The Cosmic Stories:
Beginning, Evolution, and Present Days of the Universe

Thesis by

Dmitriy Tseliakhovich

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

1891

California Institute of Technology
Pasadena, California

2012
(Defended December 16, 2011)

To my mother, Elena Tseliakhovich

Acknowledgements

The greatest achievement of my life is getting to know and work with the incredible people–people who against all odds are achieving incredible things, people whose contribution to the society will be remembered for the generations to come, people who are making this world a better place every day. I am forever grateful for the honor of association that was bestowed on me by those people.

This work would never be possible without incredible support of my friend and advisor Professor Christopher Michael Hirata. His contribution to this thesis goes far beyond the conventional role of a Ph.D. advisor. First, our collaboration led to some extremely interesting and important results, which took me to the best universities and research institutions of the nation and opened extremely interesting opportunities for me as a researcher and space scientist. These results will be described in detail throughout this thesis. More importantly, this thesis got its chance to be written because Chris recognized my dreams and ambitions that reach far beyond the realm of astrophysics and cosmology into the world of commercial and social exploration of space. Chris was able to see the value of my ideas and facilitate my work on my own research while making sure that our joint projects were moving forward at an acceptable pace. On many occasions Chris went far beyond his normal responsibilities to help me in getting through some extremely challenging moments that resulted from my natural nonconformity and desire of getting things done faster and in my way. It goes beyond words how grateful I am for Chris' support, friendship, and for being the remarkable individual he is. It has been an incredible pleasure and honor to be working with Chris!

I am completing my thesis in less than 3.5 years from the beginning of my Ph.D. work. This is possible because of the tremendous pulling force of my dream to open space for large scale commercial, social, and scientific exploration and because of extraordinary help and support from my friends who share my ideas and vision. My special thanks go to Richard F. Schaden and Andy Lee Esq for recognizing the potential of my ideas and encouraging me every step of the way. There are no words to describe my gratitude to Shaun M. Maguire and Joseph T. Meyerowitz. Because of them I was able to live through the most challenging days in the past two years and move forward, knowing that there is no challenge that we will not be able to overcome together. Shaun and Joe exemplify the absolute climax of integrity, service to the nation, and unmatched personal standards that serve as a gleaming beacon in my pursuit of personal growth. I am deeply humbled and honored by the opportunity to be associated with these extraordinary individuals!

I would like to thank Professor Harry Atwater for all his help and support. Harry serves as a unique example of productivity, efficiency, and unmatched personal standards. Learning from him and observing his leadership style has been one of the most stimulating experiences of the past two years. Harry runs one of the most extended and broad operations at Caltech with more than 50 graduate students and postdocs. And yet, he found time to listen to my ideas and agreed to help me in building my vision and dream. This is the trait of a true visionary, and I am deeply indebted for his support. Through his work Harry is creating the future of American leadership in the fields of alternative energy, novel materials, and technology in general. I really hope that our future collaboration and joint research will result in exciting discoveries that will benefit Harry's group, Caltech and the nation.

Getting a doctoral degree from Caltech is a profound milestone in my life. In fact, it is a dream I have had for more than 10 years now. Nothing in my life would be the way it is, including the completion of this degree, without two giants whose impact on my life can never be properly assessed and to whom I am indebted beyond a slightest chance of ever returning the favor. My life has been forever changed when Dr. Mario Livio of the Space Telescope Science Institute first read my letter asking to accept me as a summer student at the Space Telescope Science Institute. Since then Mario has mentored me and helped me in getting to Caltech with his support, advice and recommendations. Everything I have achieved in the field of astrophysics started with his support, recognition and help. The second moment that changed the direction of my life was a 5-minute "pitch" to Dr. Peter Diamandis who became my mentor, my inspiration and a great role model for me.

I would like to extend my thanks to brilliant scientists who worked with me on multiple projects during my Ph.D. First of all, my special thanks go to Professor Avi Loeb who recognized and supported my ideas both in the field of astrophysics and commercial space exploration. Visiting Professor Loeb's Institute for Theory and Computation at Harvard-Smithsonian Center for Astrophysics was one of the most exciting and interesting experiences in my astrophysics career. I learned a lot from Avi on how to build, maintain and lead extraordinary research institutions – it was an outstanding learning experience. Moreover, Professor Loeb's help with getting time on the Odyssey Cluster at Harvard allowed me to complete my final Ph.D. project much faster than would otherwise be possible. The greatest lesson I learned from my collaboration with Avi comes from a clay tablet sitting in his office in Cambridge: "It is nice to be important, but it is more important to be nice". Professor Loeb is the very embodiment of this statement, and it is a great pleasure to know him!

Papers that I authored during my Ph.D. work would never come to life without help and invaluable contribution from my collaborators: Professor Rennan Barkana, with whom Chris and I have been working on the application of the relative velocity effect to the problem of the formation of the first galaxies, stars and minihalos; Professor Anje Slosar, who collaborated with me and Chris on a challenging topic of non-Gaussianity and the theory of inflation; Dr. Kevin Heng, who has been working with Chris and me on the topic of Balmer-dominated shocks and physics of supernovae remnants. It has been a great pleasure working and learning from these great scientists.

I would like to thank professor Richard Ellis for his support and encouragement throughout my Ph.D. work. Richard's lectures where the most enjoyable and educating part of my learning experience during the Ay120 series and later during the two semesters of my TA experience. Professor Ellis has an extraordinary talent of encouraging the best in people around him. Richard exudes proficiency and the vibe of a true astrophysicist, a person fully devoted to exploring the mysteries of our universe. Knowing Professor Ellis is one of the major reasons that makes the astrophysical community so special to me.

I would like to thank extremely efficient and professional people who helped me in dealing with logistical issues during my work at Caltech. It has been a phenomenal pleasure working with Shirley Hampton on getting all my endless receipts and traveling document in order. I greatly appreciate help rendered by Natalie Gilmore and her always supporting attitude. I would like to thank Chris Mach for the IT help and hardware support. I would also like to thank Tiffany Kimoto, for her support and for always finding a way to squeeze me into Harry Atwater's busy schedule.

I would like to thank all of my friends and peer students from the astrophysics department and specifically Ryan Trainor, Swarnima Monohar, Shriharsh Tendulkar, Tim Morton and Vera Gluscevic. It has been a tremendous pleasure to know you. As I have been saying over the past three years: “You guys are the real stars!”

My special thanks go to all my friends from the Singularity University and specifically to David Roberts and Dr. Dan Barry. It is a great honor to know you and have a chance to spend time learning from you! It is hard to overemphasize the importance of the SU family in my life, and I am looking forward to contributing my ideas and efforts to the development of the SU and helping it to fulfill its mission of positively impacting more than one billion people in the next ten years!

Finally, the most important “thank you” goes to my family and especially to my Mom, Elena Tseliakhovich. You made me who I am, you have been with me every step of the way, no matter how many miles lay between us. Your love and support kept me going through the toughest times of my life, and I am forever grateful for your love and unwaivering trust in me.

Abstract

This work presents three studies of independent astrophysical phenomena which cover a full timeline of the universe from the epoch of inflation to the present day. Along with our results we provide concise overviews of the considered phenomena and outline major open questions.

The first part of this work is focused on the epoch of inflation. We analyze the evolution of early density fluctuations which originate during inflation and connect physical fields driving inflation with observable parameters. We study several inflationary scenarios, specifically one field inflation, in which the only field present during that epoch is the inflaton field and two field inflation, in which along with the inflaton field the epoch of inflation is effected by the second scalar field–curvaton field. Single field inflationary models predict nearly Gaussian initial conditions and hence a detection of non-Gaussianity would be a signature of more complex inflationary scenarios. In this work we study the effect of primordial non-Gaussianity on the cosmic microwave background (CMB) and on large-scale structure in a two-field inflationary model in which both the inflaton and curvaton fields contribute to the primordial density fluctuations. We show that in addition to the previously described enhancement of the galaxy bias on large scales, this setup results in large-scale stochasticity. We provide joint constraints on the local non-Gaussianity parameter $\tilde{f}_{\rm NL}$ and the ratio ξ of the amplitude of primordial perturbations due to the inflaton and curvaton using WMAP and Sloan Digital Sky Survey (SDSS) data.

The second and largest part of this study is focused on the formation of the first cosmic structures and the effect of relative velocity between dark matter and baryonic fluids. In that part we discuss a very important and previously unnoticed effect which significantly changes the process of structure formation in the early universe. At the time of recombination, baryons and photons decoupled and the sound speed in the baryonic fluid dropped from relativistic, $\sim c/\sqrt{3}$, to the thermal velocity of the hydrogen atoms, $\sim 2\times10^{-5}c$. This is less than the relative velocity of baryons and dark matter computed via linear perturbation theory, so we infer that there are supersonic coherent flows of the baryons relative to the underlying potential wells created by the dark matter. As a result, the advection of small-scale perturbations (near the baryonic Jeans scale) by large-scale velocity flows is important for the formation of the first structures. This effect involves a quadratic term in the cosmological perturbation theory equations and hence has not been included in studies based on linear perturbation theory. We show that the relative motion suppresses the abundance of the first bound objects, even if one only investigates dark matter halos, and leads to qualitative changes in their spatial distribution, such as introducing scale-dependent bias and stochasticity. We further discuss possible observable implications of this effect for high-redshift galaxy clustering and reionization. Specifically we discuss in detail the effect of the relative velocity on the gas content in the early galaxies, minihaloes and the first stars. This part of the thesis also includes a concise overview of the recent studies that investigated various aspects of the relative velocity effect and showed its importance for topics ranging from star formation to precision cosmology.

The third and final part of the thesis covers interaction between expanding shocks of the supernovae explosions with the interstellar medium. The shocks of supernovae remnants represent a unique cosmic environment which allows detailed studies of plasma physics and high-energy astrophysics phenomena in conditions unreachable in the Earth-based laboratories. Specifically, shocks of supernovae remnants are associated with production of cosmic rays–the most energetic particles that we can observe. In our study we are specifically focused on the science of Balmer-dominated shocks (BDS)–a subset of collisionless, fast shocks dominated by hydrogen line emission with both broad and narrow components. The unique feature of BDS is that they are directly observable and their observations provide an opportunity for direct testing of the phase space structure and ion velocity distribution inside of shocks.

Understanding of physical phenomena occurring inside of astrophysical shocks requires precise knowledge of cross sections for high-nl proton-hydrogen collisions. Until now scientists have been using approximations for these cross sections which fall short of the precision needed for robust analysis of the observed data and can no longer satisfy needs of the astrophysical community. Guided by the demand in high-precision calculations of the cross sections we developed and implemented a robust method for direct solution of the Schroedinger partial differential equation on a grid. In this work we provide a detailed description of our computational algorithm for calculating cross sections in high-nl proton-hydrogen collisions and show results for $n \leq 4$. We describe the code we developed, show the results of consistency tests and describe possible extensions. Finally, we show how our results are applied to the studies of Balmer-dominated shocks and specifically how the precise cross sections for $n \leq 4$ can be used in computing Balmer decrement–the ratio of $H\alpha$ and $H\beta$ line intensities.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

"Space is big. Really big. You just won't believe how vastly, hugely, mind-bogglingly big it is..." Hitchhiker's Guide to the Galaxy.

1.1 Introduction

We are living in an exciting time for astrophysics and cosmology. New facilities have stimulated incredible discoveries and allowed us to see the Universe in unprecedented detail. With observational facilities like the Hubble Space Telescope and Keck Observatory we managed to peer into the time when the formation of the first stars and galaxies took place. With the results from COBE [1], WMAP [2, 3, 4], plethora of ground-based CMB projects like BICEP [5] and BOOMERANG [6], the 2dF/SDSS [7, 8] surveys and several high-redshift supernovae searches [9, 10, 11] we managed to put tight constraints on the cosmological parameters and develop a "Concordance Model" [12, 13] capable of providing a consistent framework for most of the observed cosmic phenomena. At the same time exponential growth of computing opened doors for development of advanced cosmological and astrophysical simulations (e.g., GADGET [14]) which in a short time progressed to the point where we can consistently simulate the distribution of matter and evolution processes on both horizon and galactic scales.

Yet, the current time is extremely challenging and in many ways pivotal for the field and for the astrophysics community. It is the time to realize the the old ways of doing things will no longer lead to the same results. Today, the only way to stay relevant and make progress is to constantly innovate and question the old assumptions and practices. It is becoming more and more obvious that the old funding model, old collaboration models and old ways of leading the research are not working any more. We are witnessing fundamental changes in the way scientific research is done and in the mechanisms through which the research is funded. NASA is undergoing a dramatic restructuring process which will most certainly reduce the total budget as well as the budget for fundamental research. National priorities require aligning scientific budgets with the needs of advancing technological superiority of the nation, and this should and will inevitably lead to significant restructuring of the funding mechanisms for the fundamental research fields such as astrophysics, cosmology and theoretical physics.

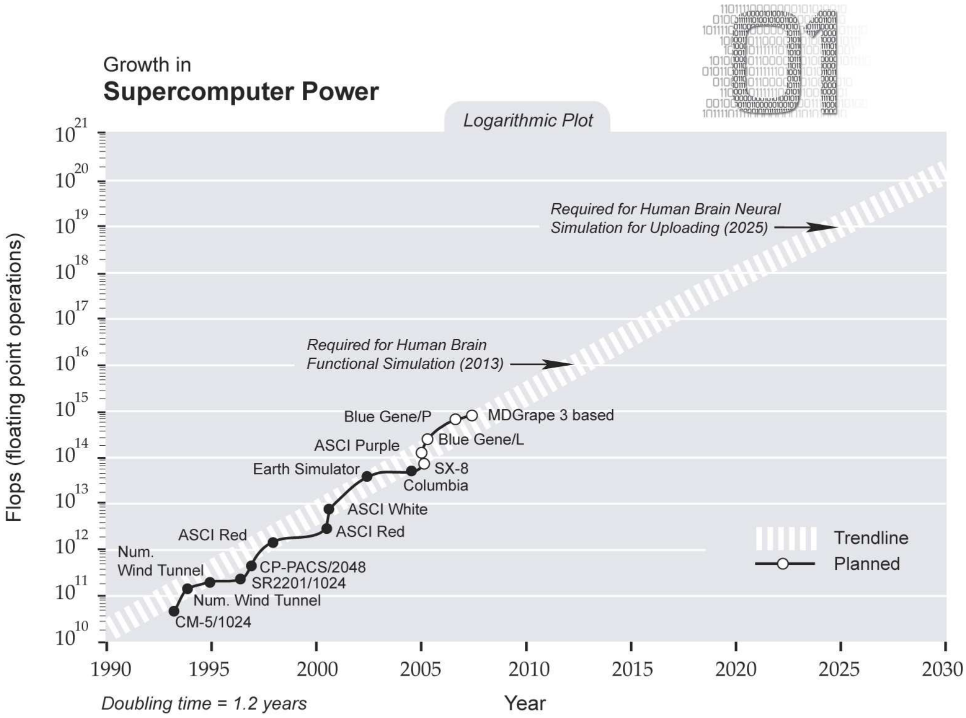


Figure 1.1: Exponential evolution of supercomputing power [15]. This plot shows explosive growth of the floating point operations performed by specific supercomputing clusters as a function of time and clearly emphasizes the trend in HPC evolution.

In the light of these changes it is extremely important to better utilize existing resources and to find new ways for leveraging exponential trends in evolution of information technologies and computing. Primarily, it is imperative to use explosive growth of high-performance computing (HPC) illustrated in figure 1.1 [15], and at the same time significantly bolster the effort on developing more effective and accurate theoretical models and algorithms. It has never been more important to think outside-the-box and come up with models which will allow continued progress in the light of rapidly changing logistical framework. It is also important to search for synergies between the needs of astrophysical community and broader needs of the nation. Such synergies clearly exist and will allow us to leverage technical expertise of scientists in addressing urgent socioeconomic needs. Furthermore, it is becoming more and more important to align the efforts of various communities of scientists that focus their efforts on far-reaching fundamental research. Specifically, the progress of fundamental science strongly depends on our ability to remove the artificial barriers between physics, astrophysics and applied mathematics establishing new mechanisms for efficient cross-pollination of ideas. In this thesis we provide three specific examples of the work that fully embraces this type of thinking and shows possible ways to implement new approaches and modes of operation.

We start by looking into the problem of the early Universe which provides a unique opportunity to bridge astrophysical observations and theoretical models with the models developed by the particle physics community. Our work is relevant to the progress of both astrophysics and particle physics and provides a great example of synergy between the two fields. Exploitation of this synergy is extremely important for a number of reasons. First, it provides an opportunity to attract funding by increasing scientific return of a given research project. Second, effective use of astrophysical observations is one of the very few possible ways for the particle physics community to make progress on the experimental front. It is clear that we are not going to see another large-scale particle physics collider in the coming decade, and hence, the early universe is the only way to study high-energy physics that goes beyond the energies accessible with the Large Hadron Collider (LHC). Finally, partnership between the two communities provides an outstanding opportunity for cross-pollination of ideas and development of new methods and tools that can only appear when the skill sets of the two groups are combined.

The very early universe is believed to have gone through a period of exponential expansion that rendered observable universe homogeneous, isotropic and flat (i.e., with vanishing curvature). While the detailed particle physics mechanism responsible for this exponential expansion, or inflation, is not known, the basic picture makes a number of predictions that have been confirmed by observation. It is believed that inflation was driven by a scalar particle or field called the inflaton. At the end of inflation inflaton decayed into lighter particles, producing extremely small matter density fluctuations with nearly Gaussian distribution (see [16] for a detailed review). This basic scenario, which was first introduced by Alan Guth [17] provides a viable explanation of the origin of our universe, however, it also raises a number of extremely interesting and important questions. Was inflaton the only particle present during inflation? How does the inflaton fit into the extended Standard Model of particle physics? Through what channels did inflaton decay into the particles (such as protons, neutrons, electrons, etc.), which comprise observable matter today?

These and many other related questions can be addressed only by a combination of astrophysical observations, theoretical work in cosmology, and theoretical work in particle physics. It would be extremely hard to reproduce the inflaton field in the laboratory, and currently available observational results are clearly insufficient to provide satisfactory answers. Making progress toward better understanding of the inflationary epoch requires robust analysis of multiple scenarios and utilization of all available data. In this work we are combining results from the large-scale structure (LSS) observations with the WMAP observations of the cosmic microwave background (CMB) to produce constraints on an inflationary scenario with two fields: inflaton, driving inflation, and curvaton, which contributes to the generation of the primordial density fluctuations. Our results extend existing constraints and provide important insights into future opportunities for the research in this field.

Our second project exhibits the importance of out-of-the-box thinking and is related to the formation of the first cosmic structures such as early galaxies and stars. In that work we critically examine existing ΛCDM model, which is known to work extremely well in most cases, and identify a weak point which comes from a set of unjustified assumptions. In correcting this weak point we introduce an important effect of relative velocity between dark matter and baryonic fluids which significantly changes our understanding of how the first structures form and evolve in the early universe. The relative velocity effect which originates at the epoch of recombination is nominally a second-order effect and thus has been consistently overlooked in the models based on the first-order perturbation theory. However, under careful examination, it appears that terms responsible for relative velocity become nonperturbative at scales where the first galaxies and stars form, and hence must be taken into account if accurate results are to be obtained. This relative velocity effect is currently actively studied by a large number of research teams in universities across the world, and its importance clearly emphasizes significance of critical reassessment of our ingrained assumptions and conventions.

Relative velocity effect is important for robust understanding of the processes of early structure formation. This topic represents an extremely exciting research direction highly prioritized by the astrophysics community. The first priority outlined in the Decadal Survey [18] is "understanding of how the first stars and galaxies formed and started to shine". Detailed and robust understanding of the mechanisms and processes which are effecting formation of the first halos, galaxies and stars is necessary for correct interpretation of data coming out of high-redshift surveys and observations studying the reionization epoch. The distribution of early objects is fundamentally important for correct extraction of cosmological parameters out of the studies that target matter distribution on various scales. Specifically, it is important for correct interpretation of the Baryonic Acoustic Oscillations (BAO) and observations of the Lyman_α forest. Furthermore, accurate knowledge of the matter distribution and evolution processes in the Universe at high redshifts ($z > 10$) is critical for interpretation of the upcoming 21 cm experiments such as Low Frequency Array (LOFAR) [19, 20], Murchinson Widefield Array (MWA) [21] and PAPER [22]. The results of this work may also turn out to be important for our attempts to understand properties of dark matter from analysis of the small-scale matter power spectrum [23, 24]. The research presented in this part of the thesis shows a clear example of the importance of subtle and often disguised effects for the future discoveries in this field.

Finally, in the third part of the thesis we utilized the power of supercomputing in addressing an extremely exciting problem of shocks that originate in the supernovae explosions and propagate into the interstellar medium (ISM). This work is yet another example of the synergy between physics and astrophysics. In this case, accurate measurements of the emission lines originating from the shocks allow detailed studies of the plasma physics phenomena which are extremely hard to reproduce in the Earth-based laboratories, yet are extremely interesting from the theoretical and experimental perspectives. In this study we combine the power of high performance computing (HPC) with the new computational algorithm to develop a solution to the problem of precision cross section calculations that was almost intractable a few years ago. The results obtained in our work will be instrumental for interpretation of the recent integral field unit (IFU) observations of the Balmer-dominated shocks (BDS) surrounding supernovae SN1006, as well as of the earlier observations of the BDS with the Hubble Space Telescope and a variety of ground-based telescopes.

The work presented in the third part of this thesis significantly extends the state of the art in precision calculations of high-nl proton-hydrogen collisions necessary for interpretation of the IFU spectrometry results and will dramatically bolster scientific output of that observational effort. Until recently most of the theoretical work in this field was based on very approximate cross sections and on the order of magnitude scaling relations. Our results will eliminate the need of using approximations and provide an important tool for analysis of physical processes inside of the supernovae shocks as well as in other astrophysical settings which require precise understanding of collisional processes between ionized and neutral hydrogen atoms. This work resulted in a large database of cross section open to scientific community as well as a code, `BDSCx`, which can be used for solving a variety of problems related to particle collisions.

1.2 Chapter Summaries

In chapter 2 we discuss physics of the very early Universe focusing on the theory of inflation and possible ways to constrain various inflationary parameters. We show various opportunities to study extremely challenging particle physics questions through astrophysical observations and discuss potential new discoveries that can come from observations of the microwave background radiation and large-scale distribution of matter. Next, in chapter 3, we consider the epoch of reionization and introduce the effect of relative velocity of dark matter and baryonic fluids. This important and previously unnoticed effect modifies the way first galaxies and stars form and introduces changes into spatial distribution of early structures. We show why this effect is important and why it was previously neglected by the astrophysics community. We further discuss consequences of the relative velocity effect and show specific examples where this effect is important from both theoretical and observational standpoints. Specifically, in chapter 4 we discuss the effect of relative velocity between dark matter and baryonic fluids on the formation of the first galaxies, minihaloes and stars. We also briefly discuss the plethora of follow-up papers that followed initial introduction of the relative velocity effect. In chapter 5 we are discussing the physics of astrophysical shocks and introduce a numerical formalism and associated code which allows precise calculations of the cross sections for proton-hydrogen collisions. We are discussing application of our results to the physics of Balmer-dominated shocks and emphasize the importance of our results for high-precision analysis of particle collisions inside of those shocks.

Chapter 2

The Nascent Universe*

2.1 Inflation and non-Gaussianity

One of the most important questions that cosmology faces today is the origin of structure in the universe. The generally accepted paradigm is that of inflation [25, 26, 27, 28], which is a period of the cosmic evolution during which the scale factor, which describes the size of the Universe, is growing exponentially. One of the major predictions of the theory of inflation is that it produces small adiabatic perturbations that evolve into the observed cosmic structure after inflation is over. The inflationary paradigm is extremely powerful as it remedies most of the problems of the original Big Bang scenario such as the flatness problem, the horizon problem, and the monopole problem [29]. The inflationary paradigm also has a set of other predictions that are well confirmed by current observations. On the other hand, although the generic predictions of inflation are quite clear, the nature of specific physical processes that govern inflation are still poorly understood.

The major obstacle in understanding inflation is that it cannot be directly observed either in the laboratory or with telescopes. This problem is at the same time a virtue of inflation as it allows to indirectly probe physics at energies and timescales that are far beyond the reach of current facilities. By comparing astrophysical observations with predictions of various inflationary models, one can expect to distinguish between different extensions of the Standard Model of particle physics [30]. Understanding of the reheating phase of the inflation, during which baryonic and dark matter particles are generated, can provide a link between scalar fields driving inflation and the observable Universe that consists of dark and baryonic matter. Studying various inflationary scenarios and their predictions opens a window into the theories beyond the Standard Model of particle physics and creates a unique, although indirect, opportunity to study particles and particle interactions at energies far exceeding those accessible at contemporary and near future collider facilities.

One of the many possible ways to deeper understand inflation is by studying the primordial density fluctuations. The usual inflationary model of a slowly rolling inflaton field requires that

---

*The material in this chapter was adapted from *Non-Gaussianity and large-scale structure in a two-field inflationary model*, D. Tseliakhovich, C. Hirata, and A. Slosar, *Phys. Rev. D 82, 043531 (2010)*. Reproduced here with the permission of the publisher, copyright (2010) by the American Physical Society. Additional material has been added into the introductory section of the chapter.

the perturbations are highly Gaussian [31, 32, 33, 34] and almost scale invariant. Therefore detection of non-Gaussianity in either the cosmic microwave background (CMB) spectrum or the large-scale structure (LSS) distribution would be a clear evidence that the physics driving inflation is more complicated than the standard inflaton scenario.

Non-Gaussianity naturally arises in inflationary models with more than one field [35, 36, 37, 38]. One of the most studied models is the curvaton model [38, 39, 40, 41, 42, 43], in which initial perturbations are generated by the curvaton field after inflation is over. In this model significant non-Gaussianity can be generated since the predicted curvature perturbation is proportional to the square of the curvaton field (as distinct from single-field inflation, where the required smoothness of the inflaton potential renders the curvature perturbation very nearly linear in the field fluctuations).

There are several important reasons for considering multi field inflationary models. The first one is a straightforward argument that today we are observing a universe comprised of multiple particles which are described by a variety of interacting fields. It is “natural” to assume that multiple fields were also present during the initial period of exponential expansion. This argument is supported by various extensions of the Standard Model of particle physics, such as supersymmetry, which allow multiple scalar fields to be present during inflation [44]. Another reason to explore theories that expand the simple inflaton scenario comes from still negative results of the experiments searching for gravitational wave signatures predicted by the simple inflaton model. It is predicted that tensor perturbations originating in the single-field inflation should produce gravitational waves which in turn imprint observable features into the cosmic microwave background (CMB) polarization maps [45, 46, 47]. The most recent observations of the CMB have not revealed the expected signal [48] and hence the search for viable alternative models is well justified.

Non-Gaussianity represents one of the key observables that could potentially discriminate among competing inflationary scenarios. Most attempts to constrain non-Gaussianity have used the so-called “local-type” or $f_{\rm NL}$ parameterization [49] in which one includes a quadratic term into the primordial potential, $\Phi = \phi + f_{\rm NL}\phi^2$. In this parametrization both linear and quadratic terms in the potential originate from the same Gaussian field, e.g., a curvaton field, and the contributions from perturbations in other fields (e.g., the inflaton field responsible for inflation itself) are negligible.

The signature of local-type non-Gaussianity in the CMB has been described at length [50]. It has also been established that $f_{\rm NL}$ has an effect on the galaxy bispectrum [51, 52, 53]. The effect on the large-scale galaxy power spectrum has been considered only recently [54, 55, 56, 57, 58], but it rapidly became clear that the method was competitive, stimulating work on N-body simulations of halo formation in non-Gaussian cosmologies [59, 60, 61, 62]. Recent constraints have been derived from the CMB bispectrum as measured by WMAP [63, 64, 65, 66, 67, 68, 69] and from large-scale structure in the Sloan Digital Sky Survey (SDSS) [55]. Recently, $\sim 3\sigma$ evidence for excess clustering consistent with non-Gaussianity has been identified in the NRAO

VLA Sky Survey (NVSS) [70].

In this study we extend the formalism to include the case where both the inflaton and curvaton contribute significantly to the curvature perturbation. Perturbations generated by the inflaton field are purely Gaussian, while curvaton fluctuations can result in non-Gaussianity if the conversion from curvaton fluctuation $\delta\sigma$ to primordial potential Φ contains quadratic terms. The ratio of inflaton to curvaton contributions ξ is arbitrary: the framework of the curvaton model allows it to take on any positive value. Usually one takes $\xi \gg 1$ since in the opposite limit ($\xi \ll 1$) the curvaton has no observable effect on the primordial perturbations. Here we investigate the consequences of general ξ–including values of order unity–for the CMB and LSS. The type of non-Gaussianity generated could be called "local-stochastic," in that it results from local nonlinear evolution of the inflaton and curvaton fields (and thus the primordial bispectrum will have the local-type configuration dependence), but that the full nonlinear potential Φ is not a deterministic function of the linear potential.

Studying non-Gaussianity is particularly important in the face of the current generation of CMB projects [45] such as the *Planck* satellite as well as ongoing and future LSS projects. To fully exploit the potential of the future probes, it is imperative to investigate theoretically the range of types of non-Gaussianities that can be produced in unconventional inflation (e.g., multi field models), and understand what effect they have on the CMB and LSS.

2.2 Non-Gaussian Initial Perturbations in Two-Field Inflationary Models

We consider a model of inflation where both the inflaton *and* the curvaton contribute to the primordial density perturbations. This configuration can exhibit a rich set of phenomenology, including both non-Gaussianity and various mixtures of adiabatic and isocurvature perturbations [71, 72, 73, 74, 75, 76]. In this study, we will restrict ourselves to the case where the dark matter decouples after the curvaton decays and its energy density is thermalized. This ensures that no dark matter isocurvature perturbation is produced, and the only observable perturbation is the curvature perturbation ζ that is conserved between curvaton decay and horizon entry.

The simplest case is that of two noninteracting scalar fields: the inflaton φ and the curvaton σ. The latter is taken to have a quadratic potential,

$$V_\sigma(\sigma) = \frac{1}{2}m^2\sigma^2. \tag{2.1}$$

During inflation, the inflaton dominates the energy density of the Universe, whereas the curvaton is effectively massless ($m \ll H$) and pinned by Hubble friction to a fixed value $\bar{\sigma}$ (aside from perturbations to be described later). Quantum fluctuations generate a spectrum of perturbations $\delta\varphi$ and $\delta\sigma$ in both the inflaton and curvaton fields:

$$P_{\delta\varphi}(k) = \frac{H_*^2}{2k^3}, \qquad \text{and} \qquad P_{\delta\sigma}(k) = \frac{H_*^2}{2k^3}, \tag{2.2}$$

where H_* is the Hubble rate evaluated at the horizon crossing for a given mode, i.e., when $k = aH$, and the post-horizon-exit field perturbations are defined on the uniform total density slice. The φ and σ perturbations are nearly Gaussian and uncorrelated.

The inflaton perturbation is parallel to the unperturbed trajectory in (φ, σ)-space and hence is an adiabatic perturbation; indeed it behaves the same way that perturbations behave in single-field inflation. The curvaton perturbation however is an isocurvature perturbation and can have complicated dynamics. In the simplest version of the curvaton scenario, the curvaton begins to oscillate after the end of inflation when the Hubble rate drops to $H \sim m$. As a massive scalar with zero spatial momentum, its energy density subsequently redshifts as $\rho_\sigma \propto a^{-3}$. Of interest to us is the fact that for quadratic potentials (equation 2.1) this energy density is also proportional to the square of the curvaton field $\sigma = \bar{\sigma} + \delta\sigma$, i.e.,

$$\delta_\sigma \equiv \frac{\delta\rho_\sigma}{\bar{\rho}_\sigma} = 2\frac{\delta\sigma}{\bar{\sigma}} + \frac{\delta\sigma^2 - \langle\delta\sigma^2\rangle}{\bar{\sigma}^2}, \tag{2.3}$$

where the subtraction of the variance arises from the $\mathcal{O}(\delta\sigma^2)$ expansion of $\bar{\rho}_\sigma \propto \bar{\sigma}^2 + \langle\delta\sigma^2\rangle$. The quadratic term allows the curvaton to generate a non-Gaussian density perturbation. In the radiation-dominated era, the curvaton's contribution to the energy density increases as $\Omega_\sigma \propto a$, thereby enhancing the importance of δ_σ. The decay of the curvaton and the thermalization of its energy density result in a non-Gaussian adiabatic perturbation.

The δN formalism [77] extended into the nonlinear regime [34] quantitatively provides the curvature perturbation to second-order in the field perturbations $(\delta\varphi, \delta\sigma)$; this is (equation 26 of Langlois et al. [75]),

$$\zeta = -\frac{H_*\delta\varphi}{\dot{\varphi}_*} + \frac{2r}{3}\frac{\delta\sigma}{\sigma_*} + \frac{2r}{9}\left(\frac{3}{2} - 2r - r^2\right)\frac{\delta\sigma^2}{\sigma_*^2}, \tag{2.4}$$

where the subscript $_*$ denotes evaluation at horizon exit, and r is related to the fraction of the energy density in the curvaton when it decays:

$$r = \left.\frac{3\rho_\sigma}{4\rho_{\mathrm{rad}} + 3\rho_\sigma}\right|_{\mathrm{decay}}. \tag{2.5}$$

The primordial potential perturbation in the Newtonian gauge is then given by the usual expression $\Phi = \frac{3}{5}\zeta$ (e.g., [78]; but note that in large-scale structure non-Gaussianity studies the opposite sign convention is adopted, so that $\Phi > 0$ corresponds to overdensities).

We may put the primordial potential in a form more closely related to that of large-scale structure non-Gaussianity studies:

$$\Phi(\mathbf{x}) = \phi_1(\mathbf{x}) + \phi_2(\mathbf{x}) + \tilde{f}_{\mathrm{NL}}[\phi_2^2(\mathbf{x}) - \langle\phi_2^2\rangle], \tag{2.6}$$

where ϕ_1 and ϕ_2 are the parts of the linear primordial potential corresponding to the inflaton and curvaton fields respectively. Their power spectra are given by

$$\frac{k^3}{2\pi^2}P_{\phi_1}(k) = \frac{9}{25}\left(\frac{H_*^2}{2\pi\dot{\varphi}_*}\right)^2, \tag{2.7}$$

and

$$\frac{k^3}{2\pi^2}P_{\phi_2}(k) = \frac{4r^2}{25}\left(\frac{H_*}{2\pi\sigma_*}\right)^2. \tag{2.8}$$

The non-Gaussianity parameter is

$$\tilde{f}_{\rm NL} = \frac{5}{6r}\left(\frac{3}{2} - 2r - r^2\right). \tag{2.9}$$

(We use the tilde since the label "$f_{\rm NL}$" is usually used to denote the non-Gaussianity parameter appearing in the primordial bispectrum.)

It is convenient to specify the relative contribution of the inflaton and curvaton fields to the primordial potential using the ratio of standard deviations $\xi = \sigma(\phi_1)/\sigma(\phi_2)$. Thus a fraction $\xi^2/(1+\xi^2)$ of the power comes from the inflaton and a fraction $1/(1+\xi^2)$ from the curvaton. This ratio is

$$\xi(k) = \left|\frac{3\sigma_* H_*}{2r\dot{\phi}_*}\right| = \frac{3}{2r}\left|\frac{\sigma_*}{(d\phi/dN)_*}\right|, \tag{2.10}$$

where N is the number of e-folds remaining in inflation. Thus the observable features of this model are specified by the primordial power spectrum $P_\Phi(k)$ and by the two new parameters $\tilde{f}_{\rm NL}$ and ξ (in principle ξ will have a scale dependence $d\ln\xi/d\ln k$ of order the slow roll parameters, but unless non-Gaussianity is detected at high statistical significance this cannot be measured). We will work with these parameters from here forward.

2.3 The CMB Bispectrum

The effect of local-type non-Gaussianity on the CMB bispectrum has a long history, both in purely adiabatic models as considered here, and in locally non-Gaussian isocurvature models [79, 80]. We evaluate the CMB bispectrum using our set of parameters here.

The CMB constraints on primordial non-Gaussianity come from the measurements of the CMB angular bispectrum [49],

$$B^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3} \equiv \langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle, \tag{2.11}$$

where $a_{\ell m}$ is the CMB temperature fluctuation expanded in spherical harmonics:

$$a_{\ell m} \equiv \int d^2\hat{\mathbf{n}} \frac{\Delta T(\hat{\mathbf{n}})}{\bar{T}} Y^*_{\ell m}(\hat{\mathbf{n}}). \tag{2.12}$$

If the primordial fluctuations are adiabatic scalar fluctuations, then $a_{\ell m}$ can be easily expressed in terms of the primordial potential Φ and the radiation transfer function $g_\ell(k)$:

$$a_{\ell m} = 4\pi(-i)^\ell \int \frac{d^3\mathbf{k}}{(2\pi)^3} \Phi(\mathbf{k}) g_\ell(k) Y^*_{\ell m}(\hat{\mathbf{k}}). \tag{2.13}$$

From equation (2.6) it follows that in the Fourier space primordial potential can be decomposed into parts associated with the linear potential perturbations ϕ_1 and ϕ_2, and with the nonlinear coupling $\tilde{f}_{\rm NL}$:

$$\Phi(\mathbf{k}) = \phi_1(\mathbf{k}) + \phi_2(\mathbf{k}) + \phi_{NL}(\mathbf{k}). \tag{2.14}$$

Here $\phi_{NL}(\mathbf{k})$ is the $\tilde{f}_{\rm NL}$-dependent part,

$$\phi_{NL}(\mathbf{k}) \equiv \tilde{f}_{\rm NL} \int \frac{d^3\mathbf{p}}{(2\pi)^3} \phi_2(\mathbf{k}+\mathbf{p}) \phi_2^*(\mathbf{p}). \tag{2.15}$$

Using Wick's theorem we calculate the bispectrum of the total potential, $B_\Phi(k_1, k_2, k_3)$. It contains one contribution from allowing each of the $\Phi(\mathbf{k}_i)$ to have a contribution from $\tilde{f}_{\rm NL}$; in the case where this is k_3:

$$\begin{gathered} \langle [\phi_1(\mathbf{k}_1) + \phi_2(\mathbf{k}_1)][\phi_1(\mathbf{k}_2) + \phi_2(\mathbf{k}_2)] \phi_{NL}(\mathbf{k}_3) \rangle = \\ 2(2\pi)^3 \delta^{(3)}(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) \frac{\tilde{f}_{\rm NL}}{(1+\xi^2)^2} P_\Phi(k_1) P_\Phi(k_2), \end{gathered} \tag{2.16}$$

where $P_\Phi(k)$ is the power spectrum of the total potential. The total bispectrum is the sum of this and the similar contributions where $\tilde{f}_{\rm NL}$ contributes to $\mathbf{k}_1$ and to $\mathbf{k}_2$:

$$\begin{aligned} B_\Phi(k_1, k_2, k_3) &= 2 \frac{\tilde{f}_{\rm NL}}{(1+\xi^2)^2} [P_\Phi(k_1) P_\Phi(k_2) + \\ & \qquad P_\Phi(k_1) P_\Phi(k_3) + P_\Phi(k_2) P_\Phi(k_3)]. \end{aligned} \tag{2.17}$$

For constraining non-Gaussianity it is convenient to introduce two new variables: $x_1 = \tilde{f}_{\rm NL}/(1+\xi^2)^2$ and $x_2 = 1/(1+\xi^2)$. Using the bispectrum of $\Phi(\mathbf{k})$, we can finally write the CMB angular bispectrum (via a calculation similar to reference [49]):

$$\begin{aligned} B^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3} &= 2\mathcal{G}^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3} \int_0^\infty r^2 dr \\ & \times [b^L_{\ell_1}(r) b^L_{\ell_2}(r) b^{NL}_{\ell_3}(r) + b^L_{\ell_1}(r) b^{NL}_{\ell_2}(r) b^L_{\ell_3}(r) + b^{NL}_{\ell_1}(r) b^L_{\ell_2}(r) b^L_{\ell_3}(r)], \end{aligned} \tag{2.18}$$

where $\mathcal{G}^{m_1 m_2 m_3}_{\ell_1 \ell_2 \ell_3}$ is the Gaunt integral, and we use

$$b^L_\ell(r) \equiv \frac{2}{\pi} \int_0^\infty k^2 dk \, P_\Phi(k) g(k) j_\ell(kr) \tag{2.19}$$

and

$$b^{NL}_\ell(r) \equiv \frac{2x_1}{\pi} \int_0^\infty k^2 dk \, g(k) j_\ell(kr). \tag{2.20}$$

We see from equation (2.17) that CMB bispectrum measurements of $f_{\rm NL}^{\rm loc}$ that assume a purely curvaton contribution ($\xi = 0$) are actually measuring x_1 in the more general case.

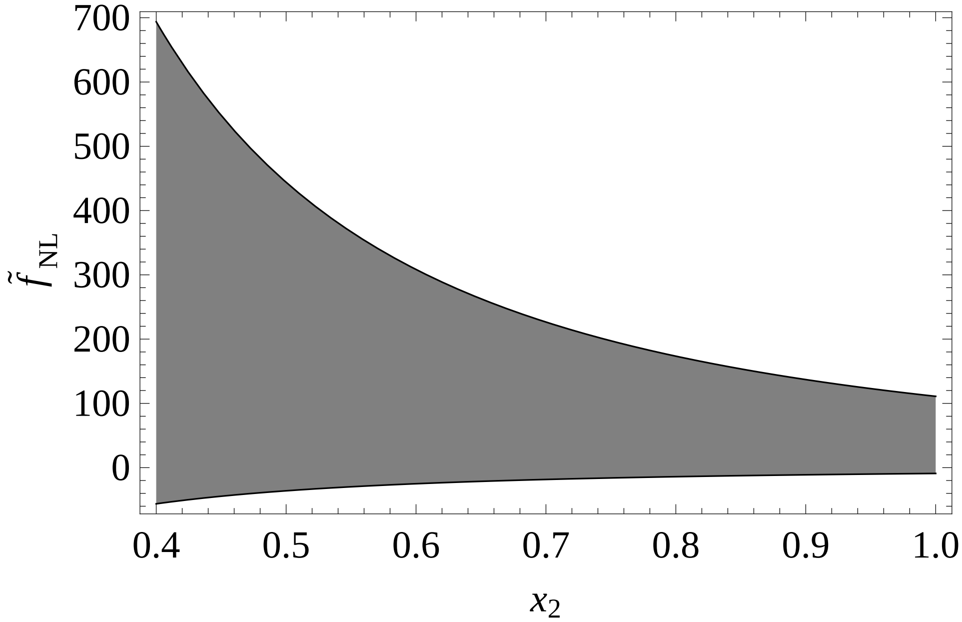


Figure 2.1: The allowed range (2σ) of $\tilde{f}_{\mathrm{NL}}$ as a function of x_2 derived from the WMAP data [66]. As discussed in the text, $\tilde{f}_{NL}$ becomes unconstrained as $x_2 \to 0$ because in this case the statistics describing the density distribution is dominated by the inflaton field and is nearly Gaussian.

Clearly, when the curvaton field dominates the perturbation power ($x_2 \approx 1$), we have $f_{\mathrm{NL}} = \tilde{f}_{\mathrm{NL}}$, and our $\tilde{f}_{\mathrm{NL}}$ constraints are identical to the constraints on models with negligible inflaton perturbation. However, as the contribution from the inflaton field increases ($x_2 \to 0$) the role of the contribution of the curvaton field to the primordial curvature perturbation becomes negligible and the statistics describing the density distribution becomes very nearly Gaussian. In that case $\tilde{f}_{\mathrm{NL}}$ becomes completely unconstrained (figure 2.1) and we can no longer make robust predictions regarding the presence of the second field on the basis of the non-Gaussianity observations alone. The CMB bispectrum is therefore not capable of breaking the degeneracy between $\tilde{f}_{\mathrm{NL}}$ and ξ. The use of other constraints is necessary. In this study, we will use large-scale structure, although we note that in principle the CMB trispectrum might also be useful for this purpose, since it scales as $\tilde{f}_{\mathrm{NL}}^2/(1+\xi^2)^3 = x_1^2/x_2$ and thus in combination with the bispectrum would allow the $(\tilde{f}_{\mathrm{NL}}, \xi)$ degeneracy to be broken.

2.4 Local non-Gaussianity in Peak-Background Split Formalism

In this section we outline the inflationary scenario in which both inflaton and curvaton fields are contributing to the initial curvature perturbation and derive the power spectrum of galaxies using the peak-background split formalism [71].

We decompose the density field into a long-wavelength and short-wavelength pieces:

$$\rho(\mathbf{x}) = \bar{\rho}(1 + \delta_l + \delta_s). \tag{2.21}$$

The linear density field here is a sum of two independent Gaussian components $\delta = \delta^{(1)} + \delta^{(2)}$ originating from the inflaton and curvaton fields, respectively.

The local Lagrangian density of halos now depends on the large scale matter perturbations of both Gaussian fields:

$$n(\mathbf{x}) = \bar{n}[1 + b_1\delta_l^{(1)} + b_2\delta_l^{(2)}]. \tag{2.22}$$

We will see that b_i can be scale-dependent (k-dependent), which means that in position space it should be thought of as a (possibly nonlocal) operator acting on the density field, i.e., $b_2\delta_l^{(2)}(\mathbf{x})$ is the convolution of $\delta_l^{(2)}(\mathbf{x})$ with the Fourier transform of $b_2(\mathbf{k})$.

We can express the bias parameters in terms of the number density function

$$b_i = \bar{n}^{-1}\frac{\partial n}{\partial \delta_l^{(i)}}. \tag{2.23}$$

It is easy to check that the bias for the δ_1 field is just the usual Lagrangian bias that applies to Gaussian cosmologies; for example, in the Press-Schechter model [81] it is $b_1 = b_g \equiv \delta_c/\sigma_\delta^2 - \delta_c^{-1}$ with $\delta_c = 1.686$ quantifying the spherical collapse linear over-density. To calculate b_2 we note that short-wavelength modes δ_s in an overdense region determined by δ_l can be written as

$$\delta_s = \alpha\left[(1 + 2\tilde{f}_{\rm NL}\phi_l^{(2)})\phi_s^{(2)} + \tilde{f}_{\rm NL}(\phi_s^{(2)})^2 + \phi_s^{(1)}\right] \equiv \alpha\left[X_1\phi_s^{(2)} + X_2(\phi_s^{(2)})^2 + \phi_s^{(1)}\right], \tag{2.24}$$

where $X_1 = 1 + 2\tilde{f}_{\rm NL}\phi_l^{(2)}$ and $X_2 = \tilde{f}_{\rm NL}$. Here α is the transfer function that converts the potential into the density field, $\delta(k) = \alpha(k)\Phi(k)$. In general one may think of α as an operator defined by its action in Fourier space, i.e., when applied to a real-space function such as $\phi(\mathbf{x})$, we have

$$\alpha\phi(\mathbf{x}) \equiv \int \frac{d^3\mathbf{k}}{(2\pi)^3}\alpha(k)e^{i\mathbf{k}\cdot\mathbf{x}}\int d^3y\, e^{-i\mathbf{k}\cdot\mathbf{y}}\phi(\mathbf{y}). \tag{2.25}$$

The specific function $\alpha(k)$ is given by equation (7) of Slosar et al. [55]:

$$\alpha(k;z) = \frac{2c^2k^2T(k)D(z)}{3\Omega_m H_0^2}, \tag{2.26}$$

where $T(k)$ is the linear transfer function with conventional normalization $T(0) = 1$, and $D(a)$ is the growth function normalized to $D(z) = (1+z)^{-1}$ at high redshift. The inverse operator α^{-1} is obtained by the replacement $\alpha(k) \to \alpha^{-1}(k) = 1/\alpha(k)$.

This shows that local number density in the non-Gaussian case depends not only on δ_l, but also on X_1, X_2, and hence b_2 becomes

$$b_2 = \bar{n}^{-1}\left[\frac{\partial n}{\partial \delta_l^{(2)}(\mathbf{x})} + 2\tilde{f}_{\rm NL}\alpha^{-1}\frac{\partial n}{\partial X_1}\right], \tag{2.27}$$

where the derivative is taken at the mean value $X_1 = 1$.

We can further simplify this expression by considering a rescaling of the power spectrum on the small scales due to the presence of non-Gaussianity. In a "local" region of some size R, and for small-scale Fourier modes $k \gg R^{-1}$ within this region, there is a local power spectrum

$$\begin{aligned} P_s^{\rm local}(k) &= \frac{\xi^2 + (1 + 2\tilde{f}_{\rm NL}\alpha^{-1}\delta_l^{(2)})^2}{1+\xi^2} P_s^{\rm global}(k) \\ &\equiv X_0 P_s^{\rm global}(k), \end{aligned} \tag{2.28}$$

from which we obtain the rescaling of σ_8:

$$\sigma_8^{\rm local} = \sigma_8 \sqrt{X_0}. \tag{2.29}$$

Using these expressions we can change the derivatives in equation (2.27) to finally obtain

$$b_2(k) = b_{\rm g} + \frac{2\tilde{f}_{\rm NL}}{1+\xi^2}\alpha^{-1}(k)\frac{\partial \ln n}{\partial \ln \sigma_8^{\rm local}}. \tag{2.30}$$

For further calculations we assume the mass function to be universal, i.e., we assume that it depends only on the significance function $\nu(M) \equiv \delta_{\rm c}^2/\sigma^2(M)$:

$$n(M, \nu) = M^{-2}\nu f(\nu)\frac{{\rm d}\ln \nu}{{\rm d}\ln M}. \tag{2.31}$$

This assumption is much more general than the Press-Schechter picture, e.g., it holds for the Sheth-Tormen mass function [82] as well. Universality implies that

$$\frac{\partial \ln n}{\partial \ln \sigma_8^{\rm local}} = \delta_{\rm c} b_{\rm g}, \tag{2.32}$$

from which we derive

$$b_2(k) = b_{\rm g} + \frac{2\delta_{\rm c}\tilde{f}_{\rm NL}}{1+\xi^2} b_{\rm g}\alpha^{-1}(k). \tag{2.33}$$

The standard Gaussian bias in Eulerian space is given by $b \equiv b_{\rm g} + 1$. The halo overdensity in Eulierian space in the non-Gaussian case is then obtained by multiplying equation (2.22) by $1 + \delta_l$; to first order,

$$\begin{aligned} \delta_{\rm h} &\equiv \frac{n(\mathbf{x})}{\bar{n}} - 1 = \delta_l + b_1\delta_l^{(1)} + b_2\delta_l^{(2)} \\ &= [1 + b_1(k)]\delta_l^{(1)} + [1 + b_2(k)]\delta_l^{(2)}. \end{aligned} \tag{2.34}$$

We can now write down the halo power spectrum in the form:

$$P_{\rm hh}(k) = \frac{(1+b_1)^2\xi^2 + [1 + b_2(k)]^2}{1+\xi^2} P^{\rm lin}(k). \tag{2.35}$$

Finally, plugging in $b_1(k) = b$ and using equation (2.33) for $b_2(k)$, we obtain

$$P_{\mathrm{hh}}(k) = \frac{\xi^2 b^2 + [b + 2(b-1)\tilde{f}_{\mathrm{NL}}\delta_c(1+\xi^2)^{-1}\alpha^{-1}(k)]^2}{1+\xi^2} P^{\mathrm{lin}}(k). \tag{2.36}$$

In the limit of $\xi \to 0$, i.e., when the contribution from the inflaton field is negligible we recover the standard expression for the power spectrum with the curvaton generated non-Gaussianity (equation 32 of [55]).

(It should be noted that the non-Gaussianity also introduces small corrections to the scale-independent part of the bias, because the small-scale fluctuations that must collapse to form a massive halo have a non-Gaussian density distribution. This effect has been seen in simulations with f_{NL}-type non-Gaussianity, where it is negative for $f_{\mathrm{NL}} > 0$, resulting in a slight reduction of the non-Gaussian bias enhancement at large k, and even a change in sign of the f_{NL} effect at very small scales [59, 61]. However, since current studies of non-Gaussianity using LSS allow the scale-independent bias to be a free parameter, they are not sensitive to this effect; it would only be important if the Gaussian contribution to the bias were inferred independently, e.g., from measurements of halo mass and the mass-bias relation.)

A further consequence of this model that does not arise in the case with only curvaton perturbations is large-scale stochasticity. In particular, the squared correlation coefficient,

$$\chi(k) = \frac{P^2_{\mathrm{hm}}(k)}{P_{\mathrm{hh}}(k)P_{\mathrm{mm}}(k)}, \tag{2.37}$$

deviates from unity on the largest scales. We can see this by writing the cross power spectrum $P_{\mathrm{hm}}(k)$ as

$$P_{\mathrm{hm}}(k) = \frac{(1+b_1)\xi^2 + [1+b_2(k)]}{1+\xi^2} P^{\mathrm{lin}}(k). \tag{2.38}$$

In the linear Gaussian theory, one would have $\chi = 1$, whereas in our case we have

$$\chi(k) = \frac{\{(1+b_1)\xi^2 + [1+b_2(k)]\}^2}{(1+\xi^2)\{(1+b_1)^2\xi^2 + [1+b_2(k)]^2\}}. \tag{2.39}$$

Note that if $\tilde{f}_{\mathrm{NL}} \neq 0$, on large scales $|b_2(k)| \gg 1, b_1$ and hence

$$\lim_{k\to 0^+} \chi(k) = \frac{1}{1+\xi^2} = x_2. \tag{2.40}$$

An example of the onset of scale-dependent bias and stochasticity is shown in figure 2.2; note that this type of stochasticity effect exists only for $x_2 \neq 1$.

It is important to note that stochasticity can arise even for $x_2 = 1$ in two ways. One is that on small scales, there is a breakdown of linear biasing. However, since our constraints on non-Gaussianity arise from the largest scales in the survey (mainly the $l < 25$ quasar data points) this effect can be neglected. The other is halo shot noise (e.g., [83]), which arises from the fact that halos containing multiple galaxies (or quasars) can produce a large "1-halo" contribution to the correlation function at small separations. When transformed to Fourier space at large scales (small k), this results in additional contribution to the power spectrum of

$$\lim_{k \to 0^+} P_{1\ \mathrm{halo}}(k) = \int 4\pi r^2 \xi_{1\ \mathrm{halo}}(r)\, dr, \tag{2.41}$$

where $\xi_{1\ \mathrm{halo}}(r)$ is one-halo correlation function. In principle since $P(k) \propto k$ on large scales, the halo shot noise term ($\propto k^0$) may become important; since it is random and not determined by the underlying long-wavelength modes of the density field, it also produces stochasticity.

However, the halo shot noise is expected to be a very small contribution for our data sets. A simple way to see this is to note that the ratio of the halo shot noise to the usual shot noise is equal to twice the ratio of 1-halo pairs to the number of galaxies (this follows from equation (2.41) and the definition of the correlation function). For the quasar sample, a simple counts-in-pixels analysis of the catalog suggests that 0.6% of the quasars are in pairs (the Healpix pixels [84] used are 3.5 arcmin in size, i.e., much larger than haloes at $z > 1$), suggesting that the halo shot noise term is $C_{l,1\mathrm{halo}} \sim 0.012/\bar{n}_{2\mathrm{D}}$. This is two orders of magnitude smaller than the error bars on the lowest-l quasar autopower point displayed in Slosar et al. [55] and hence negligible.

2.5 Galaxy Angular Power Spectrum

Additional constraints on the primordial non-Gaussianity come from the observations of the Large Scale Structure (LSS), and in particular from large galaxy surveys such as SDSS. These constraints are primarily driven by extremely large scales, corresponding to wave numbers $k < 0.01$ Mpc^{-1}. The data sets used for LSS constraints include spectroscopic and photometric Luminous Red Galaxies (LRGs) from SDSS, photometric quasars from SDSS and cross-correlation between galaxies and dark matter via Integrated Sach-Wolfe effect. Detailed description of the data used for LSS constraints can be found in [85, 55] and here we will only emphasize the redshift ranges of the most important data sets we used.

The photometric LRGs data set was constructed and discussed in detail in [86]. The sample was sliced into two redshift bins with $0.2 < z < 0.4$ and $0.4 < z < 0.6$. Our spectroscopic LRG power spectrum comes from [87] and is based on a galaxy sample that covers 4000 square degrees of sky over the redshift range $0.16 < z < 0.47$. Quasars for our constraints were photometrically selected from the SDSS quasar catalog consisting of UVX objects [88] and DR3 catalog [89]. Quasars in our sample fall into two redshift bins with $0.65 < z < 1.45$ and $1.45 < z < 2.0$.

The tightest LSS constraints on primordial non-Gaussianity involve purely photometric samples (where one observes the two-dimensional projection of the galaxy distribution with a set of color cuts) since this allows the largest volume to be probed with the highest number density. At low k, where the effects of primordial non-Gaussianity on the power spectrum are largest, there is less of an advantage to having the large number of modes ($\propto k_{\mathrm{max}}^3$ instead of k_{max}^2) achievable via spectroscopy. In order to obtain constraints on the $\tilde{f}_{\mathrm{NL}}$ and ξ parameters from LSS we need to modify equation (2.36) to give the angular power spectrum.

To obtain the angular power spectrum, we project the galaxy density field $\delta_{g,3D}$ along the

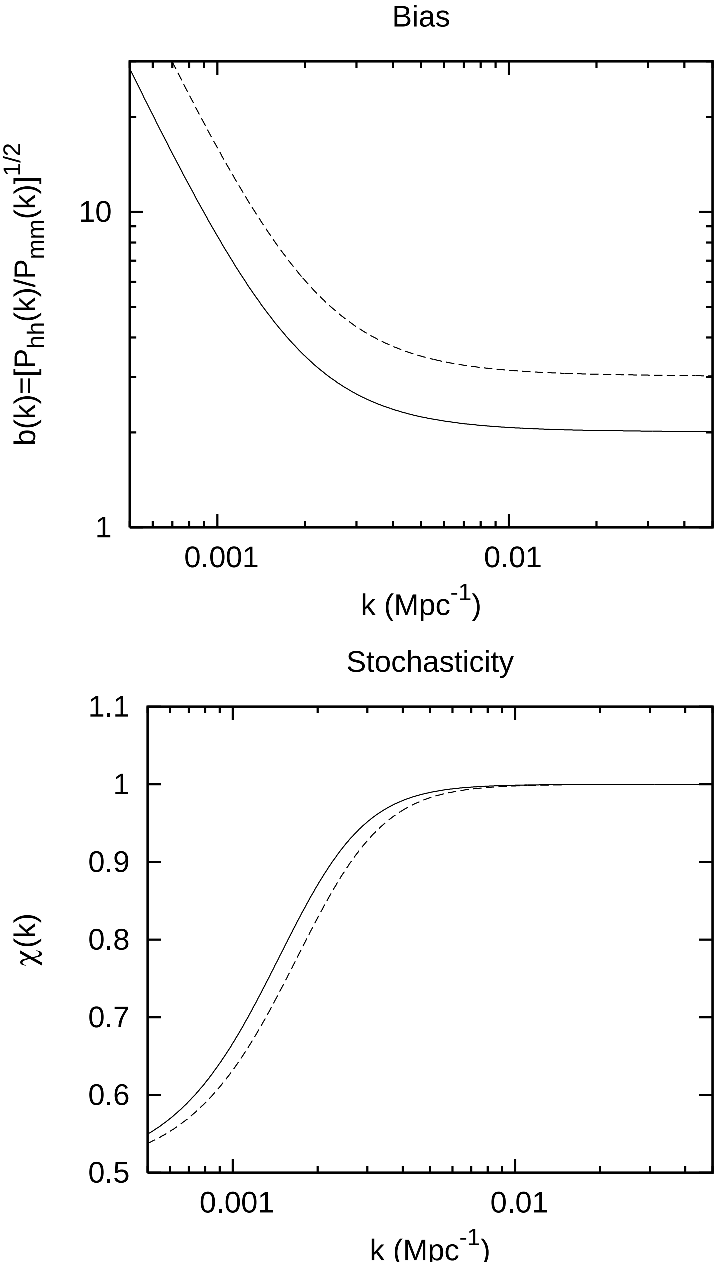


Figure 2.2: The bias and stochasticity of galaxies at $z = 1$ in a model with $x_1 = 30$ and $x_2 = 0.5$ ($\tilde{f}_{\rm NL} = 120$). The solid lines show a tracer with $b = 2$ and the dashed lines a tracer with $b = 3$. The background cosmology and power spectrum are those of WMAP5.

line of sight $\hat{\mathbf{n}}$ and take into account redshift distortions [90, 91, 92]:

$$1 + \delta_g(\hat{\mathbf{n}}) = \int dy\, f(s)[1 + \delta_{g,3D}(y, y\hat{\mathbf{n}})] \ . \tag{2.42}$$

Here, $s = y + H^{-1}\mathbf{v}\cdot\hat{\mathbf{n}}$ is the redshift distance, $f(s)$ is the normalized radial selection function, and we have explicitly written the mean contribution to the density field. We further note that peculiar velocities are generally small compared to the size of the redshift slice and hence we can Taylor expand selection function as

$$f(s) \approx f(y) + \frac{1}{aH}\frac{df}{dy}\mathbf{v}(y\hat{\mathbf{n}})\cdot\hat{\mathbf{n}} \ . \tag{2.43}$$

At large scales where the density perturbations are $\ll 1$, we may ignore second-order terms, i.e., we may ignore the product of the velocity term in equation (2.43) with the $\delta_{g,3D}(y, y\hat{\mathbf{n}})$ term in equation (2.42). This allows us to split δ_g into two terms as $\delta_g = \delta_g^0 + \delta_g^r$. In terms of the Fourier transformed fields, we can write δ_g^0 and δ_g^r as

$$\delta_g^0(\hat{\mathbf{n}}) = \int dy\, f(y) \int \frac{d^3\mathbf{k}}{(2\pi)^3}\delta_{g,3D}(y, \mathbf{k})e^{-i\mathbf{k}\cdot\hat{\mathbf{n}}y}, \tag{2.44}$$

and

$$\delta_g^r(\hat{\mathbf{n}}) = \int dy\, \frac{df}{dy} \int \frac{d^3\mathbf{k}}{(2\pi)^3}\frac{1}{aH}\mathbf{v}(y, \mathbf{k})\cdot\hat{\mathbf{n}}e^{-i\mathbf{k}\cdot\hat{\mathbf{n}}y} \ . \tag{2.45}$$

The velocity can be related to the density perturbation using linearized continuity equation:

$$H^{-1}\mathbf{v}(y, \mathbf{k}) = -i\frac{\Omega_m^{0.6}}{b}\delta_g(y, \mathbf{k})\frac{\mathbf{k}}{k^2} \ . \tag{2.46}$$

We can further transfer redshift dependence of δ_g into a growth function $D(y)$ and expand δ's in Legendre polynomials $P_\ell(x)$ using the following identity:

$$e^{-i\mathbf{k}\cdot\hat{\mathbf{n}}y} = \sum_{\ell=0}^{\infty}(2\ell + 1)i^\ell j_\ell(ky)P_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{n}}) \ , \tag{2.47}$$

where $j_\ell(s)$ is the spherical Bessel function of order ℓ. We obtain

$$\delta_g^0(\hat{\mathbf{n}}) = \int \frac{d^3\mathbf{k}}{(2\pi)^3}\sum_{\ell=0}^{\infty}(2\ell + 1)P_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{n}})\delta_{g,\ell}^0 \ , \tag{2.48}$$

where $\delta_{g,\ell}^0$ is the observed galaxy transfer function [analogous to the CMB radiation transfer function $g_\ell(k)$]:

$$\delta_{g,\ell}^0 = i^\ell \int dy\, f(y)\delta_{g,3D}(\mathbf{k})D(y)j_l(ky). \tag{2.49}$$

Similarly, we can write $\delta_{g,\ell}^r$ as

$$\delta_{g,\ell}^r = i^\ell \int dy\, \frac{df}{dy}\delta_{g,3D}(\mathbf{k})D(y)\frac{\Omega_m^{0.6}}{kb}j_l'(ky) \ . \tag{2.50}$$

Now we use equation (2.34) to express $\delta_{g,\ell}$ in terms of the overdensities generated by inflaton and curvaton fields:

$$\begin{aligned}\delta^0_{g,\ell} = i^\ell \int dy\, f(y)D(y)([1+b_1(k)]\delta^{(1)}_\ell \\ +[1+b_2(k)]\delta^{(2)}_\ell)j_\ell(ky),\end{aligned} \tag{2.51}$$

and

$$\delta^r_{g,\ell} = i^\ell \int dy\, \frac{df}{dy} D(y) \left(\delta^{(1)}_l + \delta^{(2)}_l\right) \frac{\Omega_m^{0.6}}{k} j'_l(ky) \ . \tag{2.52}$$

Using these expressions together with equation (2.36), it is straightforward to calculate angular power spectrum C_ℓ, which can be conveniently divided into three terms:

$$C_\ell = C_\ell^{gg} + C_\ell^{gv} + C_\ell^{vv}, \tag{2.53}$$

corresponding to the pure real-space galaxy density contribution (gg), the redshift space distortion term (vv), and the cross-correlation (gv). These components of the angular power spectrum are expressed in terms of the three-dimensional linear matter power spectrum Δ^2_k and x_1, x_2 parameters, introduced in the previous section:

$$\begin{aligned}
C_\ell^{gg} &= 4\pi \int \frac{\mathrm{d}k}{k}\, \Delta^2_k(|W^0_\ell(k)|^2(1-x_2) + |W^1_\ell(k)|^2 x_2),\\
C_\ell^{gv} &= 8\pi \int \frac{\mathrm{d}k}{k}\, \Delta^2_k \Big(\Re\left[W^{0*}_\ell(k)W^r_\ell(k)\right](1-x_2) + \\
&+ \Re\left[W^{1*}_\ell(k)W^r_\ell(k)\right] x_2\Big), \quad \text{and}\\
C_\ell^{vv} &= 4\pi \int \frac{\mathrm{d}k}{k}\, \Delta^2_k |W^r_\ell(k)|^2,
\end{aligned} \tag{2.54}$$

where the window functions are given by

$$\begin{aligned}
W^0_\ell(k) &= \int bD(y)f(y)j_\ell(ky)\,\mathrm{d}y,\\
W^1_\ell(k) &= \int (b+\Delta b)D(y)f(y)j_\ell(ky)\,\mathrm{d}y, \quad \text{and}\\
W^r_\ell(k) &= \int \Omega_m^{0.6}(r)D(y)f(y)\Big[\frac{2\ell^2+2\ell-1}{(2\ell-1)(2\ell+3)} j_\ell(ky)\\
&\quad -\frac{\ell(\ell-1)}{(2\ell-1)(2\ell+1)} j_{\ell-2}(ky)\\
&\quad -\frac{(\ell+1)(\ell+2)}{(2\ell+1)(2\ell+3)} j_{\ell+2}(ky)\Big]\,\mathrm{d}y.
\end{aligned} \tag{2.55}$$

Here b is the standard Gaussian bias, while Δb is a correction that applies to contributions from the curvaton field *only*:

$$\Delta b = 3\frac{x_1}{x_2}(b-1)\delta_c \frac{\Omega_m}{k^2 T(k) D(r)} \left(\frac{H_0}{c}\right)^2 . \tag{2.56}$$

2.6 Constraints

To constrain x_1 and x_2 parameters using large-scale structure we use the code developed and first implemented in Slosar et al. [55]. We included the same data: the 5-year WMAP bispectrum $x_1 = 51 \pm 30$ (1σ) [66]; and the SDSS data (spectroscopic and photometric luminous red galaxies, the photometric quasar sample, and the integrated Sachs-Wolfe effect via cross-correlation). We further included ancillary data to constrain the background cosmological model and break degeneracies with the non-Gaussianity parameters (x_1, x_2): the CMB power spectrum [93, 94, 95, 96, 97] and supernova data [98].

The Markov chain results are displayed in figure 2.3 where the probability density distribution is plotted on the (x_1, x_2) plane. Dark regions show regions with the highest likelihood and the contours outlines 68%, 95% and 99.7% confidence levels. As the role of the curvaton field in the primordial density perturbation decreases, i.e., $x_2 \to 0$ ($\xi \to \infty$), the upper limit on x_1 decreases. This is because at small x_2, LSS becomes much more sensitive to x_1, as one may see from the x_2 in the denominator of equation (2.56); the $|W^1_\ell(k)|^2$ term in equation (2.54) scales as x_1^2/x_2. In particular, if a local-type CMB bispectrum is ever robustly detected ($x_1 \neq 0$) then the nondetection of excess large-scale clustering in SDSS would immediately set a lower limit on x_2.

2.7 Discussion

This study has extended the analysis of non-Gaussianity constraints into two-field inflationary models. In most previous studies of non-Gaussianity, it was assumed that primordial density perturbations were generated either by inflaton field, in which case they are perfectly Gaussian, or only by the second field (for example curvaton) which contains quadratic part and generates non-Gaussian initial conditions. It is important, however, to realize the possibility of an intermediate case where part of the curvature perturbation is derived from quantum fluctuations of the inflaton field, while an additional part is associated with a second field and converted to an adiabatic perturbation upon its decay. This results in a peculiar type of non-Gaussian initial condition (which we may call "local-stochastic" since the field ϕ_2 entering in the nonlinear term is correlated with but not identical to the linear potential) that is both observable and distinguishable from the curvaton-only "local-deterministic" or $f_{\rm NL}$ form. This type of non-Gaussianity has two parameters: a nonlinear coupling coefficient $\tilde{f}_{\rm NL}$, and the ratio ξ of inflaton to curvaton contributions to the primordial density perturbation spectrum. We connect these parameters with parameters characterizing inflationary fields in equations (2.9) and (2.10).

Using the power spectrum and bispectrum constraints from SDSS and WMAP we are able to constrain these parameters. Adding two sets of constraints together allows us to break the degeneracy in the $(\tilde{f}_{\rm NL}, \xi)$ parameters that exists with the CMB bispectrum alone. If non-Gaussianity in the CMB is ever detected, and the bispectrum has the local configuration dependence, this will enable us to measure the relative contributions of the inflaton and curvaton.

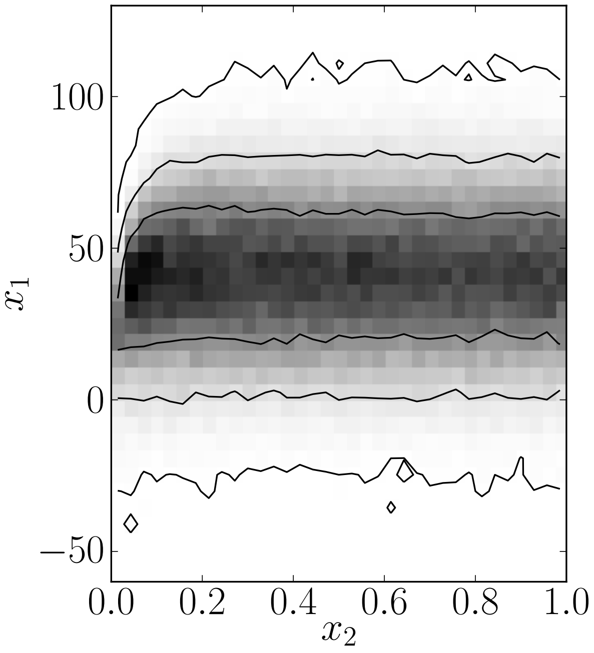


Figure 2.3: Constraints in the (x_1, x_2) plane, including both the CMB bispectrum and the galaxy power spectrum.

We have found that in contrast to the local-deterministic non-Gaussianity, whose main effect on large-scale structure is a scale-dependent increase in the bias, the local-stochastic non-Gaussianity can introduce stochasticity between the matter and halo distributions. It can also lead to relative stochasticity between halos of different masses, since equation (2.39) depends on the Gaussian bias b_g of the halos (e.g., $\chi \to 1$ if $b_\mathrm{g} \to 0$). The potential use of these effects to directly test the hypothesis of multiple fields contributing to the primordial perturbations is left to future work.

Acknowledgments

The authors are grateful to Shirley Ho for providing the large-scale structure samples used in this study. D. T. and C. H. were supported by the U.S. Department of Energy (DE-FG03-92-ER40701) and the National Science Foundation (AST-0807337). C. H. was supported by the Alfred P. Sloan Foundation. This work was supported in part by the U.S. Department of Energy under Contract No. DE-AC02-98CH10886.

Chapter 3

Relative Velocity Effect and the Formation of the First Cosmic Structures*

In this chapter we move to a much later universe and look into physical processes occurring after recombination that influence formation of the first halos, minihalos, galaxies and stars. We introduce a very important and previously unnoticed effect caused by coherent supersonic flows of dark matter and baryonic fluids after the epoch of recombination, which significantly modifies the way first cosmic structures form and evolve.

Chris Hirata (CH) and Dmitriy Tseliakhovich (DT) discovered this effect, dubbed the relative velocity or *Vbc* effect, in spring 2010 while discussing DT's primary thesis research direction. DT's long standing interest in structure formation and connection between dark matter microphysics and potentially observable features in the matter power spectrum on very small scales steered the discussion into the realm of small-scale structure and evolution of primordial density perturbations. The quest to better understand small-scale structure and early evolution of density perturbations lead the authors to look into the nonlinear effects starting at the very earliest times, and during one of the discussions CH offered to look deeper into potential consequences of the bulk velocity flows of baryons relative to dark matter which originate during the epoch of recombination. Subsequent analysis performed by DT and CH showed the importance of the discovered effect and formed a foundation for detailed analysis of its consequences by DT, CH and many other scientists.

At the time of recombination, baryons and photons decoupled and the sound speed in the baryonic fluid dropped from relativistic, $\sim c/\sqrt{3}$, to the thermal velocities of the hydrogen atoms, $\sim 2\times10^{-5}c$. This is less than the relative velocities of baryons and dark matter computed via linear perturbation theory, so we infer that there are supersonic coherent flows of the baryons relative to the underlying potential wells created by the dark matter. As a result, the advection of small-scale perturbations (near the baryonic Jeans scale) by large-scale velocity flows is

---

[*] The material in this chapter was adapted from *Relative velocity of dark matter and baryonic fluids and the formation of the first structures*, D. Tseliakhovich and C. Hirata, *Phys. Rev. D 82, 083520 (2010)*. Reproduced here with permission, copyright (2010) by the American Physical Society. Material has been significantly expanded to include a more detailed introduction and provide a broader overview of the topic of structure formation.

important for the formation of the first structures. This effect involves a quadratic term in the cosmological perturbation theory equations and hence has not been included in studies based on linear perturbation theory. However, detailed consideration of the relative velocity terms show that they become nonperturbative at scales where the very first structures form and hence must be taken into account if accurate results are to be obtained. We show that the relative motion suppresses the abundance of the first bound objects, even if one only investigates dark matter halos, and leads to qualitative changes in their spatial distribution, such as introducing scale-dependent bias and stochasticity. We conclude this chapter by discussing possible observational implications of the relative velocity effect for high-redshift galaxy clustering and reionization.

3.1 First Cosmic Structures

The early Universe was extremely homogeneous and isotropic, with small adiabatic density perturbations likely seeded during an epoch of inflation [99, 100, 101]. The subsequent evolution of the Universe is well described by a model containing baryons, cold dark matter (CDM), and a cosmological constant (Λ). This inflationary ΛCDM paradigm, with only six parameters, is simultaneously consistent with a wide range of cosmological observables [13, 102].

One of the key features of the ΛCDM scenario is the hierarchical formation of structure: since the matter power spectrum $\Delta^2(k)$ is an increasing function of wavenumber k, the smallest perturbations collapse first, followed by their assembly into larger and larger structures. The formation of the first structures has recently become a major research area: aside from the intrinsic interest in understanding the first galaxies, these objects are believed to be responsible at least partially for the reionization of the intergalactic medium [103, 104], and they are sensitive to the small-scale power spectrum of the dark matter, which is a powerful probe of dark matter microphysics.

The evolution of density perturbations in the early Universe is generally described using linear perturbation theory, which treats overdensities and velocity fields as small quantities and hence neglects second-order terms. Several previous works have extended this theory down to the postrecombination baryon Jeans scale [105, 106]. Interest in direct observations of the high-redshift Universe via absorption in the redshifted 21 cm line [107] has motivated more detailed investigation of the clustering of baryons during the epoch between recombination and reionization [108, 109], including the entropy and ionization fluctuations in the baryons [110, 111, 112]. A deficiency of linear perturbation theory is that it does not describe the collapse of perturbations to form bound halos, although analytical models such as the Press-Schechter formalism [113, 114] are often used to estimate the halo mass function and clustering. In order to go beyond linear perturbation theory, one may use spherically symmetric Lagrangian hydrodynamic models [115] or high-resolution three-dimensional hydrodynamic simulations to follow the infall of baryons into the first halos [116, 117]. Since it is inherently nonperturbative this approach can, with incorporation of appropriate chemistry and cooling processes, even be followed all the way to the formation of the first stars [118, 119, 120].

The principal purpose of this chapter and of the Physical Review paper which we published in 2010 [121] is to point out a new nonlinear effect in the growth of small-scale density perturbations that is active even at $z \sim 1000$. The idea is that prior to recombination, the baryons are tightly coupled to the photons resulting in a standing acoustic wave pattern [122]. Modern linear perturbation theory treatments including the CDM [123, 124] show a consequent relative velocity of the baryons and CDM since the latter does not suffer Thomson scattering and merely follows geodesics of the cosmic space-time. At the time of recombination, the root-mean-square (RMS) relative velocity is 30 km s^{-1}, and this is coherent over a scale of several megaparsec comoving (the Silk damping scale [125]). When the baryons recombine and are no longer tied to the photons, their sound speed drops to ~ 6 km s^{-1}, and hence there is a highly supersonic relative velocity between baryons and CDM. This means that near the baryonic Jeans scale, perturbations in the baryons and CDM are advected relative to each other in less than a Hubble time (and hence less than their growth time). This effect is investigated herein, and we find that it both suppresses the growth of small-scale structure, and leads to qualitatively new effects in the clustering of the first bound baryonic objects.

The suppression effect does not appear to have been present in previous analyses. Since it results from the coupling of large-scale and small-scale modes, it is nonlinear and hence not present in linear perturbation theory. Since the large-scale modes involved are associated with the acoustic oscillations of the photon-baryon fluid, they are not properly modeled by hydrodynamic simulations whose box size is smaller than the acoustic horizon (~ 140 Mpc comoving).

Understanding of the physical processes that determine the collapse of the first dark matter halos and subsequent accumulation of baryonic matter in those halos is of paramount importance for interpretation of future data on reionization, high-redshift galaxies, and possibly dark matter substructure. This chapter and initial paper [121] introduce the formalism and focus on the key features of the relative velocity effect, while a more detailed study of various applications, performed by the author and other research teams, as well as the future work on the relative velocity effect are discussed in the following chapter.

This chapter is organized as follows. In section 3.2 we introduce the effect and calculate the Mach number of the relative motion of dark matter and baryonic fluids at the time of recombination. We show that accounting for the relative motion leads to a suppression of the matter power spectrum near the baryon Jeans scale. In section 3.3 we compute the abundance and clustering properties of the first halos taking account of the relative motion. The treatment is simple (it uses the linear Gaussian random field model for the large-scale density and velocity perturbations in cells of size a few megaparsec, and then uses an analytic model to compute the density of small halos in each cell), but it captures the qualitative result of the relative motion effect and provides clear direction for future more detailed studies. We briefly summarize our results and outline possible future work in section 3.4.

The numerical results and plots shown in this chapter assume a cosmology with present-day baryon density $\Omega_{\mathrm{b},0} = 0.044$, CDM density $\Omega_{\mathrm{c},0} = 0.226$, dark energy density $\Omega_{\Lambda,0} = 0.73$, Hubble constant $H_0 = 71$ km s Mpc^{-1}, and adiabatic primordial perturbations of variance $\Delta_\zeta^2 = 2.42 \times 10^{-9}$.

3.2 Growth of Small-Scale Structure Including Relative Velocity of Baryons and CDM

Before recombination, baryons are tightly coupled to photons via Thomson scattering and the sound speed is that appropriate for a radiation-dominated plasma, $\sim c/\sqrt{3}$. Perturbations in the CDM component can grow, however, because the CDM experiences no drag against the radiation. As the universe expands and cools electrons recombine with protons and the universe becomes transparent [126, 127]. This also leads to a kinematic decoupling of the baryons from the radiation, so that the baryons can fall into the potential wells created by the CDM. The effective redshift of decoupling is $z_{\mathrm{dec}} \approx 1020$, which is slightly later than the surface of last scattering for microwave background photons because the baryons have lower inertia than the photons during this epoch [128].

3.2.1 Basic Setup

In the postrecombination gas, the baryonic sound speed is

$$c_{\mathrm{s}} = \sqrt{\frac{\gamma k T_{\mathrm{b}}}{\mu m_{\mathrm{H}}}}, \tag{3.1}$$

where $\gamma = 5/3$ for an ideal monoatomic gas, $\mu = 1.22$ is the mean molecular weight including a helium mass fraction of 0.24, m_{H} is the mass of the hydrogen atom, and T_{b} is the kinetic temperature of the baryons. Here T_{b} is determined by a competition between adiabatic cooling and Compton heating from the CMB; we obtain it using the Recfast code [129, 130] and parametrize it as

$$T_{\mathrm{b}}(a) = \frac{T_{\mathrm{CMB},0}}{a}\left[1 + \frac{a/a_1}{1 + (a_2/a)^{3/2}}\right]^{-1}, \tag{3.2}$$

with $a_1 = 1/119$, $a_2 = 1/115$, and $T_{\mathrm{CMB},0} = 2.726$ K.

While the baryonic velocity drops precipitously during recombination, dark matter velocity remains unaffected, and after recombination dark matter motion with respect to baryons becomes significant. The relative velocity can be written as

$$\mathbf{v}_{\mathrm{bc}}(\mathbf{k}) = \frac{\hat{\mathbf{k}}}{ik}[\theta_{\mathrm{b}}(\mathbf{k}) - \theta_{\mathrm{c}}(\mathbf{k})], \tag{3.3}$$

where $\hat{\mathbf{k}}$ is a unit vector in the direction of $\mathbf{k}$, and $\theta \equiv a^{-1}\nabla \cdot \mathbf{v}$ is the velocity divergence.

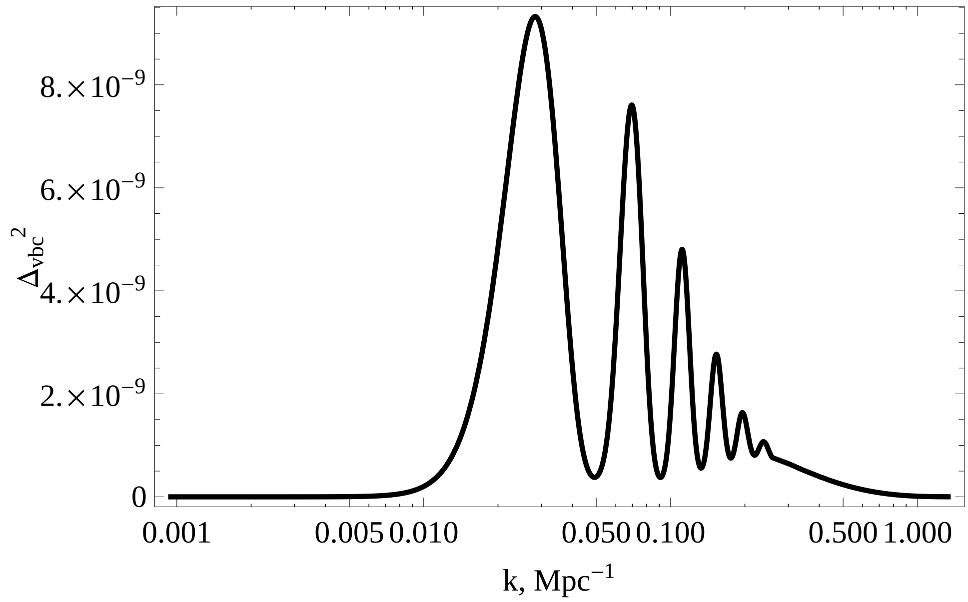


Figure 3.1: The coherence scale of $\mathbf{v}_{\rm bc}$ is determined by the range of scales over which $\Delta^2_{\rm vbc}(k)$ is nonzero. Here we plot $\Delta^2_{\rm vbc}(k)$, the variance of the relative velocity perturbation per $\ln k$, as a function of wavenumber k. The power spectrum drops rapidly at $k > 0.5\,\mathrm{Mpc}^{-1}$, indicating that the relative velocity is coherent over scales smaller than a few megaparsec comoving.

The variance of this relative velocity is

$$\begin{aligned} \langle v^2_{\rm bc}(\mathbf{x})\rangle &= \int \frac{dk}{k}\Delta^2_\zeta(k)\left[\frac{\theta_b(k)-\theta_c(k)}{k}\right]^2 \\ &= \int \frac{dk}{k}\Delta^2_{\rm vbc}(k), \end{aligned} \tag{3.4}$$

where $\Delta^2_\zeta(k) = 2.42\times10^{-9}$ is the initial curvature perturbation variance per $\ln k$ [131]. Integration of equation (3.4) at the time of recombination ($z_{rec} = 1020$) shows that dark matter moves relative to the baryons with rms velocities of $\sim 30\,\mathrm{km\,s^{-1}}$ corresponding to a Mach number of $\mathcal{M} \equiv v_{\rm bc}/c_{\rm s} \sim 5$. This supersonic relative motion allows baryons to advect out of the dark matter potential wells and significantly suppresses the growth of structure at wave numbers higher than

$$k_{\rm vbc} \equiv \left.\frac{aH}{\langle v^2_{\rm bc}\rangle^{1/2}}\right|_{\rm dec} = \frac{k_{\rm J}}{\mathcal{M}} \sim 40\,\mathrm{Mpc}^{-1}. \tag{3.5}$$

The relative contributions to $v_{\rm bc}$ from different scales are shown in figure 3.1. One can clearly see that there is no contribution to the relative velocity from the largest scales, which were outside the sound horizon at the time of decoupling, and that the dominant contribution arises from the acoustic oscillation regime, which have typical velocities of a few times $c\Delta_\zeta$ and suffer no Hubble damping at early epochs when $\rho_\gamma \gg \rho_{\rm b}$ [123]. At the smallest scales, the acoustic oscillations in the baryons are damped by photon diffusion, and the CDM velocities are suppressed by Hubble drag during the radiation era. Thus we expect that $\mathbf{v}_{\rm bc}$ contains contributions from scales ranging from the Silk damping length up to the sound horizon. This leads us to the conclusion that there is a *separation of scales*: the scales at which the first baryonic objects will form ($\sim 10\,\mathrm{kpc}$) is much smaller than the coherence length of the relative velocity field associated with acoustic oscillations (few megaparsec). This will be critical for our use of moving-background perturbation theory to follow early structure formation.

3.2.2 Fluid Equations

After recombination the small-scale inhomogeneities in the photons and neutrinos are rapidly washed out by free streaming, and the dark energy is not yet dynamically significant. Also on small scales we can ignore the general relativistic (higher order in aH/k) terms. Thus we can write the evolution equations as the pressureless Navier-Stokes equations for the CDM, the Navier-Stokes equations including pressure for the baryons, and the Poisson equation for the gravitational potential (e.g., [132]):

$$\begin{aligned}
\frac{\partial \delta_{\rm c}}{\partial t} + a^{-1}\mathbf{v}_{\rm c}\cdot\nabla\delta_{\rm c} &= -a^{-1}(1+\delta_{\rm c})\nabla\cdot\mathbf{v}_{\rm c}, \\
\frac{\partial \mathbf{v}_{\rm c}}{\partial t} + a^{-1}(\mathbf{v}_{\rm c}\cdot\nabla)\mathbf{v}_{\rm c} &= -\frac{\nabla\Phi}{a} - H\mathbf{v}_{\rm c}, \\
\frac{\partial \delta_{\rm b}}{\partial t} + a^{-1}\mathbf{v}_{\rm b}\cdot\nabla\delta_{\rm b} &= -a^{-1}(1+\delta_{\rm b})\nabla\cdot\mathbf{v}_{\rm b}, \\
\frac{\partial \mathbf{v}_{\rm b}}{\partial t} + a^{-1}(\mathbf{v}_{\rm b}\cdot\nabla)\mathbf{v}_{\rm b} &= -\frac{\nabla\Phi}{a} - H\mathbf{v}_{\rm b} - a^{-1}c_{\rm s}^2\nabla\delta_{\rm b}, \\
a^{-2}\nabla^2\Phi &= 4\pi G\bar{\rho}_{\rm m}\delta_{\rm m}.
\end{aligned} \tag{3.6}$$

Here Φ is total gravitational potential and subscripts c, b, and m stand for dark matter, baryons and total matter respectively.

A more complete treatment would also follow the baryon entropy [110] and ionization fraction [111] perturbations. We have not done this here, but we note that the moving background perturbation theory approach described here could be extended to accommodate these additional variables.

The standard way to do large-scale structure perturbation theory is to Taylor-expand in powers of the primordial perturbations, e.g., $\delta_{\rm c} = \delta_{\rm c}^{(1)} + \delta_{\rm c}^{(2)} + \delta_{\rm c}^{(3)} + \ldots$. One may then use the linear terms in the above equation to describe the behavior at order n; for $n \geq 2$, the quadratic terms in equation (3.6) may be treated as a source for the order-n perturbation, written hierarchically in terms of orders $< n$ [133, 134, 135, 136]. This approach can even be extended to include both baryons and CDM [137]. In our case, this is not appropriate: since there are relative bulk flow velocities between the baryons and CDM with Mach numbers of order 10, it follows that for perturbations at the baryonic Jeans scale the baryon and CDM components will be advected relative to each other by up to several perturbation wavelengths. Whether the standard perturbation series will converge in this case is an open question; even if it does, we expect that many orders in perturbation theory would be required. Therefore, we desire an alternative method to follow the growth of the smallest structures.

3.2.3 Moving-Background Perturbation Theory (MBPT)

Our preferred method of following the earliest structures is to do a perturbation analysis on a background where the baryons move relative to the CDM. The idea is that in the absence of density perturbations, but in the presence of a bulk relative velocity, there exists an exact solution to equation (3.6):

$$\begin{aligned}
\mathbf{v}_{\rm c}(\mathbf{x},t) &= \mathbf{v}_{\rm c}^{\rm (bg)}(t), \\
\mathbf{v}_{\rm b}(\mathbf{x},t) &= \mathbf{v}_{\rm b}^{\rm (bg)}(t), \quad \text{and} \\
\Phi &= \delta_{\rm c} = \delta_{\rm b} = 0,
\end{aligned} \tag{3.7}$$

where $\mathbf{v}_{\rm c}^{\rm (bg)}$ and $\mathbf{v}_{\rm b}^{\rm (bg)}$ are constant with position and have temporal dependence $\propto 1/a(t)$. Without loss of generality, one may boost to a different reference frame and set one of these (e.g., $\mathbf{v}_{\rm b}^{\rm (bg)}$) equal to zero.

Since the relative velocity of the baryons and CDM is coherent over scales of several comoving megaparsec, whereas the scales of direct interest for us are at a few baryonic Jeans lengths (~ 10 comoving kpc), the above moving background is an appropriate zeroeth-order solution in any small (~ 1 Mpc) region of the Universe. Thus we can imagine the Universe as composed of many individual patches, each of which has a different relative velocity $\mathbf{v}_{\rm bc}^{\rm (bg)}$. Small fluctuations on this background grow due to gravitational instability; their early stages of collapse can be modeled using linear perturbation theory around equation (3.7) using the local value of $\mathbf{v}_{\rm bc}^{\rm (bg)}$.

Perturbing around equation (3.7), and writing the perturbation variables $\mathbf{u}_{\rm b,c}$,

$$\mathbf{v}_{\rm b}(\mathbf{x},t) = \mathbf{v}_{\rm b}^{\rm (bg)}(t) + \mathbf{u}_{\rm b}(\mathbf{x},t), \tag{3.8}$$

and similarly for $\mathbf{u}_{\rm c}$, we may transform equation (3.6) into a system of equations involving $\mathbf{u}_{\rm b,c}$ instead of $\mathbf{v}_{\rm b,c}$. Working only to first order in the new perturbation variables $\{\delta_{\rm c}, \mathbf{u}_{\rm c}, \delta_{\rm b}, \mathbf{u}_{\rm b}, \Phi\}$ we find:

$$\begin{aligned}
\frac{\partial \delta_{\rm c}}{\partial t} + a^{-1}\mathbf{v}_{\rm c}^{\rm (bg)} \cdot \nabla \delta_{\rm c} &= -a^{-1}\nabla \cdot \mathbf{u}_{\rm c}, \\
\frac{\partial \mathbf{u}_{\rm c}}{\partial t} + a^{-1}(\mathbf{v}_{\rm c}^{\rm (bg)} \cdot \nabla)\mathbf{u}_{\rm c} &= -\frac{\nabla \Phi}{a} - H\mathbf{u}_{\rm c}, \\
\frac{\partial \delta_{\rm b}}{\partial t} + a^{-1}\mathbf{v}_{\rm b}^{\rm (bg)} \cdot \nabla \delta_{\rm b} &= -a^{-1}\nabla \cdot \mathbf{u}_{\rm b}, \\
\frac{\partial \mathbf{u}_{\rm b}}{\partial t} + a^{-1}(\mathbf{v}_{\rm b}^{\rm (bg)} \cdot \nabla)\mathbf{u}_{\rm b} &= -\frac{\nabla \Phi}{a} - H\mathbf{u}_{\rm b} \\
&\quad -ac_{\rm s}^2\nabla \delta_{\rm b}, \quad \text{and} \\
a^{-2}\nabla^2\Phi &= 4\pi G\bar{\rho}_{\rm m}\delta_{\rm m}.
\end{aligned} \tag{3.9}$$

It is convenient to transform these equations into Fourier space and use the last equation to eliminate Φ. We may also rewrite the velocity equations in terms of the divergence θ, with $\mathbf{u}_{\rm i}(\mathbf{k}) = -iak^{-2}\mathbf{k}\theta_{\rm i}(\mathbf{k})$ (i = b or c), since under the approximation of barotropic flow of the baryons the vorticity remains zero until the development of structure formation shocks. We may also work in the bulk baryon frame, i.e., we may set

$$\mathbf{v}_{\rm b}^{\rm (bg)} = 0, \quad \text{and} \quad \mathbf{v}_{\rm c}^{\rm (bg)} = -\mathbf{v}_{\rm bc}^{\rm (bg)}(t). \tag{3.10}$$

This reduces our system of equations to

$$\begin{aligned} \frac{\partial \delta_{\rm c}}{\partial t} &= \frac{i}{a}\mathbf{v}_{\rm bc}^{\rm (bg)} \cdot \mathbf{k}\delta_{\rm c} - \theta_{\rm c}, \\ \frac{\partial \theta_{\rm c}}{\partial t} &= \frac{i}{a}\mathbf{v}_{\rm bc}^{\rm (bg)} \cdot \mathbf{k}\theta_{\rm c} - \frac{3H^2}{2}(\Omega_{\rm c}\delta_{\rm c} + \Omega_{\rm b}\delta_{\rm b}) - 2H\theta_{\rm c}, \\ \frac{\partial \delta_{\rm b}}{\partial t} &= -\theta_{\rm b}, \text{ and} \\ \frac{\partial \theta_{\rm b}}{\partial t} &= -\frac{3H^2}{2}(\Omega_{\rm c}\delta_{\rm c} + \Omega_{\rm b}\delta_{\rm b}) - 2H\theta_{\rm b} + \frac{c_{\rm s}^2 k^2}{a^2}\delta_{\rm b}. \end{aligned} \tag{3.11}$$

Note that $\Omega_{\rm c,b}$ are evaluated at the appropriate redshift rather than taking on their present-day values. Our code evolves these equations, albeit with the scale factor a as the independent variable, which can be accomplished using the replacement $\partial/\partial t = aH\,\partial/\partial a$. It is important to note the time dependence $\mathbf{v}_{\rm bc}^{\rm (bg)} \propto 1/a(t)$ when evolving equation (3.11).

On large scales $k \ll k_{\rm vbc} \sim 40\,{\rm Mpc}^{-1}$ used for galaxy clustering and even for Lyman-α forest studies, the $\mathbf{v}_{\rm bc}$ terms in equation (3.11) are negligible. However, at $k \gtrsim k_{\rm vbc}$, the advection terms become comparable to or larger than the Hubble expansion rate, and they must be taken into account. Note that this is true even if one's interest is only in the CDM perturbations, since the baryons contribute 17% of the energy density and hence their perturbations are important in equation (3.11). (As an extreme example, below the Jeans scale $k > k_{\rm J}$, the growth of structure in the CDM switches from the "standard" $\delta \propto a$ growth to a slower growth $\delta \propto a^{\sqrt{25-24\Omega_{\rm b}}/4-1/4}$ [138, 110].)

3.2.4 Small-Scale Transfer Function and Matter Power Spectrum

The usual way to describe the small-scale distribution of matter is to derive a transfer function $T(k)$ that maps primordial to final potentials, and then to write the matter power spectrum $P_{\rm m}(k)$, equal to the primordial power spectrum times $|T(k)|^2$ (times normalization factors [29]). We may solve the transfer functions including the relative velocity effect by solving the system of equations, equation (3.11). We evolve these from the redshift of recombination, where initial conditions are determined using CMBFAST [139], to $z = 40$. The resulting transfer function, evaluated at $z = 40$, is clearly a function of the local relative velocity $\mathbf{v}_{\rm bc}^{\rm (bg)}$ and also of the angle ϑ between the direction of the wave vector $\mathbf{k}$ and $\mathbf{v}_{\rm bc}^{\rm (bg)}$.

We may determine a local *isotropically averaged* power spectrum $P_{\rm loc,m}(k; v_{\rm bc})$ by averaging over the direction of $\mathbf{k}$, i.e., we write

$$P_{\rm loc,m}(k; v_{\rm bc}) = P_\zeta(k)\,\frac{1}{2}\int_0^\pi \left|\frac{\delta_{\rm m}(\mathbf{k}; \mathbf{v}_{\rm bc})}{\zeta(\mathbf{k})}\right|^2 \sin\vartheta\, d\vartheta, \tag{3.12}$$

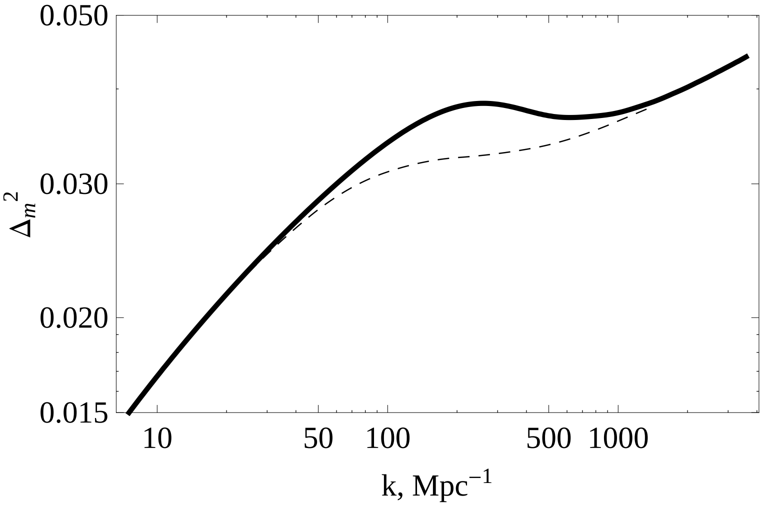


Figure 3.2: Power spectrum of matter distribution in the first order CDM model (solid line) and with the $v_{\rm bc}$ (dashed line) at the redshift of $z = 40$.

where $\zeta(\mathbf{k})$ is the primordial curvature perturbation and $P_\zeta(k)$ is its power spectrum. This power spectrum depends only on the magnitude of $v_{\rm bc}$ (we drop the superscript $^{\rm (bg)}$ to reduce clutter). In order to determine an overall effect on the small-scale matter power spectrum we need to average over a large number of coherence regions with different $v_{\rm bc}$. The latter arises from linear perturbation theory on large scales and hence is well described by a three-dimensional Gaussian distribution with variance per axis,

$$\sigma^2_{\rm vbc} = \frac{1}{3}\left\langle |\mathbf{v}_{\rm bc}(\mathbf{x})|^2 \right\rangle . \tag{3.13}$$

We may then average equation (3.12) to obtain a globally averaged matter power spectrum.

Intuitively, we expect the relative velocity effect to suppress the small-scale power spectrum, since the moving baryons have pressure $\sim \rho_{\rm b} v^2_{\rm bc}$ in the CDM frame. This suppression is shown in figure 3.2 where we plot $\Delta^2_{\rm m}(k) \equiv [k^3/(2\pi^2)]P_{\rm m}(k)$ for the cases with and without the effect of relative velocity. The power is most strongly suppressed around the Jeans scale $k_{\rm J} = aH/c_{\rm s} \sim 200$ Mpc^{-1}, where a difference of $\sim 15\%$ is computed.

The effect of v_{bc} is not limited to the suppression of power on small scales, but rather has an important implication for the distribution of the first bound structures with respect to matter distribution as well as for the number densities of the first halos. To study these effects we ran a set of simulations in which the large-scale density and velocity fields were generated according to linear perturbation theory. We then used analytical (Press-Schechter) arguments to predict the number of halos formed in each cell of our cosmological box. This hybrid approach is computationally feasible on a single desktop computer since it does not have to numerically follow the small-scale modes, and should capture the rough magnitude of the effect. However, ultimately a simulation that follows the full nonlinear evolution of the small-scale modes will be required. The key reason for using approximate methods in the present study, as opposed to a full hydrodynamic numerical simulation, is our desire to introduce the concept of relative velocity effect in the simplest and most intuitive way while allowing more detailed study to be performed by other research groups in an unbiased manner.

3.3 The Abundance and Clustering of Early Halos

We now investigate the formation of the first baryonic objects, taking account of the relative velocity effect. This is a difficult problem, which we only partially solve in this study: one has acoustic oscillations in the photon-baryon plasma that travel ~ 140 Mpc, and simultaneously one must resolve the baryon Jeans scale. We provide a computation based on the formalism described above: we generate a realization of a Gaussian random primordial perturbation on a three-dimensional grid, and then to each cell we assign an overdensity $\delta_{\rm l}$ (where the "l" referes to long-wavelength modes) using periodic boundary condition and a relative velocity $\mathbf{v}_{\rm bc}$ derived from the linear density field. Initial values of $\delta_{\rm l}$ are obtained using the linear perturbation theory, as there is no significant difference between the theory with and without relative motion effect before the time of recombination when the values of $\delta_{\rm l}$ are formed. Then, within each cell, we use the peak-background split to compute the number density of halos. The new twist is that the small-scale power spectrum is modulated by the large-scale $v_{\rm bc}$. (In some ways, this is similar to the modification of the peak-background split used for local $f_{\rm NL}$-type non-Gaussianity studies [140, 141], except that in our case the modulation of the small-scale power spectrum is a result of the advection process and arises even in standard ΛCDM cosmology with Gaussian adiabatic initial conditions.) This of course depends on an analytic model for the halo mass function; we have used the Press-Schechter model [113, 114]. The validity of Press-Schechter for any precise calculation is dubious–particularly since it is being applied here with an anisotropic local power spectrum–but we expect that the qualitative results (a scale-dependent enhancement in the bias and stochasticity at large scales, with acoustic oscillations in each) would still arise in a more accurate treatment.

3.3.1 Peak-Background Split

The collapse of the first halos can be conveniently treated in the framework of the peak-background split formalism [142], in which the growth of small-scale inhomogeneities depends on the large-scale overdensity. One can split the density field into a long-wavelength piece $\delta_{\rm l}$ and a short-wavelength piece $\delta_{\rm s}$:

$$\rho(\mathbf{x}) = \bar{\rho}\,[1 + \delta_{\rm l}(\mathbf{x}) + \delta_{\rm s}(\mathbf{x})]\,. \tag{3.14}$$

In any region, the number density of halos of any given type generally depends on the large-scale overdensity $\delta_{\rm l}$, and on the statistics of the small-scale perturbations $\delta_{\rm s}$ (in particular, their local power spectrum). In the usual case where the small- and large-scale perturbations are independent, the number density becomes purely a function of the large-scale overdensity plus a stochastic component ϵ with $\langle\epsilon(\mathbf{x})\rangle = 0$; Taylor expanding in $\delta_{\rm l}$ gives

$$n(\mathbf{x}) = \bar{n}\,[1 + b_0\delta_l(\mathbf{x})] + \epsilon(\mathbf{x}). \tag{3.15}$$

The bias is then

$$b_0 = \bar{n}^{-1}\frac{\partial n}{\partial\delta_{\rm l}}. \tag{3.16}$$

This argument leads to a generically scale-independent bias at sufficiently large scales (with the addition of a Poisson or halo-shot-noise term [143, 144]).

When the relative motion of dark matter and baryons is introduced the growth of small scale overdensities becomes dependent on the local value of the relative velocity. Equation (3.15) then generalizes to

$$n(\mathbf{x}) = n\,[\delta_{\rm l}(\mathbf{x}), v_{\rm bc}(\mathbf{x})] + \epsilon(\mathbf{x}). \tag{3.17}$$

At Mach numbers of order 10, it is not clear whether we can Taylor expand in $v_{\rm bc}$. Therefore, our strategy will be to recompute $n(\mathbf{x})$ in each cell, using the Press-Schechter conditional mass function, i.e., the number of halos per unit comoving volume per $\ln M$:

$$N(M|\delta_{\rm l}, v_{\rm bc}) = \frac{\bar{\rho}}{\sqrt{2\pi}}\frac{\delta_{\rm sc}-\delta_{\rm l}}{\sigma^2}\left|\frac{d\sigma}{dM}\right|(1+\delta_{\rm l})\exp[-\frac{(\delta_{\rm sc}-\delta_{\rm l})^2}{2\sigma^2}], \tag{3.18}$$

where $\delta_{\rm sc}$ is critical overdensity of spherical collapse, M is the halo mass and $N(M|\delta_{\rm l}, v_{\rm bc})$ has units of Mpc^{-3}.The factor $1 + \delta_{\rm l}$ is the conversion from the Lagrangian volume element (in which the Press-Schechter formalism is native) to the Eulerian volume element. Here σ^2 is the variance of the density field smoothed with the top-hat window function,

$$\sigma^2(M, \mathbf{v}_{\rm bc}) = \int \Delta_m^2(k, \mathbf{v}_{\rm bc})|W(k,R)|^2\frac{dk}{k}, \tag{3.19}$$

where for $\Delta_m^2(k, \mathbf{v}_{\rm bc})$ we use the isotropically averaged local matter power spectrum. In principle, one should follow here the formation of halos in a statistically anisotropic density field. This will ultimately require a hydrodynamic (or at least N-body) simulation to achieve results that can be used for detailed analysis. However, for the moment we use the Press-Schechter formalism; the top-hat window function in Fourier space can be written as $W(k,R) = 3j_1(kR)/(kR)$, where the smoothing scale R is determined by the halo mass $M = \frac{4}{3}\pi R^3$.

In our case–unlike the usual case of purely Gaussian density perturbations–$\sigma^2(M)$ and hence $d\sigma/dM$ are explicit functions of relative velocity and hence will change from place to place.

3.3.2 Simulation Parameters

Our fiducial box size is $1365^3\,\mathrm{Mpc}^3$. The box is divided into smaller boxes each of the size of coherence length for the relative velocity field, $\Delta = 3$ Mpc. Initial (i.e., at $z_{\rm dec}$) density and velocity distributions are generated using the CMBFAST power spectrum computations [139] and smoothed with a Gaussian window function with scale length $k_{\rm scale} = \pi/(\sqrt{3}\,\Delta)$. The smoothing is necessary to avoid aliasing and spurious effects from the finite resolution of the simulation. For each small box we generate initial values of $\mathbf{v}_{\rm bc}$ and large-scale overdensity $\delta_{\rm l}$ at the time of recombination using CMBFAST. Halo number densities in each cell are then inferred from equation (3.18).

In figure 3.3, we show an example output from this procedure. The top panel shows the matter density contrast and the bottom two panels the halo density contrast for $M_{\rm halo} = 10^6\, M_\odot$ without (middle panel) and with (bottom panel) the relative motion effect at the redshift of $z = 40$. Note that the structures in matter and halo overdensities, while correlated, are not identical. Comparison of the halo density contrasts for two different cases clearly shows the importance of relative velocity effect on the formation of first bound objects.

3.3.3 Halo Abundance

To illustrate the effect of relative velocity on the abundance of small halos, we calculate number densities of collapsed halos with and without relative velocity. The decrease in number density is quantified by

$$\Delta_N = \frac{\bar{N}_{\rm vbc} - \bar{N}_0}{\bar{N}_0}, \tag{3.20}$$

where $\bar{N}_0$ and $\bar{N}_{vbc}$ are average number densities of halos without the effect of $\mathbf{v_{bc}}$ and with it. The comparison of the two cases shows that the number density of halos is suppressed by more than 60% at the mass scale of $M \sim 10^6 M_\odot$, as can be seen in figure 3.4. Note that the strongest suppression occurs for halo masses of $10^{6.3}\, M_\odot$, corresponding to top-hat scales of 20 kpc comoving, i.e., near $k_{\rm vbc}^{-1}$. We emphasize that the results provided in the figure are based on the Press-Schecter formalism and are a good qualitative guide, but should not be interpreted quantitatively.

3.3.4 Bias, Stochasticity and the Large-Scale Distribution of Early Halos

The introduction of relative motions modifies the correlation between the first halos and the matter distributions rendering bias parameter scale dependent. Because of the nonlinear terms in the evolution equations, dark matter and baryonic matter evolve out of phase and the growth of the overdense regions become dependent on both $\delta_{\rm l}$ and $v_{\rm bc}$.

To quantify this effect we calculated halo overdensity using number densities of halos in each of the small boxes from our simulation:

$$\delta_n(M, x) = \frac{N(M, x) - \bar{N}(M)}{\bar{N}(M)}. \tag{3.21}$$

Next, we calculate power spectra of halos of various masses:

$$(2\pi)^3 \delta_{\rm D}(\mathbf{k} - \mathbf{k}') P_{\rm hh}(k|M) = \langle \delta_n(M, \mathbf{k}) \delta_n(M, \mathbf{k}')^* \rangle, \tag{3.22}$$

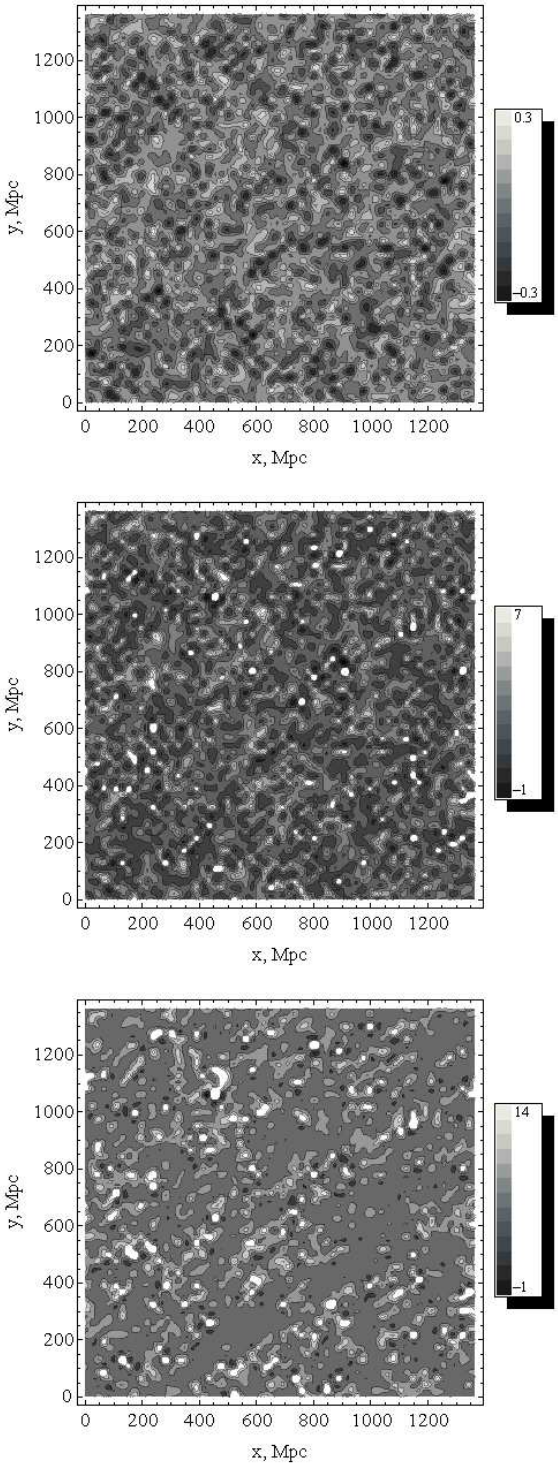


Figure 3.3: *Top panel*: The matter density contrast δ_m on a two-dimensional slice of the three-dimensional simulation box. The halo density contrast δ_n for $M_{\rm halo} = 10^6\, M_\odot$ on the same slice with $V_{bc} = 0$ (middle panel) and with $V_{bc} \neq 0$ (bottom panel). All panels are at $z = 40$.

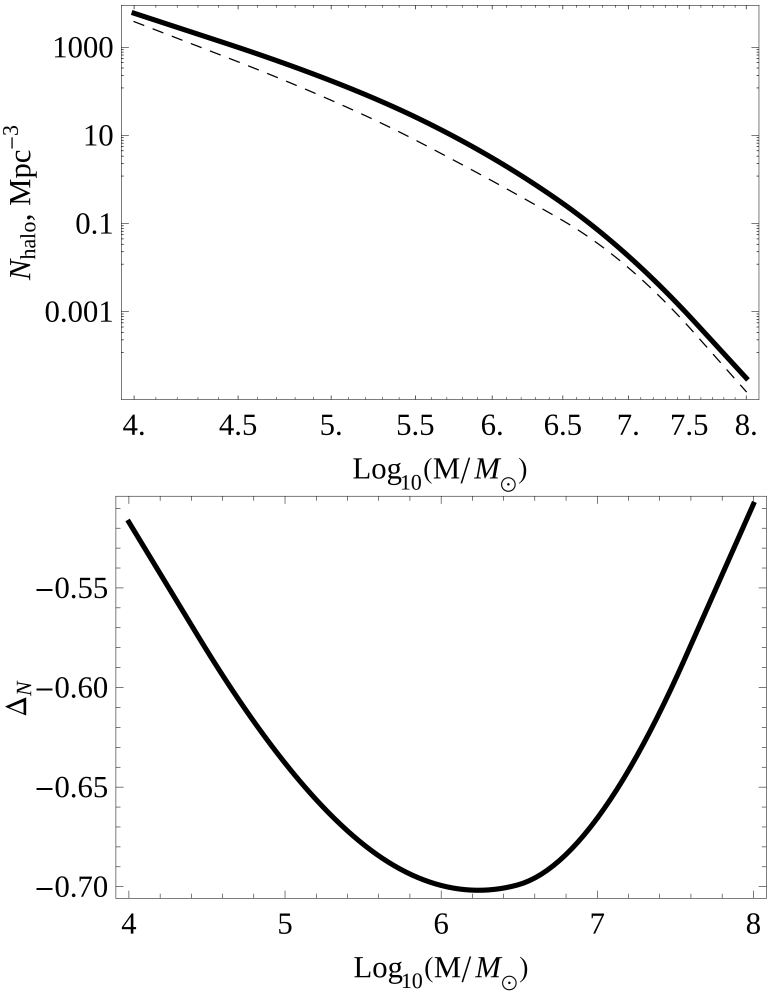


Figure 3.4: *Top panel*: The number density of dark matter halos produced in our simulation box without the effect of relative velocity (solid line) and with the effect (dashed line). *Bottom panel*: The relative decrease in the number density of halos as a function of the halo mass. Number densities in our simulation correspond to the redshift of $z = 40$.

where $\delta_{\rm D}$ is the Dirac δ-function.

The difference between this case and the case neglecting $v_{\rm bc}$ can be illustrated by defining a bias correction parameter $\Delta b(k)$:

$$P_{\rm hh}(k) = b_0^2 \left[1 + \frac{\Delta b(k)}{b_0}\right]^2 P_{\rm mm}(k), \tag{3.23}$$

where b_0 is a Gaussian scale independent bias, which in the Press-Schechter formalism is given by

$$b_0 = \frac{\delta_{\rm sc}}{\sigma^2} - \frac{1}{\delta_{\rm sc}} + 1. \tag{3.24}$$

Using these results along with the matter power spectrum, we can obtain the scale-dependent component of the bias parameter Δb which is plotted in the top panel of figure 3.5 for various halo masses. The plot shows that for halos with mass $M \sim 10^4$–$10^8 M_\odot$ there is a significant increase of the bias. The effect of $\mathbf{v}_{\rm bc}$ becomes less important for heavier and lighter masses which can be expected from the analysis of power suppression in figure 3.2. This is principally a consequence of the fact that for very massive halos the baryons advect through a distance that is only a small fraction of the halo scale R, and hence this advection does not affect the formation of the halo; whereas for the lowest-mass halos, whose scale R is smaller than the baryon Jeans length, the baryons can be treated as homogeneous irrespective of their velocity.

To further understand the importance of $v_{\rm bc}$ we calculate the stochasticity of the halos relative to the matter. In the bottom panel of figure 3.5 we plot the stochasticity χ as a function of wave number k for various halo masses. The stochasticity is defined as

$$\chi = \frac{P_{\rm hm}^2(k)}{P_{\rm hh}(k) P_{\rm mm}(k)}, \tag{3.25}$$

where the cross power spectrum $P_{\rm hm}$ is defined via $(2\pi)^3 \delta_{\rm D}(\mathbf{k} - \mathbf{k}') P_{\rm hm}(k) = \langle \delta_{\rm h}(M, \mathbf{k}) \delta_{\rm m}^*(\mathbf{k}') \rangle$. In the linear theory, without consideration of the $v_{\rm bc}$ effect one would have $\chi = 1$ (modulo Poisson corrections as described above).

We checked the convergence of our results by running the simulation with varying box sizes and varying Δ. Specifically we tried runs with $\Delta = 4$ Mpc, and found changes of less than 1% in the stochasticity and bias over the range $0.2 < k < 1\,{\rm Mpc}^{-1}$ at $M_{\rm halo} = 10^4 M_\odot$, whereas using $\Delta = 6$ Mpc produces change greater than 5% and distorts functional forms of both bias and stochastisity at $k > 0.1$ Mpc^{-1}. Similarly, increasing the box size to 2275^3 Mpc^3, with fixed Δ did not produce observable change in χ and Δb, whereas decreasing the box size to 1000^3 Mpc^3 changes our results by $\sim 5\%$ at $k < 0.1\,{\rm Mpc}^{-1}$. As a test, we repeated the analysis setting $v_{bc} = 0$ to recover the "standard" picture with a scale-independent halo bias and the stochastisity consistent with the linear theory prediction. Specifically, we found that at $k < 0.2$ Mpc^{-1} the stochastisity is $0.98 < \chi < 1$. The small deviation from unity can be explained by the fact that mapping from the overdensity $\delta_{\rm l}$ to the number density of halos $N(M|\delta_{\rm l}, v_{\rm bc})$ is not exactly linear even in Press-Schecter model.

We also would like to mention that figure 3.5 exhibits strong oscillations of Δb which corre-

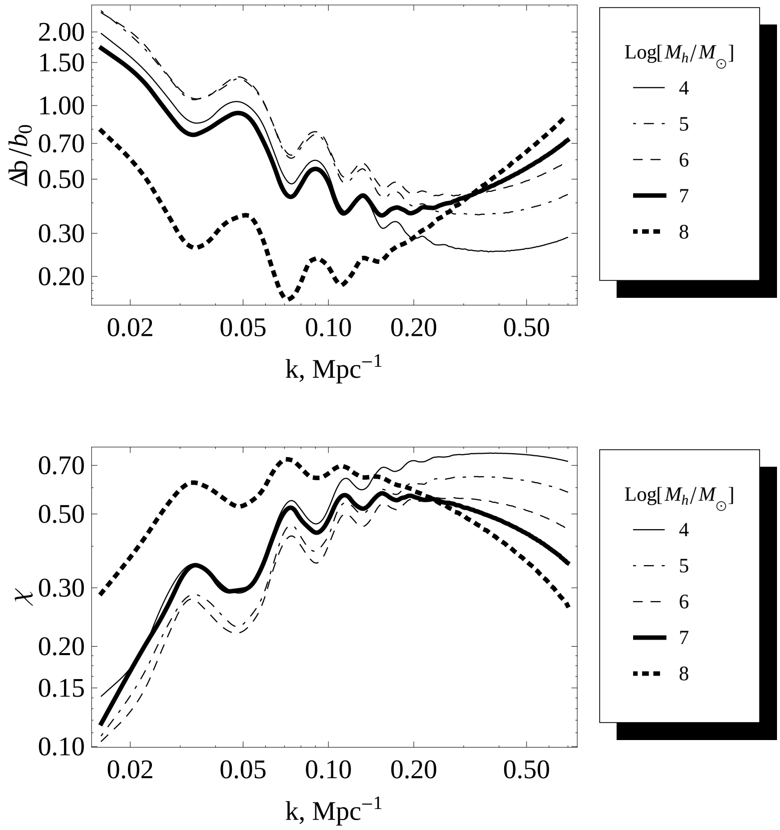


Figure 3.5: The correction to the bias parameter Δb (top panel) and the stochasticity $\chi = r_{\mathrm{hm}}^2$ (bottom panel) for various halo masses at $z = 40$. The solid curve corresponds to $M_h = 10^4 M_\odot$; the thick-solid to $M_h = 10^5 M_\odot$; the dashed to $M_h = 10^6 M_\odot$; the dot-dashed to $M_h = 10^7 M_\odot$; and the dotted to $M_h = 10^8 M_\odot$. In the first-order CDM model $\Delta b = 0$ and $\chi = 1$ on large scales. The enhancement of bias on small scales $k > 0.3\,\mathrm{Mpc}^{-1}$ is due to the nonlinear dependence of halo abundance on δ_{l}.

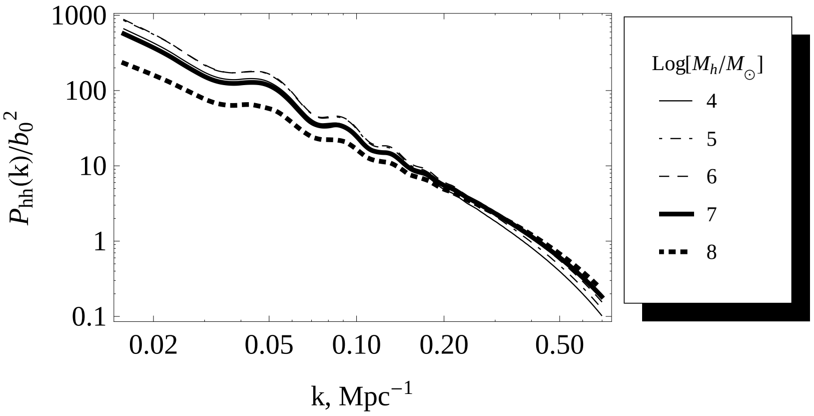


Figure 3.6: Scaled halo-halo power spectrum $P_{hh}(k)/b_0^2(M_h)$ at the redshift of $z = 40$ for various halo masses. The solid curve corresponds to $M_h = 10^4 M_\odot$; the thick-solid to $M_h = 10^5 M_\odot$; the dashed to $M_h = 10^6 M_\odot$; the dot-dashed to $M_h = 10^7 M_\odot$; and the dotted to $M_h = 10^8 M_\odot$.

spond to the BAO in the matter power spectrum. This means that the signal of the BAO in the halo power spectrum for these halo masses is very different from that of the dark matter. To illustrate this point we plot the actual scaled halo power spectrum $P_{hh}(k)/b_0^2(M_h)$ (figure 3.6) for different halo masses covering the range from $M_h = 10^4$ to $M_h = 10^8$ that shows the behavior of the BAO signal. Although these are very low-mass halos compared to those probed by BAO surveys ($M > 10^{11} M_\odot$), they are the seeds of present day galaxies, and their subsequent evolution might alter the BAO signal in the galaxy power spectrum at lower redshifts. As with other interesting applications of the relative velocity effect we relegate detailed analysis of the this problem to a future study.

3.4 Conclusions and Prospects

We have shown that the relative velocity of baryonic and dark matter fluids plays an important role in the formation and evolution of small-scale structure of the early universe. In light of the increasing interest in reionization, high-redshift galaxy clustering, and dark matter substructure, it is imperative to understand the evolution of small-scale structure and all physical effects that contribute to this evolution. Here we discuss the possible implications and next steps in exploring the relative velocity effect.

Early galaxies may be observed in the next decade, either directly via the James Webb Space Telescope or indirectly through the near-infrared background and its fluctuations. Due to the relative velocity effect, the large-scale clustering of these galaxies should show scale-dependent bias and (if sufficient statistics are available to split the galaxy population into multiple samples) relative stochasticity between different samples of galaxies. For example, in our calculation at $z = 15$, halos of mass $10^8\,M_\odot$ show an increase of $\Delta b = 0.73$ over the no-velocity result of $b_0 = 4.6$ at $k = 0.02\,\mathrm{Mpc}^{-1}$. Whether this effect will be detectable depends on the as-yet-unknown luminosity function of the highest-redshift galaxies, and whether the relative velocity effect can be separated from the scale-dependent bias produced by reionization [145, 146].

Since reionization is believed to be driven by the formation of the first halos massive enough to contain gas and produce stellar photoionizing radiation [147], the relative velocity effect will delay reionization. However, this effect is degenerate with the mapping from halo mass to the number of massive stars formed, and given the modest (factor of ~ 2) effects investigated here we do not expect that the effects can be disentangled from the reionization history alone. More interesting would be an investigation of the spatial structure of reionization and of the related high-redshift 21 cm signal, which has been investigated analytically and in simulations [148, 149, 150, 151, 152]. The scale-dependent bias and stochasticity we find here for early halos may have a significant effect on the structure (power spectrum, topology, and correlation with the matter density field) of early reionization bubbles.

If the *pre*-reionization 21 cm signal [107] is ever observed, and cosmological information extracted, the relative velocity effect will be very important: the smallest scale fluctuations in the baryons are modified at the tens of percents level. Indeed, since the 21 cm signal is nonlinear in the baryon density (in the limit where the hydrogen spin temperature is closely coupled to the CMB temperature, the signal is proportional to n^2 times the temperature-dependent collision cross section [153]), it is likely that even the large-scale fluctuations would be affected because regions with increased small-scale baryon power spectra will show more absorption. The locally anisotropic nature of the small-scale baryonic perturbations would also represent an issue for weak lensing of the 21 cm field [154, 155, 156, 157] and/or non-Gaussianity searches [158, 159]. A full analysis of the effect on the 21 cm power spectrum and non-Gaussian statistics is deferred to future work.

Finally, the ΛCDM cosmology predicts that early dark matter halos in the affected range of scales (mainly $\sim 10^4$–$10^8\,M_\odot$) are assimilated into larger structures. Some of these early halos may still be present today as dark matter substructure, which has attracted a great deal of interest since the subhalo mass function is in principle sensitive to the primordial small-scale CDM power spectrum and hence to possible deviations from "vanilla" CDM behavior (e.g., warm dark matter, or particles that are kinetically coupled to baryons at high redshift). Unfortunately, the overall power suppression effect we describe is probably *not* detectable via substructure since the transition from an initial CDM power spectrum through the formation and survival of substructure is still not quantitatively understood (e.g., [160, 161, 162]). However, the power suppression effect is modulated by the relative velocity field, which comes primarily from large-scale modes in the primordial density fluctuations ($k \sim 0.1\,\mathrm{Mpc}^{-1}$) and hence can be reconstructed from large-scale structure surveys. Therefore it would be valuable for future work to investigate whether the $v_{\rm bc}$ effect can be detected by differential measurements that compare the substructure abundance in strong lens systems where the lens halos have similar mass but different reconstructed $\mathbf{v}_{\rm bc}$. It is also important to mention that the suppression of the formation of the first halos, which seed present day galaxies, and the decrease in the high-k power spectrum might help alleviating the known problem of the overabundance of substructure of dark matter halos in the ΛCDM model (the missing satellite problem). This effect might also be important for predictions of the annihilation signal from dark matter particles. We relegate detailed investigation of these questions to the future study.

In summary, we have shown that in the post recombination Universe, there are bulk relative motions between the baryons and dark matter that are supersonic and are coherent over scales of several comoving megaparsecs. The combined growth of small-scale structure (between the baryon Jeans length $\lambda_{\rm J}$ and $\mathcal{M}\lambda_{\rm J}$, where $\mathcal{M}$ is the Mach number of relative motion) is suppressed due to the baryons advecting out of the potential wells created by the dark matter. We find at lower redshifts (e.g., $z \sim 40$) a suppression of the power spectrum by $\sim 10\%$ on scales of 50–500 $\rm Mpc^{-1}$ that is highly spatially variable. The suppression results in some reduction in the abundance of early halos, but more importantly changes their spatial structure, leading to scale-dependent bias and stochasticity of the first halos. These latter effects may be large for early low-mass, high-bias halos, e.g., we find squared correlation coefficients $\chi = r_{\rm hm}^2$ as small as ~ 0.2 at $z = 40$. Whether this unusual clustering pattern affects the spatial morphology of reionization depends on the importance of low-mass halos and hence is unknown at this time, although we note that future 21 cm observations combined with simulations that distribute their sources of ionizing radiation in different ways may shed some light on this issue. Farther in the future, the power suppression effect would certainly be significant for the interpretation of any pre-reionization 21 cm signal. In any case, our analysis highlights the importance of reconsidering "standard" notions of structure formation (e.g., linear bias of halos on large scales) as we enter new physical regimes at high redshift.

Acknowledgments

We thank Neal Dalal, Eiichiro Komatsu, Kiyoshi Masui, Leonidas Moustakas, Michael Kuhlen and Ue-Li Pen for helpful conversations. Work of D.T. and C.H. on this project was supported by the U.S. Department of Energy (DE-FG03-92-ER40701) and the National Science Foundation (AST-0807337). Work of C.H. was also supported by the Alfred P. Sloan Foundation.

Chapter 4

Applications of the Relative Velocity Effect*

The relative velocity effect introduced in the previous chapter has drawn significant attention from the astrophysics community and resulted in a large number of follow up projects which investigated various applications of the effect. In this chapter we discuss a number of applications studied by our group as well as others. Specifically we focus on the effect of the relative velocity on the evolution of the first galaxies, minihalos and the first stars in the early Universe.

We study the gas content and analyze the effect of relative velocity on the properties of halos over a wide range of halo masses and redshifts. We calculate accurately the linear evolution of the baryon and dark matter fluctuations, and quantify the resulting effect on halos based on an analytical formalism that has been carefully checked with simulations in the case with no relative velocity. We estimate the effect on the abundance of early halos and the gas fraction in them. We find that the relative velocity effect causes several changes: (i) the characteristic mass that divides gas-rich and gas-poor halos is increased by roughly an order of magnitude, from $2 \times 10^4\, M_\odot$ to about $2 \times 10^5\, M_\odot$; (ii) this characteristic mass has a large scatter (full width at half maximum is $\sim 1.5 \times 10^5 M_\odot$ at $z = 20$); (iii) the fraction of baryons in star-less minihalos is suppressed by a factor of 3.3 at $z = 20$; (iv) the fraction of baryons in halos that can cool and form stars is suppressed by a factor of 1.5 at $z = 20$; and (v) there are enhanced spatial variations of these various fractions.

Further we study the impact of this relative motion on the distribution of the star-forming halos and on the formation redshift of the very first stars. We discuss recent efforts by a number of research groups to simulated the relative velocity effect and include a new aspect of the relative velocity effect found in the simulations by fitting their results to obtain the spatially-varying minimum halo mass needed for molecular cooling. We emphasize that the relative velocity effect has three separate components: suppression of the halo abundance, suppression of the gas content within each halo, and boosting of the minimum cooling mass. We show that the gas suppression primarily affects the smaller minihalos that cannot form stars, while the other

---

[*] The material in this chapter was adapted from *Suppression and spatial variation of early galaxies and minihalos*, D. Tseliakhovich, R. Barkana and C. Hirata, *Mon. Not. R. Astron. (2011).* Reproduced here with permission of the publisher. Material in this chapter is significantly expanded compared to the original paper to include more detailed explanations, and additional plots. We also extended the discussion section to include the most recent results on the topic.

two effects together produce order unity fluctuations in stellar density. In addition, we estimate the redshift of the first star to be $z \sim 65$, which includes a delay of $\Delta z \sim 5$ due to the relative velocity.

4.1 Relative Velocity and the Formation of the First Cosmic Structures

One of the most important questions in astrophysics today is understanding the formation and evolution of the first bound structures. This question is actively studied through theoretical and observational work as well as through advanced numerical simulations. Significant theoretical and observational efforts are devoted to understanding properties of the first galaxies and minihalos, at what redshifts they form and how they influence the epoch of reionization. Observations, most notably of the cosmic microwave background (CMB), have established the basic parameters for the initial conditions for structure formation [163], thus providing a foundation for theoretical work on the first structures. At the same time advances in computation have made it possible to simulate the formation of the first stars [164, 165, 166, 167, 168, 169]. Meanwhile, several efforts are underway to probe the structure of the intergalactic medium (IGM) during the reionization epoch using the 21 cm line of hydrogen, and second-generation experiments may be able to explore the early stages of reionization.

In studying the formation of the first structures it is convenient to identify two major classes of the early-type objects. The first class consists of large halos in which the gas can cool and form stars; these are the presumed sites of the first dwarf galaxies, which represent the first sources of metals in the Universe, and provide ultraviolet photons that begin the decoupling of the hydrogen spin temperature from the CMB [170] and eventually start the epoch of reionization. The second class consists of smaller halos ("minihalos") that are too small for molecular cooling, but still affect the epoch of reionization by acting as sinks for ionizing photons [171, 172, 173, 174] and may generate a 21 cm signal from collisional excitation of H I (e.g., [175, 176]). It is important to understand both the abundance and distribution of halos, as well as the precise boundaries separating halos that undergo cooling and star formation, those that collect baryons in their potential wells but do not cool, and the lightest halos that exist only as dark matter structures and do not collect gas.

In this work we study the formation of the first galaxies and minihalos in the light of an important effect of the relative velocity of dark matter and baryonic fluids [121] that was previously overlooked. This effect leads to power suppression on scales that correspond to the first bound halos between $10^4\ M_\odot$ and $10^8\ M_\odot$ and delays the formation of the first objects. More importantly this effect introduces scale-dependent bias and stochasticity, leading to significant qualitative changes in the distribution of the first objects. The relative velocity effect is especially important on the small scales where the first stars and galaxies form. Introduction of this effect dramatically changes the gas distribution inside the first halos and changes the characteristic mass of gas-rich objects. N. Dalal, U. L. Pen, and U. Seljak [177] recently calculated analytically the effect on the gas content of halos and found a large effect on the fluctuations of the Lyman-α background at high redshifts. Their analysis, however, was based on a very simplified model of which halos can form stars and in what abundance. In this study we carry out a detailed analytical study of the distribution of gas and stars in the first halos.

Furthermore, in this study we use a more robust set of initial conditions compared to [121] and remove a number of simplifying assumptions. We emphasize important role of proper initial conditions following a detailed study by [178], who showed that three commonly used initial condition setups lead to significantly different abundances and properties of the first star-forming gas clouds as well as the first minihalos.

The rest of this chapter is organized as follows. Section 4.2 reviews the relative velocity effect (section 4.2.1) and improves the analysis of [121] to account for spatial variation of the sound speed (section 4.2.2). Section 4.3 then investigates the early halos and their gas content, focusing on computation of the filtering mass (section 4.3.1) and then examining the fraction of baryons in minihalos and in larger halos that can cool, including an analysis of spatial variations in the baryon budget (sections 4.3.2 and 4.3.3). We summarize our results in section 4.4 and provide an overview of the latest results on the relative velocity effect including simulations and advanced analytical studies.

The numerical results and plots shown in this chapter assume a cosmology with present-day baryon density $\Omega_{\rm b,0} = 0.044$, CDM density $\Omega_{\rm c,0} = 0.226$, dark energy density $\Omega_{\Lambda,0} = 0.73$, Hubble constant $H_0 = 71\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$, and adiabatic primordial perturbations of variance $\Delta^2_\zeta(k_\star) = 2.42 \times 10^{-9}$ at $k_\star = 0.002\ \mathrm{Mpc}^{-1}$, and with slope $n_s = 0.96$.

4.2 Initial Conditions for Halo Formation

In this section we detail the formalism necessary for generation of correct initial conditions taking into account two important effects that are often overlooked in the literature. First of all we introduce the effect of relative velocity of dark matter and baryonic fluids after recombination. This effect, first studied in [121], is nominally a second-order effect in the perturbation theory of density evolution and hence has been ignored in studies based on the linear theory. Secondly, we emphasize the importance of a correct treatment of the sound speed variations in the time between recombination and $z \sim 200$ due to residual Compton heating of the electrons on the CMB photons. As we show later in the chapter, both effects play a significant role during the epoch of first halo formation and dramatically impact gas fractions in the first bound objects.

4.2.1 Relative Velocity of Dark Matter and Baryonic Fluids

Before recombination, the baryons are tightly coupled to the photons and the sound speed is highly relativistic, $\sim c/\sqrt{3}$. As the universe expands and cools the electrons recombine with the protons and the universe becomes transparent, leading to a kinematic decoupling of the baryons from the radiation around $z_{\rm dec} \approx 1000$. After recombination the sound speed of the baryons drops precipitously down to thermal velocities, whereas the dark matter velocity remains unaffected, and after recombination the dark matter motion with respect to the baryons becomes supersonic. The relative velocity can be written as

$$\mathbf{v}_{\rm bc}(\mathbf{k}) = \frac{\hat{\mathbf{k}}}{ik}[\theta_{\rm b}(\mathbf{k}) - \theta_{\rm c}(\mathbf{k})], \tag{4.1}$$

where $\hat{\mathbf{k}}$ is a unit vector in the direction of $\mathbf{k}$, and $\theta \equiv a^{-1}\nabla \cdot \mathbf{v}$ is the velocity divergence (we use comoving coordinates).

The variance of this relative velocity is

$$\begin{aligned} \langle v_{\rm bc}^2(\mathbf{x})\rangle &= \int \frac{dk}{k}\Delta_\zeta^2(k)\left[\frac{\theta_b(k) - \theta_c(k)}{k}\right]^2 \\ &= \int \frac{dk}{k}\Delta_{\rm vbc}^2(k), \end{aligned} \tag{4.2}$$

where $\Delta_\zeta^2(k) = 2.42 \times 10^{-9}$ is the initial curvature perturbation variance per $\ln k$. Integration of equation (4.2) at the time of recombination[†] ($z_{\rm rec} = 1020$) shows that the dark matter moves relative to the baryons with root-mean-square velocity of $\sim$30 km s^{-1} corresponding to a Mach number of $\mathcal{M} \equiv v_{\rm bc}/c_{\rm s} \sim 5$. This supersonic relative motion allows baryons to advect out of the dark matter potential wells and significantly suppresses the growth of structure at wave numbers higher than

$$k_{\rm vbc} \equiv \left.\frac{aH}{\langle v_{\rm bc}^2\rangle^{1/2}}\right|_{\rm dec} = \frac{k_{\rm J}}{\mathcal{M}} \sim 40\,{\rm Mpc}^{-1}, \tag{4.3}$$

were $k_{\rm J}$ is the Jeans wave number.

As shown in [121] the relative velocity of the baryons and cold dark matter (CDM) is coherent over scales of several comoving Mpc and the velocity in each coherence region is well described by a three-dimensional Gaussian probability distribution with variance

$$\sigma_{\rm vbc}^2 = \left\langle |\mathbf{v}_{\rm bc}(\mathbf{x})|^2 \right\rangle. \tag{4.4}$$

(Note that this is the *total* variance, i.e.,, including velocities in all 3 directions; the variance per axis is smaller by a factor of 3.)

To see how the relative motion of baryons and dark matter affect the formation of the first objects, we need to solve a system of evolution equations that incorporate this effect. The system of equations describing a high-k perturbation mode in the presence of a background

---

[†] Technically the effective redshift of kinematic decoupling [179], since recombination is an extended process.

relative velocity is

$$\begin{aligned}
\frac{\partial \delta_{\rm c}}{\partial t} &= \frac{i}{a}\mathbf{v}_{\rm bc}^{\rm (bg)} \cdot \mathbf{k}\delta_{\rm c} - \theta_{\rm c}, \\
\frac{\partial \theta_{\rm c}}{\partial t} &= \frac{i}{a}\mathbf{v}_{\rm bc}^{\rm (bg)} \cdot \mathbf{k}\theta_{\rm c} - \frac{3H^2}{2}(\Omega_{\rm c}\delta_{\rm c} + \Omega_{\rm b}\delta_{\rm b}) - 2H\theta_{\rm c}, \\
\frac{\partial \delta_{\rm b}}{\partial t} &= -\theta_{\rm b}, \text{ and} \\
\frac{\partial \theta_{\rm b}}{\partial t} &= -\frac{3H^2}{2}(\Omega_{\rm c}\delta_{\rm c} + \Omega_{\rm b}\delta_{\rm b}) - 2H\theta_{\rm b} + \frac{c_{\rm s}^2 k^2}{a^2}\delta_{\rm b}.
\end{aligned} \tag{4.5}$$

The $v_{\rm bc}$ terms are nominally second-order in perturbation theory, and hence one may wonder why they, rather than other second-order terms, are included. The reason is that the expansion parameter for these terms is not the density perturbation δ, but rather the ratio of the advection terms (e.g., $v_{\rm bc}^{\rm (bg)} k\delta_c/a$ in the δ_c equation) to the linear terms (e.g., $\partial\delta_c/\partial t \sim \delta_c/H$). This ratio is

$$\frac{v_{\rm bc}^{\rm (bg)} k}{aH}. \tag{4.6}$$

One can see that this expansion parameter increases as one goes to smaller scales and is of order unity at $k \sim k_{\rm vbc}$. Thus the $v_{\rm bc}$ terms become nonperturbative at small scales $k > k_{\rm vbc}$, and when treating these small scales one must keep these terms even if they are formally higher order in the perturbation theory.

4.2.2 Complete Heating Model

The system of equations of equation (4.5) assumes a spatially uniform sound speed which is a good first-order approximation. However, as shown in [180], it underestimates baryon density fluctuations by more than 30 percent at $z = 100$ and 10 percent at $z = 20$ for large wavenumbers. A fully correct treatment of baryon density evolution requires analysis of the Compton heating from the CMB on the sound speed and fluctuations in the temperature distribution. Following [180], we rewrite the sound speed term of the last equation of equation (4.5) as

$$\frac{c_{\rm s}^2 k^2}{a^2}\delta_{\rm b} \rightarrow \frac{k^2}{a^2}\frac{k_B \bar{T}}{c^2 \mu m_H}(\delta_{\rm b} + \delta_{\rm T}), \tag{4.7}$$

where $\delta_{\rm T}$ is the temperature perturbation which evolves as

$$\frac{d\delta_T}{dt} = \frac{2}{3}\frac{d\delta_{\rm b}}{dt} + \frac{x_e(t)}{t_\gamma}a^{-4}\left\{\delta_\gamma\left(\frac{\bar{T}_\gamma}{\bar{T}} - 1\right) + \frac{\bar{T}_\gamma}{\bar{T}}\left(\delta_{T_\gamma} - \delta_T\right)\right\}. \tag{4.8}$$

The second term on the right-hand side accounts for the Compton scattering of the CMB photons on the residual electrons from recombination. Here $x_e(t)$ is the electron fraction relative to the total number density of gas particles[‡], $\bar{T}_\gamma = [2.725 \text{ K}]/a$ is the mean CMB temperature, and

$$t_\gamma^{-1} \equiv \frac{8}{3}\rho_\gamma^0 \frac{\sigma_T\, c}{m_e} = 8.55 \times 10^{-13} \text{yr}^{-1}, \tag{4.9}$$

[‡] This is different from the recombination literature, which often takes x_e to be normalized to the number of hydrogen nuclei. At low redshifts these differ by 8 per cent due to the presence of helium.

where $\sigma_{\rm T}$ is the Thomson scattering cross section, ρ_γ^0 is the photon energy density at $z = 0$, and $\bar{T}$ is the average temperature of the baryons, which can be calculated using the first law of thermodynamics:

$$\frac{d\bar{T}}{dt} = -2H\bar{T} + \frac{x_e(t)}{t_\gamma}\,(\bar{T}_\gamma - \bar{T})\,a^{-4}. \tag{4.10}$$

Accounting for Compton heating of the residual electrons by the CMB photons is especially important on small scales ($k > 1$ Mpc^{-1}), which are also impacted by the relative motion effect.

4.3 First Halos and Their Gas Content

Both of the effects discussed above have a significant impact on the evolution of density perturbations on small scales and affect the formation of the first dark matter halos, as well as the subsequent accretion of the baryons and the formation of the first stars. We investigate the specific effects by studying the change in the characteristic mass scale that divides gas-rich and gas-poor halos produced by the relative velocity of the dark matter and baryonic fluids.

4.3.1 Filtering Mass

In the ΛCDM universe, virialized dark matter halos form hierarchically on extremely small scales at very early times and start accreting baryons into their potential wells. If halos are heavy enough, accretion proceeds to the point where baryons start cooling through molecular line emission, condensing into the first stars and galaxies. This accretion is counteracted by the bulk motion of baryons with respect to dark matter as well as by the thermal gas pressure. The combination of the two effects leads to the presence of the minimal halo mass scale at which baryons are still able to effectively accrete onto a halo.

To study the effect of halo formation and baryonic accretion it is convenient to consider a large ensemble of patches of the size of the relative velocity coherence scale (~ 3 Mpc across). Each patch has certain mean density and bulk velocity, determined by linear initial conditions at the time of recombination. Bulk velocity for each patch is drawn from a Gaussian random distribution with the variance determined by equation (4.4), whereas mean density is generated using CMBFAST ([139]) power spectrum. We follow the evolution of density perturbations in each patch from the time of recombination to some later redshift z, including the spatial variation of the baryonic speed of sound due to Compton heating from the CMB [180].

By evolving the system of equations (4.5) with the correct sound speed term of equation (4.7) in each patch, characterized by a fixed value of $v_{\rm bc}$, we calculate the baryonic and dark matter power spectra. Their ratio is constant on large scales (small k), and drops at high k due to the suppression of growth by the baryonic pressure. Gneding and Hui [181] originally defined a "filtering" scale (essentially a time-averaged Jeans scale) that they used to identify the largest scale on which the baryon fluctuations are substantially suppressed compared to those of the dark matter. We use the generalized definition from [182], in which the baryon-to-total ratio is

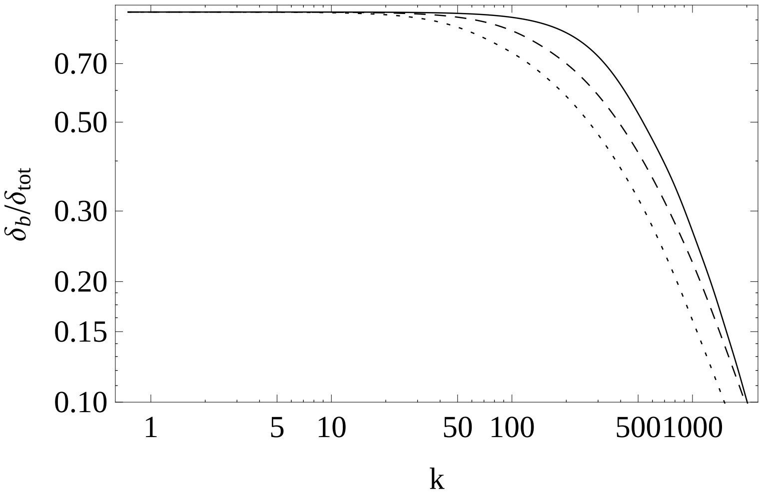


Figure 4.1: Perturbation ratio $\delta_b/\delta_{\rm tot}$ vs. comoving wavenumber k evaluated at $z = 20$ for the cases of $v_{\rm bc} = 0$ (solid curve), $v_{\rm bc} = 1\sigma_{vbc}$ (dashed curve), and $v_{\rm bc} = 2\sigma_{vbc}$ (dotted curve). In all cases overdensities are isotropically averaged over the direction of $\mathbf{k}$ with respect to $\mathbf{v}_{\rm bc}$.

expanded to linear order in k^2, and written in the following form:

$$\frac{\delta_{\rm b}}{\delta_{\rm tot}} = 1 - \frac{k^2}{k_F^2} + r_{\rm LSS}, \tag{4.11}$$

where the total density perturbation $\delta_{\rm tot} = f_{\rm b}\delta_{\rm b} + f_{\rm dm}\delta_{\rm dm}$ (in terms of the cosmic baryon and dark matter mass fractions $f_{\rm b}$ and $f_{\rm dm}$), and the k-independent $r_{\rm LSS}$ term (which is negative) describes the relative baryon-to-total difference in the limit of large-scale structure, i.e.,, where both the $v_{\rm bc}$ effect and the thermal pressure of the gas are negligible (and restricted also to scales below the baryon acoustic oscillations).

In figure 4.1 we plot isotropically averaged perturbation ratio $\delta_b/\delta_{\rm tot}$ by averaging over the direction of $\mathbf{k}$ with respect to $\mathbf{v}_{\rm bc}$. On large scales the ratio is very close to constant, and using equation (4.11) we can deduce the value of $r_{\rm LSS}$ at various redshifts. For example, $r_{\rm LSS} = -0.054$ at $z = 20$. Here we note that since $r_{\rm LSS}$ is determined by the values of $\delta_b/\delta_{\rm tot}$ at large scales (low values of k) its value remains unaffected by the relative velocity effect. We plot the evolution of the $r_{\rm LSS}$ with redshift in figure 4.2.

The filtering scale k_F is obtained by fitting equation (4.11) to the calculated values of the ratio $\delta_b/\delta_{\rm tot}$ from figure 4.1. This allows us to define the filtering mass in terms of the filtering wavenumber:

$$M_F = \frac{4\pi}{3}\bar{\rho}_0 \left(\frac{\pi}{k_F}\right)^3, \tag{4.12}$$

where $\bar{\rho}_0$ is the mean matter density today. We note that this relation is $\frac{1}{8}$ of the definition originally used by [183], who also used a nonstandard definition of the Jeans mass.

The filtering mass plays an extremely important role in understanding the evolution of the first halos, as it provides a good approximation for the boundary between gas-rich halos and halos that do not contain substantial quantities of gas. Traditionally one would assume that

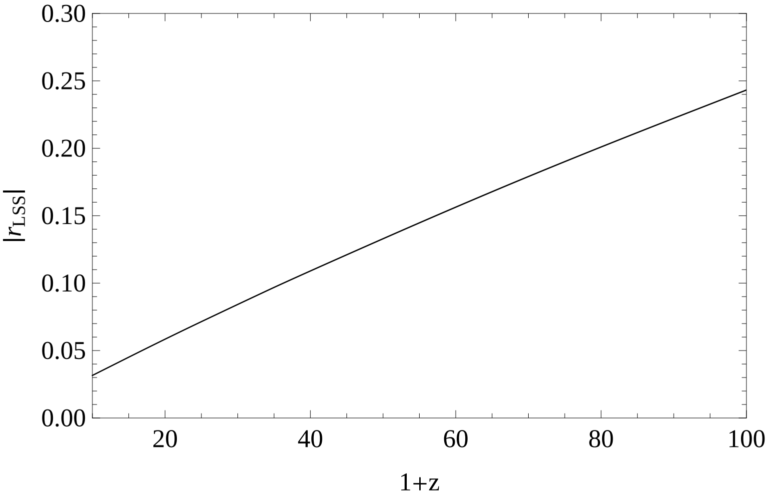


Figure 4.2: Evolution of the $|r_{\mathrm{LSS}}|$ with redshift.

the separation between gas-rich and gas-poor objects is represented by the Jeans scale, which is the minimum scale on which a small gas perturbations will grow due to gravity overcoming the pressure gradient. However, the Jeans scale is related only to the evolution of the perturbations at a given point in time and does not account for significant variation of the Jeans mass with time. The filtering mass on the other hand reflects the baryonic pressure effects integrated over the entire history of the Universe, and provides a much better approximation to the boundary between gas-rich and gas-poor halos.

An extensive study of the filtering mass properties and evolution history in models without the relative velocity effect was performed in [180] and [182]. The properties of the filtering mass, however, are significantly modified in the regions where the bulk motion of baryons with respect to dark matter potential wells is significant. In regions with high values of v_{bc} baryons tend to advect out of the collapsing dark matter halos, significantly increasing the filtering mass. We demonstrate this in figure 4.3 where we plot the evolution of the filtering mass with redshift in the regions with $v_{\mathrm{bc}}/\sigma_{\mathrm{vbc}} = 0$, 1, and 2. We also show the globally averaged case by integrating the filtering mass over the full probability distribution of the relative velocity, given by:[§]

$$P_{\mathrm{vbc}}(v) = \left(\frac{3}{2\pi\sigma_{\mathrm{vbc}}^2}\right)^{3/2} 4\pi v^2 \exp\left(-\frac{3v^2}{2\sigma_{\mathrm{vbc}}^2}\right). \tag{4.13}$$

As noted earlier, the variance per axis is $\sigma_{\mathrm{vbc}}^2/3$.

In figure 4.3 we also compare the filtering mass with the Jeans mass defined as

$$M_J = \frac{4\pi}{3}\bar{\rho}_0\left(\frac{\pi}{k_J}\right)^3, \tag{4.14}$$

where $k_J = \sqrt{2/3}aH/c_{\mathrm{s}}$ is the Jeans scale (defined by setting the right-hand side of equation (4.5) to zero, without the relative velocity term, and neglecting here the correction of

---

[§] This is the distribution of the magnitude of v_{bc}, where the vector $\mathbf{v}_{\mathrm{bc}}$ is the result of linear perturbations and hence is drawn from a multivariate Gaussian. It thus happens to be the same as the Maxwell-Boltzmann distribution, even though the bulk velocities of baryons have nothing to do with thermal motions of particles.

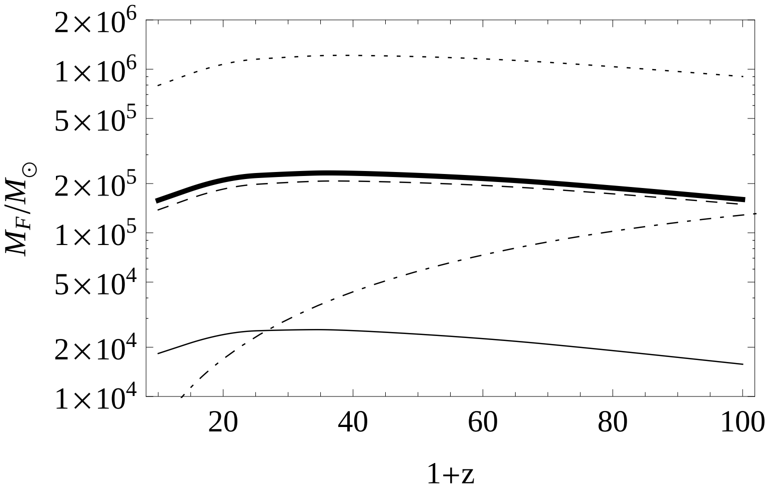


Figure 4.3: Evolution of the filtering mass with redshift in the regions with $v_{\rm bc} = 0$ (thin solid line), $v_{\rm bc} = 1\sigma_{vbc}$ (dashed line), $v_{\rm bc} = 2\sigma_{vbc}$ (dotted line) and global average case (thick solid line). We also show the evolution of the Jeans mass M_J (dot-dashed line).

$v_{\rm bc}/\sigma_{v_{\rm bc}}$	$P(> v_{\rm bc})$	k_F (Mpc^{-1})	M_F ($M_\odot$)
4	2.1×10^{-10}	85	7.75×10^6
3	5.9×10^{-6}	113	3.37×10^6
2	7.4×10^{-3}	166	1.07×10^6
1	0.392	298	1.85×10^5
0	1	591	2.39×10^4

Table 4.1: Filtering scale and filtering mass for the isotropic averaging of the direction of $\mathbf{k}$ with respect to $\mathbf{v}_{\rm bc}$ at $z = 20$.

equation (4.7)). Figure 4.3 shows that the filtering mass reaches a maximum value at redshift $z \sim 40$ (and generally varies only slightly throughout the plotted redshift range), whereas the Jeans mass continuously decreases with time due to the drop in the sound speed of the gas as the Universe cools.

The filtering mass represents a time averaged Jeans mass and hence it decreases at the low redshifts, however, right after recombination baryonic perturbations on small scales are highly suppressed and they only catch up gradually, causing the filtering mass to increase from low initial values. We emphasize that in the regions with a high value of the relative velocity the filtering mass is significantly larger than in the regions with small values of $v_{\rm bc}$ and hence the formation of gas-rich objects in those regions proceeds quite differently than in the regions with $v_{\rm bc} \sim 0$. The filtering scale and mass (from equations (4.11) and (4.12)) in regions with varying values of $v_{\rm bc}$ are given in table 4.1, and the filtering masses at $z = 20$ and $z = 40$ are plotted in figure 4.4. In figure 4.5, we also show the dependence of the filtering mass on the angle θ between the direction of $\mathbf{v_{bc}}$ and that of the wavevector $\mathbf{k}$ in regions where $v_{\rm bc} = 2\sigma_{vbc}$ at $z = 20$; the plot shows that the main contribution to the filtering mass comes from the regions where the wavenumber $\mathbf{k}$ and the relative velocity vector $\mathbf{v_{bc}}$ are parallel (or anti parallel).

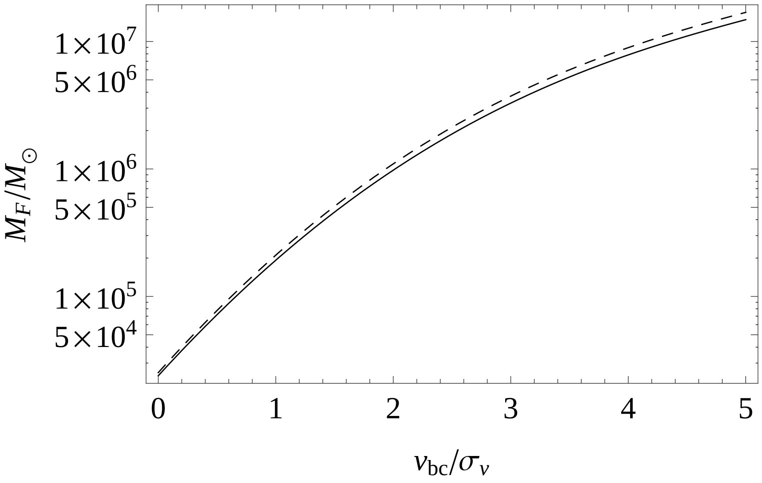


Figure 4.4: This figure shows the filtering mass M_F as a function of the relative velocity of the dark matter and baryonic fluids at $z = 20$ (solid line) and $z = 40$ (dashed line).

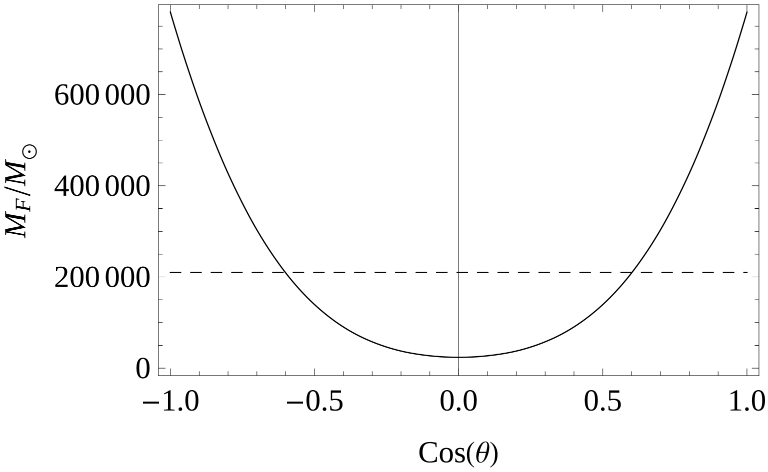


Figure 4.5: Dependence of the filtering mass on the angle θ between the directions of $\mathbf{v_{bc}}$ and the wave-vector $\mathbf{k}$ (solid line), and the isotropically averaged value of M_F (dashed horizontal line), in regions where $v_{\rm bc} = 1\sigma_{vbc}$ at $z = 20$.

4.3.2 Gas Content of the First Galaxies and Minihalos

We now investigate the amount of gas that falls into early halos, and how much of this gas is capable of cooling. Here we use analytical approximations–namely the relation between the gas mass fraction $f_{\rm g}$ and filtering mass M_F, and the [184] mass function–that have been calibrated against simulations with statistically isotropic initial conditions and no bulk relative velocity. In our case with $v_{\rm bc} \neq 0$ the power spectra are both reduced and slightly anisotropic, but we expect these approximations to still be a useful guide since statistical anisotropy (e.g., θ-dependent filtering mass) can only appear at second-order in scalars such as the halo mass function or gas content.

There is no *a priori* reason to suppose that the filtering mass, which is defined based on linear perturbations, can also accurately describe properties of highly nonlinear, virialized objects. Qualitatively we may argue that if pressure significantly opposes gravity during the halo formation process (which for some time is accurately described by linear theory) then it will significantly suppress the amount of gas in the final virialized halo. N. Gnedin [183] suggested based on simulations during cosmic reionization that the filtering mass accurately fits the mass for which a halo contains half the mean cosmic baryon fraction f_b, and fitted the simulation results to the following formula:

$$f_{\rm g} = f_{\rm b,0}\left[1 + \left(2^{\alpha/3} - 1\right)\left(\frac{M_F}{M}\right)^{\alpha}\right]^{-3/\alpha}, \tag{4.15}$$

where $f_{\rm b,0}$ is the gas fraction in the high-mass limit. In this function, a higher α causes a sharper transition between the high-mass (constant $f_{\rm g}$) limit and the low-mass limit (assumed to be $f_{\rm g} \propto M^3$). This formula has subsequently been found to agree with hydrodynamic simulations [185, 178] if we set $\alpha \approx 0.7$ and $f_{\rm b,0} = f_{\rm b}(1+3.2 r_{\rm LSS})$ [186], and use the filtering mass as defined in equation (4.12) (which differs from [183], as noted earlier). Thus, at each redshift, in each patch of the Universe we may calculate the local value of M_F and from it the gas fraction in halos of various total mass. In figure 4.6, we plot the gas fraction as a function of halo mass in regions with varying values of relative velocity at $z = 20$. It is clear that halos that would be gas-rich in the Universe with no $v_{\rm bc}$ effect become gas-poor in the regions where the relative velocity is high. We also see that on average, introduction of the $v_{\rm bc}$ effect significantly lowers the gas fraction in all halos with $M_h < 10^7 M_\odot$.

In order to find the total amount of gas in galaxies, we must integrate over the halo mass function in each patch. We start with the $v_{\rm bc} = 0$ case. Standard models for halo formation are based on spherical collapse calculations, in which the linear overdensity must reach a critical threshold $\delta_c(z)$ for the corresponding region to form a collapsed halo at redshift z. The halo abundance depends on the statistics of fluctuations on various scales, which can be parameterized by the function $S(R)$, the variance of fluctuations in spheres of radius R (S is also a function of

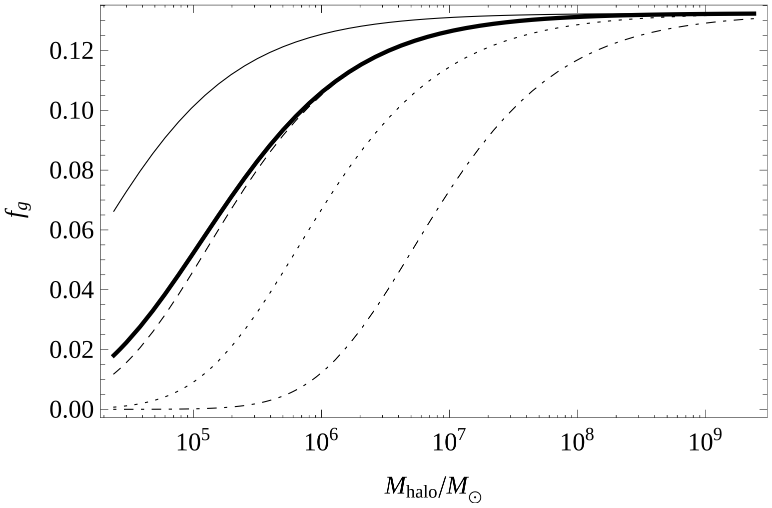


Figure 4.6: Change in the gas fraction f_g at $z = 20$ as a function of halo mass for regions with $v_{\rm bc} = 0$ (thin solid line), $v_{\rm bc} = 1\sigma_{vbc}$ (dashed line), $v_{\rm bc} = 2\sigma_{vbc}$ (dotted line), $v_{\rm bc} = 4\sigma_{vbc}$ (dash-dotted line) and isotropically averaged case (thick solid line, nearly coincident with the dashed line above $M_{halo} \sim 10^6 M_\odot$).

redshift). $S(R)$ can be written as

$$S(R) = \int \Delta_m^2(k)|W(k,R)|^2 \frac{dk}{k}, \tag{4.16}$$

where $\Delta_m^2(k)$ is the isotropically averaged local matter power spectrum and $W(k,R)$ is the window function corresponding to spheres of radius R. In this work we use top-hat window function which can be written as

$$W(k,R) = 3\left[\frac{sin(kR)}{(kR)^3} - \frac{cos(kR)}{(kR)^2}\right] \equiv 3\frac{j_1(kR)}{(kR)}. \tag{4.17}$$

We define $f(\delta_c(z), S)\, dS$ to be the mass fraction contained at z within halos with mass in the range corresponding to S to $S + dS$. We convert between halo mass M and (initial comoving) radius R using the cosmic mean density: $M = \frac{4}{3}\pi R^3 \bar{\rho}_0$. The halo abundance is then

$$\frac{dn}{dM} = \frac{\bar{\rho}_0}{M}\left|\frac{dS}{dM}\right| f(\delta_c(z), S), \tag{4.18}$$

where dn is the comoving number density of halos with masses in the range M to $M + dM$. In the model of [187],

$$f_{\rm PS}(\delta_c(z), S) = \frac{1}{\sqrt{2\pi}} \frac{\nu}{S} {\rm e}^{-\nu^2/2}, \tag{4.19}$$

where $\nu = \delta_c(z)/\sqrt{S}$ is the number of standard deviations that the critical collapse overdensity represents on the mass scale M corresponding to the variance S.

However, the Press-Schechter mass function fits numerical simulations only qualitatively, and in particular it substantially underestimates the abundance of the rare halos that host galaxies at high redshift. The halo mass function of [184], which fits numerical simulations much more

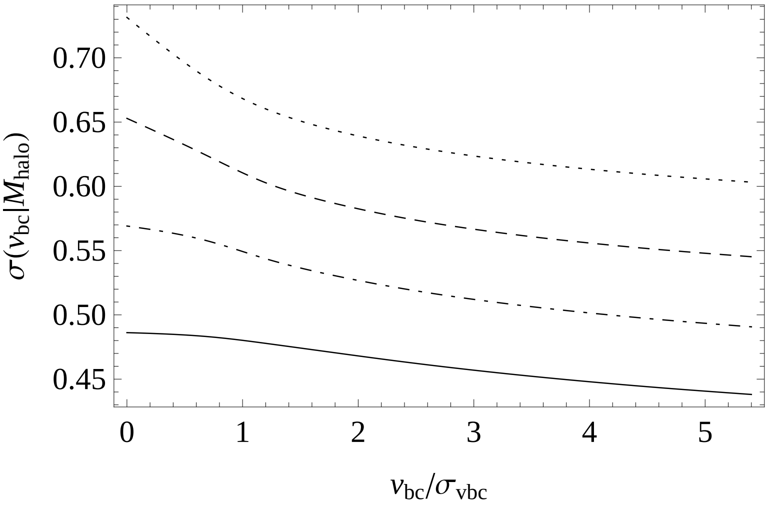


Figure 4.7: Dependence of $\sigma(v_{\rm bc}|M_h)$ on the relative velocity $v_{\rm bc}$ at $z = 20$ for a fixed mass of the collapsed halos: $M_h = 10^7 M_\odot$ (solid line), $M_h = 10^6 M_\odot$ (dot-dashed line), $M_h = 10^5 M_\odot$ (dashed line), and $M_h = 10^4 M_\odot$ (dotted line).

accurately, is given by

$$f_{\rm ST}(\delta_c(z), S) = A' \frac{\nu}{S} \sqrt{\frac{a'}{2\pi}} \left[1 + \frac{1}{(a'\nu^2)^{q'}}\right] {\rm e}^{-a'\nu^2/2}, \tag{4.20}$$

with best-fit parameters [188] $a' = 0.75$ and $q' = 0.3$, and where the normalization to unity is ensured by taking $A' = 0.322$.

The critical density of collapse $\delta_c(z)$ is independent of mass and equals 1.69 in the Einstein de-Sitter limit, valid over a wide range of redshifts. Its value decreases at low redshift due to the cosmological constant, but more relevant for this study is the decrease at very high redshift due to the effects of the baryons and radiation. The decrease is by $\sim 0.05(1+z)$ per cent from the Einstein de-Sitter value [189, 182], a small effect that is however greatly amplified by the fact that galaxies at the highest redshifts form in halos that correspond to very rare density fluctuations.

In our approach we must calculate how the halo mass function varies in different regions. The Press-Schechter model has been extended [190] to describe the variation of the halo abundance in regions of various density, and we can generalize this to include the bulk velocity by including the variation of the function $S(R)$ with $v_{\rm bc}$. To demonstrate this dependence we plot in figure 4.7 $\sqrt{S} \equiv \sigma(v_{\rm bc}|M_h)$ as a function of the relative velocity at a fixed halo mass M_h. We see that the variance of the density perturbations decreases with increasing $v_{\rm bc}$, leading to a delay in the collapse of dark matter halos. We also find that the change in σ is larger for halos of low mass since the power spectrum on scales much larger than the filtering mass is unaffected by the relative velocity.

We can summarize the important effects of the change in halo abundance and halo gas content at a given mass scale by calculating various gas fractions at each redshift. This can be done by using the Sheth-Tormen mass function and accounting for the changes due to the relative velocity effect. We calculate the fraction of the total matter density in halos above a certain mass scale by

$$f_{\rm tot}(>M_h) = \int_{M_h}^{\infty} \frac{M}{\bar{\rho}_0} \frac{dn}{dM} dM \ , \tag{4.21}$$

and the fraction of the baryon density contained in those halos using equation (4.15):

$$f_{\rm gas}(>M_h) = \int_{M_h}^{\infty} \frac{M}{\bar{\rho}_0} \frac{dn}{dM} \frac{f_g}{f_b} dM \ . \tag{4.22}$$

We plot both fractions in figure 4.8 at $z = 20$ for $v_{\rm bc}/\sigma_{\rm vbc} = 0$, 1, and 2, and for the globally averaged case, where we take into consideration the global distribution of the $v_{\rm bc}$. The plot clearly shows that in regions with high relative velocity the gas fraction in halos is dramatically suppressed. The global average (which comes out very close to the $v_{\rm bc}/\sigma_{\rm vbc} = 1$ case) gives a suppression by a factor of 2.3 of the total gas fraction in halos. In order to separate out the various effects, we plot one case in which we use the correct halo mass function (as it varies with $v_{\rm bc}$) but fix the filtering mass to the $v_{\rm bc} = 0$ value. We find that the suppression arises from comparable contributions from the change in halo numbers (about a factor of 1.8) and from the reduction in the internal halo gas fractions (about a factor of 1.5).

Stars are understood to form at high redshift out of gas that cooled and subsequently condensed to high densities in the cores of dark matter halos. Since metals are absent in the prestellar universe, the earliest available coolant is molecular hydrogen (H_2), and thus the minimum halo mass that can form a star is set by requiring the infalling gas to reach a temperature of several hundred Kelvin required for exciting H_2 to the $J \geq 2$ rotational levels [191]. This has been confirmed with high-resolution numerical simulations containing gravity, hydrodynamics, and chemical processes in the primordial gas [164, 192, 193, 194, 195]. These simulations imply a minimum halo circular velocity $V_c \sim 4.5$ km s^{-1} for forming a star, where $V_c = \sqrt{GM/R}$ in terms of the halo virial radius R. Using this approach we calculate minimum halo mass for gas cooling and plot it as a function of redshift in figure 4.9.

The simulations show that in a halo above the minimum mass (which at $z = 20$ is $M_{\rm min} \approx 6 \times 10^5 M_\odot$), the gas cools in the dense center and forms at least one star very quickly; this is understood theoretically since both the cooling time and the dynamical time are a small fraction of the cosmic age at that time. We are thus interested in the total gas fraction in halos above this cooling threshold; if there is a fixed star formation efficiency in these halos, then this gas fraction is directly proportional to the stellar density in each region.

We plot in figure 4.10 the evolution of various gas and total mass fractions as a function of redshift. Even without the relative velocity effect, there is some (spatially uniform) suppression predicted for the gas fraction in halos that can cool (i.e.,, a suppression of the overall star formation) by a factor of 1.2 at $z = 20$ and 1.5 at $z = 40$ (relative to the cosmic baryon fraction); this is due to the fact that the baryon perturbations are still catching up to the dark matter perturbations at these redshifts, even on large scales (beyond the filtering mass), and simulations suggest that nonlinear halo formation amplifies the remaining differences [178, 186].

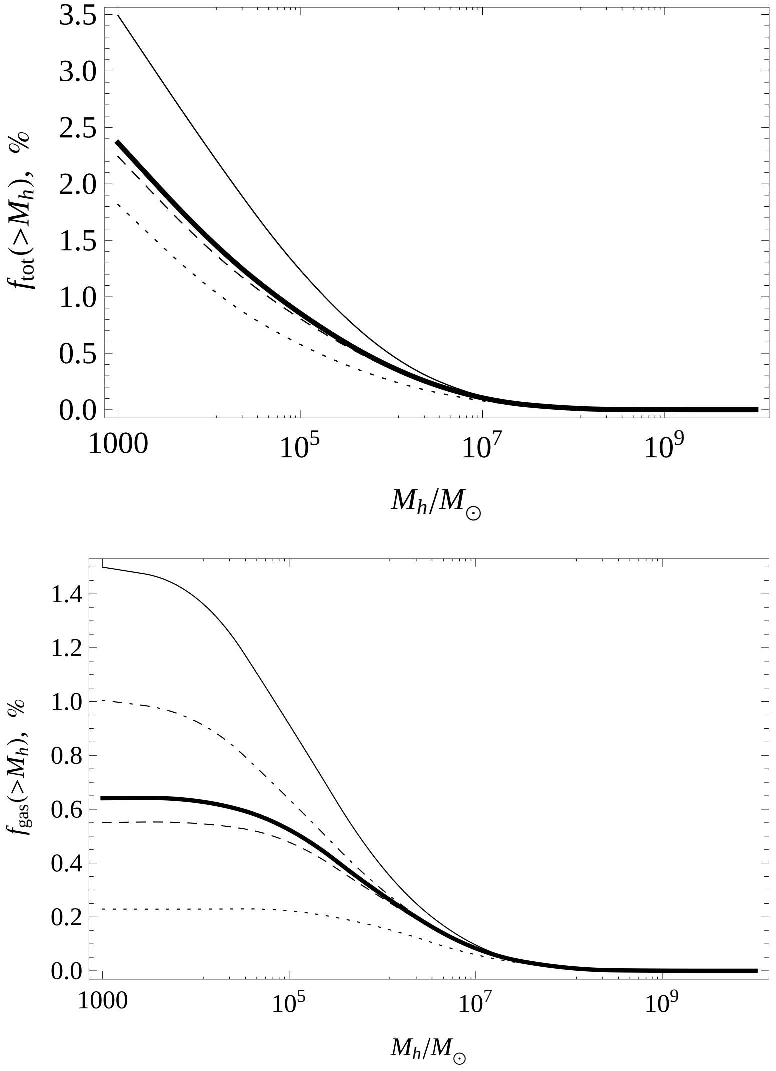


Figure 4.8: The mass fraction in halos above M_h (upper panel) and the gas fraction in halos of all masses (lower panel) at redshift $z = 20$ for the case $v_{\rm bc}/\sigma_{\rm vbc} = 0$ (thin solid line), 1 (dashed line), 2 (dotted line), and for the globally averaged case (thick solid line). In the lower panel we also show the case where we fix the value of M_F as calculated for $v_{\rm bc} = 0$ and use the correct globally averaged halo mass function including the variation with $v_{\rm bc}$ (dot-dashed line).

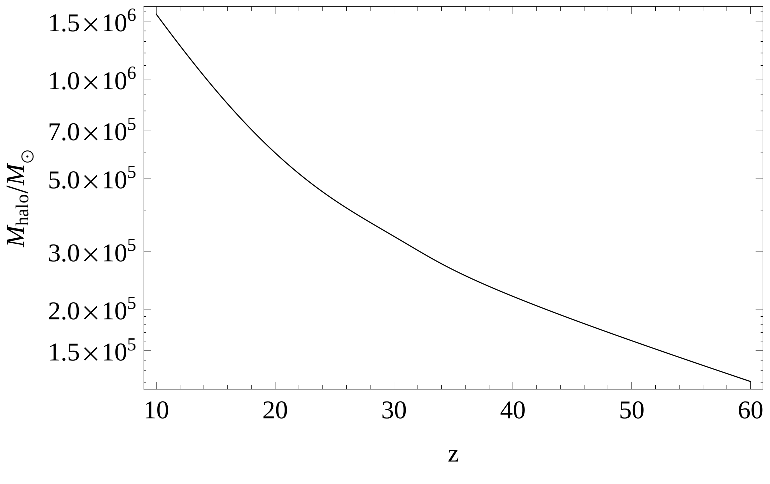


Figure 4.9: Minimum halo mass which allows efficient gas cooling and subsequent star formation.

The relative velocity effect adds an additional suppression of cosmic star formation by a factor of 1.4 at $z = 20$ and 2.1 at $z = 40$. The relative velocities have a larger effect on the gas in minihalos, the smaller halos that accrete gas that cannot cool. Since the total mass fraction in halos continues to increase as we consider smaller and smaller halos masses (figure 4.8), the total amount of gas in minihalos is very sensitive to the filtering mass, which is what produces the (gradual) low-mass cutoff in gas accretion onto halos. In the absence of the relative velocities, the total gas fraction in halos at $z = 20$ is 1.5×10^{-2}, consisting of 1.0×10^{-2} in minihalos and 5×10^{-3} in galaxies. At $z = 40$, these gas fractions are 2.4×10^{-5}, 2.3×10^{-5}, and 1×10^{-6}, respectively. The relative velocities, in the global average, reduce these fractions to 6.4×10^{-3}, 3×10^{-3}, and 3.4×10^{-3} at $z = 20$, and 2.8×10^{-6}, 2.4×10^{-6}, and 4×10^{-7} at $z = 40$. Note that the gas fraction above the H_2 cooling mass is really an upper limit to the gas fraction that undergoes star formation. Any significant feedback effect will raise the effective threshold for star formation, making the total gas fraction in halos correspond almost completely to starless halos (see discussion in Sec. 4.4).

4.3.3 Probability Distribution Functions

In addition to plotting mean values of various quantities, it is interesting to consider their variation in different patches across the sky. Above, we have explicitly varied $v_{\rm bc}$ but averaged over the density fluctuations (note that the density and velocity are uncorrelated at a given point). However, in order to calculate the full amount of variation of different quantities, i.e.,, the probability distribution function (PDF), we must explicitly vary both the mean density and the value of $v_{\rm bc}$ in each region.

In the absence of relative velocities, the extended Press-Schechter model gives the variation of the Press-Schechter halo mass function in regions of various mean density. No analytical generalization of this formalism is known for the more accurate Sheth-Tormen model, but [196] suggested a hybrid prescription that adjusts the abundance in various regions based on the extended Press-Schechter formula, and showed that it fits a broad range of simulation results. Generalizing this prescription to include the effect of relative velocity, we set

$$\begin{aligned} f_{\rm bias}(\delta_c(z), \bar{\delta}_{R_{LS}}, R_{LS}, M, v_{\rm bc}) = &\left[\frac{f_{\rm ST}(\delta_c(z), S'(R))}{f_{\rm PS}(\delta_c(z), S'(R))}\right] \\ &\times f_{\rm PS}\left(\delta_c(z) - \bar{\delta}_{R_{LS}}, S'(R) - S'(R_{LS})\right), \end{aligned} \tag{4.23}$$

where the mean overdensity in the patch is $\bar{\delta}_{R_{LS}}$, and for a given halo mass, the variance $S'(R)$ is calculated using the power spectrum modified by the local bulk velocity. The subtraction of $S'(R_{LS})$ accounts for the fact that $\bar{\delta}_{R_{LS}}$ arises from density modes on scales larger than the patch size, leaving only the remaining variance $S'(R) - S'(R_{LS})$ for fluctuation modes within the patch to supply the additional density needed to reach $\delta_c(z)$ and thus form a halo. We clarify that the difference in variance is computed in a similar way as the variance from the equation (4.16):

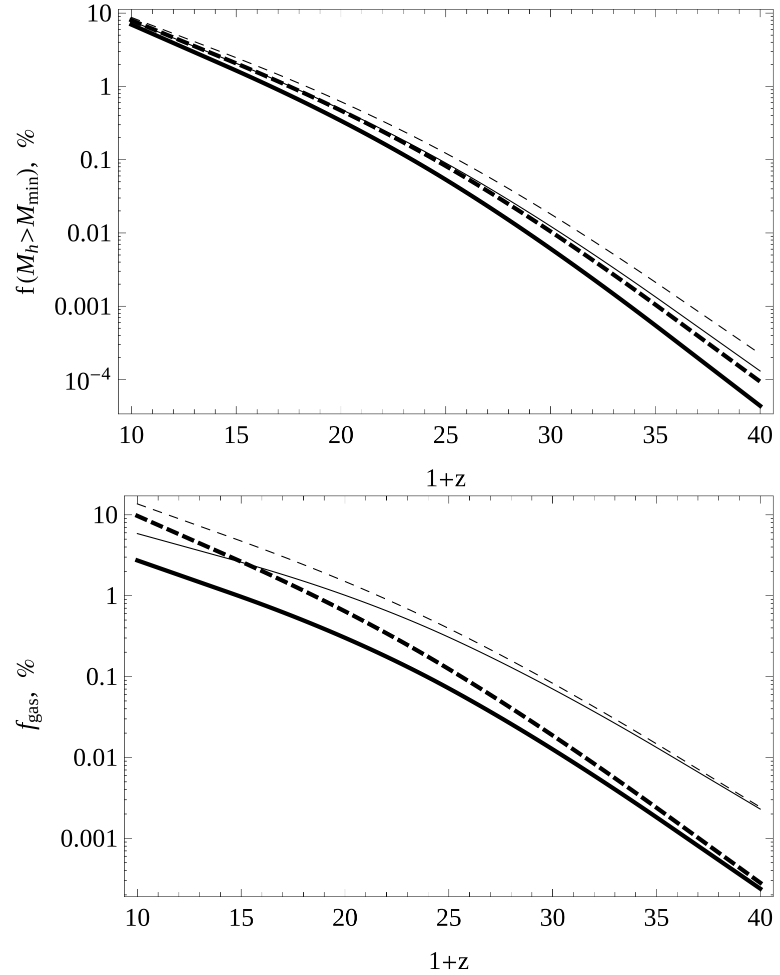


Figure 4.10: In the left panel we plot the total mass fraction in halos above the cooling mass (dashed lines) and total gas fraction in halos above the minimum cooling mass (solid lines). In the right panel we show total gas fraction in halos (dashed lines) and gas fraction in minihalos, i.e.,, in halos below minimum cooling mass (solid lines). All plots show two cases: no $v_{\rm bc}$ (regular lines) and the correct case where the $v_{\rm bc}$ effect is taken into account and the isotropic averaging is performed (thick lines).

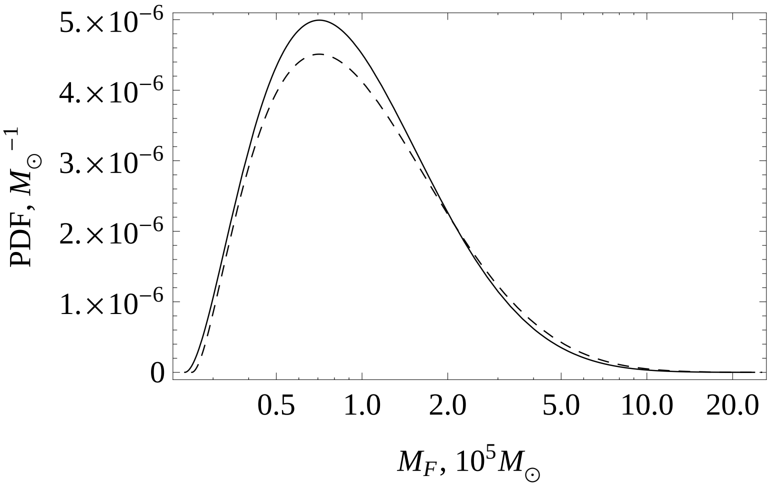


Figure 4.11: The PDF of the filtering mass M_F at $z = 20$ (solid line) and $z = 40$ (dashed line). We consider patches of radius $R_{LS} = 3$ Mpc, and include the variation of $v_{\rm bc}$ as well as the mean density in each cell.

$$S'(R) - S'(R_{LS}) = \int \Delta_m^2(k, v_{\rm bc})(|W(k,R)|^2 - |W(k,R_{LS})|^2)\frac{dk}{k}, \tag{4.24}$$

where now $\Delta_m^2(k, v_{\rm bc})$ is the isotropically averaged local matter power spectrum modified by the relative velocity effect.

In our case, the patches in which we will compute the baryon collapse fraction PDF will be spheres of radius $R_{LS} = 3$ Mpc (comoving). Note that if used to compute a mass function, the above formula gives the Lagrangian halo number density, while the Eulerian density is larger by a factor of $1 + \bar{\delta}_{R_{LS}}$; however no such transformation is necessary to compute the local *fraction* of gas in halos.

We start by calculating the PDF for the filtering mass M_F. In the scenario without $v_{\rm bc}$ we would have a universal value of M_F, however, since various regions of space have different values of relative velocity of baryonic and dark matter fluids this produces a variation in M_F. The distribution of relative velocities is given by equation (4.13) and it translates into the distribution of M_F using

$$P_{M_F}(M_F) = P_{vbc}(v_{\rm bc})\frac{dv_{\rm bc}}{dM_F}. \tag{4.25}$$

The PDF of the filtering mass at $z = 20$ and $z = 40$ is plotted in figure 4.11. These distribution functions are essentially determined by the distribution of the relative velocity and exhibit clear peaks which correspond to values of the filtering mass around the maximum of the $v_{\rm bc}$ distribution, which occurs at $v_{\rm bc} \approx 1.2\sigma_{vbc}$. As noted before (Figs. 4.3 and 4.4), the filtering mass does not vary much in this redshift range, but at $z = 20$ it is slightly more sharply peaked while the PDF at $z = 40$ extends more toward high values of M_F. Filtering mass has a rather significant scatter with the full width at half maximum of $\sim 1.5 \times 10^5 M_\odot$ at the redshift of $z = 20$.

To better understand global properties of the first objects we calculate probability distribution functions of the total gas fraction in halos as well as the gas fraction in halos above the minimum cooling mass $M_{\rm min}$. As noted earlier, these gas fractions are affected by the distribution of relative velocities as well as the distribution of large-scale overdensities $\bar{\delta}_{R_{LS}}$. We consider the PDF of gas fractions inside spherical patches ("cells") of radius $R_{LS} = 3$ Mpc, which are small enough that $v_{\rm bc}$ can be treated as roughly constant over a cell. We obtain the PDFs by running a Monte Carlo simulation that generates random values of $v_{\rm bc}$ using equation (4.13) and of the large-scale overdensity within the cell $\bar{\delta}_{R_{LS}}$ using a Gaussian of variance $S'(R_{LS})$. In figure 4.12 we show the PDFs of the gas fractions in halos above and below the minimum cooling mass $M_{\rm min} \approx 6 \times 10^5 M_\odot$ at $z = 20$. We also show the same distributions for the case with no $v_{\rm bc}$ effect. The figure shows how minihalos would be dominant at $z = 20$ (by a factor of 2 compared to galaxies), but since $v_{\rm bc}$ has a larger effect on the minihalos, it makes the gas content roughly equal between galaxies and minihalos at that redshift. Each PDF has a non-Gaussian extension towards high fractions (in fact, the distribution is approximately lognormal). Thus, the peak of the PDF is significantly lower than the mean value; without the relative velocity it is 0.002 for galaxies and 0.008 for minihalos, and $v_{\rm bc}$ moves it to ~ 0.0015 for both. Also, the relative velocities reduce the full width at half maximum from 0.004 (galaxies) and 0.008 (minihalos) to 0.003 for both.

4.4 Discussion and Overview of the Recent Results on the Relative Velocity Effect

In this chapter we have shown that the relative velocity of baryons and dark matter has a significant impact on the properties of the first bound objects and has to be considered in detailed studies of the epoch of reionization and especially earlier epochs. The supersonic motion of the baryonic fluid relative to the underlying potential wells created by the dark matter causes advection of small-scale perturbations by large-scale velocity flows, leading to a significant suppression of gas accretion during halo formation and dramatically increasing the characteristic mass of gas-rich objects at high redshifts ($z > 10$). In particular, instead of this characteristic filtering mass being close to the Jeans mass of $2 \times 10^4 M_\odot$ at $z = 20$, it varies among various regions from this value up to $\sim 10^6 M_\odot$, with a 1σ value (and global average) around $M_F = 2 \times 10^5 M_\odot$, i.e.,, an order of magnitude higher than without the relative velocity effect.

In our detailed treatment, we included the spatial variation of the baryonic sound speed, the suppression of baryonic perturbations on large scales, and the effect of the relative velocity, through the modified power spectrum, both on the halo mass function and the internal gas fractions in halos. In order to gauge the induced spatial variability, we further calculated the full probability distribution functions of the characteristic mass and of gas fractions inside of the first collapsed halos. These results are important for understanding of the relative velocity effect on large scales, and clearly emphasize the need for more detailed studies of the effect.

Our results significantly extend the work done recently by Dalal, Pen, and Seljak [177],

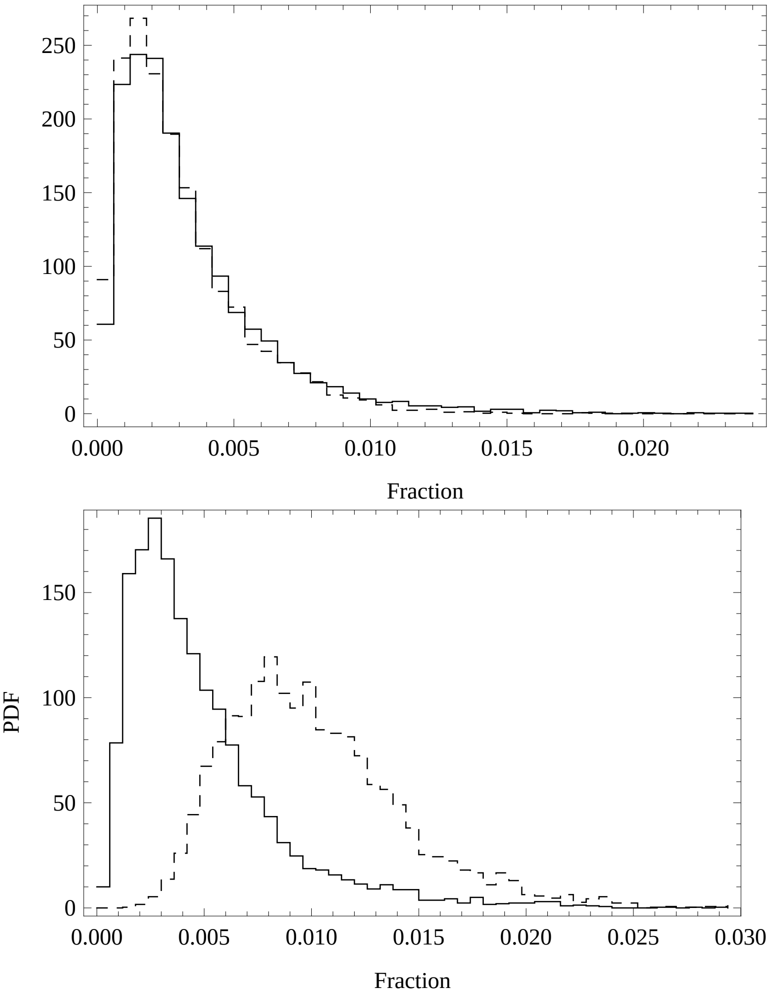


Figure 4.12: The PDFs of the gas fraction in halos above the minimum cooling mass (solid line), and the gas fraction in minihalos, i.e.,, halos below the minimum cooling mass (dashed line) at $z = 20$ for the case with the $v_{\rm bc}$ effect (upper panel) and without the effect (lower panel).

who showed that the relative velocity effect imprints a characteristic shape in the clustering power spectrum of the earliest structures, with significant power on 100 Mpc scales featuring highly pronounced baryon acoustic oscillations (BAO). This feature in the power spectrum can potentially allow probes of physics on kpc scales using observations of the 100 Mpc scales. Dalal, Pen, and Seljak also argued that the amplitude of the BAO is orders of magnitude larger at z=20 than previously expected and that this characteristic signature can allow distinguishing the effects of minihalos on intergalactic gas at times preceding and during reionization.

Our work used a more detailed treatment of the halo formation and gas accretion physics compared to the work of Dalal, Pen,and Seljak and found a suppression factor due to $v_{\rm bc}$ at $z = 20$ of 1.6 and 3.3, for star-forming halos and minihalos, respectively. In their approach they did not separate these two categories, and found a factor of 2.5 suppression in the collapsed fraction, which under their approximation can be interpreted as a suppression of star formation. Our work removed this and many other approximations used in [177]. Comparing to our work, we expect that their calculation of Lyman-α flux fluctuations is qualitatively correct but may be somewhat overestimated and requires a more detailed analysis.

4.4.1 Formation of the First Stars

The relative velocity effect also modifies the star formation history, delaying star formation and causing significant spatial fluctuations. We participated in two detailed studies of the modifications to the star formation caused by the relative velocity effect led by Rennan Barkana, Eli Visbal and Anastasia Fialkov [197, 198]. These studies significantly improved our understanding of the relative velocity effect and showed that effect of relative velocity on the star formation is not limited to the suppression of halo number density and reduction of gas content in early halos, but also includes significant boosting of the minimal cooling mass [197].

The formation of the very first stars is expected to be relatively simple compared to the star formation processes in the present-day universe due to the absence of heavy elements, simplified gas dynamics in the absence of dynamically relevant magnetic fields and feedback from luminous objects [191, 199, 164]. The first stars form in halos which become massive enough to accrete sufficient amount of gas necessary for molecular cooling [191]. The threshold for the formation of the first stars is characterized by the minimum cooling mass (M_{cool}) or equivalently by the minimum circular velocity $V_c = \sqrt{GM_{halo}/R_{vir}}$, where R_{vir} is the virial radius. In a patch with no relative motion, the mass drops rapidly with redshift, since at higher redshift the gas density is higher and a given halo mass heats the infalling gas to a higher virial temperature. The drop in M_{min} with increasing redshift in patches without the relative velocity effect is illustrated in figure 4.9. The results of our investigation presented in previous sections of this chapter and published in [200] rely on the approximation in which minimum cooling mass does not depend on relative velocity and is constant at a fixed redshift. The subsequent work led by A. Fialkov [197] removed that approximation, showing that in regions with large relative velocity M_{cool} and corresponding V_c are significantly increased. Specifically it was shown that in a region at the root-mean-square value of v_{bc} the higher bulk velocity at high redshift implies that a higher halo mass is needed for efficient molecular cooling. In particular, at redshift $z = 20$ a patch with $v_{bc} = 0$ will form stars in $3.6 \times 10^5 M_\odot$ halos, while a patch with the root-mean-square value of v_{bc} has a minimum cooling mass of $6.0 \times 10^5 M_\odot$. At $z = 60$ these numbers become $7.2 \times 10^4 M_\odot$ and $7.0 \times 10^5 M_\odot$ respectively. In patches with low bulk velocity we expect stars to form earlier, since the halos with lower masses are more abundant and form earlier in the hierarchical picture of structure formation.

In order to quantify the effect of relative velocity on the minimum cooling mass we analyzed the results of recent small scale numerical simulations of structure formation which we describe in detail in section 4.4.2. Small-scale simulations performed by Greif et al. [201], Maio et al. [202], and Stacy et al. [167] allowed fitting the minimum circular velocity necessary for a given halo to form a bound object as a function of the relative velocity. We introduced a simple, yet physically well motivated ansatz, describing the dependence of the V_c on the magnitude of relative velocity in a given region of space:

$$V_{cool}(z) = \left(V_{cool,0}^2 + [\alpha v_{bc}]^2\right)^{1/2}, \tag{4.26}$$

where $V_{cool,0}$ is the minimum halo circular velocity required for molecular cooling in the absence of the relative velocity.

The best fit to numerical simulations, which we refer to as “optimal fit” gives us $V_{cool,0} =$ 3.714 km/sec and $\alpha = 4.015$. The results of the fit are illustrated in figure 4.13. These results clearly show a substantially increased minimum cooling mass in regions with a significant relative velocity and emphasize the importance of proper consideration of the relative velocity effect in the analysis of structure formation. The effect of boosting the minimum cooling mass that changes the way first structures form is a different effect from the suppression of the amount of gas, which we introduced in section 4.3.1 and which implies a smaller number of stars in the halo at a given time; instead in this case there is a substantial delay in the formation of the first stars within halos. Moreover, this effect is not simply related to the total amount of accreted gas, since in the cases with a bulk velocity, even if we wait for the halo to accrete the same total gas mass as its no-velocity counterpart, it still does not form a star (even within the now deeper potential of a more massive host halo); the delay is substantially longer than would be expected based on a fixed total mass of accreted gas. The explanation lies with the internal density and temperature profiles of the gas, which are strongly affected by the presence of the streaming motion. The resulting delay in star formation can be explained by the fact that the first stars form from the gas that would have accreted early and formed the dense central cores in which stars form; this gas tries to accrete early (when v_{bc} is still very large) into a still-small halo progenitor, so it is affected most strongly by the suppression of gas accretion due to the bulk velocity.

The central conclusion of the work presented in [197] is that the early star formation is significantly delayed by three different manifestations of the relative velocity effect: suppression of the halo abundance, suppression of the gas content within each halo, and boosting of the

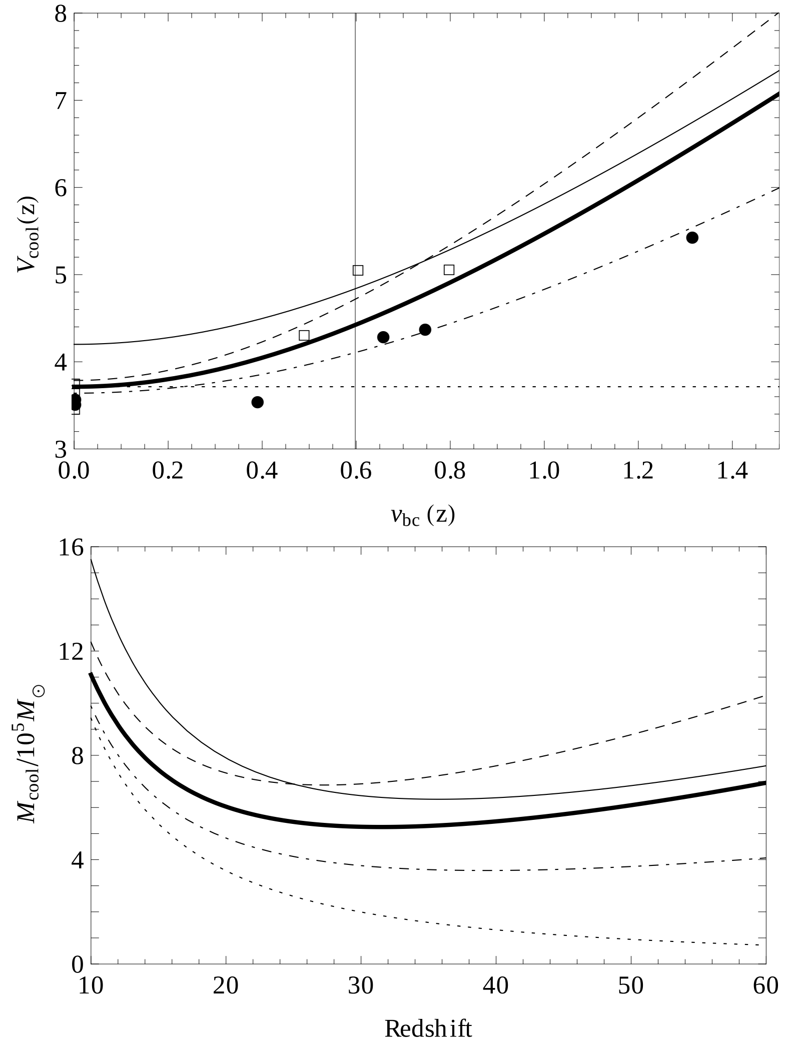


Figure 4.13: **Top panel**: The minimum halo circular velocity for gas cooling via molecular hydrogen versus the bulk velocity $v_{bc}(z)$ when the halo virializes. Data are taken from [167] and [201]. We show our fits to each set of simulation results (dot-dashed and dashed, respectively). We also show our "optimal" fit to SPH simulations (thick solid line), the "fit" to AMR simulations (regular solid line), and the case of no streaming velocity (dotted line, based on our optimal fit). The vertical solid line marks the root-mean-square value of $v_{bc}(z)$ at $z = 20$. **Bottom panel**: We show the minimum halo mass for molecular cooling versus redshift, in a patch with the root-mean-square value of $v_{bc}(z)$ at each redshift z, for each of the fits from the top panel; in particular, we show (dotted line) the case of no relative motion based on our optimal fit (i.e.,, $V_{cool} = V_{cool,0} = 3.714$ km/sec). This figure is adopted from [197] with the permission of the first author.

minimum cooling mass. Out of these three manifestations cooling mass boost and the halo abundance cut effect the star formation in the most significant way. The importance of the relative velocity on the star formation is best illustrated by considering the formation of the very first star in the observable universe. By combining our formalism with the theoretical models developed by Naoz and Barkana we show that the very first star in the observable universe is delayed by $\Delta z \sim 5$ and would form around the redshift $z = 65$ instead of the redshift $z \sim 70$ as predicted by models without the relative velocity effect.

Second collaboration led by E. Visbal also produced exciting results [198] related to the spatial distribution of the first stars and early galaxies, showing that the very first objects where formed in a cosmic web which is significantly different from the one expected in simulations without the relative velocity. By using a fully consistent hybrid approach to produce maps of the first stars and we showed unique features in the spatial distribution of the first sources of reionizing radiation and discovered that correct treatment of the initial conditions significantly modifies large-scale clustering of the first stars and produces clear observable features in all radiation fields affected by the early stars including 21 cm radiation. Figure 4.14 prepared by E. Visbal using an expanded version of the code described in chapter 3 compares distribution of the very first cosmic structures in cases with and without relative velocity effect and provides a clear intuitive picture of the impact rendered by the relative velocity.

4.4.2 Simulations of the Relative Velocity Effect

Following our initial publication [121], which introduced relative velocity effect, a number of groups attempted to advance the understanding of the implications of the effect using numerical simulations. Major simulation efforts include studies by Greif et al. [201], Maio et al. [202], Stacy et al. [167], and Naoz, Yoshida, and Gnedin [203]. These simulations significantly expand our understanding of the relative velocity effect and serve as a great foundation for larger and more detailed future simulations.

Numerical simulations face a great difficulty at high redshift, since they must resolve the then-typical tiny galaxies while at the same time capture the global galaxy distribution which is characterized by strong fluctuations on very large scales [204]. The relative velocities are correlated up to scales above 100 Mpc [198], and they are important at high redshifts where star formation is dominated by very small halos. Cosmological simulations that cover this range of scales are not currently feasible, and therefore various approximations and simplifications are necessary to obtain results that capture the importance of the relative velocity effect. One of the possible solutions is to consider boxes of the size of coherence scale of the relative velocity field. This approach allows simulations of halo formation in small patches of uniform relative velocity, and this is the route taken by most simulations groups that studied relative velocity effect so far. In table 4.2 we summarize simulation parameters from several studies.

These simulations produced extremely interesting results and clearly demonstrated suppression of the star formation. Their results, however, vary significantly in estimating the importance

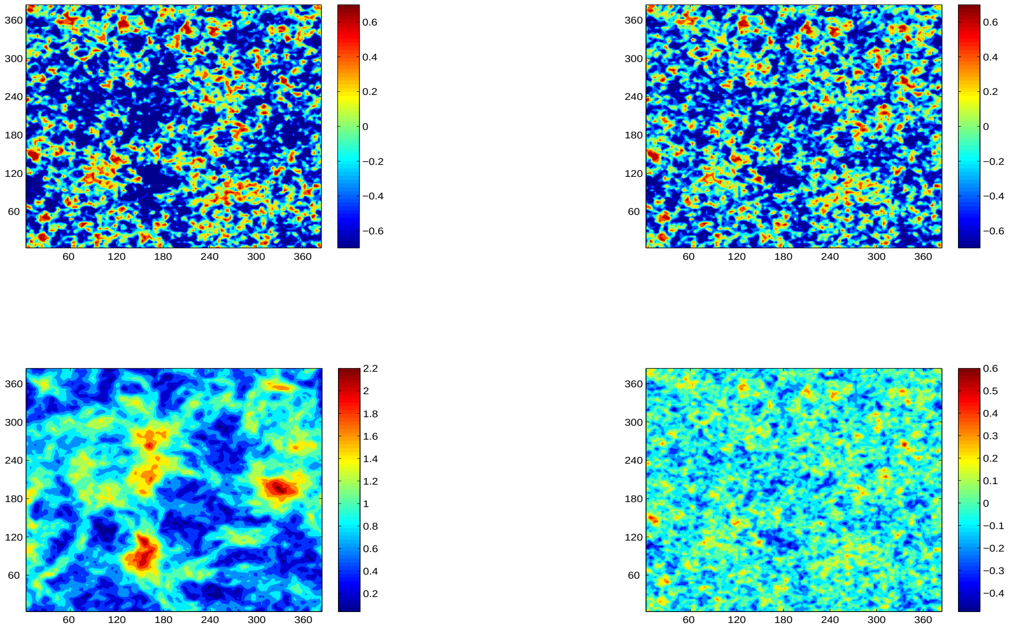


Figure 4.14: A two-dimensional slice of a volume of ~384 Mpc on a side at $z = 20$. We show the overdensity (bottom right panel), the magnitude of the relative baryon to dark-matter velocity (bottom left panel), and the gas fraction in star-forming halos, including the effect of relative velocity (upper left panel) or with the effect of density only (upper right panel). The relative velocity is given in units of the root-mean-square value. For the gas fraction, the colors correspond to the logarithm of the fraction normalized by the mean values, 0.0012 and 0.0021 for the case with and without the velocity effect, respectively; for ease of comparison, the scale in each plot ranges from 1/5 to 5 times the mean. The no-velocity gas fraction map is a biased version of the density map, while the velocity effect increases the large-scale power and the map's contrast, producing larger, emptier voids in the distribution of stars. This figure was adopted from [198] with the permission of the first author.

Study	Box Size	M_{min} of the Gas Particle	Number of particles
Maio et al. [202]	$5/h$ & $0.7/h$ Mpc	100 $M_\odot$	2×320^3
Stacy et al. [167]	$100/h$ kpc	9 $M_\odot$	2×128^3
Greif et al. [201]	500 kpc	272 $M_\odot$	2×256^3

Table 4.2: Parameters of the simulations of the relative velocity effect produced by various groups [167], [201], [202]

of the relative velocity effect and appear at first glance to show a smaller suppression effect than we predict using analytical tools. This difference is not surprising if we note that these simulation papers focus on star-forming halos around $z \approx 15$, while the largest effects that we find occur for star-less minihalos at higher redshifts. At $z > 20$, Maio et al. [202] find tens of percents difference in the gas fractions, although the statistical errors are large. Stacy et al. [167] find a delay in gas collapse by $\Delta a/a = 0.14$ for $v_{\rm bc}/\sigma_{\rm vbc} = 1$ and [201] observed a delay in star formation by $\Delta z \sim 4$ for objects forming around $z \sim 15 - 30$. Difference in the results is caused also by different numerical techniques and a set of approximations used by the authors. Specifically, the difference may come in part due to different treatment of the cooling process and implementation of molecular and atomic cooling processes. We also note that the choice of initial conditions should be carefully considered: standard initial condition codes do not properly treat the separate baryonic and dark matter perturbations or the gas temperature perturbations, leading to a filtering mass that is too high by a factor of ~ 2 at $v_{\rm bc} = 0$ [178]; as such they may underestimate the effect of relative velocities.

A detailed comparison between theoretical predictions of gas distribution in the first galaxies and minihalos and results of extensive numerical simulations has been performed by Naoz, Yoshida, and Gnedin [203]. This study statistically quantified the effect of relative velocity using large cosmological simulations with different resolutions, particle numbers, and values of the v_{bc}. Naoz, Yoshida and Gnedin concluded that the total number density of halos is suppressed by $\sim 20\%$ at $z = 25$ for $v_{bc} = 1\sigma_{vbc}$, while for $v_{bc} = 3.4\sigma_{vbc}$ the relative suppression at the same redshift reaches 50%, remaining at or above the 30% level all the way to $z = 11$. They also found high abundance of "empty halos", i.e.,, halos that have gas fraction below half of the cosmic mean baryonic fraction $\bar{f}_b$. Specifically, they showed that for $v_{bc} = 1\sigma_{vbc}$ *all* halos below $10^5\ M_\odot$ are empty at $z \geq 19$.

Along with important conclusions and insights simulation papers produced pictures and maps which allow visualization of the relative velocity effect and make our understanding of the underlying processes much more intuitive. A great example of the simulation results is provided in figure 4.15, which shows the delay in star formation caused by relative velocity effect.

Finally, we note that both simulations and theoretical studies of the relative velocity effect are complicated by a large number of degeneracies and various feedback effects which may reduce or mask some of the effect of the relative velocity. For galaxies, local feedback from star formation may effectively raise the minimum halo mass for star formation (except for the very first generation of stars). The possibilities include supernova feedback as well as radiative feedback acting via photoheating and photoevaporation or suppression of H_2 formation, although "positive" feedback due to X-ray ionization enhancing H_2 formation has also been suggested [205, 206]. For minihalos, astrophysical heating, e.g.,, from an early X-ray background, may heat the gas and raise the filtering mass above the value due to $v_{\rm bc}$. There are many unknowns, but these various effects could begin to be significant by $z \sim 20$, and very likely by the time of significant cosmic reionization. Still, the relative velocity between baryons and dark matter

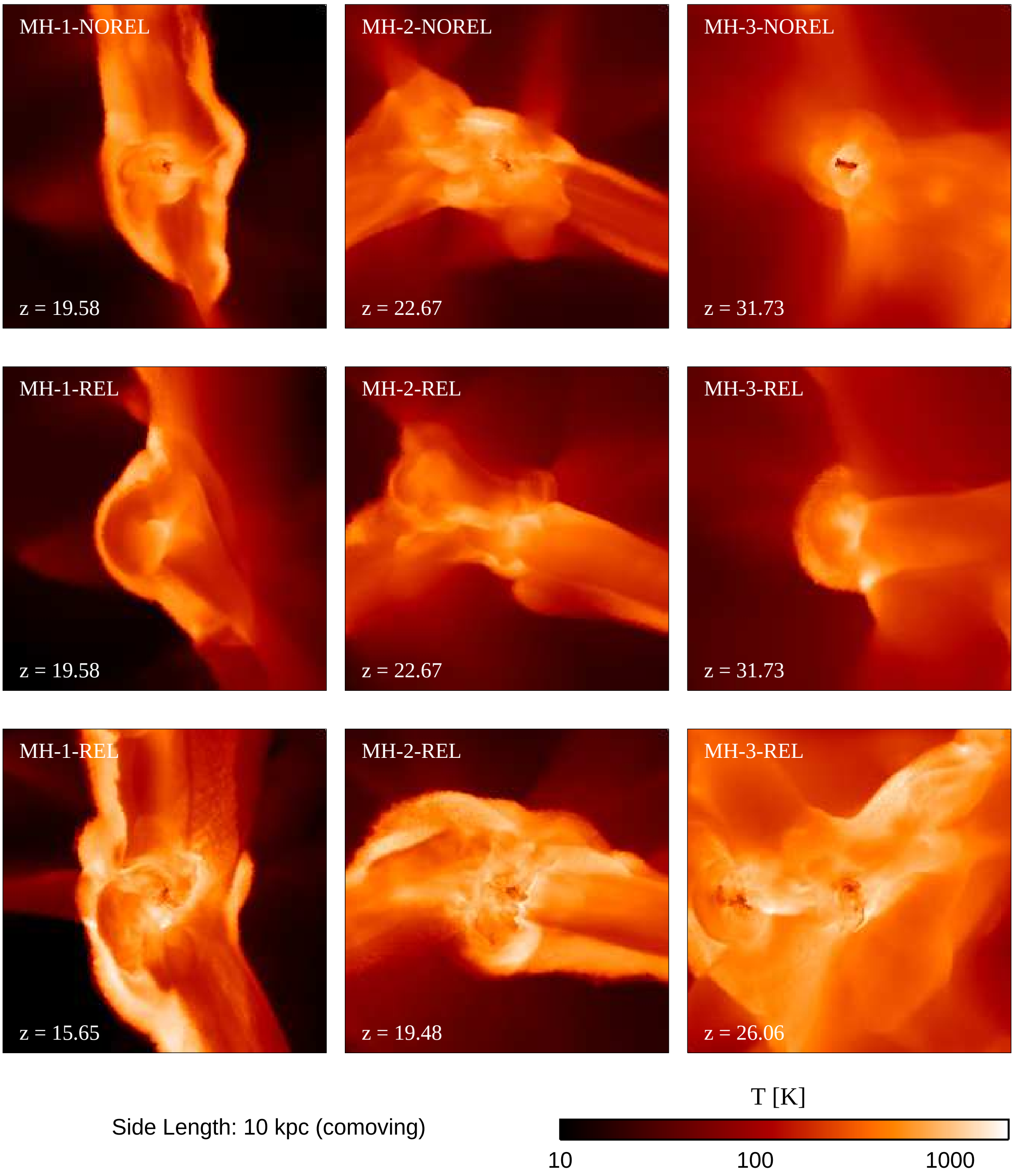


Figure 4.15: A comparison of three statistically independent minihalos with no streaming velocity (top panels), and with an initial streaming velocity of $3\,\mathrm{km\,s^{-1}}$ applied at $z = 99$ from left to right (middle and bottom panels). The figure shows the density-squared weighted gas temperature projected along the line of sight when the hydrogen density in the center has just exceeded $n_{\mathrm{H}} = 10^9\,\mathrm{cm^{-3}}$ (top and bottom panels), and when the streaming case has evolved to the same redshift as the no-streaming case (middle panels). In the presence of streaming velocities, the effective Jeans mass of the gas is increased. The underlying dark matter halo therefore becomes more massive before the gas can cool, which delays the onset of collapse. This figure was adopted from [201] with the permission of the authors.

is the main determinant of the gas content of halos at the highest redshifts. Clearly, detailed analysis of these feedback effects and consequences of the relative velocity effect for observational cosmology represents an exciting direction of modern astrophysics and will provide extremely interesting results in the coming years.

Chapter 5

Proton-Hydrogen Collisions and Physics of Balmer-Dominated Shocks*

In this chapter we look into a unique cosmic environment which allows detailed investigation of plasma physics phenomena in conditions unreachable in the Earth-based laboratories. This unique environment is a shock of a supernovae remnant (SNR) propagating into the interstellar medium (ISM) with densities around $0.1-1$ cm^{-3}. The uniquely low densities of the ISM, which are typically about 1000 times lower than density of the most high-quality vacuum produced in terrestrial laboratories, render astrophysical shocks collisionless, meaning that the shocks are not formed by two-body Coulomb collisions, since the mean free path for these collisions is too large. Collisionless shocks from the supernovae remnants are associated with production of cosmic rays which are believed to be accelerated by the Fermi acceleration mechanism [207, 208]. For that reason SNR shocks represent an exceptional resource for the particle physics and high-energy astrophysics communities.

In this study we concentrate on a subset of shocks in supernovae remnants called Balmer-dominated shocks (BDS). These shocks represent a unique laboratory for studying astrophysical shocks, including processes such as electron heating and cosmic ray acceleration that are not observationally accessible in other astrophysical settings. BDS are fast (200 – 10000 km/s) astrophysical shocks with the presence of a strong hydrogen emission with both narrow and broad components and almost complete absence of forbidden metal lines. BDS are observed around energetic cosmic pistons such as supernova remnants, pulsars, wind nebulae and novae; they may also be relevant in gas accretion in young, high-redshift galaxies (see [209] for a review).

Specific observations of the BDS include emission spectra from several supernovae remnants, such as SN 1006 [210, 211, 212, 213], Kepler Remnant SN 1604 [214, 215, 216], Cygnus Loop [217, 218, 219], Tycho SN 1572 [211, 212, 219], RCW 86 [220, 221, 219] and SN 1987A (see [222] and references therein). In the case of SNR 1987A hydrogen line emission originates in the reverse

---

*The material in this chapter is based on *Excitation and charge transfer of hydrogen atoms: an improved algorithm for $n \geq 4$ and $E \sim 10$ keV*, D. Tseliakhovich, C. Hirata, and K. Heng, to be submitted to *Mon. Not. R. Astron.* in January 2012. Material in this chapter is significantly expanded compared to the paper to include broader overview of the shock physics.

shock propagating into the rapidly expanding supernovae debris, rather than into stationary ISM gas as in most of the other observed cases. Balmer-dominated shocks have also been observed in the pulsar wind nebulae [223].

Investigation of the physical properties of Balmer-dominated shocks represents an exciting frontier of astrophysics which draws a lot of attention from both the astronomy and physics communities. The observational importance of Balmer-dominated shocks stems from the fact that collisional excitation and charge exchange processes occurring inside of those shocks lead to emission of atomic hydrogen lines with profiles sensitive to the ion velocity distribution. For that reason BDS can be used for testing the equipartition of energy between ions and electrons, yielding important insights into the plasma physics of these shock fronts. Furthermore, observation and modeling of the broad line emission and of the broad-to-narrow intensity ratio can be effectively used to determine the ratio of electron to proton temperature, $\beta \equiv T_e/T_p$, which is critically important for understanding physical mechanisms at work in collisionless plasmas [209, 210].

Electron-proton temperature equilibration inside of shocks represents an important open question in plasma physics studies. In the case when there is no energy transfer between the two species, the temperature ratio equals the mass ratio $\beta \equiv T_e/T_p = m_e/m_p \sim 1/2000$. In most cases however, effects such as plasma waves, magneto-hydrodynamic turbulence and electrostatic instabilities at the shock front can increase the electron temperature, with the final ratio ranging from $T_e/T_p \sim 0.1$ to $T_e/T_p = 1$. One of the goals and motivations behind detailed BDS studies is to constrain the value of β and fully understand the mechanisms responsible for temperature equilibration.

Connecting plasma physics phenomena occurring inside of BDS with observed hydrogen lines requires detailed understanding of high-nl proton-hydrogen collisions. At relative velocities of $v_{\rm rel} \gtrsim 1000$ km s^{-1}, the excitation of hydrogen atoms becomes dominated by interactions with protons rather than electrons [224]. However, the lack of precise cross sections at these velocities, either from theory or experiment, has hampered a more accurate diagnostic of the physical parameters of Balmer-dominated shocks. In particular, the Balmer decrement, which is the ratio of the Hα to Hβ line intensities, requires the knowledge of accurate proton-hydrogen excitation and charge transfer cross sections to the excited hydrogen states characterized by principal quantum number $n = 4$. Furthermore, the use of higher-order Balmer lines (e.g., Hγ) or the infrared Paschen and Brackett line series as additional diagnostics of these shocks is not possible until $n > 4$ cross sections have been computed.

Recent observations of Balmer-dominated shocks [211, 219, 225] (see also [209] for a detailed review) in supernovae remnants create a clear demand for high precision cross sections for high-nl proton-hydrogen collisions. These observations, as well as results which are becoming available in the near future motivated us to develop a novel algorithm for computing cross sections for the $H + H^+$ collisional system, including excitation and charge transfer to the $n \geq 2$ levels of hydrogen atom. In this chapter we develop and implement a code named `BDSCx` which directly solves the Schrödinger equation with variable (but nonadaptive) resolution and utilizes a hybrid spatial-Fourier grid. Our novel hybrid grid reduces the number of grid points needed from $\sim 4000n^6$ (for a "brute force", Cartesian grid) to $\sim 2000n^4$ and speeds up the computation by a factor ~ 50 for calculations going up to $n = 4$. We present (l, m)-resolved results for charge-transfer and excitation final states for $n = 2$–4 and for projectile energies of 5–80 keV, as well as fitting functions for the cross sections. We use our results to calculate Hβ line intensities and the Balmer decrement (the ratio of Hα to Hβ line intensities). We also intend to use our results in a detailed study of the recent observations of the BDS in SN-1006 performed in collaboration with S. Nikolic, K. Heng, and G. van de Ven [226].

The rest of this chapter is organized as follows. In section 5.1 we discuss general properties of shocked media and specific properties of the Balmer-dominated shocks; we also provide a concise explanation of how one could compute observable parameters of the BDS and emphasize the importance of knowing precise cross sections of proton-hydrogen collisions. In section 5.2 we provide a detailed description of the theoretical model behind our analysis of high-nl proton-hydrogen collisions and describe major constraints driving the development of the code for cross section calculations. Section 5.3 describes the code developed for our calculations and shows the results of extensive consistency tests. We discuss capabilities and limitations of the code and provide guidance on how this code can be expanded and used by other groups. In section 5.4, as well as in appendixes A and B, we provide results of cross section calculations and compare our results with earlier studies. In section 5.5 we briefly discuss astrophysical applications of the obtained cross sections with specific focus on Balmer-dominated shocks. Our results are summarized in section 5.6.

5.1 Physics of Balmer-Dominated Shocks

Shocks are generated wherever matter is accelerated past the sound speed of the medium in which they are propagating and are present in a large variety of physical media (see [227] for a review). For example, in astrophysics shocks are encountered in stellar winds, planetary nebulae, nova ejection, supernova explosions, and accretion of hot intergalactic gas onto galaxies.

To better understand the physics of shocks we consider a simple model in which a shock front, defined as the region over which the velocity, density, and pressure undergo sudden changes, moves into a stationary gas medium with the velocity v_s. In this case we define the speed of the gas relative to the shock front as u. When the preshocked gas enters the shock front, its velocity, density, and pressure changes discretely from u_0, ρ_0, P_0 to u_1, ρ_1, P_1. This transition is schematically illustrated in figure 5.1. Equations connecting initial values and the values after the shock front transition are known as *jump conditions across a shock* and can be derived from the mass, momentum, and energy conservation requirements. For ideal gas jump conditions can be written as [228]

$$\frac{\rho_1}{\rho_0} = \frac{u_0}{u_1} = \frac{(\gamma+1)M_0^2}{(\gamma+1) + (\gamma-1)(M_0^2-1)}, \tag{5.1}$$

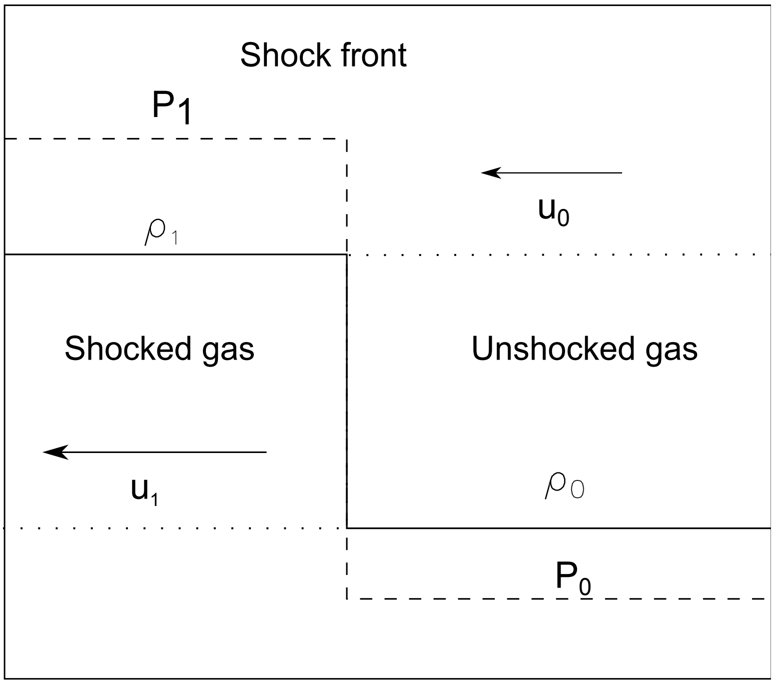


Figure 5.1: A schematic illustration of a shock wave in which discrete jump in density, velocity, and pressure occurs at the shock front with zero thickness.

$$\frac{P_1}{P_0} = \frac{2\gamma}{\gamma+1}M_0^2 - \frac{\gamma-1}{\gamma+1}. \tag{5.2}$$

Here, γ is the adiabatic constant ($\gamma = 5/3$ for adiabatic monoatomic gas), and M_0 is the Mach number of the shock wave, which is the ratio of the shock speed to the sounds speed of the gas. Equation 5.1 clearly shows that for strong shocks ($M_0 \to \infty$) propagating into an ideal gas velocity jump is $u_0/u_1 \to 4$.

Jump conditions along with ideal gas law allow complete characterization of the shock in this approximation. For example, it is straightforward to derive temperature ratio for pre- and postshock gas and to show that for shock velocities $v_s \sim 1000$ km/s the postshock gas can be heated to over 10^7 K. This simple approximation allows one to clearly see that shock represents an extremely effective environment for heating of the ISM and provides an opportunity to study a large variety of phenomena which can not be easily reproduced on Earth. Specifically, astrophysical shocks represent the most likely environment for cosmic ray acceleration and hence observations and theoretical studies of the shocks are central for the high-energy astrophysics as well as for the particle physics communities.

Balmer-dominated shocks (BDS) are specific subset of astrophysical shocks characterized by strong Balmer and Lyman line emission. BDS are high velocity ($\gtrsim$ 200–2000 km/s) collisionless, nonradiative shocks that are most commonly observed around supernovae remnants. The age of typical BDS observed in supernovae remnants is $\sim$ 100–1000 years, which is much less than typical cooling and recombination timescales, and hence BDS do not produce radiation that can significantly impact the dynamics of the shock fronts. The collisionless nature of Balmer-dominated shocks is determined by the fact that the length scale of atomic collisions between particles inside of the shock is much larger than typical scale of gyration in the ISM's magnetic field. This is easy to see by writing down the corresponding length scales:

$$
\begin{aligned}
l_{col} &\sim \frac{1}{n\sigma_{col}} \sim 10^{15}(\frac{n}{1cm^{-3}})^{-1} \ \ \text{cm}, \\
l_{gyro} &\sim \frac{v_s m_p}{qB_{ISM}} \sim 10^{6}(\frac{v_s}{1000km/s})(\frac{B_{ISM}}{1\mu G})^{-1} \ \ \text{cm},
\end{aligned}
\tag{5.3}
$$

where $\sigma_{col} \sim 10^{-15}$ cm^2 is dominated by charge transfer and excitation for shock velocities $\sim$ 1000 km/s. Because of the collisionless nature of the Balmer-dominated shocks, the temperature ratio between electrons and protons is determined by collective electromagnetic interactions (such as electrostatic interactions in the shock front and magneto-hydrodynamic turbulence), and not by physical collisions [209].

Hydrogen line emission in Balmer-dominated shocks consists of both broad and narrow lines which come from two distinct populations of hydrogen atoms. As the shock impinges upon the ISM, some of the cold hydrogen atoms present in the upstream gas get excited through collisions with energetic protons from the shock and emit a narrow line ($\sim$ 10 km/s) characteristic of the thermal velocity of the ISM hydrogen. The second population contributing to the BDS line emission is called "broad neutrals" and consists of hydrogen atoms produced in charge transfer reactions between neutral hydrogen of the ISM and energetic protons from the shock. Emission from these "broad neutrals" gives rise to a broad hydrogen line ($\sim$ 1000 km/s) [229, 210]. Other important, although not unique, observational characteristics of the BDS include absence of forbidden lines of metals and lack of nonthermal X-ray emission.

The observable quantities of Balmer-dominated Shocks, that allow comparison between theory and observation, are the profiles and intensities of broad and narrow lines, and the velocity shifts between the two components. An illustration of the BDS from the combined optical and X-ray images of the supernova remnant RCW86, which emphasizes the observable hydrogen lines, is provided in figure 5.2. Observable features of the BDS, such as ratio of the broad to narrow lines, the line profiles and ratios of various hydrogen lines depend on the equilibration of electron and ion temperatures, on the velocity of the shock, and on optical properties of the ambient medium (specifically optical depth of the ISM for Lyman photons). Theoretical explanation and reproduction of these observable characteristics is a primary goal behind development of theoretical models and numerical codes for studies of the BDS. Such models had been under development for several decades [210, 219, 224, 230, 231, 232, 233, 234] and still represent an active area of research [209, 226].

Below we describe a general mechanism for calculation of the line intensities and show the importance of knowing cross sections for proton-hydrogen collisions, which are calculated later in this chapter.

5.1.1 Line Strength and Profiles

Intensities and profiles of emission lines can be calculated by multiplying the rate at which atoms are excited by a weighting factor determined by a probability of a given line being emitted in

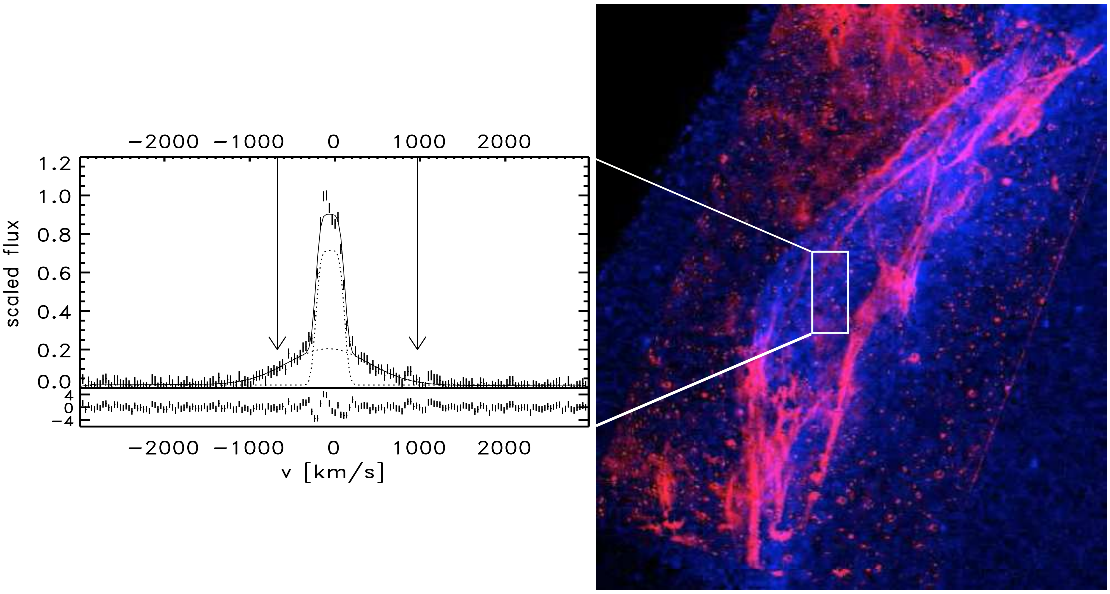


Figure 5.2: The image shows combined optical and X-ray image of northeastern RCW86 (a Galactic supernova remnant), adapted from [235]. The blue and pink colors represent the Chandra broad-band (0.110 keV) and Very Large Telescope (VLT) narrow-band $H\alpha$ emission, respectively. The image also illustrates one of the observables in the BDS studies–the two-component $H\alpha$ line profile from a with clear broad and narrow components. Images courtesy of Eveline Helder.

that reaction. For H_α emission, the narrow line emission rate per atom can be written as [224]

$$I_{narrow}(\mathrm{H}\alpha) = \frac{C_{32}}{1 - P_{E_0}} \sum_{n=3}^{m} R_{E_0,n}\, C_{n3}. \tag{5.4}$$

Here the $C_{nn'}$ are the cascade matrices which give the probability that a hydrogen atom excited to state n will get deexcited into the state $n' < n$ via all possible routes (emission of a photon or collisional deexcitation). $R_{E_0,n}$ is the rate coefficient for excitations to the n-state before charge transfer reaction, and P_{E_0} is the net probability that the atom will be excited into any level n. This net excitation probability is determined by the total reaction rates per atom for ionization (R_{I_0}), excitation (R_{E_0}), and charge transfer (R_{T_0}) and can be written as

$$P_{E_0} = \frac{R_{E_0}}{R_{I_0} + R_{E_0} + R_{T_0}}. \tag{5.5}$$

The broad line emission is composed of all excitations after the first charge transfer and also from charge transfer reactions to excited states:

$$I_{broad}(\mathrm{H}\alpha) = \frac{C_{32}}{1 - P_{E_0}} \sum_{n=3}^{m} \left[\frac{P_{T_0}}{P_I} \left(R_{E,n} + R_{T^*,n}\right) + R_{T_0^*,n} \right] C_{n3}, \tag{5.6}$$

where $R_{E,n}$ is the rate of excitations to state n after any number $i \geq 1$ of charge transfer reactions and $R_{T^*,n}$ is the corresponding rate of charge transfer reactions to excited states [224]. Here again P_{E_0} represents the net probability of excitation, however in this case the particle of interest is “broad neutral” rather than cold hydrogen atom from the preshocked gas. We note that Lyman line emission profiles can be computed in a similar way.

In general, the rate (in units of s^{-1}) at which a hydrogen atom from the ISM will have a reaction X (where X stands for excitation (E), ionization (I), or charge transfer (T)) after a collision with ions from the shock propagating at the velocity $\mathbf{v}_{sh}$ can be written as

$$R(\mathbf{v}_{sh};\sigma_X) = n_a \int\int f_a(\mathbf{v}_a)\ f_b(\mathbf{v}_b)\ \sigma_X(|\mathbf{v}_a - \mathbf{v}_b|)\ |\mathbf{v}_a - \mathbf{v}_b|\ d^3v_a\ d^3v_b. \tag{5.7}$$

In the case of Balmer-dominated shocks, the ions causing the reaction to occur come from the shock and can be either protons, electrons or alpha particles. Moving forward with this integral requires a set of assumptions regarding the content of the shock front and the velocity distributions of various ionic species inside of the shock. Here we assume incoming ions to have mass m_s, number density n_s, temperature T_s, and a Maxwellian velocity distribution function:

$$f_M(\mathbf{v}) = (m_s/2\pi kT_s)^{3/2}\ \exp\left(-\frac{m_s v^2}{2kT_s}\right), \tag{5.8}$$

To simplify the derivation, the cold, preshocked hydrogen atoms from the ISM can be described by a delta function:

$$f_0(\mathbf{v}_H, \mathbf{v}_{sh}) = \delta(\mathbf{v}_H - \mathbf{v}_{sh}). \tag{5.9}$$

With these assumption we can write down the rate of the reaction X between incoming ions of the shock front and hydrogen atoms of the ISM as

$$\begin{aligned} R_{X_0}(\mathbf{v}_{sh};\sigma_X) &= n_s f_{M,0} \int\int \exp\left(-\frac{m_s v^2}{2kT_s}\right)\ \delta(\mathbf{v}_H - \mathbf{v}_{sh})\ \sigma_X(|\mathbf{v}_H - \mathbf{v}|)\ |\mathbf{v}_H - \mathbf{v}|\ d^3v_H\ d^3v \\ &= 2\pi n_s f_{M,0} \int_{-\infty}^{\infty}\int_0^{\infty}\ \exp\left[-\frac{m_s(v_r^2 + v_z^2)}{2kT_s}\right]\ \Delta v\ \sigma_X(\Delta v)\ v_r\ dv_r\ dv_z \end{aligned} \tag{5.10}$$

In case of Balmer-dominated shocks we are mainly interested in production of Balmer line photons which originate in transitions from $3s$ and $3d$ to $2p$ levels and from $3p$ to $2s$. In the case of the $3p$ to $2s$ atomic transitions, one needs to take into account possibility of direct transition from $3p$ to $1s$ with the emission of Ly_β photon. This possibility is taken into account by using appropriate branching ratio coefficient and writing the total H_α cross section in the following way:

$$\sigma_{H_\alpha} = \sigma_{3s} + \sigma_{3d} + B_{3p,2s}\sigma_{3p}. \tag{5.11}$$

The branching ratio depends on the physical conditions inside of the shock. If the shock medium is optically thick to Ly_β photons, than all those photons get reabsorbed by neutral hydrogen atoms, and the only photons that can escape are those contributing to the Balmer line. In this case $B_{3p,2s} \approx 1$. On the other hand, if the shock is optically thin, the branching ratio is $B_{3p,2s} \approx 0.12$.

Observable line intensities are sensitive to the velocity of the shock, distribution of the hydrogen atoms and protons in the shock and to the temperatures of different species inside of shocks. To extract these quantities it is important to measure not only the broad and narrow H_α lines but also higher lines of hydrogen and different ratios such as Balmer decrement, H_α/H_β.

These additional lines and ratios are critical for accurate reconstruction of ion temperature and velocity distributions and phase space structure of the shock. From the above equations we see that reconstruction of lines with higher values of quantum number n require excitation and charge transfer cross sections for high-nl proton hydrogen collisions. Specifically, for analysis of H_β line and Balmer decrement we need to know cross section for $4s$, $4p$ and $4d$ levels of hydrogen.

Reaction cross sections for excitations and charge transfer in hydrogen atoms are among the most important quantities in determining reaction rates and thus the observable line intensities in BDS emission. A glaring and surprising gap in the existing literature, either from theory or experiment, is that cross sections for the reaction,

$$\mathrm{H(1s)} + \mathrm{H}^+ \rightarrow \mathrm{H}\,(n \geq 4) + \mathrm{H}^+, \tag{5.12}$$

are essentially nonexistent at the velocities relevant to Balmer-dominated shocks. In the absence of these cross sections, some researchers have resorted to using inaccurate scaling laws such as

$$\sigma_{nl} = \left(\frac{n_0}{n}\right)^3 \sigma_{n_0 l}, \tag{5.13}$$

where $n_0 < n$, to extrapolate for cross sections with $n \geq 4$ using available ones with $n_0 \leq 3$. This scaling law is derived using Born approximation and is only approximately valid at intermediate energies [236]. Besides the inaccuracy associated with such extrapolation, it also leaves open the question of how to obtain cross sections for levels with l values which do not exist for $n_0 \leq 3$ (e.g.,, 4f).

Initial attempts to calculate and measure cross sections of hydrogen collisions date back to the 1960s [237, 238, 239, 240]. On the theoretical side, the success of these efforts has been hindered by the high computational cost of numerical simulations. On the experimental side, it has been limited by stringent requirements of creating high vacuum states and the high costs of preparing and characterizing atomic hydrogen targets.

In calculating cross sections for high-nl proton-hydrogen collisions it is important to consider several distinct cases. At low velocities, $V << 1$, collisions between a hydrogen atom and a proton lead to large deflections of colliding particles and in this case, as in the case of unbound H_2^+ molecule, electron wave function deforms adiabatically during the collisional time which significantly exceeds atomic timescale. In such situations the initial configuration of the collision system is considerably modified during the collision, and the process must be treated in a way which reflects the interplay between various quantum states of the electron's wave function. This problem has been addressed with the close-coupling approximation which assumes that, during the atomic collision, the electron wave function transitions between a certain number of configurations which form the "basis set" of functions (see [230] for a detailed review). The dominant outcome of the collisions at low energies is the charge transfer between colliding particles.

On the other hand at high energies, $V >> 1$, the collision is relativistic, and the colliding particles follow undeflected straight-line trajectories. This case is well described by Born approximation in which the incoming proton is seen as a small perturbation of the electron's wave function [231]. The dominant outcome of the collision in relativistic regime is excitation or ionization of the hydrogen atom.

The third case corresponds to intermediate velocities $V \sim 1$, where collisional times are of the order of atomic timescale, and therefore perturbative treatment of the problem is no longer possible. This is precisely the regime relevant to the Balmer-dominated shocks, and in this case behavior of the electron wave function is more complicated than at low energies or relativistic energies. At the intermediate energies ($\sim 10 - 100$ keV) there is no dominant outcome for a collision: charge transfer, collisional excitation and ionization are all important and interconnected. There is no clear intuitive picture of the mechanism for populating various electron quantum states, and therefore the use of close-coupling approximation is challenging and requires development of multiple basis sets and extensive convergence tests. Analysis of the proton-hydrogen collisions at intermediate energies in the close-coupling approximation has been an active area of research [230, 232, 241, 242, 243, 244, 245, 246, 247]; however, accuracy of the obtained results is still not fully determined as convergence tests of these methods are extremely hard, especially if extended over a large range of energies [242, 243].

An alternative method to addressing the problem of proton-hydrogen collisions at intermediate energies is direct solution of the Schrödinger differential equation on a numerical grid [248, 249, 250]. In fact, the grid-based method can be thought of as a finite basis set method with one basis function for each point on the grid. Over the past decade, several groups successfully attempted this approach [251, 252], producing results for energies ranging from 10 to 100 keV. Numerical grid methods are easier to convergence test, however they are notoriously computationally demanding due to the long-range nature of the Coulomb electrostatic force. This is especially true if one needs to accurately represent states of high n and for that reasons previous studies with this methods were limited to $n \leq 3$.

In most cases, the results obtained in previous studies measure and/or calculate cross sections of hydrogen collisions in the velocity range $\sim$ 100–1000 km s^{-1} only with a precision $\sim$ 10%–30% (see [209] and references therein). There is also a substantial disagreement between experimental results and theoretical calculations (see, for example [241, 253], and references within [209]). In cases of final states with $n > 3$, robust theoretical or experimental cross sections in the energy range relevant to Balmer-dominated shocks studies do not exist at all.

5.2 Computational Model for Cross Section Calculations

5.2.1 Initial Setup for Precise Cross Section Calculations

Our objective is to determine the cross sections for reactions of the form:

$$\mathrm{H_A}(1s) + \mathrm{H_B^+} \to X. \tag{5.14}$$

The hydrogen nuclei are assumed to be very massive so that their motion can be treated classically and one only has to follow the evolution of the electron wave function in the potential created by the two nuclei. Except for extremely small impact parameter $b \leq m_p^{-1} v^{-2}$ all potential energies are negligible compared to the nuclear kinetic energy, so we may use an undeflected (straight line, constant velocity) trajectory for the nuclei. In this limit one can distinguish the two nuclei, hence their description as $\mathrm{H_A}$ and $\mathrm{H_B}$. We define the time of closest approach to be $t = 0$.

The initial electronic state is that of the $1s$ orbital of atom A, i.e., $|1s_\mathrm{A}\rangle$. The final states X under consideration correspond to (i) no reaction ($|1s_\mathrm{A}\rangle$); (ii) excitation ($|nlm_\mathrm{A}\rangle$, $n \geq 2$); (iii) charge transfer ($|nlm_\mathrm{B}\rangle$); and (iv) ionization (everything else). All of these are of interest, even for high n levels, for example for the Hβ lines the upper level is $n = 4$ and for the Brα lines the upper level is $n = 5$.

We choose a coordinate system such that the relative velocity lies along the z-axis, $\mathbf{V} = V\hat{\mathbf{e}}_z$, and the nuclear separation vector lies in the xz-plane. That is, the relative separation is

$$\mathbf{r}_\mathrm{A} - \mathbf{r}_B = b\hat{\mathbf{e}}_x + Vt\hat{\mathbf{e}}_z. \tag{5.15}$$

Note that the electron wave function is always symmetric under reflection across the xz-plane. We will choose the origin of the coordinate system in the x-direction such that $x_\mathrm{A} = \frac{1}{2}b$, and $x_\mathrm{B} = -\frac{1}{2}b$. The choice of origin in the z-direction will be discussed later.

The cross section to produce a particular final state X is given by

$$\sigma_X = \lim_{T\to\infty} \int_0^\infty 2\pi b \left|\langle X|\hat{S}(-T,T)|1s_\mathrm{A}\rangle\right|^2 \mathrm{d}b, \tag{5.16}$$

where $\hat{S}(t_\mathrm{i}, t_\mathrm{f})$ is the time evolution operator. The S-matrix element can in principle be obtained by evolving $|1s_\mathrm{A}\rangle$ forward in time, $|X\rangle$ backward in time, or some combination of both. For example, one could evolve both states to $t = 0$, that is one could factor $\hat{S}$ as $\hat{S}(-T,T) = \hat{S}(0,T)\hat{S}(-T,0)$ and have $\hat{S}(0,T)$ back-operate on $\langle X|$. The results in the present study are based on evolving $|1s_\mathrm{A}\rangle$ forward as this is the most efficient way to generate cross sections for large numbers of final states.

Computation of the matrix elements requires us to solve the Schrödinger equation. Two general types of methods could be considered: finite basis set methods, and direct solution of the Schrödinger equation on a grid (this is not an exhaustive list). A grid-based method can in

Dimension	Name	Expression	Value in SI
Length	Bohr radius	$a_0 = \hbar/(m_e c\alpha)$	5.29×10^{-11} m
Energy	Hartree	$E_h = \alpha^2 m_e c^2$	4.36×10^{-18} J
Velocity	–	αc	2.19×10^{6} m s^{-1}
Electric field	–	$E_h/(ea_0)$	5.14×10^{11} V m^{-1}

Table 5.1: Connection between the atomic units and the SI units

fact be thought of as a finite basis set method, with one basis function for each point on the grid. In this work we focus on grid methods because they allow easier convergence tests and can be applied over a wide range of energies without significant modifications. Numerical grid methods also allow direct visualization of wave function evolution during the collision.

Throughout this chapter we use atomic units for all quantities, i.e., energy in Hartrees, length in Bohr radii, velocity in units of $\alpha c = 2190$ km s^{-1}, and mass in electron masses. Conversions between the Système Internationale (SI) units and atomic units are provided in table 5.1.

5.2.2 Grid Choice

The most obvious way to implement a grid method is to choose a spacing Δx and a box size L. The number of grid points is then $N \approx (L/\Delta x)^3$. Unfortunately this will be computationally prohibitive: if we want to consider a highly excited state of hydrogen nl, then the grid must go out to at least a radius of $2n^2$, and preferably much more, so $L > 4n^2$. On the other hand, to resolve the $1s$ state properly a fine spacing (below ~ 0.2) is needed. This leads us to the conclusion that we need $N \sim 4000n^6$ grid points, which is prohibitive for states above $n \sim 4$.

Clearly, we will need a type of grid that puts resolution where we need it: high resolution near the protons, and more modest resolution far away. To avoid the complexity of developing an adaptive code, we will insist on high resolution near the trajectories of the proton and lower resolution elsewhere. This immediately suggests developing a generalized cylindrical coordinate system, i.e., introducing a mapping $(u, v) \leftrightarrow (x, y)$ and using as our fundamental coordinates (u, v, z) instead of (x, y, z). A constant grid spacing $\Delta u = \Delta v$ can then correspond to a variable spacing in the xy-plane, in accordance with the Jacobian of the transformation. Note that high resolution (several grid points per Bohr radius) is also required in the region in between the protons in order to correctly model the $1s_\mathrm{A} \leftrightarrow 1s_\mathrm{B}$ tunneling that is primarily responsible for charge exchange at low and intermediate velocities.

A minimal criterion for such a grid is that it should be able to adequately sample all bound wave functions with several grid points per cycle. The momentum of a bound wave function can be as large as $\sqrt{2/\rho}$ where ρ is the minimum separation from the nucleus. Therefore the grid spacing should be at most $\sim \sqrt{\rho/2}$ (and preferably better). This requirement could be relaxed if $\rho \leq 1$, where the classical intuition concerning the "maximum momentum of a bound electron" is invalid; in this regime instead all that is required is to have at least a few grid points per Bohr radius.

After considering and rejecting several other choices[†] we decided on the coordinate system

$$x = u\sqrt{1+\frac{u^2}{u_s^2}}; \quad y = v\sqrt{1+\frac{v^2}{4}}, \tag{5.17}$$

where u_s is a parameter. The Jacobian is

$$\frac{\partial x}{\partial u} = \frac{1+2u^2/u_s^2}{\sqrt{1+u^2/u_s^2}} \to \begin{cases} 1 & |u|,|x| \ll u_s, \\ 2u/u_s \approx \sqrt{8x/u_s} & |u|,|x| \gg u_s, \end{cases} \tag{5.18}$$

and

$$\frac{\partial y}{\partial v} = \frac{1+v^2/2}{\sqrt{1+v^2/4}} \to \begin{cases} 1 & |v|,|y| \ll 1, \\ v \approx \sqrt{2y} & |v|,|y| \gg 1. \end{cases} \tag{5.19}$$

This satisfies our resolution criteria if Δu is no more than a few tenths, and u_s is at least as large as $\sim \max(1, b)$.

We illustrate the proposed approach in figure 5.3 by using a sample case of $\Delta_u = \Delta_z = 0.2$ and comparing our proposed grid to a brute force Cartesian grid with $\Delta_x = \Delta_z = 0.2$. This figure shows every third point in both x-direction and z-direction, and clearly illustrates benefits of putting high resolution in the region surrounding colliding particles while reducing the resolution far from the collision region. The resolution in the physical x-space between colliding particles is close to $\Delta_x \approx \Delta_u = 0.2$, while at larger separation the spacing between the grid points in x-direction significantly increases reaching $\Delta_x \sim 1.6$ at $x \sim 16$.

By using a grid spacing of, e.g., 0.2 in (u, v, z) space, and noting that we only have to go out to maximum values of $u_{\max} \sim \sqrt{2u_s}n$, $v_{\max} \sim \sqrt{2}n$, the number of grid points necessary would be

$$N \sim \frac{1}{2}(10\sqrt{2u_s}n)(10\sqrt{2}n)(20n^2) = 2000u_s^{1/2}n^4. \tag{5.20}$$

This is much more manageable than $4000n^6$ as found earlier but is probably still too large.

We can make yet another improvement by Fourier-transforming the z-direction. We suppose that we have a box of size L_z in the z-direction and at each point in the (u, v)-plane we do a Fourier transform,

$$\Psi(x, y, z) = L_z^{-1} \sum_q \Psi_q(u, v) \mathrm{e}^{2\pi \mathrm{i} q / L_z}, \tag{5.21}$$

where q is an integer. The maximum required value of q is now $p_{z,\max}L_z/2\pi$. Here $p_{z,\max}$ can be set to a large value ($p_{z,\max} \sim 20$) within a few Bohr radii of the atoms (i.e., $|u|, |v|$ less than a few), but a smaller value ($p_{z,\max} \sim 4$) at large separations. For the results presented in this study (cross sections with $n = 4$ final principal quantum number) it is computationally feasible to do the calculations at a single $q_{\max}$. However the code we have written, `BDSCx`, supports the use of two values of $q_{\max}$ in different regions of the (u, v) plane. The latter capability will be required for cross sections to the $n \geq 5$ levels of hydrogen. We are currently tuning the parameters of the code for this purpose, but this chapter presents the $n \leq 4$ results with a single

---

[†] For example, parabolic cylinder coordinates would have provided the desired resolution for head-on collisions, but at large impact parameter would have difficulty providing resolution at the locations of both protons.

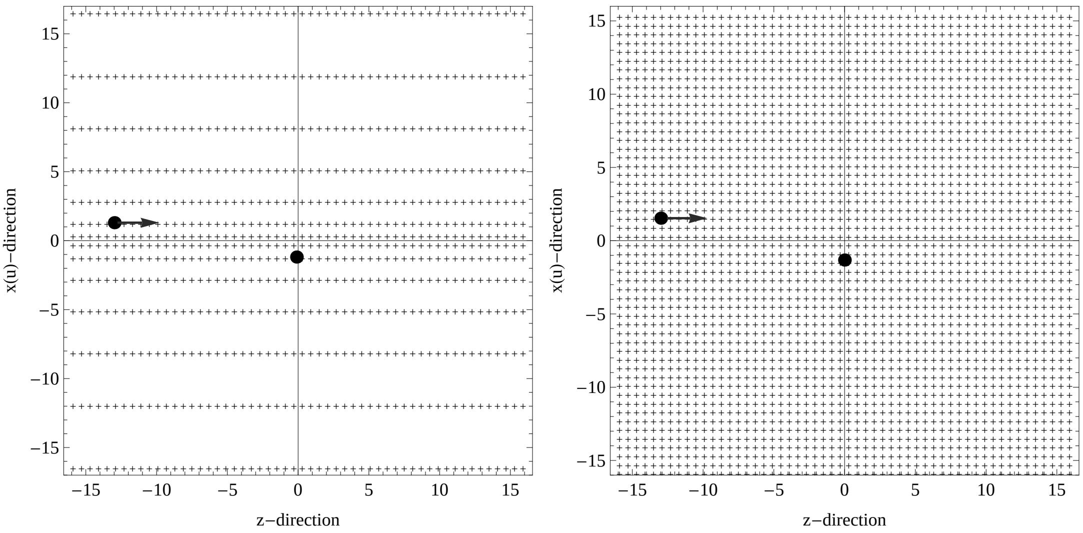


Figure 5.3: An example of a grid in xz-plane used in our analysis (left panel) with every third point plotted in both x- and z-direction, so that the aspect ratio is illustrated correctly. We also show positions of the moving hydrogen atom (thick dot with the vector) and stationary proton at $z = 0$. Distances between grid points in x-direction correspond to equidistant intervals ($\Delta_u = 0.2$ in u-space), so that in the region close to the interacting particles $\Delta_x \approx \Delta_u = 0.2$, while at large separations spacing in x-direction is significantly increased reaching, for example, $\Delta_x \sim 1.6$ at $x \sim 16$. The right panel shows a "brute force" Cartesian grid with $\Delta_x = \Delta_z = 0.2$ (again only every third point is plotted) and allows to clearly see the advantage of our grid choice.

$q_{\max}$ in order to make these available to the community in a more expedient way.

A peculiar property of this setup is that since we can only track momenta out to $p_{z,\max}$ it actually matters which nucleus we take as moving and which as fixed. One would expect that the best results would be obtained by taking the H atom as fixed and the H$^+$ ion as moving, but for charge exchange reactions the definitions of "fixed" and "moving" change. We prefer to handle charge exchange by applying a boost operation by an amount V (i.e., increment the values of q by $VL_z/2\pi$) to the wave function of the electron. An alternative would be to evolve final states $|nlm_{\mathrm{B}}\rangle$ backward from $t = T \to 0$, apply a boost, and compute an inner product with the forward-evolved $|1s_{\mathrm{A}}\rangle$ state. For Hermitian discretized Hamiltonians these two methods are equivalent.

To properly apply boost operation we consider two reference frames: first, a stationary frame K with our initial coordinates (u, v, z) and the second is the restframe of a moving atom K$'$, which moves with the velocity V in the $+z$ direction. The transformation is

$$\Psi(\mathbf{r}, t) = \Psi'(\mathbf{r} - \mathbf{V}t, t)\mathrm{e}^{\mathrm{i}(\mathbf{V}\cdot\mathbf{r} - V^2 t/2)}. \qquad (5.22)$$

5.2.3 Operators

Here we restrict ourselves to orthogonal coordinate systems, i.e., where $\nabla u \cdot \nabla v = 0$.

We next need a method to compute inner products, and Hermitian discretizations of the kinetic and potential operators. The inner product is simply

$$\langle\varphi|\psi\rangle = L_z^{-1}\sum_{quv}\varphi_q^*(u,v)\psi_q(u,v)\frac{\Delta u\,\Delta v}{\sqrt{g^{uu}g^{vv}}}, \tag{5.23}$$

where $g^{uu} = |\nabla u|^2$, and $g^{vv} = |\nabla v|^2$.

The kinetic energy operator is the sum of the operators along the x-, y-, and z-axes. The z-operator is trivial, being simply a multiplication by $2\pi^2 q^2/L_z^2$. The x- and y-operators are trickier; fortunately, they commute with the Fourier transform in the z-direction so we may implement them independently on each q-slice of the wave function. We recall that the x- and y-components of the kinetic energy operator can be written as

$$\begin{gathered}\langle\varphi|\hat{T}_{xy}|\psi\rangle = \frac{1}{2}\int\left(\frac{\partial\varphi^*}{\partial x}\frac{\partial\psi}{\partial x} + \frac{\partial\varphi^*}{\partial y}\frac{\partial\psi}{\partial y}\right)\mathrm{d}^3\mathbf{r} = \\ \frac{1}{2L_z}\sum_q\int\left(g^{uu}\frac{\partial\varphi^*}{\partial u}\frac{\partial\psi}{\partial u} + g^{vv}\frac{\partial\varphi^*}{\partial v}\frac{\partial\psi}{\partial v}\right)\frac{\mathrm{d}u\,\mathrm{d}v}{\sqrt{g^{uu}g^{vv}}}.\end{gathered} \tag{5.24}$$

Thus we see that the u and v parts of the kinetic energy operator are simply additive, i.e., we can write $\hat{T} = \hat{T}_u + \hat{T}_v + \hat{T}_z$. The u-part can be recast by discretizing the partial derivative as

$$\frac{\partial\psi}{\partial u} = \frac{\psi(u+\Delta u/2) - \psi(u-\Delta u/2)}{\Delta u}. \tag{5.25}$$

Note that this partial derivative is measured not on the grid points, but halfway in between (i.e., $\Delta u/2$ to the "right" of each grid point, or alternatively along each grid segment). Then we may write $\langle\varphi|\hat{T}_u|\psi\rangle$ as a sum over such grid segments. This gives an approximation to $\hat{T}_u$,

$$\begin{aligned}\hat{T}_u\psi(u) &= \frac{\sqrt{g^{uu}(u)g^{vv}(v)}}{2\Delta u^2} \\ &\quad\times\left\{-\sqrt{\frac{g^{uu}\left(u+\frac{\Delta u}{2}\right)}{g^{vv}(v)}}[\psi(u+\Delta u) - \psi(u)]\right. \\ &\quad\left.+\sqrt{\frac{g^{uu}\left(u-\frac{\Delta u}{2}\right)}{g^{vv}(v)}}[\psi(u) - \psi(u-\Delta u)]\right\}.\end{aligned} \tag{5.26}$$

A simple calculation shows that with the discretized inner product of equation (5.23), this kinetic energy operator is exactly Hermitian. A similar equation holds for $\hat{T}_v$. Off-grid points are assumed to have $\psi = 0$, corresponding to Dirichlet boundary conditions. We note, that versions of this operator with higher-order accuracy can be constructed by using more than 2 points in the derivative, equation (5.25). An intelligent boundary condition would have to be chosen at the endpoints, however only the ionized electrons will reach the boundary and they will reflect off. The higher-order derivatives can be used to reduce spurious oscillations between $ns \leftrightarrow np \leftrightarrow nd$ states. We relegate the study of the use of the higher-order derivatives to a future study.

The potential energy operator is local in three-dimensional position space, but not in (u, v, q)-

space. The potential at any position is given by

$$V(\mathbf{r}) = -\frac{1}{\sqrt{(x-b/2)^2+y^2+(z-z_\mathrm{A})^2}} - \frac{1}{\sqrt{(x+b/2)^2+y^2+(z-z_\mathrm{B})^2}}. \tag{5.27}$$

The most efficient way to implement the potential operator is to FFT $\psi_q(u,v)$ in the z-direction, multiply by $V(x,y,z)$, and perform an inverse FFT. For N_z points this implies $\sim lnN_z$ operations per grid point, which is manageable especially since in most cases N_z is small because of the q-drop procedure.

We note, that this choice of potential term in the Hamiltonian is associated with numerical difficulties due to its divergent nature near the proton. We eliminate these difficulties by capping the potential with a continuous function near the origin. We choose

$$V(r) = \begin{cases} -\frac{1}{4}R_0^{-1}(9-5R_0^{-2}r^2) & r < R_0, \\ -r^{-1} & r \geq R_0, \end{cases} \tag{5.28}$$

which was chosen so that the volumetric integral $\int(V_\mathrm{capped} - V_\mathrm{true})\mathrm{d}^3\mathbf{r} = 0$. The advantage of this is that spurious features in the potential near the origin will result in a spurious interaction Hamiltonian between any two states ψ_1 and ψ_2 given by $\int \psi_1^*\psi_2(V_\mathrm{capped} - V_\mathrm{true})\mathrm{d}^3\mathbf{r}$, and hence setting the volumetric integral to zero should yield improved behavior over, e.g., imposing a simple floor on V. The capping radius is chosen to be $R_0 = 0.2$ for the results shown here. The capping procedure was tested extensively and shown not to introduce any spurious deviations from the results with uncapped potential. Details of the testing procedures are discussed in subsection 5.3.3 and are illustrated in figure 5.6. For all relevant quantities calculated using our code the difference between capped and uncapped potential is $<< 1\%$.

5.3 The Code

5.3.1 Grid Parameters

To implement the proposed algorithm we developed a grid-based code `BDSCx`, which computes cross sections of hydrogen-proton collisions with hydrogen starting in $|1s\rangle$ state:

$$|1s\rangle = \Psi_{100} = \frac{1}{\sqrt{\pi}}e^{-r}. \tag{5.29}$$

We start and end the collision when the the two particles A and B are separated by a distance sufficient to fully resolve wave functions of interest. Distance from the edges of the box is also determined by the requirement to properly resolve nlm states of interest at the beginning and at the end of the collision. Our choice of the box size parameters is guided by the charge distribution in the states of interest. In figure 5.4 we plot the charge density $r^2R_{nl}^2(r)$ for hydrogen states with $n = 1, 2, 3$. Here, $R_{nl}(r)$ are normalized radial eigenfunctions of the

n	b	u_s	$L_x(L_u)$	$L_y(L_v)$	L_z	N_u	N_v	N_z
2	1	1	31(8)	30(11)	65	46	62	362
2	5	5	35(18)	30(11)	65	100	62	362
4	1	1	101(15)	100(20)	205	84	112	1140
4	5	5	105(32)	100(20)	205	178	112	1140
5	1	1	143(18)	142(24)	285	100	134	1584
5	5	5	147(38)	142(24)	285	212	134	1584

Table 5.2: Examples of the simulation parameters for collisions involving $n = 2$, $n = 4$ and $n = 5$ states

hydrogen atom:

$$R_{nl}(r) = \sqrt{\frac{(n-l-1)!}{4n^4[(n+l)!]^3}}\, \mathrm{e}^{-r/n} \left(\frac{2r}{n}\right)^l L^{2l+1}_{n-l-1}\left(\frac{2r}{n}\right), \tag{5.30}$$

where $L^{2l+1}_{n-l-1}(2r/n)$ are the generalized Laguerre polynomials.

We choose the length of the box in the z-direction by requiring charge density decrease of more than 2 orders of magnitude relative to its maximum value for the state of interest. For example, to properly resolve the $1s$ state a particle should be separated from the edge of the box by more than 5 Bohr radii, whereas for the $n = 2$ we need more than 15 Bohr radii of separation. These conservative resolution requirements lead to the size of the box in z direction given by $L_z = 5 + 15 \times 4 = 65$ in the case when only $n = 2$ states are of interest.

We also use the same conservative requirements for the box size in the x (u)- and y (v)-directions. The size in the y-direction is fixed as soon as we decide on the upper n state of interest, while the size in x-direction also depends on the impact parameter b so that $L_x = L_y + b$. We further require high resolution near the particles so that $\Delta u \sim \Delta v \sim \Delta z \sim 0.18$. This resolution requirement was tested and found to converge with the difference between $\Delta \sim 0.18$ and $\Delta \sim 0.16$ being less than 0.1 percent. Several examples of input parameters required for accurate cross section results are provided in table 5.2.

One of the important advantages of the grid approach chosen in our calculations is the ability to visualize the evolution of the wave functions during the collision. In figure 5.5 we show the time evolution of electron's probability density function as the hydrogen atom moves past the stationary proton with impact parameter $b = 2$ and velocity $V = 1$ a.u. We clearly see that after the collision part electron's probability density is spread between the two atoms, indicating possible charge transfer during the impact.

5.3.2 Collisions with Large Impact Parameter

At large values of the impact parameter b the number of required grid points in the u and v directions increases dramatically. Fortunately, at large separations ($b \geq 5$) the cross section results are well described by the Born approximation if the collision energies are reasonably high ($E_{\mathrm{col}} \geq 1$ keV). To reduce the amount of numerical computations without sacrificing precision of the cross section results we use Born approximation to obtain transition probabilities above $b = 5$. It is easy to show that in this approximation the transition probability into the X final

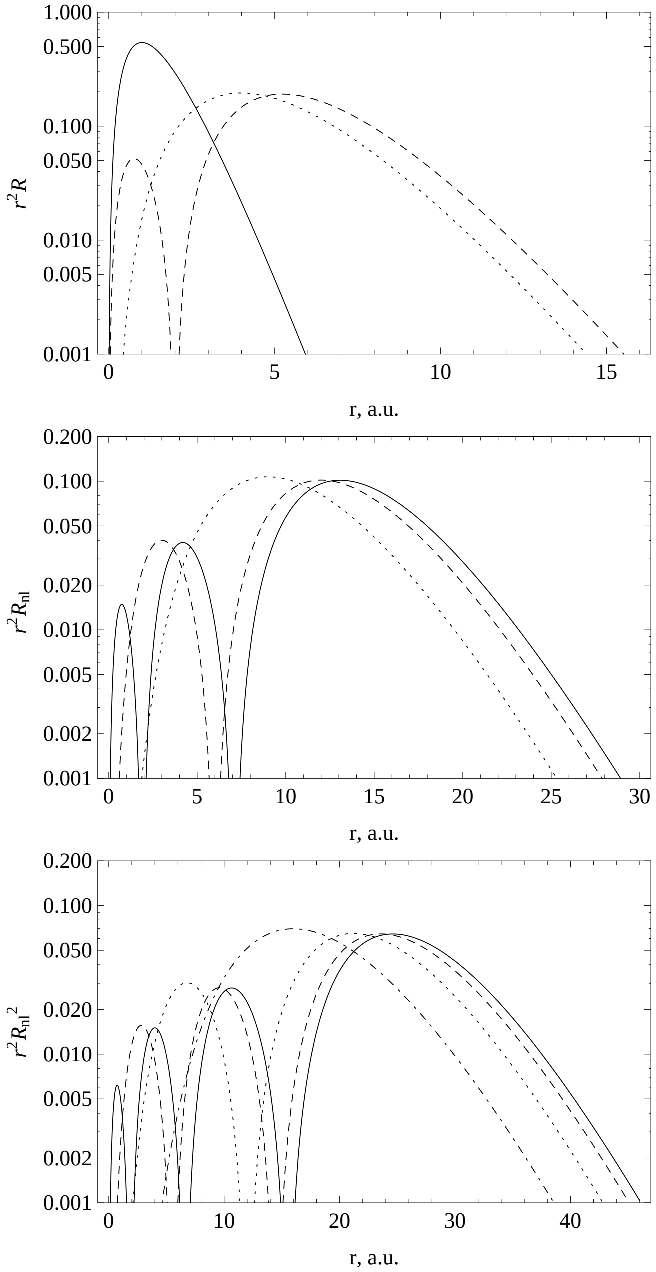


Figure 5.4: Plot on the left shows charge density distribution $r^2R_{nl}^2(r)$ for 1*s* (solid), 2*s* (dashed), and 2*p* (dotted) states. Plot on the right shows charge distribution for 3*s* (solid), 3*p* (dashed) and 3*d* (dotted) states.

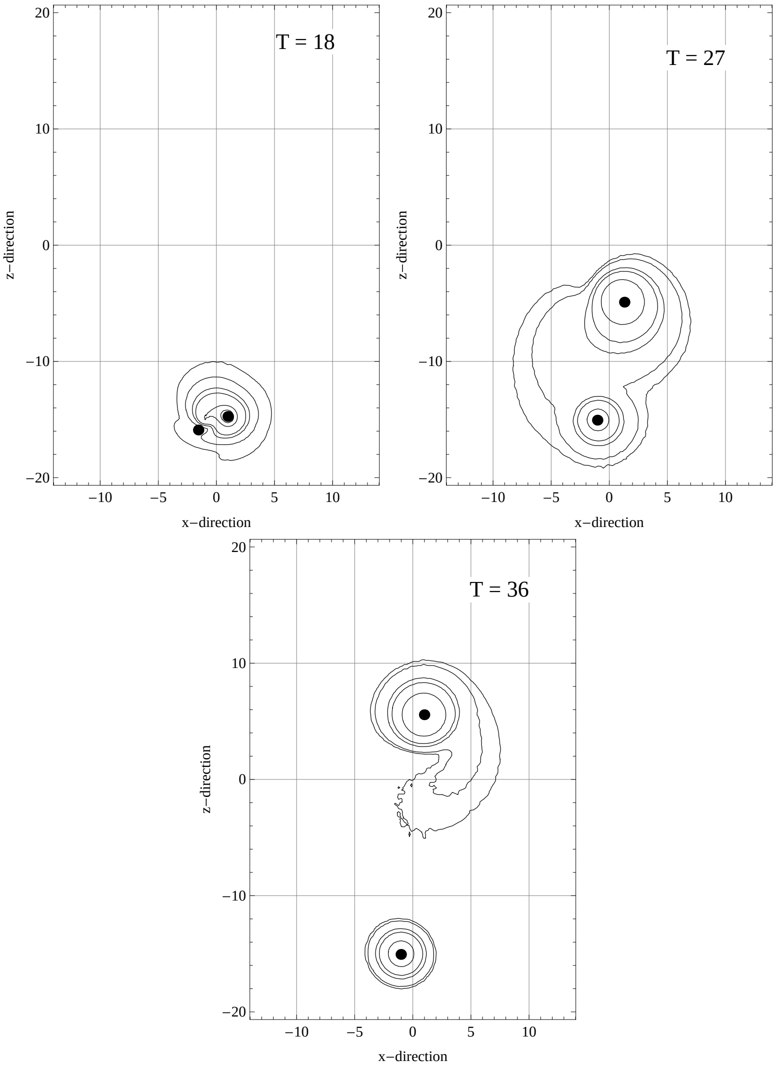


Figure 5.5: Contour plots showing time evolution of the electron probability density as hydrogen atom moves with $V = 1$ and $b = 2$ past the proton at rest. Time in the panels is given in atomic units.

state can be written as

$$|\langle X|\hat{S}(-T,T)|1s\rangle|^2 = \left|\int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}(E_X-E_0)t}\langle X|\hat{W}|1s\rangle\,\mathrm{d}t\right|^2, \tag{5.31}$$

where $\hat{W}$ is the perturbation Hamiltonian operator, which in our case is given by

$$\hat{W} = \frac{1}{R} - \frac{1}{|\mathbf{R}-\mathbf{r}|}, \tag{5.32}$$

and $\mathbf{R}$ is the vector separating the two protons. If one approximates the Hamiltonian as a dipole, $\hat{W} \approx -\mathbf{r}\cdot\mathbf{R}/|R|^3$, as appropriate at large impact parameter, then an analytic solution for the transition probability is obtained. In the first order, probability of excitation into $|X>$ can be written as

$$|\langle X|\hat{S}(-\infty,\infty)|1s\rangle|^2 \approx |-i\int_{-\infty}^{\infty} e^{-i(E_f-E_0)\xi/v}\times \tag{5.33}$$
$$(\frac{b}{(b^2+\xi^2)^{3/2}}\langle X|x|1s\rangle + \frac{\xi}{(b^2+\xi^2)^{3/2}}\langle X|z|1s\rangle)\frac{d\xi}{v}|^2.$$

For example, the first order probability of transition into $2p_0$ state becomes

$$|\langle 2p_0|\hat{S}(-\infty,\infty)|1s\rangle|^2 \approx \left(\frac{32\sqrt{2}K_0(\frac{3b}{8v})}{81v^2}\right)^2, \tag{5.34}$$

where $K_0(r)$ is a Bessel function of the second kind.

Note that the Born approximation only allows transitions into the $|np_\mathrm{A}\rangle$ final states (and to ionized final states).

At collision energies $E_\mathrm{col} > 1$ keV and impact parameters $b > 5$ the results produced by using Born approximation differ from the results obtained by our code by less than 5 percent, and because of the small value of the transition probabilities at large bs the difference introduced into cross section results is $\ll 1$ per cent.

5.3.3 Convergence and Consistency Tests

To test the proper functionality of our code and fully understand the dependence of the results on the input parameters such as grid resolution, the size of the box within which we keep a large number of point in q-space, box size and others we ran a large number of tests most important of which are

1. Time-dependent evolution of an electron in the field of a single proton at rest. In this case we would expect that the state of a system remains unchanged, i.e., if a system starts in $|1s\rangle$ state it remains there and probabilities of transitions into states with $n = 2$ vanish within the numerical precision of a discretized Hamiltonian system. Figure 5.6 shows the result of system's evolution over a few dynamical times ($T_{dyn} = 2\pi n^3$). We clearly see that the system exhibits small oscillations associated with discretization of the

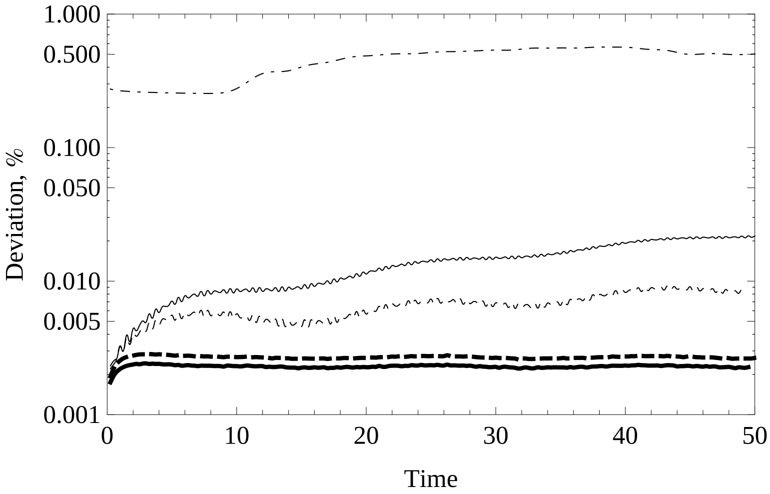


Figure 5.6: Deviation of $|\langle 1s|\hat{S}(T)|1s\rangle|$ from unity as a function of time in a test runs without the second particle. Thick curves show the evolution of a stationary hydrogen atom located at $x = 0.5$, $y = 0$ and $z = -20$ for the cases with unmodified $1/r$ potential (solid) and modified potential of equation 5.28 (dashed). Regular curves show the evolution of a hydrogen moving in the positive z-direction with $V = 1$ with unmodified potential (solid) and capped potential (dashed). Dash-dotted curve shows deviation from the exact energy expectation $\langle H_0\rangle = -1/(2n^2)$ for stationary hydrogen.

Hamiltonian, however, the results remain close to the expected values for $|\langle 1s|S(T)|1s\rangle|^2$ and for the energy expectation $\langle H_0\rangle$. For example, $|\langle 1s|\hat{S}(T)|1s\rangle|^2 \approx 1$ over the full time period with the precision better than 0.05 percent; energy expectation deviates from the theoretical prediction $\langle H_0\rangle = -\frac{1}{2}n^{-2}$ by less than 0.5%. The probability of $|1s\rangle$ to $|2s\rangle$ transition remains small and oscillates with the period $T_{\rm osc} \sim 3\pi/4$ determined by the energy splitting between $n = 1$ and $n = 2$ levels. Furthermore, the precision of the results increases as we increase the resolution of the grid, allowing us to achieve desired accuracy of the final results through convergence.

2. The next test incorporates motion of the hydrogen atom. In this case initial wave function of the electron is phase-shifted by the factor $\mathrm{e}^{\mathrm{i}Vz}$ as discussed at the end of section 5.2.2. The results of these tests are also shown in figure 5.6. We clearly see that the normalization of the $|1s\rangle$ is properly conserved. We note, however, that the motion of a particle introduces small additional oscillations into the shapes of the curves.

3. We tested convergence of our cross section results by running our code with three different resolution values: $\Delta = 0.175, 0.19$ and 0.22 at collision energy $E_{col} = 80$ keV. The results converge, with cross section difference between the $\Delta = 0.175$ and $\Delta = 0.19$ cases being less than 3%. We also checked that increasing the box size at a fixed resolution does not modify our results, which indicates that the results of the boundary interactions are minimal for the set of parameters used in our runs.

It is important to note that while our cross section results converge within the limits of simulation accuracy, they still contain inherent uncertainty associated with our discretization procedure, finite resolution of the numerical grid, finite size of the collision box, approxima-

Ψ_{nlm}	Expression	Ψ_{nlm}	Expression
Ψ_{100}	$\frac{e^{-r}}{\sqrt{\pi}}$	Ψ_{200}	$\frac{e^{-r/2}(2-r)}{4\sqrt{2\pi}}$
Ψ_{210}	$\frac{e^{-r/2}r\text{Cos}(\theta)}{4\sqrt{2\pi}}$	Ψ_{211}	$\mp\frac{e^{\pm i\phi}e^{-r/2}r\text{Sin}[\theta]}{8\sqrt{\pi}}$
Ψ_{300}	$\frac{e^{-r/3}(27-18r+2r^2)}{81\sqrt{3\pi}}$	Ψ_{310}	$\frac{e^{-r/3}(4-2r/3)r\text{Cos}(\theta)}{27\sqrt{2\pi}}$
Ψ_{311}	$\frac{e^{i\phi-r/3}(4-2r/3)r\text{Sin}(\theta)}{54\sqrt{\pi}}$	Ψ_{320}	$\frac{e^{-r/3}r^2(-1+3\text{Cos}(\theta)^2)}{81\sqrt{6\pi}}$
Ψ_{321}	$\frac{e^{i\phi-r/3}r^2\text{Cos}(\theta)\text{Sin}(\theta)}{81\sqrt{\pi}}$	Ψ_{322}	$\frac{e^{2i\phi-r/3}r^2\text{Sin}(\theta)^2}{162\sqrt{\pi}}$
Ψ_{400}	$\frac{e^{-r/4}(192-144r+24r^2-r^3)}{1536\sqrt{\pi}}$	Ψ_{410}	$\frac{e^{-r/4}(80-20r+r^2)r\text{Cos}(\theta)}{512\sqrt{5\pi}}$
Ψ_{411}	$\frac{e^{i\phi-r/4}(80-20r+r^2)r\text{Sin}(\theta)}{512\sqrt{10\pi}}$	Ψ_{420}	$\frac{e^{-r/4}\left(6-\frac{r}{2}\right)r^2\left(-1+3\text{Cos}[\theta]^2\right)}{1536\sqrt{\pi}}$
Ψ_{421}	$\frac{e^{i\phi-r/4}(6-r/2)r^2\text{Cos}(\theta)\text{Sin}(\theta)}{256\sqrt{6\pi}}$	Ψ_{422}	$\frac{e^{2i\phi-r/4}(6-r/2)r^2\text{Sin}(\theta)^2}{512\sqrt{6\pi}}$
Ψ_{430}	$\frac{e^{-r/4}r^3(-3\text{Cos}(\theta)+5\text{Cos}(\theta)^3)}{3072\sqrt{5\pi}}$	Ψ_{431}	$\frac{e^{i\phi-r/4}r^3(-1+5\text{Cos}(\theta)^2)\text{Sin}(\theta)}{2048\sqrt{15\pi}}$
Ψ_{432}	$\frac{e^{2i\phi-r/4}r^3\text{Cos}(\theta)\text{Sin}(\theta)^2}{1024\sqrt{6\pi}}$	Ψ_{433}	$\frac{e^{3i\phi-r/4}r^3\text{Sin}(\theta)^3}{6144\sqrt{\pi}}$

Table 5.3: The list of normalized hydrogen eigenfunctions Ψ_{nlm} up to $n = 4$.

tion of straight line trajectory for colliding protons, discretization of the transition probability on a grid of impact parameters b, and use of Born approximation at high values of impact parameters. Numerical errors associated with these sources are extremely hard to quantify precisely, although for each individual source the error has been tested and minimized with the convergence procedure to be below 1%.

In testing we also compared our predictions for low n ($n \leq 3$) cross sections with the results of previous studies and showed that our results are fully consistent with the results presented in [243], [252], and [241]. We note, however, that the major goal of this study is to introduce the computational algorithm and the code that allow accurate cross section calculations at high values of n as well as illustrate the importance of these calculations in studying Balmer-dominated shocks. Therefore, detailed comparison of our results to the results of other studies is not performed here and will be presented in future papers.

5.4 Cross Section Results

In this section, we present results for the cross sections for excitation and charge transfer final states, and (where possible) compare our results to previous computations.

We obtain cross sections by integrating equation (5.16), with $|X\rangle = |nlm\rangle$. The final state wave functions in the rest frame of the final atom (A or B) are given by

$$|nlm\rangle = \Psi_{nlm}(r, \theta, \phi) = R_{nl}(r)Y_l^m(\theta, \phi). \tag{5.35}$$

Here, $R_{nl}(r)$ are normalized radial eigenfunctions (equation 5.30) and $Y_l^m(\theta, \phi)$ are spherical harmonics. We provide explicit expressions for normalized hydrogen eigenfunctions up to $n = 4$ in table 5.3. If the final atom is moving, then a boost must be applied to the relevant wave function.

The first results obtained by running our code correspond to collisions involving only $n = 1$

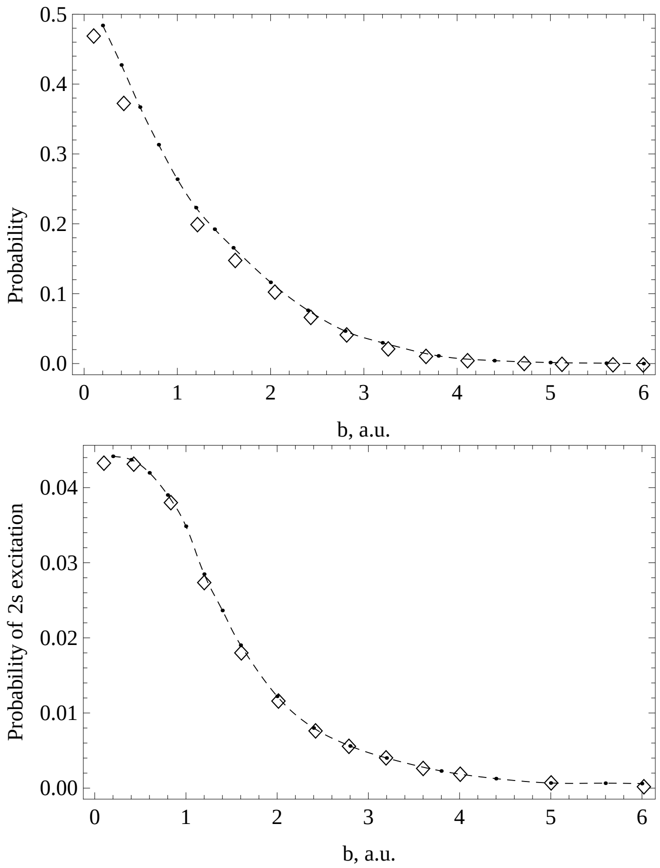


Figure 5.7: Upper plot shows probability of charge transfer into 1s state as a function of impact parameter for $E_{col} = 40$ keV, and the lower plot shows probability of excitation into 2s state. Diamonds show corresponding results from Kolakowska et al., 98 [251].

and $n = 2$ states. This problem is the least computationally demanding, and it has a large amount of data produced by previous studies, allowing for further tests of the code. In figure 5.7 we plot probabilities of excitation to $2s$ state and charge transfer into $1s$ state in collisions with impact energy $E_{\mathrm{col}} = 40$ keV. We compare our results with the results obtained in [251] and show good agreement over the full range of b with slight deviation at small values of impact parameter caused by higher resolution used in our runs. Our deviation in the case of charge transfer into $1s$ state causes higher value of the overall cross section, which is consistent with the results obtained by other groups–see for example [241] and references therein.

A significant advantage of using direct solution of the Schrödinger equation on a grid is the ability to see the convergence of the probability results. In figure 5.8 we show an example of probability evolution for excitations into $2s$ states. This method clearly allows to see the convergence of the numerical calculation and allows direct comparison of runs with various box parameters.

In the current study we focus on the introduction of the formalism and the code for high-n final states as well as on the preliminary illustration of how our results apply to the studies of astrophysical shocks. We also note that our results are fully consistent with the results obtained by other groups for low n cases. For example, in figure 5.9 we plot cross section for charge transfer into $1s$ state for various collision energies and compare our results with other

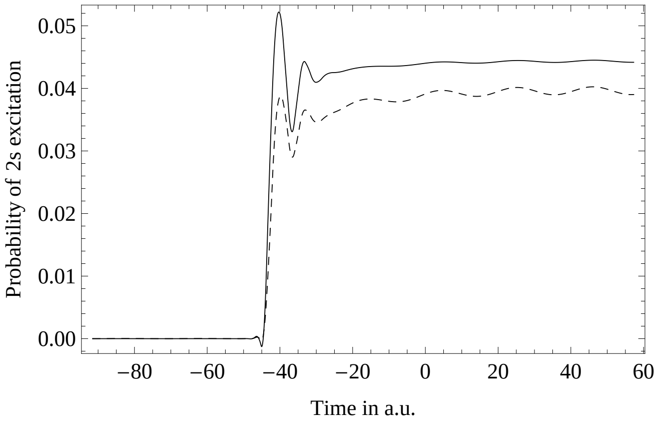


Figure 5.8: Probability of excitations into $2s$ states as a function of time for 2 different values of impact parameter $b = 0.2$ a.u. (solid curve) and $b = 0.8$ a.u. (dashed curve).

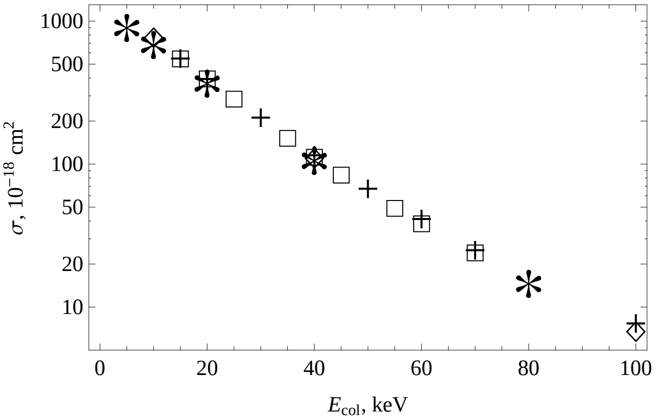


Figure 5.9: Excitation cross sections (in units of 10^{-18} cm^2) for charge transfer into $1s$ state. Stars show results of our calculations, diamonds correspond to Kolakowska (1998), squares show results from Winter (2009) and plus signs correspond to the results from Kuang and Lin (1996).

theoretical calculations [243, 251, 241].

In this study we limited our cross section calculation to $n = 4$, which allows us to calculate Balmer decrement discussed in the next section. Our results, showing cross sections of charge transfer and excitations are provided in tables 5.4 and 5.5. In appendixes A and B we also provide Chebyshev polynomial fits to our results for excitations and charge transfer into $3s$, $3p$, $3d$, $4s$, $4p$, $4d$, and $4f$ states, and plot obtained cross sections along with the results from [251] and [241]. We note, that the `BDSCx` code produces the cross sections for transitions into states with various n, l, and m, allowing for studies of polarization dependent signal. Calculation of higher n states, as well as more detailed analysis of the q-drop procedure are relegated to the future work.

σ/V	$2s$	$2p_0$	$2p_{\pm1}$	$3s$	$3p_0$	$3p_{\pm1}$	$3d_0$	$3d_{\pm1}$	$3d_{\pm2}$	
5 keV	6	3	13	0.35	0.4	0.65	0.18	1.3	0.01	
10 keV	4.1	5.7	6.9	0.8	0.6	1.3	0.3	0.6	0.009	
20 keV	12.0	15.1	10.4	2.2	1.9	1.8	0.8	0.6	0.07	
40 keV	15.8	26.5	19.5	3.7	4.3	3.3	1.1	0.7	0.15	
80 keV	11.5	27	23	2.7	4.4	4.1	0.6	0.4	0.18	

	$4s$	$4p_0$	$4p_{\pm1}$	$4d_0$	$4d_{\pm1}$	$4d_{\pm2}$	$4f_0$	$4f_{\pm1}$	$4f_{\pm2}$	$4f_{\pm3}$
5 keV	0.1	0.15	0.08	0.1	0.35	0.002	0.03	0.2	0.004	2×10^{-4}
10 keV	0.39	0.17	0.38	0.16	0.34	.0.003	0.04	0.08	8×10^{-4}	4×10^{-5}
20 keV	0.8	0.75	0.58	0.35	0.29	0.03	0.07	0.04	0.005	8×10^{-4}
40 keV	1.35	1.5	1.1	0.6	0.4	0.07	0.06	0.04	0.007	8×10^{-4}
80 keV	1.1	1.6	1.5	0.3	0.2	0.08	0.015	0.008	0.003	0.001

Table 5.4: Cross section results for excitation transitions into $n = 2,\ 3$ levels.

σ/V	$1s$	$2s$	$2p_0$	$2p_{\pm1}$	$3s$	$3p_0$	$3p_{\pm1}$	$3d_0$	$3d_{\pm1}$	$3d_{\pm2}$
5 keV	1092	5.8	2.1	11.5	0.25	0.35	0.6	0.2	1.0	0.008
10 keV	795	18	4.9	13	1.5	0.9	1.5	0.3	1.1	0.05
20 keV	425	39	7.9	6.5	8.7	2.5	1.7	0.35	0.28	0.03
40 keV	120	22	3.6	1.5	6.5	1.3	0.4	0.1	0.03	0.005
80 keV	17	3.5	0.49	0.15	1.1	0.2	0.05	0.01	0.004	5×10^{-4}

	$4s$	$4p_0$	$4p_{\pm1}$	$4d_0$	$4d_{\pm1}$	$4d_{\pm2}$	$4f_0$	$4f_{\pm1}$	$4f_{\pm2}$	$4f_{\pm3}$
5 keV	0.07	0.2	0.1	0.2	0.3	0.03	0.05	0.2	0.001	2×10^{-5}
10 keV	0.74	0.39	0.34	0.11	0.50	0.02	0.04	0.02	0.009	1×10^{-4}
20 keV	2.9	1.2	0.6	0.3	0.17	0.02	0.03	0.02	0.003	6×10^{-4}
40 keV	2.7	0.65	0.17	0.18	0.02	0.004	0.005	0.0015	1×10^{-4}	1×10^{-5}
80 keV	0.5	0.1	0.02	0.007	0.002	0.0003	2×10^{-4}	6×10^{-5}	1×10^{-5}	1×10^{-6}

Table 5.5: Cross section results (in units of 10^{-18} cm^2) for charge transfer transitions into $n = 2,\ 3$ and 4 levels.

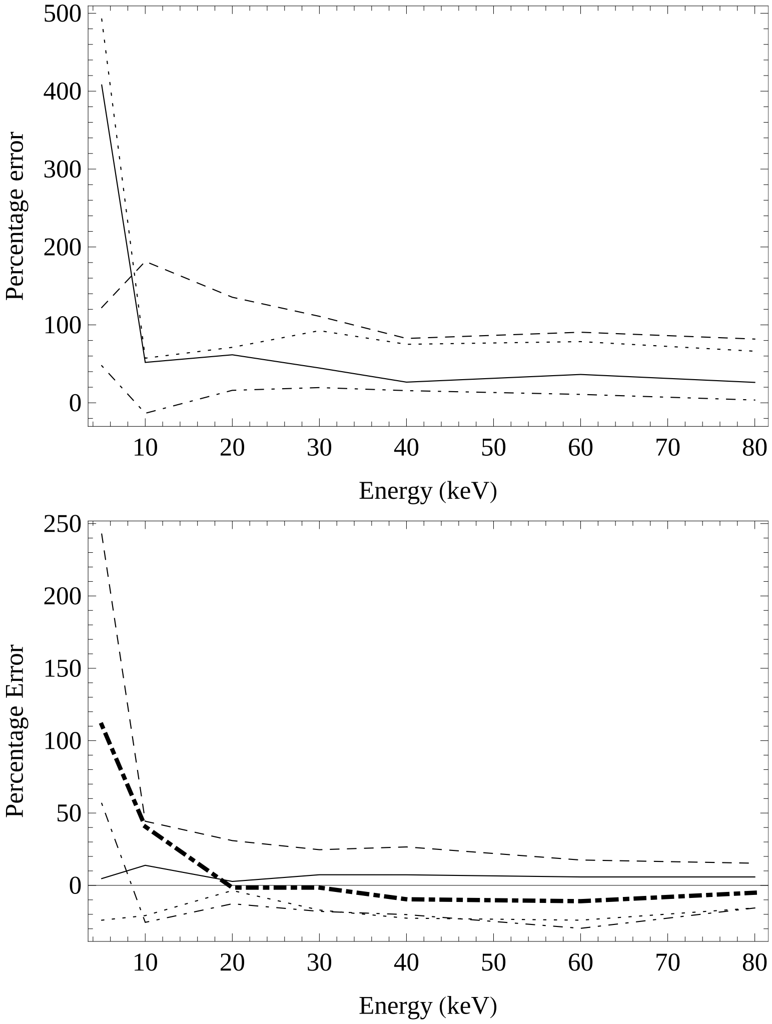


Figure 5.10: Errors associated with the calculated versus extrapolated (via equation 5.13) cross sections as a function of the impact energy. Upper panel shows percentage errors for $3s$ (solid curve), $3p_0$ (dashed curve), $3p_1$ (dotted curve), and $4s$ (dot-dashed curve) excitations. Lower panel shows errors for $4p_0$ (solid curve), $4p_1$ (dashed curve), $4d_0$ (dotted curve), $4d_1$ (dot-dashed curve), and $4d_2$ (thick dot-dashed curve) transitions.

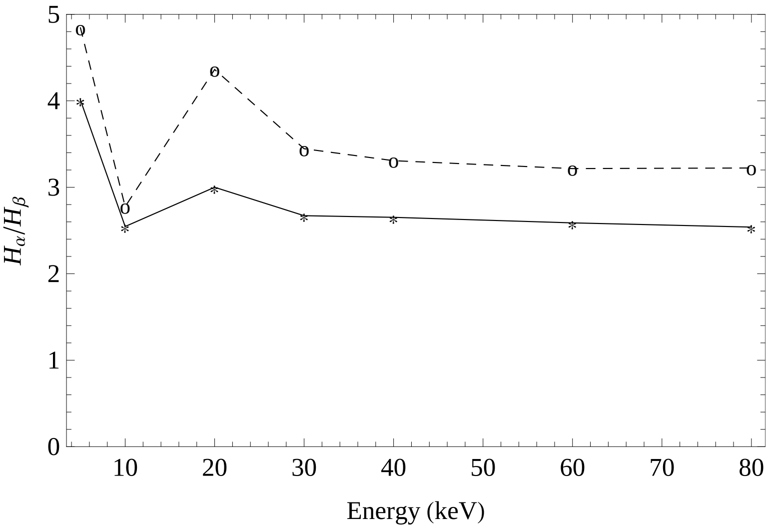


Figure 5.11: The Balmer decrement (Hα/Hβ) as a function of the impact energy for collisional excitation of hydrogen atoms by protons. Solid line corresponds to Case B, while dashed line shows Case A.

5.5 Astrophysical Applications

5.5.1 Errors Associated with Extrapolating from Cross Sections with Lower n-values

In the absence of available cross sections, one is often forced to extrapolate from known cross sections with lower principal quantum numbers n (e.g.,, [224]), thus generating errors in the calculation of line profiles and intensities which are unquantifiable. With the benefit of now being able to calculate excitation cross sections for the H-H$^+$ collisional system, we quantify the error associated with using the scaling law in equation (5.13) for obtaining cross sections with $n \geq 3$. In figure 5.10, we see that the errors associated with extrapolating for $n = 4$ cross sections from $n_0 = 3$ is typically a factor ~ 2. The less necessary extrapolation of obtaining $n = 3$ cross sections from $n_0 = 2$ (since data for $n = 3$ is available) results in errors of a factor ~ 5. Since these errors are nonnegligible, they probably dominate any uncertainty associated with a numerical integration technique used to compute line profiles and intensities. A direct calculation, such as the one we have performed in this study, is necessary in order to obtain accurate cross sections and in turn perform a spectral analysis of lines such as Hβ at energies $\gtrsim 5$ keV.

5.5.2 The Balmer Decrement

The Balmer decrement generally refers to the ratio of different lines in the Balmer series: Hα/Hβ, Hβ/Hγ, etc. It is somewhat insensitive to the electron temperature and (low) densities. For example, Hα/Hβ has a value of about 2–3 [254], unless collisional excitation of hydrogen atoms by electrons dominate in which case its value is as high as 8 [255], which only occurs at electron densities $\sim 10^4$ cm^{-3} or higher. The relative insensitivity of the Balmer decrement to the atomic processes makes it an appropriate diagnostic for the presence of dust extinction, since the bluer lines in the Balmer series (e.g.,, Hβ) are subjected to increased extinction and thus the Balmer decrement attains a value larger than, e.g.,, 2–3 in the case of Hα/Hβ. A high value of the Balmer decrement may also be attained if the population of excited hydrogen (e.g.,, 2s) is sufficient to cause self-absorption in the Balmer lines [256, 257]. As examples, the Balmer decrement has been used as a diagnostic in the study of supernovae (e.g.,, [258]), active galactic nuclei (e.g.,, [259]) and the *Sloan Digital Sky Survey (SDSS)* sample of galaxies (e.g.,, [260]).

Since our work was originally motivated by the study of fast astrophysical shocks, we use them as an example in the calculations presented in this subsection. In particular, we use the example of Balmer-dominated shocks, which are fast ($\sim$ 1000 km s^{-1}) shocks impinging upon tenuous media ($\sim$ 1 cm^{-3}). As Balmer-dominated shocks are mostly observed around young ($\lesssim$ 1000 years), Galactic supernova remnants, the production of photons via radiative recombination is unimportant since the recombination time is $\sim 10^4$ years. The calculation of line intensities then requires knowledge of how the various atomic levels of hydrogen are populated via collisions as well as their subsequent rates of radiative decay. For a strong shock (Mach number greatly exceeding unity), the relative velocity between the electrons/protons and hydrogen atoms is $\delta v = 3v_s/4$, where v_s denotes the shock velocity. Thus, the shock velocity can be related to the interaction energy E via

$$v_s = \frac{4}{3}\sqrt{\frac{2E}{m_{\rm H}}} \approx 1300 \text{ km s}^{-1} \left(\frac{E}{5 \text{ keV}}\right)^{1/2}, \tag{5.36}$$

where $m_{\rm H}$ is the mass of the hydrogen atom. The cross sections presented in the present study are thus relevant to shocks with $v_s \approx$ 1300–5200 km s^{-1}.

Formally, evaluating the rate at which a given nl level is populated by collisions requires the calculation of the rate coefficient [224],

$$\mathcal{R} = \iiint \iiint F_1 \left(\mathbf{v}_1\right) \ F_2 \left(\mathbf{v}_2\right) \ \Delta v \ \sigma \left(\Delta v\right) \ d^3 v_1 \ d^3 v_2. \tag{5.37}$$

The preceding, six-dimensional integral is evaluated over all of the relative velocities (Δv) between the velocity distributions of the atoms (F_1) and electrons/protons (F_2), weighted by the relevant cross section (σ) of the atomic process being considered.

To gain an intuition for the functional dependence of the Balmer decrement on the interaction energy E, it is sufficient to consider either individual pairs of particles or particles in Delta-function distributions. In this case, the rate coefficient reduces to $\mathcal{R} = \sigma \Delta v$. The Balmer decrement is then the sum of the cross sections for the collisional population of each nl level weighted by the appropriate branching ratio,

$$\text{H}\alpha/\text{H}\beta = \frac{\sigma\left(3\text{s}\right) + B_{3\text{p},2\text{s}}\sigma\left(3\text{p}\right) + \sigma\left(3\text{d}\right)}{B_{4\text{s},2\text{p}}\sigma\left(4\text{s}\right) + B_{4\text{p},2\text{s}}\sigma\left(4\text{p}\right) + B_{4\text{d},2\text{p}}\sigma\left(4\text{d}\right)}. \tag{5.38}$$

The branching ratio is simply the Einstein A-coefficient for a given transition normalized by the

Einstein A-coefficients of all of the transitions allowed by the electric dipole selection rule. For example, $B_{3\mathrm{p},2\mathrm{s}} = A_{3\mathrm{p},2\mathrm{s}}/(A_{3\mathrm{p},2\mathrm{s}} + A_{3\mathrm{p},1\mathrm{s}})$.

Two extreme limiting cases are typically considered: Case A and Case B [261, 262, 263]. Case B occurs when the neutral hydrogen column density is large enough for Lyman lines to be optically thick. They undergo multiple scatterings and are eventually degraded into a Balmer line and Lyα or two-photon emission. Case A occurs when the neutral hydrogen column density is small enough for Lyman lines to be optically thin and so freely escape the cloud. In a Case A scenario, we have (e.g.,, [264]),

$$\begin{aligned} B_{3\mathrm{p},2\mathrm{s}} &\approx 0.1183, \\ B_{4\mathrm{s},2\mathrm{p}} &\approx 0.5841, \\ B_{4\mathrm{p},2\mathrm{s}} &\approx 0.1191, \\ B_{4\mathrm{d},2\mathrm{p}} &\approx 0.7456. \end{aligned} \tag{5.39}$$

In a Case B scenario, these branching ratios are essentially unity, since $B_{4\mathrm{p},3\mathrm{s}} \sim 10^{-2}$ and $B_{4\mathrm{p},3\mathrm{s}} \sim 10^{-3}$.

In figure 5.11, we show calculations of the Balmer decrement, for both Case A and B, using equation (5.38). For simplicity, we consider only the collisional excitation (and not charge transfer) of hydrogen atoms by protons; collisional excitation by electrons is subdominant at these energies. It is apparent that Hα/Hβ remains somewhat constant at values of 2–3 and depends weakly on whether Case A or B applies. This implies that even in situations where Balmer-dominated shocks may contribute to the overall signal of an astrophysical object, the Balmer decrement remains a useful diagnostic for constraining the degree of dust extinction.

5.6 Summary

In this work we introduced a new formalism for computing precise cross sections for high-nl proton-hydrogen collisions and developed a numerical code which implements our formalism. We further used our code to obtain accurate cross sections for collisions between protons and hydrogen atoms which start in the ground state. Our cross section results focused on the energy range of direct interest for the studies of Balmer-dominated shocks, and we are currently using the produced results to prepare a detailed analysis of various physical processing occurring inside of the shocks. Specifically, we are working to extend a formalism developed in [224] to include the effects of cosmic rays. By cosmic rays, we refer to energetic protons and electrons produced in the postshock gas which are able to travel upstream and impart some of their energy to the preshock gas. The goal is to construct a phenomenological model with only one extra parameter describing this process.

The code introduced in this chapter has a large number of potential applications in atomic physics. Using the q-drop procedure and curvilinear coordinates proposed in our work enables relatively inexpensive calculations of charge exchange and excitation cross sections for proton-hydrogen collision with $n \leq 7$. In this chapter we focused on the formalism and numerical implementation of the proposed cross section calculations, while detailed tests of the q-drop procedure and the cross section results for $n > 4$ will be reported in the future papers.

Acknowledgements

We are grateful to Avi Loeb for his help with getting the time on the Odyssey cluster supported by the FAS Science Division Research Computing Group at Harvard University. We are also grateful to Mark Scheel for helping to get the time on the SHC cluster at Caltech. D. T. and C. H. are supported by the U.S. Department of Energy (DE-FG03-92-ER40701) and the National Science Foundation (AST-0807337). C. H. is supported by the David and Lucile Packard Foundation.

A Fitting Functions for Computed Cross Sections

In order to facilitate broader use of our cross section results, we are providing a fitting functions which allow a more straightforward utilization of our results. We fit the obtained cross sections for excitations and charge transfer into $3s$, $3p$, $3d$, $4s$, $4p$, $4d$, and $4f$ states with a series of Chebyshev orthogonal polynomials following the practice proposed in [264]. The fitting function is of the form:

$$\mathcal{F}\left(x;\mathbf{A}\right) = \exp\left(\frac{A_0}{2} + \sum_{i=1}^{4} A_i \, \mathcal{C}_i\left(x\right)\right), \tag{A40}$$

where the coefficients $\mathbf{A} = A_i$ for $0 \leq i \leq 4$ are the fitting parameters. The quantities $\mathcal{C}_i$ are the Chebyshev orthogonal polynomials:

$$\mathcal{C}_1\left(x\right) = x, \tag{A41}$$

$$\mathcal{C}_2\left(x\right) = 2x^2 - 1, \tag{A42}$$

$$\mathcal{C}_3\left(x\right) = 4x^3 - 3x, \tag{A43}$$

$$\mathcal{C}_4\left(x\right) = 8\left(x^4 - x^2\right) + 1. \tag{A44}$$

The fitting variable x is defined as

$$x = \frac{\ln\left(E/E_{\text{min}}\right) - \ln\left(E_{\text{max}}/E\right)}{\ln\left(E_{\text{max}}/E_{\text{min}}\right)}, \tag{A45}$$

where E is the relative energy between the proton and hydrogen atom; E_{min} and E_{max} are the respective minimum and maximum energies in our simulation. We use the Newton fitting algorithm which provides local-optimal fit to the array of available cross section results.

The fitting parameters are provided in tables A6 and A7.

	A0	A1	A2	A3	A4
3s	0.704603	1.19152	-0.408371	-0.16998	0.0277847
3p	3.37863	1.06547	-0.094358	-0.0800694	-0.0789251
3d	1.48837	0.0532829	-0.213154	-0.285436	0.266437
4s	-1.16343	1.2132	-0.440247	-0.0142552	-0.0816767
4p	-0.153539	0.125901	-0.0884664	-0.0897404	-0.0217473
4d	0.942133	1.35938	-0.234833	-0.0107613	-0.0587963
4f	-3.6574	-0.874456	-0.10575	-0.325656	-0.109631

Table A6: Fitting coefficients A_i corresponding to excitation transitions into $n = 3$ and $n = 4$ states.

B Comparison of Computed Cross Sections with Previous Work

This appendix provides figures of cross sections for excitation and charge transfer transitions into $3s$, $3p$, $3d$, $4s$, $4p$, $4d$, and $4f$ states. Our results are compared with the results of several other theoretical studies. Along with discrete cross section data points we are providing the

	A0	A1	A2	A3	A4
3s	1.08785	0.982647	-1.40441	-0.241845	0.21499
3p	1.1467	-0.753756	-1.07891	-0.0673582	0.122695
3d	-1.60258	-2.49308	-0.778485	0.117194	-0.00765272
4s	-0.656038	1.08682	-1.37046	-0.103767	0.0222725
4p	-2.42829	-1.9747	-0.959062	-0.178264	-0.13058
4d	-0.922396	-0.375844	-1.15834	-0.149067	0.178331
4f	-7.64717	-3.27305	-0.887953	-0.37366	0.371682

Table A7: Fitting coefficients A_i corresponding to charge transfer into $n = 3$ and $n = 4$ states.

results of our Chebyshev polynomial fits described in appendix A.

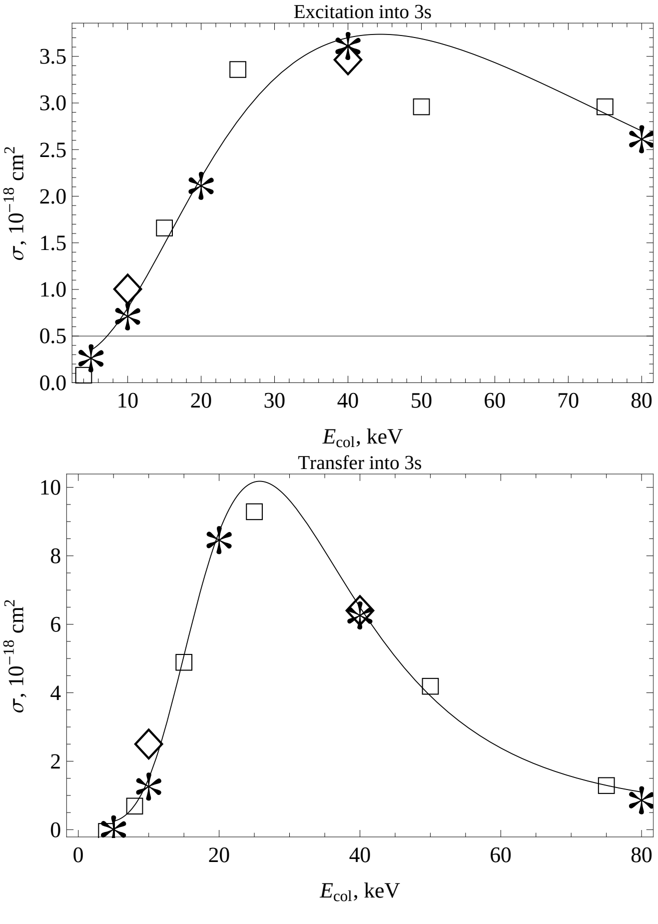


Figure 5.12: Cross sections for excitation (left panel) and charge transfer (right panel) into $3s$ state. Stars show results of our calculations, diamonds correspond to Kolakowska (1998), and squares show results from Winter (2009). Solid line corresponds to our Chebyshev polynomial fit.

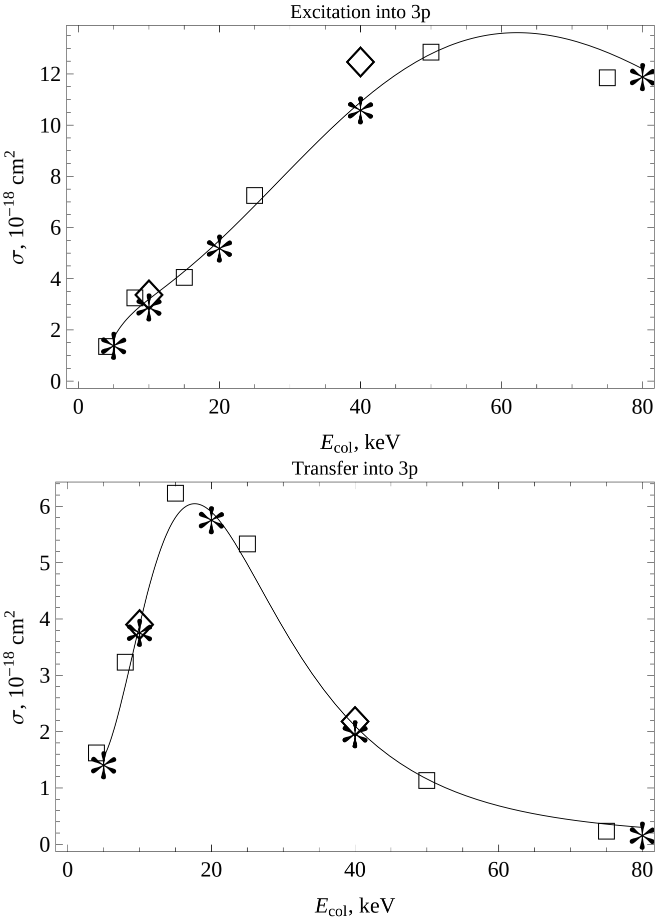


Figure 5.13: Cross sections for excitation (upper panel) and charge transfer (lower panel) into $3p$ state. Stars show results of our calculations, diamonds correspond to Kolakowska (1998), and squares show results from Winter (2009). Solid line corresponds to our Chebyshev polynomial fit.

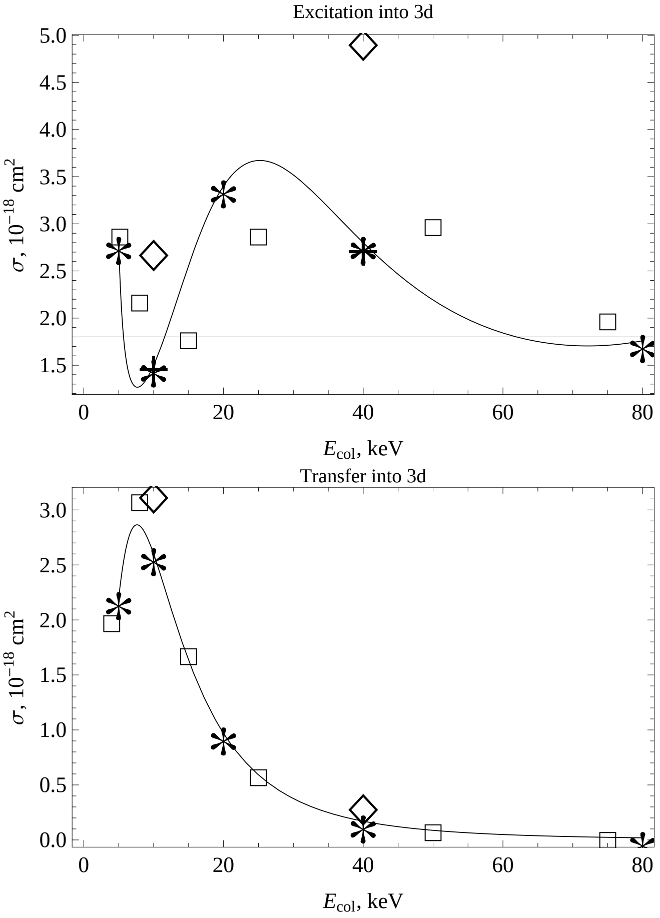


Figure 5.14: Cross sections for excitation (upper panel) and charge transfer (lower panel) into $3d$ state. Stars show results of our calculations, diamonds correspond to Kolakowska (1998), squares show results from Winter (2009) and plus signs correspond to the results from Kuang and Lin (1996). Solid line corresponds to our Chebyshev polynomial fit.

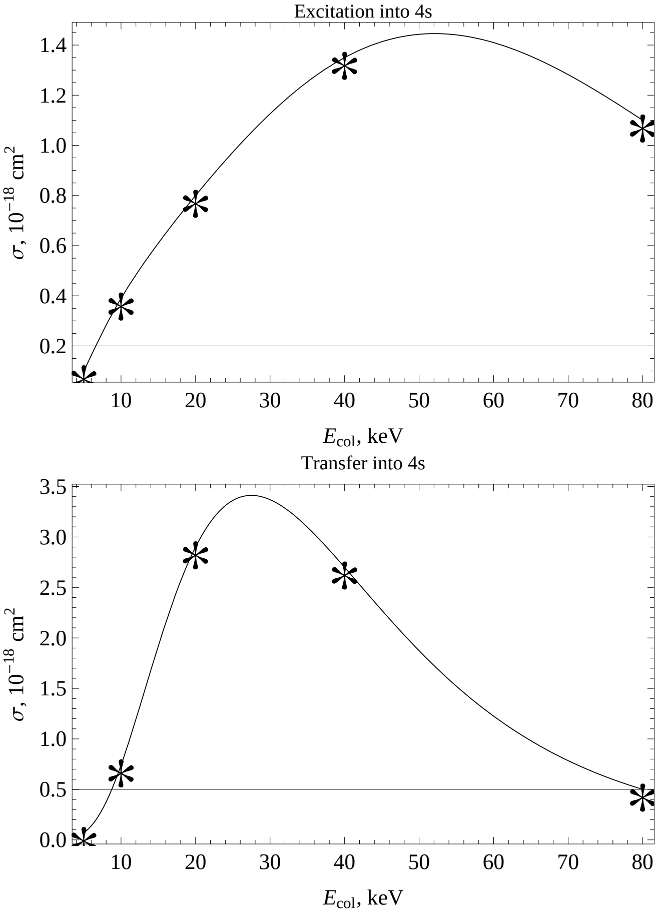


Figure 5.15: Cross sections for excitation (upper panel) and charge transfer (lower panel) into $4s$ state. Stars show results of our calculations. Solid line corresponds to our Chebyshev polynomial fit.

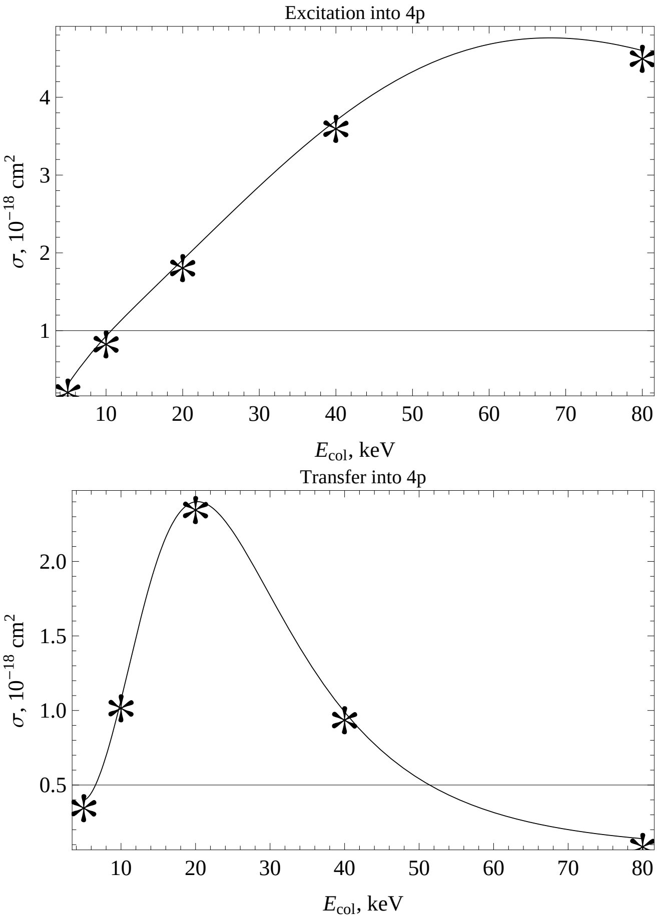


Figure 5.16: Cross sections for excitation (upper panel) and charge transfer (lower panel) into $4p$ state. Stars show results of our calculations. Solid line corresponds to our Chebyshev polynomial fit.

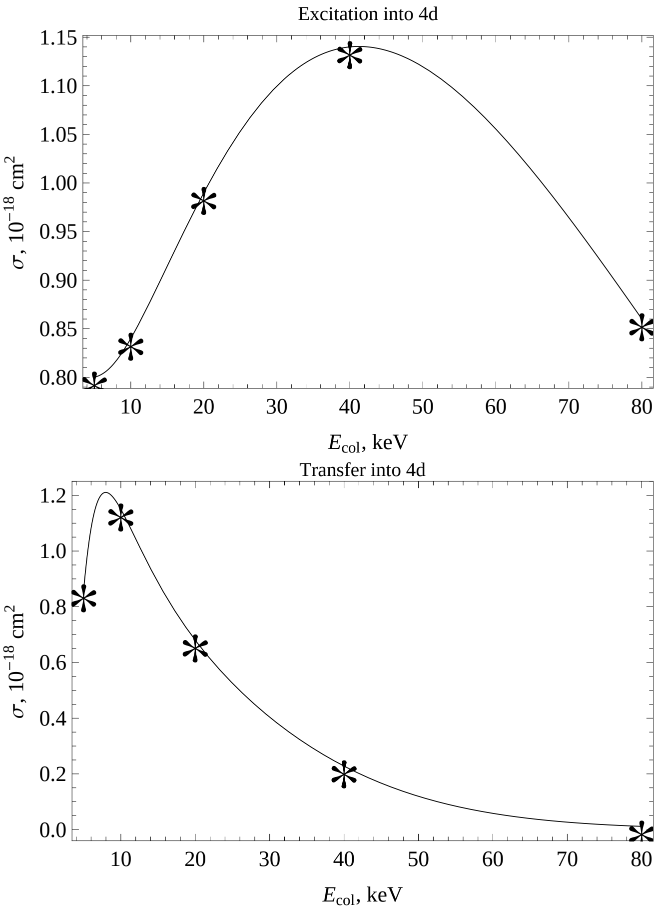
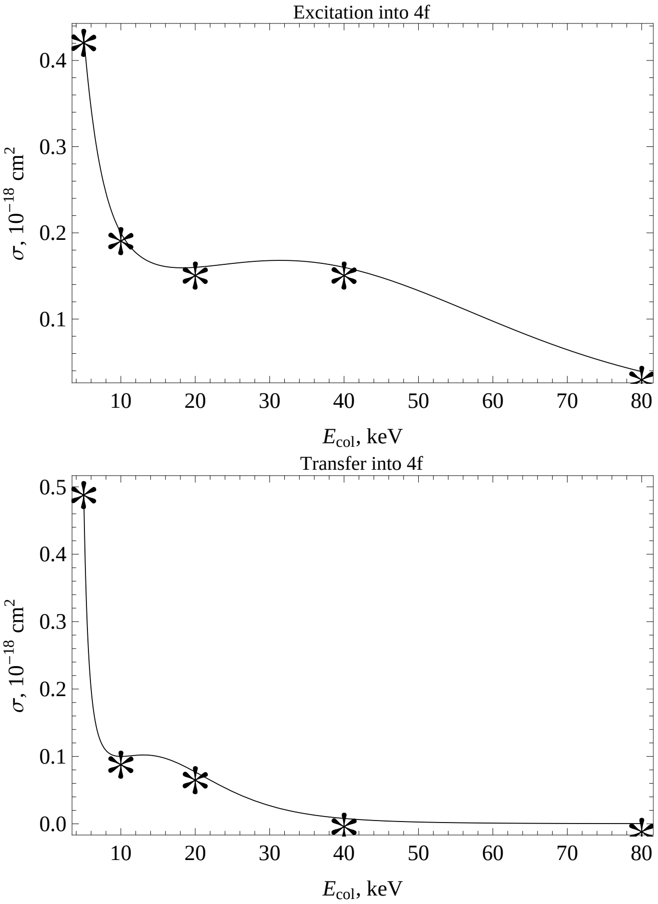


Figure 5.17: Cross sections for excitation (upper panel) and charge transfer (lower panel) into $4d$ state. Stars show results of our calculations. Solid line corresponds to our Chebyshev polynomial fit.



Excitation into 4f

Transfer into 4f

Figure 5.18: Cross sections for excitation (upper panel) and charge transfer (lower panel) into $4f$ state. Stars show results of our calculations. Solid line corresponds to our Chebyshev polynomial fit.

Bibliography

[1] E. L. Wright et al., Astrophys. J. Lett. **396**, L13 (1992).

[2] D. N. Spergel et al. (WMAP), Astrophys. J. Supp. **148**, 175 (2003).

[3] D. N. Spergel et al. Astrophys. J. Supp. **170**, 377 (2007).

[4] G. Hinshaw et al., Astrophys. J. Supp. **180**, 225 (2009).

[5] Y. D. Takahashi et al., Astrophys. J. **711**, 1141, (2010).

[6] A. E. Lange et al. (Boomerang Collaboration), Phys. Rev. D **63**, 042001 (2001).

[7] M. Tegmark et al. (SDSS), Phys. Rev. D 69, 103501 (2004).

[8] M. Tegmark et al., Astrophys. J. **606**, 702 (2004).

[9] A. G. Riess et al., Astron. J. **116**, 1009 (1998).

[10] B. P. Schmidt et al., Astrophys. J. **507**, 46 (1998).

[11] S. Perlmutter et al., Astrophys. J. **517**, 565 (1999).

[12] M. Tegmark et al., Phys. Rev. D **69**, 103501 (2004).

[13] D. N. Spergel et al., ApJS 170, 377 (2007).

[14] V. Springel, N. Yoshida, S. White, New Astron., **6**, 79 (2001).

[15] R. Kurzweil, *The Singularity is Near*, Penguin Publishing Group, (2005).

[16] A. R. Liddle and D. H. Lyth, *Cosmological Inflation and Large-Scale Structure*, Cambridge University Press, (2000).

[17] A. H. Guth, Phys. Rev. D **23**, 347 (1981).

[18] R. Blanford et al. (Decadal Survey), New Worlds New Horizons, The National Academies Press, (2010).

[19] M. F. Morales and J. N. Hewitt, Astrophys. J. **615**, 7 (2004).

[20] V. Jelic et al., Mon. Not. R. Astron. Soc. **389**, 1319 (2008).

[21] J. D. Bowman, M. F. Morales, and J. N. Hewitt, Astrophys. J. **695**, 18 (2009).

[22] S. R. Furlanetto, S. P. Oh, F. H. Briggs, Phys. Rep. **433**, 181 (2006).

[23] E. Bertschinger, Phys. Rev. D **74**, 063509 (2006).

[24] A. Loeb, M. Zaldarriaga, Phys.Rev.D **71**, 103520 (2005).

[25] A. A. Starobinskii, Soviet Journal of Experimental and Theoretical Physics Letters **30**, 682 (1979).

[26] A. H. Guth, Phys. Rev. D **23**, 347 (1981).

[27] A. D. Linde, Physics Letters B **108**, 389 (1982).

[28] A. Albrecht and P. J. Steinhardt, Physical Review Letters **48**, 1220 (1982).

[29] S. Dodelson, *Modern Cosmology*, Academic Press, (2003).

[30] D. Lyth and A. Riotto Phys. Rep. **314**, 1 (1999).

[31] J. Maldacena, Journal of High Energy Physics **5**, 13 (2003).

[32] V. Acquaviva, N. Bartolo, S. Matarrese, and A. Riotto, Nucl. Phys. **B667**, 119 (2003).

[33] P. Creminelli, JCAP **0310**, 003 (2003).

[34] D. H. Lyth and Y. Rodriguez, Phys. Rev. Letters **95**, 121302 (2005).

[35] T. J. Allen, B. Grinstein and M. B. Wise, Phys. Lett. **B197**, 66 (1987).

[36] L. A. Kofman and D. Y. Pogosian, Phys. Lett. **B214**, 508 (1988).

[37] D. S. Salopek, J. R. Bond and J. M. Bardeen, Phys. Rev. **D40**, 1753 (1989).

[38] A. D. Linde and V. F. Mukhanov, Phys. Rev. **D56**, R535 (1997).

[39] S. Mollerach, Phys. Rev. **D42**, 313 (1990).

[40] D. Lyth and D. Wands, Phys. Rev. Letters **B524**, 5 (2002).

[41] D. H. Lyth, C. Ungarelli, and D. Wands, Phys. Rev. D **67**, 023503 (2003).

[42] L. Boubekeur, D. H. Lyth, Phys. Rev. D **73**, 021301(R) (2006).

[43] Q. Huang, J. Cosm. Astropart. Phys. **0811**, 005 (2008).

[44] L. Randal, Perspectives on Supersymmetry, Advanced Series on Directions in High Energy Physics, Vol. 18, Edited by G. L. Kane, p. 293 (1998).

[45] J. Bock et al., *Task Force on Cosmic Microwave Background Research*, (2005).

[46] U. Seljak and M. Zaldarriaga, Phys. Rev. Letters **78**, 205 (1997).

[47] M. Kamionkowski, S. Kosowsky, and A. Stebbins, Phys. Rev. Letters **78**, 2058 (1997).

[48] Chiang et al., Astrophys. J. , **711**, 1123 (2010).

[49] E. Komatsu and D. N. Spergel, Phys. Rev. D **63**, 063002 (2001), `[astro-ph/0005036]`.

[50] L. Verde et al., Mon. Not. Roy. Astron. Soc. **313**, L141 (2000).

[51] L. Verde, L. Wang, A. Heavens, M. Kamionkowski, Mon. Not. R. Astron. Soc. **313**, 141 (2000).

[52] E. Sefusatti, E. Komatsu, Phys. Rev. D **76**, 083004 (2007).

[53] D. Jeong, E. Komatsu, Astrophys. J. **703**, 1230 (2009).

[54] N. Dalal, O. Doré, D. Huterer, and A. Shirokov, Phys. Rev. D **77**, 123514 (2008).

[55] A. Slosar et al., J. Cosm. Astropart. Phys. **08**, 031 (2008).

[56] C. Carbone, L. Verde, S. Matarrese, Astrophys. J. **684**, L1 (2008).

[57] N. Afshordi, A. J. Tolley, Phys. Rev. D **78**, 123507 (2008).

[58] P. McDonald, Phys. Rev. D **78**, 123519 (2008).

[59] V. Desjacques, U. Seljak, I. Iliev, Mon. Not. R. Astron. Soc. **396**, 85 (2009).

[60] M. Grossi et al., Mon. Not. R. Astron. Soc. **398**, 321 (2009).

[61] A. Pillepich, C. Porciani, O. Hahn, Mon. Not. R. Astron. Soc. **402**, 191 (2010).

[62] B. Reid, L. Verde, K. Dolag, S. Matarrese, L. Moscardini, J. Cosm. Astropart. Phys. **07**, 013 (2010).

[63] E. Komatsu *et al.*, Astrophys. J. Supp. **148**, 119 (2003).

[64] D. Spergel et al., Astrophys. J. Supp. **170**, 377 (2007).

[65] A. P. S. Yadav, B. D. Wandelt, Phys. Rev. Letters **100**, 181301 (2008).

[66] E. Komatsu et al., Astrophys. J. Supp. **180**, 330 (2009).

[67] E. Komatsu, et al., Astrophys. J. Supp. **192**, 18 (2011).

[68] K. Smith, L. Senatore, M. Zaldarriaga, J. Cosm. Astropart. Phys. **09**, 006 (2009).

[69] C. Hikage, K. Koyama, T. Matsubara, T. Takahashi and M. Yamaguchi Mon. Not. R. Astron. Soc. **398**, 2188 (2009).

[70] J. Xia, M. Viel, C. Baccigalupi, G. de Zotti, S. Matarrese, L. Verde, Astrophys. J. Lett. **717L**, 17 (2010).

[71] S. Cole and N. Kaiser, Mon. Not. R. Astron. Soc. **237**, 1127 (1989).

[72] D. Langlois and F. Vernizzi, Phys. Rev. D **70**, 063522 (2004).

[73] F. Ferrer, S. Rasanen, and J. Valiviita, J. Cosm. Astropart. Phys. **04**, 010 (2004).

[74] G. Lazarides, R. RuizdeAustri and R. Trotta, Phys. Rev. D **70**, 123527 (2004).

[75] D. Langlois, F. Vernizzi and D. Wands, J. Cosm. Astropart. Phys. **0812**, 004 (2008).

[76] D. Wands, Class. and Quant. Grav. **27**, 124002 (2010).

[77] M. Sasaki and E. D. Stewart, Prog. Theor. Phys. **95**, 71 (1996).

[78] A. R.Liddle and D. H. Lyth, *Cosmological inflation and large-scale structure*, Univ. Press (2000).

[79] M. Kawasaki, K. Nakayama, T. Sekiguchi, T. Suyama and F. Takahashi, J. Cosm. Astropart. Phys. **0811**, 019 (2008).

[80] M. Kawasaki, K. Nakayama, T. Sekiguchi, T. Suyama and F. Takahashi, J. Cosm. Astropart. Phys. **0901**, 042 (2009).

[81] W. H. Press and P. Schechter, Astrophys. J. **187**, 425 (1974).

[82] R. Sheth and G. Tormen, Mon. Not. R. Astron. Soc. **308**, 119 (1999).

[83] U. Seljak, Mon. Not. R. Astron. Soc. **318** 203 (2000).

[84] K. Gorski et al., Astrophys. J. **622**, 759 (2005).

[85] S. Ho, C. Hirata, N. Padmanabhan, U. Seljak, and N. Bahcall, Phys. Rev. D **78**, 043519 (2008).

[86] N. Padmanabhan et al., Mon. Not. R. Astron. Soc. **359**, 237 (2005).

[87] M. Tegmark et al., Phys. Rev. D **74**, 123507 (2006).

[88] G. T. Richards et al., Astron. J. **133**, 2222 (2007).

[89] G. T. Richards et al., Astron. J. **131**, 2766 (2006).

[90] Huterer D., Knox L., Nichol R. C., Astrophys. J. **555**, 547 (2001).

[91] Tegmark M. et al., Astrophys. J. **571**, 191 (2002).

[92] N. Padmanabhan et al., Mon. Not. R. Astron. Soc. **378**, 852 (2007).

[93] G. Hinshaw et al., Astrophys. J. Supp. **170**, 288 (2007).

[94] L. Page et al., Astrophys. J. Supp. **170**, 335 (2007).

[95] K. Grainge et al., Mon. Not. R. Astron. Soc. **341**, L23 (2003).

[96] C.-L. Kuo et al., Astrophys. J. **664**, 687 (2007).

[97] A. Readhead et al., Astrophys. J. **609**, 498 (2004).

[98] P. Astier et al., Astron. Astrophys. **447**, 31 (2006).

[99] A. H. Guth, Phys. Rev. D **23**, 347 (1981).

[100] A. D. Linde, Phys. Lett. B **108**, 389 (1982).

[101] A. Albrecht and P. J. Steinhardt, Phys. Rev. Letters **48**, 1220 (1982).

[102] E. Komatsu, et al., Astrophys. J. Supp. **192**, 18 (2011).

[103] M. Fukugita and M. Kawasaki, Mon. Not. R. Astron. Soc. **269**, 563 (1994).

[104] J. P. Ostriker and N. Y. Gnedin, Astrophys. J. Lett. **472**,L63 (1996).

[105] K. Yamamoto, N. Sugiyama, and H. Sato, Phys. Rev. D **56**, 7566 (1997).

[106] K. Yamamoto, N. Sugiyama, and H. Sato, Astrophys. J. **501**, 442 (1998).

[107] A. Loeb and M. Zaldarriaga, Phys. Rev. Letters **92**, 211301 (2004).

[108] R. Barkana and A. Loeb, Mon. Not. R. Astron. Soc. **363**, L36 (2005).

[109] S. Naoz and R. Barkana, Mon. Not. R. Astron. Soc. **377**, 667 (2007).

[110] S. Naoz and R. Barkana, Mon. Not. R. Astron. Soc. **362**, 1047 (2005).

[111] A. Lewis, Phys. Rev. D **76**, 063001 (2007).

[112] A. Lewis and A. Challinor, Phys. Rev. D **76**, 083005 (2007).

[113] W. H. Press and P. Schechter, Astrophys. J. **187**, 425 (1974).

[114] J. R. Bond, S. Cole, G. Efstathiou, and N. Kaiser, Astrophys. J. **379**, 440 (1991).

[115] Z. Haiman, A. A. Thoul, and A. Loeb, Astrophys. J. **464**, 523 (1996).

[116] V. Bromm, P. S. Coppi, and R. B. Larson, Astrophys. J. **564**, 23 (2002).

[117] S. Naoz, R. Barkana, and A. Mesinger, Mon. Not. R. Astron. Soc. **399**, 369 (2009).

[118] T. Abel, G. L. Bryan, and M. L. Norman, Science **295**, 93 (2002).

[119] B. W. OShea, T. Abel, D. Whalen, and M. L. Norman, Astrophys. J. Lett. **628**, L5 (2005).

[120] N. Yoshida, K. Omukai, and L. Hernquist, Science **321**, 669 (2008).

[121] D. Tseliakhovich and C. M. Hirata, Phys. Rev. D **82**, 083520 (2010).

[122] R. A. Sunyaev and Y. B. Zeldovich, Astrophys. Space Sci. **7**, 3 (1970).

[123] W. Hu and N. Sugiyama, Astrophys. J. **444**, 489 (1995).

[124] C. Ma and E. Bertschinger, Astrophys. J. **455**, 7 (1995).

[125] J. Silk, Astrophys. J. **151**, 459 (1968).

[126] P. J. E. Peebles, Astrophys. J. Astrophys. J. 153, 1 (1968).

[127] Y. B. Zeldovich, V. G. Kurt, and R. A. Syunyaev, Zhurnal Eksperimental noi i Teoreticheskoi Fiziki **55**, 278 (1968).

[128] D. J. Eisenstein and W. Hu, Astrophys. J. **496**, 605 (1998).

[129] S. Seager, D. D. Sasselov, and D. Scott, Astrophys. J. Lett. **523**, L1 (1999).

[130] S. Seager, D. D. Sasselov, and D. Scott, Astrophys. J. Supp. **128**, 407 (2000).

[131] J. Dunkley, et al., Astrophys. J. Supp. **180**, 306 (2009).

[132] F. Bernardeau, S. Colombi, E. Gaztanaga, and R. Scoccimarro, Phys. Rep. **367**, 1 (2002).

[133] J. N. Fry, Astrophys. J. **279**, 499 (1984).

[134] M. H. Goroff, B. Grinstein, S. Rey, and M. B. Wise, Astrophys. J. **311**, 6 (1986).

[135] B. Jain and E. Bertschinger, Astrophys. J. **431**, 495 (1994).

[136] J. A. Peacock and S. J. Dodds, Mon. Not. R. Astron. Soc. **267**, 1020 (1994).

[137] M. Shoji and E. Komatsu, Astrophys. J. **700**, 705 (2009).

[138] W. Hu and D. J. Eisenstein, Astrophys. J. **498**, 497 (1998).

[139] U. Seljak and M. Zaldarriaga, Astrophys. J. **469**, 437 (1996).

[140] N. Dalal, O. Dore, D. Huterer, and A. Shirokov, Phys. Rev. D **77**, 123514 (2008).

[141] A. Slosar, et al. J. Cosm. Astropart. Phys. **8**, 31 (2008).

[142] S. Cole and N. Kaiser, Mon. Not. R. Astron. Soc. 237, 1127 (1989).

[143] U. Seljak, Mon. Not. R. Astron. Soc. **318**, 203 (2000).

[144] P. McDonald, Phys. Rev. D **74**, 103512 (2006).

[145] S. R. Furlanetto, M. Zaldarriaga, and L. Hernquist, Mon. Not. R. Astron. Soc. **365**, 1012 (2006).

[146] M. Dijkstra, A. Lidz, and J. S. B. Wyithe, Mon. Not. R. Astron. Soc. **377**, 1175 (2007).

[147] R. Barkana and A. Loeb, Phys. Rep. **349**, 125 (2001).

[148] I. T. Iliev, E. Scannapieco, H. Martel, and P. R. Shapiro, Mon. Not. R. Astron. Soc. **341**, 81 (2003).

[149] M. Zaldarriaga, S. R. Furlanetto, and L. Hernquist, Astrophys. J. **608**, 622 (2004).

[150] S. R. Furlanetto, M. Zaldarriaga, and L. Hernquist, Astrophys. J. **613**, 1 (2004).

[151] B. Ciardi et al., Mon. Not. R. Astron. Soc. **366**, 689 (2006).

[152] O. Zahn, et al., Astrophys. J. **654**, 12 (2007).

[153] B. Zygelman, Astrophys. J. **622**, 1356 (2005).

[154] U. Pen, New Astronomy **9**, 417 (2004).

[155] K. S. Mandel and M. Zaldarriaga, Astrophys. J. **647**, 719 (2006).

[156] O. Zahn and M. Zaldarriaga, Astrophys. J. **653**, 922 (2006).

[157] R. B. Metcalf and S. D. M. White, Mon. Not. R. Astron. Soc. **381**, 447 (2007).

[158] A. Cooray, Phys. Rev. Letters **97**, 261301 (2006).

[159] A. Cooray, C. Li, and A. Melchiorri, Phys. Rev. D **77**, 103506 (2008).

[160] A. R. Zentner and J. S. Bullock, Astrophys. J. **598**, 49 (2003).

[161] J. Diemand, M. Kuhlen, and P. Madau, Astrophys. J. **667**, 859 (2007).

[162] V. Springel, et al., Mon. Not. R. Astron. Soc. **391**, 1685 (2008).

[163] C. Bennett et al., Astrophys. J. Supp. **148**, 1 (2003).

[164] T. L. Abel, G. L. Bryan, and M. L. Norman, Science **295**, 93 (2002).

[165] B. O'Shea et al., Astrophys. J. Lett. **628**, L5 (2005).

[166] N. Yoshida, K. Omukai, L. Hernquist, Science **321**, 669 (2008).

[167] A. Stacy, V. Bromm, and A. Loeb, Astrophys. J. **730L**, 1 (2011).

[168] A. Stacy, T. Greif, V. Bromm, AIP Conference Proceedings, **1294**, 289 (2010).

[169] P. Clark, et al., Science **331**, 1040 (2011).

[170] P. Madau, A. Meiksin, M. J. Rees, Astrophys. J. **475**, 429 (1997).

[171] Z. Haiman, T. Abel, P. Madau, Astrophys. J. **551**, 599 (2001).

[172] R. Barkana, A. Loeb, Astrophys. J. **578**, 1 (2002).

[173] I. Iliev, E. Scannapieco, P. Shapiro, Astrophys. J. **624**, 491 (2005).

[174] B. Ciardi et al., Mon. Not. R. Astron. Soc. **366**, 689 (2005).

[175] I. Iliev, E. Scannapieco, H. Martel, P. Shapiro, Mon. Not. R. Astron. Soc. **341**, 81 (2003).

[176] S. Furlanetto, S. P. Oh, Astrophys. J. **652**, 849 (2006).

[177] N. Dalal, U. L. Pen, U. Seljak, J. Cosm. Astropart. Phys. **11**, 007 (2010).

[178] S. Naoz, N. Yoshida, R. Barkana, Mon. Not. R. Astron. Soc. **416**, 232 (2011).

[179] D. Eisenstein, W. Hu, Astrophys. J. **496**, 605 (1998).

[180] S. Naoz, R. Barkana, Mon. Not. R. Astron. Soc. **362**, 1047 (2005).

[181] N. Y. Gnedin, L. Hui, Mon. Not. R. Astron. Soc. **296**, 44 (1998).

[182] S. Naoz, R. Barkana, Mon. Not. R. Astron. Soc. **377**, 667 (2007).

[183] N. Y. Gnedin, Astrophys. J. **542**, 535 (2000).

[184] R. K. Sheth, G. Tormen, Mon. Not. R. Astron. Soc. **308**, 119 (1999).

[185] S. Naoz, R. Barkana, A. Mesinger, Mon. Not. R. Astron. Soc. **377**, 667 (2009).

[186] R. Barkana, A. Loeb, submitted to Mon. Not. R. Astron. Soc. **415**, 3113 (2011).

[187] W. H. Press, P. Schechter, Astrophys. J. **187**, 425 (1974).

[188] R. K. Sheth, G. Tormen, Mon. Not. R. Astron. Soc. **329**, 61 (2002).

[189] S. Naoz, S. Noter, R. Barkana, Mon. Not. R. Astron. Soc. : Letters **373**, L98 (2006).

[190] J. R. Bond, S. Cole, G. Efstathiou, and N. Kaiser, Astrophys. J. **379**, 440 (1991).

[191] M. Tegmark et al., Astrophys. J. **474**, 1 (1997).

[192] V. Bromm, P. S. Coppie, R. B. Larson, Astrophys. J. **564**, 23 (2002).

[193] T. M. Fuller, H. M. P. Couchman, Astrophys. J. **544**, 6 (2000).

[194] N. Yoshida et al., Astrophys. J. **598**, 73 (2003).

[195] D. S. Reed et al., Mon. Not. R. Astron. Soc. **363**, 393 (2005).

[196] R. Barkana, A. Loeb, Astrophys. J. **609**, 474 (2004).

[197] A. Fialkov, R. Barkana, D. Tseliakhovich, and C. M. Hirata, Mon. Not. R. Astron. Soc. submitted (arXiv1110.2111) (2011).

[198] E. Visbal, R. Barkana, A. Fialkov, D. Tseliakhovich, and C. M. Hirata, submitted to Nature, (2012).

[199] R. Barkana, A. Loeb, Phys. Rept. **349**, 125 (2001).

[200] D. Tseliakhovich, R. Barkana, and C. M. Hirata, Mon. Not. R. Astron. Soc. **418**, 906 (2011).

[201] T. Greif, S. White, R. Klessen, and V. Springel, Astrophys. J. **736**, 147 (2011).

[202] U. Maio, L. V. E. Koopmans, and B. Ciardi, Mon. Not. R. Astron. Soc. **412L**, 40 (2011).

[203] S. Naoz, N. Yoshida, and N. Y. Gnedin, submitted to Mon. Not. R. Astron. Soc. arXiv:1108.5176, (2011).

[204] R. Barkana and A. Loeb, Astrophys. J. **609**, 474 (2004).

[205] Z. Haiman, M. J. Rees, and A. Loeb, Astrophys. J. **467**, 522 (1996).

[206] Z. Haiman, M. J. Rees, and A. Loeb, Astrophys. J. **476**, 458; erratum, 484, 985 (1997).

[207] E. Fermi, Phys. Rev. **75**, 1169 (1949).

[208] M. A. Lieberman and A. J. Lichtenberg, Phys. Rev. A **5**, 1852 (1972).

[209] K. Heng, PASA **27**, 23 (2010).

[210] R. A. Chevalier, R. P. Kirshner and J. C. Raymond, Astrophys. J. **235**, 186 (1980).

[211] R. Kirshner, P. F. Winkler, and R. A. Chevalier, Astrophys. J. **315**, L135 (1987).

[212] R.C. Smith, R. P. Kirshner, W. P. Blair, and P. R. Winkler, Astrophys. J. **375**, 652 (1991).

[213] P. Ghavamian, P. R. Winkler, J. C. Raymond,and K.S. Long, Astrophys. J. **572**, 888 (2002).

[214] R. A. Fesen, R. H. Becker, W. P. Blair, and K.S. Long, Astrophys. J. **338**, L13 (1989).

[215] W. P. Blair, K. S. Long, and O. Vancura, Astrophys. J. **366**, 484 (1991).

[216] J. Sollerman et al., A&A **407**, 249 (2003).

[217] J. C. Raymond, W. P. Blair, R. A. Fesen, and T. R. Gull, Astrophys. J. **275**, 636 (1983).

[218] R. A. Fesen, and H. Itoh, Astrophys. J. **295**, 43 (1985).

[219] P. Ghavamian, J. Raymond, R. C. Smith, and P. Hartigan, Astrophys. J. **547**, 995 (2001).

[220] K.S. Long, and W. P. Blair, Astrophys. J. **358**, L13 (1990).

[221] R.C. Smith, Astron. J. **114**, 2664 (1997).

[222] K. Heng et al., Astrophys. J. **644**, 959 (2006).

[223] N. Bucciantini, A&A **393**, 629 (2002).

[224] K. Heng and R. McCray, Astrophys. J. **654**, 923 (2007).

[225] P. Ghavamian, J. M. Laming, and C. E. Rakowski, Astrophys. J. **654**, L69 (2007).

[226] S. Nikolic, K. Heng, G. van de Ven, D. Tseliakhovich, and C. M. Hirata, in preparation (2012).

[227] Ya. B. Zel'dovich, and Yu. P. Raizer, *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena*, Vols. 1 and 2, New York: Dover Publications, (1966).

[228] S. Kwok, Physics and Chemistry of the Interstellar Medium, University Science Books, (2007).

[229] R. A. Chevalier and J. C. Raymond, Astrophys. J. **225**, L27 (1978).

[230] W. Fritsch and C. D. Lin, Phys. Rep. **202**, 1 (1991).

[231] D. R. Bates and G. Griffing, Proc. Phys. Soc., London, Sect. A **66**, 961 (1953).

[232] F. Martin, J. Phys. B **32**, 501 (1999).

[233] D. R. Schultz, P. S. Krstic, T. G. Lee, and J. C. Raymond. Astrophys. J. **678**, 950 (2008).

[234] R. Fathi, E. Ghanbari-Adivi, M. A. Bolorizadeh, F. Shojaei, and M. J. Brunger, J. Phys. B **42**, 1 (2009).

[235] E. A. Helder, et al., Science **325**, 719 (1998).

[236] A. Salin, *High Energy IonAtom Collisions* (Berlin: Springer), p. 245 (1988).

[237] R. F. Stebbings et al., Phys. Rev. **138**, A1312 (1965).

[238] L. Wilets, and D. F. Gallaher, Phys. Rev. **147**, 13 (1966).

[239] G. Ryding, A. B. Wittkower, and H. B. Gilbody, Proc. Phys. Soc. **89**, 547 (1966).

[240] J. E. Bayfield, Phys. Rev. **185**, 105 (1969).

[241] T. G. Winter, Phys. Rev. A **80**, 032701 (2009).

[242] A. L. Ford, J. F. Reading, and K. A. Hall, J. Phys. B **26**, 4537 (1993).

[243] J. Kuang, and C. Lin, J. Phys. B **29**, 5443 (1996).

[244] B. M. McLaughlin, T. G. Winter, and J. F. McCann, J. Phys. B **30**, 1043 (1997).

[245] N. Toshima, Phys. Rev. A **59**, 1981 (1999).

[246] D. S. F. Crothers and L. J. Dube, Adv. At., Mol., Opt. Phys. **30**, 287 (1992).

[247] G. J. N. Brown and D. S. F. Crothers, Phys. Rev. Letters **76**, 392 (1996).

[248] V. Maruhn-Reswani, N. Grun, and W. Scheid, Phys. Rev. Letters **43**, 512 (1979).

[249] C. Bottcher, Phys. Rev. Letters **48**, 85 (1982).

[250] K. C. Kulander, K. R. Sandhya Devi, and S. E. Koonin, Phys. Rev. A **25**, 2968 (1982).

[251] A. Kolakowska et al., Phys. Rev. A **58**, 2872 (1998).

[252] A. Kolakowska, M. S. Pindzola, and D. R. Schultz, Phys. Rev. A **59**, 3588 (1999).

[253] E. Sidky and C. Lin, Phys. Rev. A **65**, 012711 (2001).

[254] D. E. Osterbrock, Astron. J. **73**, 904 (1968).

[255] W. Adams and V. Petrosian, Astrophys. J. **192**, 199 (1974).

[256] E. Capriotti, Astrophys. J. **140**, 632 (1964).

[257] H. Netzer, Mon. Not. R. Astron. Soc. **171**, 395 (1975).

[258] G. Aldering, et al., Astrophys. J. **650**, 510 (2006).

[259] X. Dong et al., Mon. Not. R. Astron. Soc. **383**, 581 (2008).

[260] B. Groves, J. Brinchmann, and C. J. Walcher, Mon. Not. R. Astron. Soc. in press (arXiv:1109.2597) (2011).

[261] J. Baker and D. Menzel, Astrophys. J. **88**, 5 (1938).

[262] M. Seaton, Rep. Prog. Phys. **23**, 31 (1960).

[263] D. Osterbrock, Astrophysics of Gaseous Nebulae and Active Galactic Nuclei, Univ. Science Books, Mill Valley (1989).

[264] K. Heng and R. A. Sunyaev, A&A **481**, 117 (2008).