: evidence for mergers. MNRAS, 413:1383–1394, May 2011. doi: 10.1111/j.1365-2966.2011.18221.x.
- C. DeGraf, T. Di Matteo, N. Khandai, R. Croft, J. Lopez, and V. Springel. Early black holes in cosmological simulations: luminosity functions and clustering behaviour. MNRAS, 424:1892–1898, August 2012. doi: 10.1111/j.1365-2966.2012.21294.x.

- W. Dehnen and H. Aly. Improving convergence in smoothed particle hydrodynamics simulations without pairing instability. MNRAS, 425:1068–1082, September 2012. doi: 10.1111/j.1365-2966.2012.21439.x.
- A. Dekel and J. Silk. The origin of dwarf galaxies, cold dark matter, and biased galaxy formation. ApJ, 303:39–55, April 1986. doi: 10.1086/164050.
- A. Dekel, Y. Birnboim, G. Engel, J. Freundlich, T. Goerdt, M. Mumcuoglu, E. Neistein, C. Pichon, R. Teyssier, and E. Zinger. Cold streams in early massive hot haloes as the main mode of galaxy formation. Nature, 457:451–454, January 2009. doi: 10.1038/nature07648.
- T. Di Matteo, V. Springel, and L. Hernquist. Energy input from quasars regulates the growth and activity of black holes and their host galaxies. Nature, 433:604–607, February 2005. doi: 10.1038/nature03335.
- T. Di Matteo, J. Colberg, V. Springel, L. Hernquist, and D. Sijacki. Direct Cosmological Simulations of the Growth of Black Holes and Galaxies. ApJ, 676:33–53, March 2008. doi: 10.1086/524921.
- T. Di Matteo, N. Khandai, C. DeGraf, Y. Feng, R. A. C. Croft, J. Lopez, and V. Springel. Cold Flows and the First Quasars. ApJ, 745:L29, February 2012. doi: 10.1088/2041-8205/745/2/L29.
- S. Dodelson. Modern cosmology. 2003.
- K. Dolag, M. Meneghetti, L. Moscardini, E. Rasia, and A. Bonaldi. Simulating the physical properties of dark matter and gas inside the cosmic web. MNRAS, 370: 656–672, August 2006. doi: 10.1111/j.1365-2966.2006.10511.x.
- Y. Dubois, J. Devriendt, A. Slyz, and R. Teyssier. Self-regulated growth of supermassive black holes by a dual jet-heating active galactic nucleus feedback mechanism: methods, tests and implications for cosmological simulations. MNRAS, 420:2662– 2683, March 2012. doi: 10.1111/j.1365-2966.2011.20236.x.
- Y. Dubois, C. Pichon, J. Devriendt, J. Silk, M. Haehnelt, T. Kimm, and A. Slyz. Blowing cold flows away: the impact of early AGN activity on the formation of a brightest cluster galaxy progenitor. MNRAS, 428:2885–2900, February 2013. doi: 10.1093/mnras/sts224.
- D. J. Eisenstein, I. Zehavi, D. W. Hogg, R. Scoccimarro, M. R. Blanton, R. C. Nichol, R. Scranton, H.-J. Seo, M. Tegmark, Z. Zheng, S. F. Anderson, J. Annis, N. Bahcall, J. Brinkmann, S. Burles, F. J. Castander, A. Connolly, I. Csabai, M. Doi, M. Fukugita, J. A. Frieman, K. Glazebrook, J. E. Gunn, J. S. Hendry, G. Hennessy, Z. Ivezić, S. Kent, G. R. Knapp, H. Lin, Y.-S. Loh, R. H. Lupton, B. Margon,

T. A. McKay, A. Meiksin, J. A. Munn, A. Pope, M. W. Richmond, D. Schlegel, D. P. Schneider, K. Shimasaku, C. Stoughton, M. A. Strauss, M. SubbaRao, A. S. Szalay, I. Szapudi, D. L. Tucker, B. Yanny, and D. G. York. Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies. ApJ, 633:560–574, November 2005. doi: 10.1086/466512.

- X. Fan, V. K. Narayanan, R. H. Lupton, M. A. Strauss, G. R. Knapp, R. H. Becker, R. L. White, L. Pentericci, S. K. Leggett, Z. Haiman, J. E. Gunn, Ž. Ivezić, D. P. Schneider, S. F. Anderson, J. Brinkmann, N. A. Bahcall, A. J. Connolly, I. Csabai, M. Doi, M. Fukugita, T. Geballe, E. K. Grebel, D. Harbeck, G. Hennessy, D. Q. Lamb, G. Miknaitis, J. A. Munn, R. Nichol, S. Okamura, J. R. Pier, F. Prada, G. T. Richards, A. Szalay, and D. G. York. A Survey of z¿5.8 Quasars in the Sloan Digital Sky Survey. I. Discovery of Three New Quasars and the Spatial Density of Luminous Quasars at z~6. AJ, 122:2833–2849, December 2001. doi: 10.1086/324111.
- X. Fan, M. A. Strauss, D. P. Schneider, R. H. Becker, R. L. White, Z. Haiman, M. Gregg, L. Pentericci, E. K. Grebel, V. K. Narayanan, Y.-S. Loh, G. T. Richards, J. E. Gunn, R. H. Lupton, G. R. Knapp, Ž. Ivezić, W. N. Brandt, M. Collinge, L. Hao, D. Harbeck, F. Prada, J. Schaye, I. Strateva, N. Zakamska, S. Anderson, J. Brinkmann, N. A. Bahcall, D. Q. Lamb, S. Okamura, A. Szalay, and D. G. York. A Survey of z¿5.7 Quasars in the Sloan Digital Sky Survey. II. Discovery of Three Additional Quasars at z¿6. AJ, 125:1649–1659, April 2003. doi: 10.1086/368246.
- X. Fan, J. F. Hennawi, G. T. Richards, M. A. Strauss, D. P. Schneider, J. L. Donley, J. E. Young, J. Annis, H. Lin, H. Lampeitl, R. H. Lupton, J. E. Gunn, G. R. Knapp, W. N. Brandt, S. Anderson, N. A. Bahcall, J. Brinkmann, R. J. Brunner, M. Fukugita, A. S. Szalay, G. P. Szokoly, and D. G. York. A Survey of z¿5.7 Quasars in the Sloan Digital Sky Survey. III. Discovery of Five Additional Quasars. AJ, 128:515–522, August 2004. doi: 10.1086/422434.
- X. Fan, M. A. Strauss, R. H. Becker, R. L. White, J. E. Gunn, G. R. Knapp, G. T. Richards, D. P. Schneider, J. Brinkmann, and M. Fukugita. Constraining the Evolution of the Ionizing Background and the Epoch of Reionization with z<sup>6</sup> Quasars. II. A Sample of 19 Quasars. AJ, 132:117–136, July 2006a. doi: 10.1086/504836.
- X. Fan, M. A. Strauss, G. T. Richards, J. F. Hennawi, R. H. Becker, R. L. White, A. M. Diamond-Stanic, J. L. Donley, L. Jiang, J. S. Kim, M. Vestergaard, J. E. Young, J. E. Gunn, R. H. Lupton, G. R. Knapp, D. P. Schneider, W. N. Brandt, N. A. Bahcall, J. C. Barentine, J. Brinkmann, H. J. Brewington, M. Fukugita, M. Harvanek, S. J. Kleinman, J. Krzesinski, D. Long, E. H. Neilsen, Jr., A. Nitta, S. A. Snedden, and W. Voges. A Survey of z¿5.7 Quasars in the Sloan Digital Sky Survey. IV. Discovery of Seven Additional Quasars. AJ, 131:1203–1209, March 2006b. doi: 10.1086/500296.

- N. Fanidakis, A. V. Macciò, C. M. Baugh, C. G. Lacey, and C. S. Frenk. The most luminous quasars do not live in the most massive dark matter haloes at any redshift. MNRAS, 436:315–326, November 2013. doi: 10.1093/mnras/stt1567.
- Y. Feng, R. A. C. Croft, T. Di Matteo, N. Khandai, R. Sargent, I. Nourbakhsh, P. Dille, C. Bartley, V. Springel, A. Jana, and J. Gardner. Terapixel Imaging of Cosmological Simulations. ApJS, 197:18, December 2011. doi: 10.1088/0067-0049/197/2/18.
- Y. Feng, R. A. C. Croft, T. Di Matteo, and N. Khandai. Growth and anisotropy of ionization fronts near high-redshift quasars in the MassiveBlack simulation. MN-RAS, 429:1554–1563, February 2013. doi: 10.1093/mnras/sts447.
- Y. Feng, T. Di Matteo, R. Croft, and N. Khandai. High-redshift supermassive black holes: accretion through cold flows. MNRAS, 440:1865–1879, May 2014. doi: 10.1093/mnras/stu432.
- Yu Feng. Brute-force pair counting and correlation function with kd-tree pruning, 2013a.
- Yu Feng. An extended python multi-processing library easy interface to inherited shared memory segments, 2013b.
- *Fine Intl. Conf. on Gigapixel Imaging for Science*, 2010. Fine Intl. Conf. on Gigapixel Imaging for Science.
- A. Font-Ribera, P. McDonald, and J. Miralda-Escudé. Generating mock data sets for large-scale Lyman-α forest correlation measurements. J. Cosmology Astropart. Phys., 1:001, January 2012. doi: 10.1088/1475-7516/2012/01/001.
- A. Font-Ribera, D. Kirkby, N. Busca, J. Miralda-Escudé, N. P. Ross, A. Slosar, J. Rich, É. Aubourg, S. Bailey, V. Bhardwaj, J. Bautista, F. Beutler, D. Bizyaev, M. Blomqvist, H. Brewington, J. Brinkmann, J. R. Brownstein, B. Carithers, K. S. Dawson, T. Delubac, G. Ebelke, D. J. Eisenstein, J. Ge, K. Kinemuchi, K.-G. Lee, V. Malanushenko, E. Malanushenko, M. Marchante, D. Margala, D. Muna, A. D. Myers, P. Noterdaeme, D. Oravetz, N. Palanque-Delabrouille, I. Pâris, P. Petitjean, M. M. Pieri, G. Rossi, D. P. Schneider, A. Simmons, M. Viel, C. Yeche, and D. G. York. Quasar-Lyman α forest cross-correlation from BOSS DR11: Baryon Acoustic Oscillations. J. Cosmology Astropart. Phys., 5:027, May 2014. doi: 10.1088/1475-7516/2014/05/027.
- National Center for Computational Applications. Bluewaters users guide, 2014. URL https://bluewaters.ncsa.illinois.edu/user-guide.
- W. L. Freedman and B. F. Madore. The Hubble Constant. ARA&A, 48:673–710, September 2010. doi: 10.1146/annurev-astro-082708-101829.

- M. M. Friedrich, G. Mellema, I. T. Iliev, and P. R. Shapiro. Radiative transfer of energetic photons: X-rays and helium ionization in C<sup>2</sup>-RAY. MNRAS, 421:2232– 2250, April 2012. doi: 10.1111/j.1365-2966.2012.20449.x.
- Matteo Frigo and Steven G. Johnson. The design and implementation of FFTW3. *Proceedings of the IEEE*, 93(2):216–231, 2005. Special issue on "Program Generation, Optimization, and Platform Adaptation".
- S. R. Furlanetto and S. J. Stoever. Secondary ionization and heating by fast electrons. MNRAS, 404:1869–1878, June 2010. doi: 10.1111/j.1365-2966.2010.16401.x.
- S. R. Furlanetto, M. Zaldarriaga, and L. Hernquist. The Growth of H II Regions During Reionization. ApJ, 613:1–15, September 2004. doi: 10.1086/423025.
- L. Gao, S. D. M. White, A. Jenkins, C. S. Frenk, and V. Springel. Early structure in ACDM. MNRAS, 363:379–392, October 2005. doi: 10.1111/j.1365-2966.2005.09509.x.
- M. Gaspari, C. Melioli, F. Brighenti, and A. D'Ercole. The dance of heating and cooling in galaxy clusters: three-dimensional simulations of self-regulated active galactic nuclei outflows. MNRAS, 411:349–372, February 2011. doi: 10.1111/j.1365-2966.2010.17688.x.
- P. M. Geil and J. S. B. Wyithe. The impact of a percolating IGM on redshifted 21-cm observations of quasar HII regions. MNRAS, 386:1683–1694, May 2008. doi: 10.1111/j.1365-2966.2008.13159.x.
- J. M. Gelb and E. Bertschinger. Cold dark matter. 1: The formation of dark halos. ApJ, 436:467–490, December 1994a. doi: 10.1086/174922.
- J. M. Gelb and E. Bertschinger. Cold dark matter. 2: Spatial and velocity statistics. ApJ, 436:491–508, December 1994b. doi: 10.1086/174923.
- R. A. Gingold and J. J. Monaghan. Smoothed particle hydrodynamics Theory and application to non-spherical stars. MNRAS, 181:375–389, November 1977.
- M. L. Giroux and P. R. Shapiro. The Reionization of the Intergalactic Medium and Its Observational Consequences. ApJS, 102:191, February 1996. doi: 10.1086/192257.
- N. Y. Gnedin and N. Hollon. Cooling and Heating Functions of Photoionized Gas. ApJS, 202:13, October 2012. doi: 10.1088/0067-0049/202/2/13.
- N. Y. Gnedin and L. Hui. Probing the Universe with the Lyalpha forest I. Hydrodynamics of the low-density intergalactic medium. MNRAS, 296:44–55, May 1998. doi: 10.1046/j.1365-8711.1998.01249.x.

- N. Y. Gnedin, K. Tassis, and A. V. Kravtsov. Modeling Molecular Hydrogen and Star Formation in Cosmological Simulations. ApJ, 697:55–67, May 2009. doi: 10.1088/0004-637X/697/1/55.
- Y. N. Gnedin. Ifrit, 2011. URL http://sites.google.com/site/ifrithome/Home.
- Henry Gomersall. Python binding of fftw library, 2014.
- K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann. HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. ApJ, 622:759–771, April 2005. doi: 10.1086/427976.
- F. Governato, A. Zolotov, A. Pontzen, C. Christensen, S. H. Oh, A. M. Brooks, T. Quinn, S. Shen, and J. Wadsley. Cuspy no more: how outflows affect the central dark matter and baryon distribution in Λ cold dark matter galaxies. MNRAS, 422: 1231–1240, May 2012. doi: 10.1111/j.1365-2966.2012.20696.x.
- A. H. Guth and D. I. Kaiser. Inflationary Cosmology: Exploring the Universe from the Smallest to the Largest Scales. *Science*, 307:884–890, February 2005. doi: 10.1126/science.1107483.
- Salman Habib, Vitali Morozov, Hal Finkel, Adrian Pope, Katrin Heitmann, Kalyan Kumaran, Tom Peterka, Joe Insley, David Daniel, Patricia Fasel, Nicholas Frontiere, and Zarija Lukić. The universe at extreme scale: Multi-petaflop sky simulation on the bg/q. In Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis, SC '12, pages 4:1–4:11, Los Alamitos, CA, USA, 2012. IEEE Computer Society Press. ISBN 978-1-4673-0804-5. URL http://dl.acm.org/citation.cfm?id=2388996.2389002.
- Salman Habib, Vitali Morozov, Nicholas Frontiere, Hal Finkel, Adrian Pope, and Katrin Heitmann. Hacc: Extreme scaling and performance across diverse architectures. In Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis, SC '13, pages 6:1-6:10, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2378-9. doi: 10.1145/2503210.2504566. URL http://doi.acm.org/10.1145/2503210.2504566.
- J. C. Hayes, M. L. Norman, R. A. Fiedler, J. O. Bordner, P. S. Li, S. E. Clark, A. ud-Doula, and M.-M. Mac Low. Simulating Radiating and Magnetized Flows in Multiple Dimensions with ZEUS-MP. ApJS, 165:188–228, July 2006. doi: 10.1086/504594.
- L. Hernquist and N. Katz. TREESPH A unification of SPH with the hierarchical tree method. ApJS, 70:419–446, June 1989. doi: 10.1086/191344.

- S. Hilbert, J. Hartlap, S. D. M. White, and P. Schneider. Ray-tracing through the Millennium Simulation: Born corrections and lens-lens coupling in cosmic shear and galaxy-galaxy lensing. A&A, 499:31–43, May 2009. doi: 10.1051/0004-6361/200811054.
- G. Hinshaw, D. Larson, E. Komatsu, D. N. Spergel, C. L. Bennett, J. Dunkley, M. R. Nolta, M. Halpern, R. S. Hill, N. Odegard, L. Page, K. M. Smith, J. L. Weiland, B. Gold, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, E. Wollack, and E. L. Wright. Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results. ApJS, 208:19, October 2013. doi: 10.1088/0067-0049/208/2/19.
- A. Hobbs, C. Power, S. Nayakshin, and A. R. King. Modelling supermassive black hole growth: towards an improved sub-grid prescription. MNRAS, 421:3443–3449, April 2012. doi: 10.1111/j.1365-2966.2012.20563.x.
- P. F. Hopkins. A general class of Lagrangian smoothed particle hydrodynamics methods and implications for fluid mixing problems. MNRAS, 428:2840–2856, February 2013. doi: 10.1093/mnras/sts210.
- P. F. Hopkins and E. Quataert. How do massive black holes get their gas? MNRAS, 407:1529–1564, September 2010. doi: 10.1111/j.1365-2966.2010.17064.x.
- P. F. Hopkins, G. T. Richards, and L. Hernquist. An Observational Determination of the Bolometric Quasar Luminosity Function. ApJ, 654:731–753, January 2007. doi: 10.1086/509629.
- P. F. Hopkins, D. Narayanan, and N. Murray. The meaning and consequences of star formation criteria in galaxy models with resolved stellar feedback. MNRAS, 432: 2647–2653, July 2013. doi: 10.1093/mnras/stt723.
- L. Hui and N. Y. Gnedin. Equation of state of the photoionized intergalactic medium. MNRAS, 292:27, November 1997.
- J. D. Hunter. Matplotlib: A 2d graphics environment. Computing In Science & Engineering, 9(3):90−95, 2007.
- K. Husband, M. N. Bremer, E. R. Stanway, L. J. M. Davies, M. D. Lehnert, and L. S. Douglas. Are z~5 quasars found in the most massive high-redshift overdensities? MNRAS, 432:2869–2877, July 2013. doi: 10.1093/mnras/stt642.
- S. Ikeuchi and J. P. Ostriker. Evolution of the intergalactic medium What happened during the epoch Z = 3-10? ApJ, 301:522–543, February 1986. doi: 10.1086/163921.

- I. T. Iliev, B. Ciardi, M. A. Alvarez, A. Maselli, A. Ferrara, N. Y. Gnedin, G. Mellema, T. Nakamoto, M. L. Norman, A. O. Razoumov, E.-J. Rijkhorst, J. Ritzerveld, P. R. Shapiro, H. Susa, M. Umemura, and D. J. Whalen. Cosmological radiative transfer codes comparison project - I. The static density field tests. MNRAS, 371:1057–1086, September 2006. doi: 10.1111/j.1365-2966.2006.10775.x.
- A. Jenkins, C. S. Frenk, F. R. Pearce, P. A. Thomas, J. M. Colberg, S. D. M. White, H. M. P. Couchman, J. A. Peacock, G. Efstathiou, and A. H. Nelson. Evolution of Structure in Cold Dark Matter Universes. ApJ, 499:20, May 1998. doi: 10.1086/305615.
- M. Jeon, A. H. Pawlik, T. H. Greif, S. C. O. Glover, V. Bromm, M. Milosavljević, and R. S. Klessen. The First Galaxies: Assembly with Black Hole Feedback. ApJ, 754:34, July 2012. doi: 10.1088/0004-637X/754/1/34.
- L. Jiang, X. Fan, F. Bian, J. Annis, K. Chiu, S. Jester, H. Lin, R. H. Lupton, G. T. Richards, M. A. Strauss, V. Malanushenko, E. Malanushenko, and D. P. Schneider. A Survey of z ~ 6 Quasars in the Sloan Digital Sky Survey Deep Stripe. II. Discovery of Six Quasars at z <sub>AB</sub> ¿ 21. AJ, 138:305–311, July 2009. doi: 10.1088/0004-6256/138/1/305.
- P. H. Johansson, T. Naab, and A. Burkert. Termination of star formation by BH feedback in equal- and unequal-mass mergers of disk and elliptical galaxies. Astronomische Nachrichten, 329:956, December 2008. doi: 10.1002/asna.200811064.
- Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, 2001–. URL http://www.scipy.org/.
- N. Kaiser. On the spatial correlations of Abell clusters. ApJ, 284:L9–L12, September 1984. doi: 10.1086/184341.
- N. Kaiser. Clustering in real space and in redshift space. MNRAS, 227:1–21, July 1987.
- N. Katz, L. Hernquist, and D. H. Weinberg. Galaxies and gas in a cold dark matter universe. ApJ, 399:L109–L112, November 1992. doi: 10.1086/186619.
- N. Katz, D. H. Weinberg, and L. Hernquist. Cosmological Simulations with TreeSPH. ApJS, 105:19, July 1996. doi: 10.1086/192305.
- S. M. Kent. Dark matter in spiral galaxies. II Galaxies with H I rotation curves. AJ, 93:816–832, April 1987. doi: 10.1086/114366.
- D. Kereš, N. Katz, D. H. Weinberg, and R. Davé. How do galaxies get their gas? MNRAS, 363:2–28, October 2005. doi: 10.1111/j.1365-2966.2005.09451.x.

- N. Khandai and J. S. Bagla. A modified TreePM code. Research in Astronomy and Astrophysics, 9:861–873, August 2009. doi: 10.1088/1674-4527/9/8/002.
- N. Khandai, Y. Feng, C. DeGraf, T. Di Matteo, and R. A. C. Croft. The formation of galaxies hosting z ~ 6 quasars. MNRAS, 423:2397–2406, July 2012. doi: 10.1111/j.1365-2966.2012.21047.x.
- N. Khandai, T. Di Matteo, R. Croft, S. M. Wilkins, Y. Feng, E. Tucker, C. De-Graf, and M.-S. Liu. The MassiveBlack-II Simulation: The Evolution of Halos and Galaxies to z<sup>~0</sup>. ArXiv e-prints, February 2014.
- S. Kim, M. Stiavelli, M. Trenti, C. M. Pavlovsky, S. G. Djorgovski, C. Scarlata, D. Stern, A. Mahabal, D. Thompson, M. Dickinson, N. Panagia, and G. Meylan. The Environments of High-Redshift Quasi-Stellar Objects. ApJ, 695:809–817, April 2009. doi: 10.1088/0004-637X/695/2/809.
- A. A. Klypin and S. F. Shandarin. Three-dimensional numerical model of the formation of large-scale structure in the Universe. MNRAS, 204:891–907, September 1983.
- S. M. Koushiappas, J. S. Bullock, and A. Dekel. Massive black hole seeds from low angular momentum material. MNRAS, 354:292–304, October 2004. doi: 10.1111/j.1365-2966.2004.08190.x.
- M. A. Latif, D. R. G. Schleicher, W. Schmidt, and J. Niemeyer. Black hole formation in the early Universe. MNRAS, 433:1607–1618, August 2013. doi: 10.1093/mnras/stt834.
- J. M. Le Goff, C. Magneville, E. Rollinde, S. Peirani, P. Petitjean, C. Pichon, J. Rich, C. Yeche, E. Aubourg, N. Busca, R. Charlassier, T. Delubac, J. C. Hamilton, N. Palanque Delabrouille, I. Pâris, and M. Vargas Magaña. Simulations of BAO reconstruction with a quasar Ly-α survey. A&A, 534:A135, October 2011. doi: 10.1051/0004-6361/201117736.
- K.-G. Lee, S. Bailey, L. E. Bartsch, W. Carithers, K. S. Dawson, D. Kirkby, B. Lundgren, D. Margala, N. Palanque-Delabrouille, M. M. Pieri, D. J. Schlegel, D. H. Weinberg, C. Yèche, É. Aubourg, J. Bautista, D. Bizyaev, M. Blomqvist, A. S. Bolton, A. Borde, H. Brewington, N. G. Busca, R. A. C. Croft, T. Delubac, G. Ebelke, D. J. Eisenstein, A. Font-Ribera, J. Ge, J.-C. Hamilton, J. F. Hennawi, S. Ho, K. Honscheid, J.-M. Le Goff, E. Malanushenko, V. Malanushenko, J. Miralda-Escudé, A. D. Myers, P. Noterdaeme, D. Oravetz, K. Pan, I. Pâris, P. Petitjean, J. Rich, E. Rollinde, N. P. Ross, G. Rossi, D. P. Schneider, A. Simmons, S. Snedden, A. Slosar, D. N. Spergel, N. Suzuki, M. Viel, and B. A. Weaver. The BOSS Lyα Forest Sample from SDSS Data Release 9. AJ, 145:69, March 2013. doi: 10.1088/0004-6256/145/3/69.

- K.-G. Lee, J. P. Hennawi, D. N. Spergel, D. H. Weinberg, D. W. Hogg, M. Viel, J. S. Bolton, S. Bailey, M. M. Pieri, W. Carithers, D. J. Schlegel, B. Lundgren, N. Palanque-Delabrouille, N. Suzuki, D. P. Schneider, and C. Yeche. IGM Constraints from the SDSS-III/BOSS DR9 Ly-alpha Forest Flux Probability Distribution Function. ArXiv e-prints, May 2014.
- M. Levi, C. Bebek, T. Beers, R. Blum, R. Cahn, D. Eisenstein, B. Flaugher, K. Honscheid, R. Kron, O. Lahav, P. McDonald, N. Roe, D. Schlegel, and representing the DESI collaboration. The DESI Experiment, a whitepaper for Snowmass 2013. *ArXiv e-prints*, August 2013.
- Antony Lewis, Anthony Challinor, and Anthony Lasenby. Efficient computation of CMB anisotropies in closed FRW models. Astrophys. J., 538:473–476, 2000.
- Y. Li, L. Hernquist, B. Robertson, T. J. Cox, P. F. Hopkins, V. Springel, L. Gao, T. Di Matteo, A. R. Zentner, A. Jenkins, and N. Yoshida. Formation of z<sup>~6</sup> Quasars from Hierarchical Galaxy Mergers. ApJ, 665:187–208, August 2007. doi: 10.1086/519297.
- B. Little, D. H. Weinberg, and C. Park. Primordial fluctuations and non-linear structure. MNRAS, 253:295–306, November 1991.
- G. Lodato and P. Natarajan. Supermassive black hole formation during the assembly of pre-galactic discs. MNRAS, 371:1813–1823, October 2006. doi: 10.1111/j.1365-2966.2006.10801.x.
- A. Loeb. The race between stars and quasars in reionizing cosmic hydrogen. J. Cosmology Astropart. Phys., 3:22, March 2009. doi: 10.1088/1475-7516/2009/03/022.
- A. Loeb and S. R. Furlanetto. The First Galaxies in the Universe. 2013.
- L. B. Lucy. A numerical approach to the testing of the fission hypothesis. AJ, 82: 1013–1024, December 1977. doi: 10.1086/112164.
- R. Lupton, M. R. Blanton, G. Fekete, D. W. Hogg, W. O'Mullane, A. Szalay, and N. Wherry. Preparing Red-Green-Blue Images from CCD Data. PASP, 116:133– 137, February 2004. doi: 10.1086/382245.
- S. Majumdar, T. R. Choudhury, and S. Bharadwaj. Constraining Quasar Parameters Through Bubble Detection in Redshifted 21-cm Maps. ArXiv e-prints, November 2011.
- A. Maselli, A. Ferrara, and B. Ciardi. CRASH: a radiative transfer scheme. MNRAS, 345:379–394, October 2003. doi: 10.1046/j.1365-8711.2003.06979.x.

- A. Maselli, S. Gallerani, A. Ferrara, and T. R. Choudhury. On the size of HII regions around high-redshift quasars. MNRAS, 376:L34–L38, March 2007. doi: 10.1111/j.1745-3933.2007.00283.x.
- A. Maselli, B. Ciardi, and A. Kanekar. CRASH2: coloured packets and other updates. MNRAS, 393:171–178, February 2009. doi: 10.1111/j.1365-2966.2008.14197.x.
- L. Mayer, S. Kazantzidis, A. Escala, and S. Callegari. Direct formation of supermassive black holes via multi-scale gas inflows in galaxy mergers. Nature, 466: 1082–1084, August 2010. doi: 10.1038/nature09294.
- I. G. McCarthy, J. Schaye, R. G. Bower, T. J. Ponman, C. M. Booth, C. Dalla Vecchia, and V. Springel. Gas expulsion by quasar-driven winds as a solution to the overcooling problem in galaxy groups and clusters. MNRAS, 412:1965–1984, April 2011. doi: 10.1111/j.1365-2966.2010.18033.x.
- P. McDonald, J. Miralda-Escudé, M. Rauch, W. L. W. Sargent, T. A. Barlow, R. Cen, and J. P. Ostriker. The Observed Probability Distribution Function, Power Spectrum, and Correlation Function of the Transmitted Flux in the Lyα Forest. ApJ, 543:1–23, November 2000. doi: 10.1086/317079.
- P. McDonald, U. Seljak, S. Burles, D. J. Schlegel, D. H. Weinberg, R. Cen, D. Shih, J. Schaye, D. P. Schneider, N. A. Bahcall, J. W. Briggs, J. Brinkmann, R. J. Brunner, M. Fukugita, J. E. Gunn, Ž. Ivezić, S. Kent, R. H. Lupton, and D. E. Vanden Berk. The Lyα Forest Power Spectrum from the Sloan Digital Sky Survey. ApJS, 163:80–109, March 2006. doi: 10.1086/444361.
- I. D. McGreer, L. Jiang, X. Fan, G. T. Richards, M. A. Strauss, N. P. Ross, M. White, Y. Shen, D. P. Schneider, A. D. Myers, W. N. Brandt, C. DeGraf, E. Glikman, J. Ge, and A. Streblyanska. The z = 5 Quasar Luminosity Function from SDSS Stripe 82. ApJ, 768:105, May 2013. doi: 10.1088/0004-637X/768/2/105.
- A. A. Meiksin. The physics of the intergalactic medium. Reviews of Modern Physics, 81:1405–1469, October 2009. doi: 10.1103/RevModPhys.81.1405.
- M. Milosavljević, S. M. Couch, and V. Bromm. Accretion Onto Intermediate-Mass Black Holes in Dense Protogalactic Clouds. ApJ, 696:L146–L149, May 2009. doi: 10.1088/0004-637X/696/2/L146.
- J. J. Monaghan. Smoothed particle hydrodynamics. ARA&A, 30:543–574, 1992. doi: 10.1146/annurev.aa.30.090192.002551.
- J. J. Monaghan. SPH without a Tensile Instability. Journal of Computational Physics, 159:290–311, April 2000. doi: 10.1006/jcph.2000.6439.

- Kenneth Moreland. Diverging color maps for scientific visualization. In Proceedings of the 5th International Symposium on Advances in Visual Computing: Part II, ISVC '09, pages 92–103, Berlin, Heidelberg, 2009. Springer-Verlag. ISBN 978-3-642-10519-7.
- E. Morganson, G. De Rosa, R. Decarli, F. Walter, K. Chambers, I. McGreer, X. Fan, W. Burgett, H. Flewelling, J. Greiner, K. Hodapp, N. Kaiser, E. Magnier, P. Price, H.-W. Rix, B. Sweeney, and C. Waters. The First High-redshift Quasar from Pan-STARRS. AJ, 143:142, June 2012. doi: 10.1088/0004-6256/143/6/142.
- J. P. Morris. A study of the stability properties of smooth particle hydrodynamics. PASA, 13:97–102, January 1996.
- D. J. Mortlock, S. J. Warren, B. P. Venemans, M. Patel, P. C. Hewett, R. G. McMahon, C. Simpson, T. Theuns, E. A. Gonzáles-Solares, A. Adamson, S. Dye, N. C. Hambly, P. Hirst, M. J. Irwin, E. Kuiper, A. Lawrence, and H. J. A. Röttgering. A luminous quasar at a redshift of z = 7.085. Nature, 474:616–619, June 2011. doi: 10.1038/nature10159.
- F. Nakamura and M. Umemura. On the Initial Mass Function of Population III Stars. ApJ, 548:19–32, February 2001. doi: 10.1086/318663.
- G. S. Novak, J. P. Ostriker, and L. Ciotti. Radiative transfer and radiative driving of outflows in active galactic nuclei and starbursts. MNRAS, 427:2734–2756, December 2012. doi: 10.1111/j.1365-2966.2012.21844.x.
- T. Okamoto, C. S. Frenk, A. Jenkins, and T. Theuns. The properties of satellite galaxies in simulations of galaxy formation. MNRAS, 406:208–222, July 2010. doi: 10.1111/j.1365-2966.2010.16690.x.
- T. Padmanabhan. Advanced Topics in Cosmology: A Pedagogical Introduction. In S. Daflon, J. Alcaniz, E. Telles, and R. de la Reza, editors, *Graduate School in Astronomy: X*, volume 843 of *American Institute of Physics Conference Series*, pages 111–166, June 2006. doi: 10.1063/1.2219327.
- I. Pâris, P. Petitjean, É. Aubourg, S. Bailey, N. P. Ross, A. D. Myers, M. A. Strauss, S. F. Anderson, E. Arnau, J. Bautista, D. Bizyaev, A. S. Bolton, J. Bovy, W. N. Brandt, H. Brewington, J. R. Browstein, N. Busca, D. Capellupo, W. Carithers, R. A. C. Croft, K. Dawson, T. Delubac, G. Ebelke, D. J. Eisenstein, P. Engelke, X. Fan, N. Filiz Ak, H. Finley, A. Font-Ribera, J. Ge, R. R. Gibson, P. B. Hall, F. Hamann, J. F. Hennawi, S. Ho, D. W. Hogg, Ž. Ivezić, L. Jiang, A. E. Kimball, D. Kirkby, J. A. Kirkpatrick, K.-G. Lee, J.-M. Le Goff, B. Lundgren, C. L. MacLeod, E. Malanushenko, V. Malanushenko, C. Maraston, I. D. McGreer, R. G. McMahon, J. Miralda-Escudé, D. Muna, P. Noterdaeme, D. Oravetz, N. Palanque-Delabrouille, K. Pan, I. Perez-Fournon, M. M. Pieri, G. T. Richards, E. Rollinde,

E. S. Sheldon, D. J. Schlegel, D. P. Schneider, A. Slosar, A. Shelden, Y. Shen, A. Simmons, S. Snedden, N. Suzuki, J. Tinker, M. Viel, B. A. Weaver, D. H. Weinberg, M. White, W. M. Wood-Vasey, and C. Yèche. The Sloan Digital Sky Survey quasar catalog: ninth data release. A&A, 548:A66, December 2012. doi: 10.1051/0004-6361/201220142.

- K. Park and M. Ricotti. Accretion onto Intermediate-mass Black Holes Regulated by Radiative Feedback. I. Parametric Study for Spherically Symmetric Accretion. ApJ, 739:2, September 2011. doi: 10.1088/0004-637X/739/1/2.
- K. Park and M. Ricotti. Accretion onto Black Holes from Large Scales Regulated by Radiative Feedback. II. Growth Rate and Duty Cycle. ApJ, 747:9, March 2012. doi: 10.1088/0004-637X/747/1/9.
- A. H. Pawlik, J. Schaye, and E. van Scherpenzeel. Keeping the Universe ionized: photoheating and the clumping factor of the high-redshift intergalactic medium. MNRAS, 394:1812–1824, April 2009. doi: 10.1111/j.1365-2966.2009.14486.x.
- P. J. E. Peebles. The Effect of a Lumpy Matter Distribution on the Growth of Irregularities in an Expanding Universe. A&A, 32:391, June 1974.
- P. J. E. Peebles. Renormalization group computation of the mass distribution in an expanding universe. I - Method. ApJ, 297:350–360, October 1985. doi: 10.1086/163534.
- Dmitry Pekurovsky. P3dfft: A framework for parallel computations of fourier transforms in three dimensions. SIAM J. Scientific Computing, 34(4), 2012.
- F. I. Pelupessy, T. Di Matteo, and B. Ciardi. How Rapidly Do Supermassive Black Hole "Seeds" Grow at Early Times? ApJ, 665:107–119, August 2007. doi: 10.1086/519235.
- W. J. Percival and M. White. Testing cosmological structure formation using redshiftspace distortions. MNRAS, 393:297–308, February 2009. doi: 10.1111/j.1365-2966.2008.14211.x.
- W. J. Percival, B. A. Reid, D. J. Eisenstein, N. A. Bahcall, T. Budavari, J. A. Frieman, M. Fukugita, J. E. Gunn, Ž. Ivezić, G. R. Knapp, R. G. Kron, J. Loveday, R. H. Lupton, T. A. McKay, A. Meiksin, R. C. Nichol, A. C. Pope, D. J. Schlegel, D. P. Schneider, D. N. Spergel, C. Stoughton, M. A. Strauss, A. S. Szalay, M. Tegmark, M. S. Vogeley, D. H. Weinberg, D. G. York, and I. Zehavi. Baryon acoustic oscillations in the Sloan Digital Sky Survey Data Release 7 galaxy sample. MNRAS, 401:2148–2168, February 2010. doi: 10.1111/j.1365-2966.2009.15812.x.

- S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom, I. M. Hook, A. G. Kim, M. Y. Kim, J. C. Lee, N. J. Nunes, R. Pain, C. R. Pennypacker, R. Quimby, C. Lidman, R. S. Ellis, M. Irwin, R. G. McMahon, P. Ruiz-Lapuente, N. Walton, B. Schaefer, B. J. Boyle, A. V. Filippenko, T. Matheson, A. S. Fruchter, N. Panagia, H. J. M. Newberg, W. J. Couch, and T. S. C. Project. Measurements of Ω and Λ from 42 High-Redshift Supernovae. ApJ, 517:565–586, June 1999. doi: 10.1086/307221.
- P. Petitjean, J. K. Webb, M. Rauch, R. F. Carswell, and K. Lanzetta. Evidence for structure in the H I column density distribution of QSO absorbers. MNRAS, 262: 499–505, May 1993.
- Michael Pippig and Daniel Potts. Parallel three-dimensional nonequispaced fast Fourier transforms and their application to particle simulation. SIAM J. Sci. Comput., 35:C411 – C437, 2013.
- Planck Collaboration, P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday, and et al. Planck 2013 results. XVI. Cosmological parameters. *ArXiv e-prints*, March 2013.
- Thomas Porter and Tom Duff. Compositing digital images. *SIGGRAPH Comput. Graph.*, 18(3):253–259, January 1984. ISSN 0097-8930. doi: 10.1145/964965.808606. URL http://doi.acm.org/10.1145/964965.808606.
- C. Power and A. Knebe. The impact of box size on the properties of dark matter haloes in cosmological simulations. MNRAS, 370:691–701, August 2006. doi: 10.1111/j.1365-2966.2006.10562.x.
- D. J. Price. splash: An Interactive Visualisation Tool for Smoothed Particle Hydrodynamics Simulations. PASA, 24:159–173, October 2007. doi: 10.1071/AS07022.
- D. J. Price. Modelling discontinuities and Kelvin Helmholtz instabilities in SPH. Journal of Computational Physics, 227:10040–10057, December 2008. doi: 10.1016/j.jcp.2008.08.011.
- D. J. Price. Smoothed particle hydrodynamics and magnetohydrodynamics. Journal of Computational Physics, 231:759–794, February 2012. doi: 10.1016/j.jcp.2010.12.011.
- R. S. Priddey, R. J. Ivison, and K. G. Isaak. Environments of z ¿ 5 quasars: searching for protoclusters at submillimetre wavelengths. MNRAS, 383:289–296, January 2008. doi: 10.1111/j.1365-2966.2007.12536.x.
- Michael J. Quinn. *Parallel Programming in C with MPI and OpenMP*. McGraw-Hill Education Group, 2003. ISBN 0071232656.

- M. Rauch. The Lyman Alpha Forest in the Spectra of QSOs. ARA&A, 36:267–316, 1998. doi: 10.1146/annurev.astro.36.1.267.
- M. Rauch, J. Miralda-Escudé, W. L. W. Sargent, T. A. Barlow, D. H. Weinberg, L. Hernquist, N. Katz, R. Cen, and J. P. Ostriker. The Opacity of the Ly $\alpha$  Forest and Implications for  $\Omega_b$  and the Ionizing Background. ApJ, 489:7–20, November 1997.
- J. I. Read, T. Hayfield, and O. Agertz. Resolving mixing in smoothed particle hydrodynamics. MNRAS, 405:1513–1530, July 2010. doi: 10.1111/j.1365-2966.2010.16577.x.
- B. A. Reid, W. J. Percival, D. J. Eisenstein, L. Verde, D. N. Spergel, R. A. Skibba, N. A. Bahcall, T. Budavari, J. A. Frieman, M. Fukugita, J. R. Gott, J. E. Gunn, Ž. Ivezić, G. R. Knapp, R. G. Kron, R. H. Lupton, T. A. McKay, A. Meiksin, R. C. Nichol, A. C. Pope, D. J. Schlegel, D. P. Schneider, C. Stoughton, M. A. Strauss, A. S. Szalay, M. Tegmark, M. S. Vogeley, D. H. Weinberg, D. G. York, and I. Zehavi. Cosmological constraints from the clustering of the Sloan Digital Sky Survey DR7 luminous red galaxies. MNRAS, 404:60–85, May 2010. doi: 10.1111/j.1365-2966.2010.16276.x.
- A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, R. P. Kirshner, B. Leibundgut, M. M. Phillips, D. Reiss, B. P. Schmidt, R. A. Schommer, R. C. Smith, J. Spyromilio, C. Stubbs, N. B. Suntzeff, and J. Tonry. Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. AJ, 116:1009–1038, September 1998. doi: 10.1086/300499.
- Jens Schneider Roland Fraedrich and Rüdiger Westermann. Exploring the "millennium run" - scalable rendering of large-scale cosmological datasets. *IEEE Trans*actions on Visualization and Computer Graphics (Proceedings Visualization / Information Visualization 2009), 15(6):1251–1258, November-December 2009.
- E. Romano-Diaz, I. Shlosman, M. Trenti, and Y. Hoffman. The Dark Side of QSO Formation at High Redshifts. ApJ, 736:66, July 2011. doi: 10.1088/0004-637X/736/1/66.
- V. C. Rubin, D. Burstein, W. K. Ford, Jr., and N. Thonnard. Rotation velocities of 16 SA galaxies and a comparison of Sa, Sb, and SC rotation properties. ApJ, 289: 81–98, February 1985. doi: 10.1086/162866.
- J. Schaye and C. Dalla Vecchia. On the relation between the Schmidt and Kennicutt-Schmidt star formation laws and its implications for numerical simulations. MN-RAS, 383:1210–1222, January 2008. doi: 10.1111/j.1365-2966.2007.12639.x.

- I. Schuessler and D. Schmitt. Comments on smoothed particle hydrodynamics. A&A, 97:373–379, April 1981.
- U. Seljak, A. Makarov, P. McDonald, S. F. Anderson, N. A. Bahcall, J. Brinkmann, S. Burles, R. Cen, M. Doi, J. E. Gunn, Ž. Ivezić, S. Kent, J. Loveday, R. H. Lupton, J. A. Munn, R. C. Nichol, J. P. Ostriker, D. J. Schlegel, D. P. Schneider, M. Tegmark, D. E. Berk, D. H. Weinberg, and D. G. York. Cosmological parameter analysis including SDSS Lyα forest and galaxy bias: Constraints on the primordial spectrum of fluctuations, neutrino mass, and dark energy. Phys. Rev. D, 71(10): 103515, May 2005. doi: 10.1103/PhysRevD.71.103515.
- U. Seljak, A. Slosar, and P. McDonald. Cosmological parameters from combining the Lyman-α forest with CMB, galaxy clustering and SN constraints. J. Cosmology Astropart. Phys., 10:014, October 2006. doi: 10.1088/1475-7516/2006/10/014.
- P. R. Shapiro and M. L. Giroux. Cosmological H II regions and the photoionization of the intergalactic medium. ApJ, 321:L107–L112, October 1987. doi: 10.1086/185015.
- P. R. Shapiro, I. T. Iliev, M. A. Alvarez, and E. Scannapieco. Relativistic Ionization Fronts. ApJ, 648:922–935, September 2006. doi: 10.1086/506242.
- Y. Shen, C. K. McBride, M. White, Z. Zheng, A. D. Myers, H. Guo, J. A. Kirkpatrick, N. Padmanabhan, J. K. Parejko, N. P. Ross, D. J. Schlegel, D. P. Schneider, A. Streblyanska, M. E. C. Swanson, I. Zehavi, K. Pan, D. Bizyaev, H. Brewington, G. Ebelke, V. Malanushenko, E. Malanushenko, D. Oravetz, A. Simmons, and S. Snedden. Cross-correlation of SDSS DR7 Quasars and DR10 BOSS Galaxies: The Weak Luminosity Dependence of Quasar Clustering at z ~ 0.5. ApJ, 778:98, December 2013. doi: 10.1088/0004-637X/778/2/98.
- M.-S. Shin, H. Trac, and R. Cen. Cosmological H II Bubble Growth during Reionization. ApJ, 681:756–770, July 2008. doi: 10.1086/588247.
- J. M. Shull and M. E. van Steenberg. X-ray secondary heating and ionization in quasar emission-line clouds. ApJ, 298:268–274, November 1985. doi: 10.1086/163605.
- D. Sijacki, V. Springel, T. Di Matteo, and L. Hernquist. A unified model for AGN feedback in cosmological simulations of structure formation. MNRAS, 380:877–900, September 2007. doi: 10.1111/j.1365-2966.2007.12153.x.
- D. Sijacki, V. Springel, and M. G. Haehnelt. Growing the first bright quasars in cosmological simulations of structure formation. MNRAS, 400:100–122, November 2009. doi: 10.1111/j.1365-2966.2009.15452.x.

- A. Slosar, A. Font-Ribera, M. M. Pieri, J. Rich, J.-M. Le Goff, É. Aubourg, J. Brinkmann, N. Busca, B. Carithers, R. Charlassier, M. Cortês, R. Croft, K. S. Dawson, D. Eisenstein, J.-C. Hamilton, S. Ho, K.-G. Lee, R. Lupton, P. Mc-Donald, B. Medolin, D. Muna, J. Miralda-Escudé, A. D. Myers, R. C. Nichol, N. Palanque-Delabrouille, I. Pâris, P. Petitjean, Y. Piškur, E. Rollinde, N. P. Ross, D. J. Schlegel, D. P. Schneider, E. Sheldon, B. A. Weaver, D. H. Weinberg, C. Yeche, and D. G. York. The Lyman-α forest in three dimensions: measurements of large scale flux correlations from BOSS 1st-year data. J. Cosmology Astropart. Phys., 9:001, September 2011. doi: 10.1088/1475-7516/2011/09/001.
- A. Sokasian, T. Abel, L. Hernquist, and V. Springel. Cosmic reionization by stellar sources: Population II stars. MNRAS, 344:607–624, September 2003. doi: 10.1046/j.1365-8711.2003.06847.x.
- V. Springel. The cosmological simulation code GADGET-2. MNRAS, 364:1105–1134, December 2005. doi: 10.1111/j.1365-2966.2005.09655.x.
- V. Springel and L. Hernquist. Cosmological smoothed particle hydrodynamics simulations: the entropy equation. MNRAS, 333:649–664, July 2002. doi: 10.1046/j.1365-8711.2002.05445.x.
- V. Springel and L. Hernquist. Cosmological smoothed particle hydrodynamics simulations: a hybrid multiphase model for star formation. MNRAS, 339:289–311, February 2003. doi: 10.1046/j.1365-8711.2003.06206.x.
- V. Springel, M. White, and L. Hernquist. Hydrodynamic Simulations of the Sunyaev-Zeldovich Effect(s). ApJ, 549:681–687, March 2001. doi: 10.1086/319473.
- V. Springel, T. Di Matteo, and L. Hernquist. Modelling feedback from stars and black holes in galaxy mergers. MNRAS, 361:776–794, August 2005a. doi: 10.1111/j.1365-2966.2005.09238.x.
- V. Springel, S. D. M. White, A. Jenkins, C. S. Frenk, N. Yoshida, L. Gao, J. Navarro, R. Thacker, D. Croton, J. Helly, J. A. Peacock, S. Cole, P. Thomas, H. Couchman, A. Evrard, J. Colberg, and F. Pearce. Simulations of the formation, evolution and clustering of galaxies and quasars. Nature, 435:629–636, June 2005b. doi: 10.1038/nature03597.
- J. A. Srbinovsky and J. S. B. Wyithe. Constraining the quasar contribution to the reionization of cosmic hydrogen. MNRAS, 374:627–633, January 2007. doi: 10.1111/j.1365-2966.2006.11173.x.
- A. R. H. Stevens, M. Martig, D. J. Croton, and Y. Feng. Where do galaxies end? A study of hydrodynamic-simulation galaxies and their integrated properties. ArXiv e-prints, April 2014.

- R. S. Sutherland and M. A. Dopita. Cooling functions for low-density astrophysical plasmas. ApJS, 88:253–327, September 1993. doi: 10.1086/191823.
- A. S. Szalay, T. Matsubara, and S. D. Landy. Redshift-Space Distortions of the Correlation Function in Wide-Angle Galaxy Surveys. ApJ, 498:L1–L4, May 1998. doi: 10.1086/311293.
- T. Szalay, V. Springel, and G. Lemson. GPU-Based Interactive Visualization of Billion Point Cosmological Simulations. *ArXiv e-prints*, November 2008.
- T. Tanaka and Z. Haiman. The Assembly of Supermassive Black Holes at High Redshifts. ApJ, 696:1798–1822, May 2009. doi: 10.1088/0004-637X/696/2/1798.
- T. L. Tanaka, M. Li, and Z. Haiman. The effect of baryonic streaming motions on the formation of the first supermassive black holes. MNRAS, 435:3559–3567, November 2013. doi: 10.1093/mnras/stt1553.
- M. Tegmark, M. A. Strauss, M. R. Blanton, K. Abazajian, S. Dodelson, H. Sandvik, X. Wang, D. H. Weinberg, I. Zehavi, N. A. Bahcall, F. Hoyle, D. Schlegel, R. Scoccimarro, M. S. Vogeley, A. Berlind, T. Budavari, A. Connolly, D. J. Eisenstein, D. Finkbeiner, J. A. Frieman, J. E. Gunn, L. Hui, B. Jain, D. Johnston, S. Kent, H. Lin, R. Nakajima, R. C. Nichol, J. P. Ostriker, A. Pope, R. Scranton, U. Seljak, R. K. Sheth, A. Stebbins, A. S. Szalay, I. Szapudi, Y. Xu, J. Annis, J. Brinkmann, S. Burles, F. J. Castander, I. Csabai, J. Loveday, M. Doi, M. Fukugita, B. Gillespie, G. Hennessy, D. W. Hogg, Ž. Ivezić, G. R. Knapp, D. Q. Lamb, B. C. Lee, R. H. Lupton, T. A. McKay, P. Kunszt, J. A. Munn, L. O'Connell, J. Peoples, J. R. Pier, M. Richmond, C. Rockosi, D. P. Schneider, C. Stoughton, D. L. Tucker, D. E. vanden Berk, B. Yanny, and D. G. York. Cosmological parameters from SDSS and WMAP. Phys. Rev. D, 69(10):103501, May 2004. doi: 10.1103/Phys-RevD.69.103501.
- A. Tenneti, R. Mandelbaum, T. Di Matteo, Y. Feng, and N. Khandai. Galaxy shapes and intrinsic alignments in the MassiveBlack-II simulation. MNRAS, 441:470–485, June 2014. doi: 10.1093/mnras/stu586.
- R. Teyssier. Cosmological hydrodynamics with adaptive mesh refinement. A new high resolution code called RAMSES. A&A, 385:337–364, April 2002. doi: 10.1051/0004-6361:20011817.
- R. M. Thomas and S. Zaroubi. Time-evolution of ionization and heating around first stars and miniqsos. MNRAS, 384:1080–1096, March 2008. doi: 10.1111/j.1365-2966.2007.12767.x.
- H. Trac and R. Cen. Radiative Transfer Simulations of Cosmic Reionization. I. Methodology and Initial Results. ApJ, 671:1–13, December 2007. doi: 10.1086/522566.

- H. Y. Trac and N. Y. Gnedin. Computer Simulations of Cosmic Reionization. Advanced Science Letters, 4:228–243, February 2011. doi: 10.1166/asl.2011.1214.
- M. J. Turk, B. D. Smith, J. S. Oishi, S. Skory, S. W. Skillman, T. Abel, and M. L. Norman. yt: A Multi-code Analysis Toolkit for Astrophysical Simulation Data. ApJS, 192:9, January 2011. doi: 10.1088/0067-0049/192/1/9.
- A. Venkatesan and A. Benson. X-rays and hard ultraviolet radiation from the first galaxies: ionization bubbles and 21-cm observations. MNRAS, 417:2264–2275, November 2011. doi: 10.1111/j.1365-2966.2011.19407.x.
- M. Viel, M. G. Haehnelt, and V. Springel. Inferring the dark matter power spectrum from the Lyman  $\alpha$  forest in high-resolution QSO absorption spectra. MNRAS, 354: 684–694, November 2004. doi: 10.1111/j.1365-2966.2004.08224.x.
- Z. Vlah, U. Seljak, T. Okumura, and V. Desjacques. Distribution function approach to redshift space distortions. Part V: perturbation theory applied to dark matter halos. J. Cosmology Astropart. Phys., 10:053, October 2013. doi: 10.1088/1475-7516/2013/10/053.
- M. Vogelsberger, S. Genel, D. Sijacki, P. Torrey, V. Springel, and L. Hernquist. A model for cosmological simulations of galaxy formation physics. MNRAS, 436: 3031–3067, December 2013. doi: 10.1093/mnras/stt1789.
- M. Volonteri and M. J. Rees. Rapid Growth of High-Redshift Black Holes. ApJ, 633: 624–629, November 2005. doi: 10.1086/466521.
- J. W. Wadsley, J. Stadel, and T. Quinn. Gasoline: a flexible, parallel implementation of TreeSPH. New A, 9:137–158, February 2004. doi: 10.1016/j.newast.2003.08.004.
- J. W. Wadsley, G. Veeravalli, and H. M. P. Couchman. On the treatment of entropy mixing in numerical cosmology. MNRAS, 387:427–438, June 2008. doi: 10.1111/j.1365-2966.2008.13260.x.
- C. Weniger. A tentative gamma-ray line from Dark Matter annihilation at the Fermi Large Area Telescope. J. Cosmology Astropart. Phys., 8:007, August 2012. doi: 10.1088/1475-7516/2012/08/007.
- M. White. The Zel'dovich approximation. MNRAS, 439:3630–3640, April 2014. doi: 10.1093/mnras/stu209.
- M. White, A. D. Myers, N. P. Ross, D. J. Schlegel, J. F. Hennawi, Y. Shen, I. McGreer, M. A. Strauss, A. S. Bolton, J. Bovy, X. Fan, J. Miralda-Escude, N. Palanque-Delabrouille, I. Paris, P. Petitjean, D. P. Schneider, M. Viel, D. H. Weinberg, C. Yeche, I. Zehavi, K. Pan, S. Snedden, D. Bizyaev, H. Brewington, J. Brinkmann, V. Malanushenko, E. Malanushenko, D. Oravetz, A. Simmons, A. Sheldon, and

B. A. Weaver. The clustering of intermediate-redshift quasars as measured by the Baryon Oscillation Spectroscopic Survey. MNRAS, 424:933–950, August 2012. doi: 10.1111/j.1365-2966.2012.21251.x.

- S. D. M. White, C. S. Frenk, and M. Davis. Clustering in a neutrino-dominated universe. ApJ, 274:L1–L5, November 1983. doi: 10.1086/184139.
- R. P. C. Wiersma, J. Schaye, T. Theuns, C. Dalla Vecchia, and L. Tornatore. Chemical enrichment in cosmological, smoothed particle hydrodynamics simulations. MN-RAS, 399:574–600, October 2009. doi: 10.1111/j.1365-2966.2009.15331.x.
- S. M. Wilkins, A. Bunker, W. Coulton, R. Croft, T. D. Matteo, N. Khandai, and Y. Feng. Interpreting the observed UV continuum slopes of high-redshift galaxies. MNRAS, 430:2885–2890, April 2013a. doi: 10.1093/mnras/stt096.
- S. M. Wilkins, W. Coulton, J. Caruana, R. Croft, T. D. Matteo, N. Khandai, Y. Feng, A. Bunker, and H. Elbert. Theoretical predictions for the effect of nebular emission on the broad-band photometry of high-redshift galaxies. MNRAS, 435:2885–2895, November 2013b. doi: 10.1093/mnras/stt1471.
- S. M. Wilkins, T. Di Matteo, R. Croft, N. Khandai, Y. Feng, A. Bunker, and W. Coulton. Confronting predictions of the galaxy stellar mass function with observations at high redshift. MNRAS, 429:2098–2103, March 2013c. doi: 10.1093/mnras/sts480.
- C. J. Willott, A. Omont, and J. Bergeron. Redshift 6.4 Host Galaxies of 10<sup>8</sup> Solar Mass Black Holes: Low Star Formation Rate and Dynamical Mass. ApJ, 770:13, June 2013. doi: 10.1088/0004-637X/770/1/13.
- J. Wurster and R. J. Thacker. A comparative study of AGN feedback algorithms. MNRAS, 431:2513–2534, May 2013. doi: 10.1093/mnras/stt346.
- J. S. B. Wyithe and A. Loeb. A characteristic size of ~10Mpc for the ionized bubbles at the end of cosmic reionization. Nature, 432:194–196, November 2004. doi: 10.1038/nature03033.
- J. S. B. Wyithe and A. Loeb. Smooth boundaries to cosmological HII regions from galaxy clustering. MNRAS, 374:960–964, January 2007. doi: 10.1111/j.1365-2966.2006.11201.x.
- X. Xu, A. J. Cuesta, N. Padmanabhan, D. J. Eisenstein, and C. K. McBride. Measuring  $D_A$  and H at z=0.35 from the SDSS DR7 LRGs using baryon acoustic oscillations. MNRAS, 431:2834–2860, May 2013. doi: 10.1093/mnras/stt379.
- N. Yoshida, A. Sokasian, L. Hernquist, and V. Springel. Early Structure Formation and Reionization in a Cosmological Model with a Running Primordial Power Spectrum. ApJ, 598:73–85, November 2003. doi: 10.1086/378852.

- K. Yoshikawa and T. Fukushige. PPPM and TreePM Methods on GRAPE Systems for Cosmological N-Body Simulations. PASJ, 57:849–860, December 2005. doi: 10.1093/pasj/57.6.849.
- O. Zahn, A. Lidz, M. McQuinn, S. Dutta, L. Hernquist, M. Zaldarriaga, and S. R. Furlanetto. Simulations and Analytic Calculations of Bubble Growth during Hydrogen Reionization. ApJ, 654:12–26, January 2007. doi: 10.1086/509597.

Joe Zuntz. Python binding of camb, 2013.